

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Tímový projekt – tím č. 14

Dokumentácia

Email: teamproject14@googlegroups

Webová stránka: team14-21.studenti.fiit.stuba.sk

Názov témy: Vytvorenie inteligentného model-based agenta (umelá inteligencia na báze Knowledge grafov) pre tvorbu komplexných dátových štruktúr a vzťahov pre aplikovaný výskum v klinickej onkológii

Vedúci tímu: Ing. Juraj Vincúr

Názov a číslo tímu: OncoAgents, tím č. 14

Členovia: Ivana Baleková, Martina Billichová, Martin Budinský, Miloš Galeta, Ivana Háberová, Patrícia Huďanová, Ján Huďan

Akademický rok: 2021/2022

Obsah

Riadenie projektu	9
Úvod	9
Role členov tímu a podiel práce	9
Členovia tímu:	9
Sumarizácie Šprintov	10
Zimný semester	10
Šprint 1 - The one where it all began (07.10.2021 – 18.10.2021)	10
Šprint 2 - The one where we had call during holiday (18.10.2021 – 01.11.2021)	12
Šprint 3 - The one with all the documentation (01.11.2021 – 15.11.2021)	13
Šprint 4 - The one where we tried NLP (15.11.2021 – 29.11.2021)	15
Šprint 5 - The one where we created our own path (29.11.2021 – 13.12.2021)	17
Letný semester	19
Šprint 6 - The one with new project (28.02.2022 – 07.03.2022)	19
Šprint 7 - The one where we created base project (07.03.2022 – 21.03.2022)	20
Šprint 8 - The one where server fell (21.03.2022 – 04.04.2022)	22
Šprint 9 - The one where we created base project (04.04.2022 – 25.04.2022)	24
Šprint 10 - The one where we finished (25.04.2022 – 09.05.2022)	26
Retrospektíva	28
Zimný semester	28
Letný semester	30
Motivačný dokument	32
Kto sme	32
O nás	32
Zoradenie tém podľa priority	33
Rozvrh tímu	33
Metodiky	34
Metodika komunikácie	34
Komunikačné kanály	34
Metodika dokumentácie	35
Pravidlá pre písanie dokumentácie:	35
Zverejňovanie dokumentácie	35

Revízie dokumentácie	35
Šablóny opakujúcich sa dokumentov	35
Šablóna zápisnice	35
Šablóna retrospektívy	36
Metodika pre správu úloh	36
Story	37
Subtasky	37
Vytváranie Story	37
Vytváranie Subtasku	37
Metodika verziovania	38
Repozitár	38
Základné pojmy	38
Branch	38
Pomenovanie vetiev	38
Spájanie vetiev	38
Commit	39
Commit message	39
Metodika testovania	39
UNIT testy	39
Nepretržité testovanie (Continuous testing)	39
Inžinierske dielo - Zimný semester	40
Big picture	40
Ciele projektu na zimný semester	40
Celkový pohľad na vyhotovené časti projektu	40
Dokumentácia k no-coding platforme BigRing	40
Modelovanie	40
Model 1	41
Model 2	41
Analýza medicínskej domény	42
Analýza NLP	42
Získavanie odborných článkov	43
Inžinierske dielo - Letný semester (zmena témy)	44
Big picture	44

Ciele projektu na letný semester	44
Celkový pohľad na vyhotovené časti projektu	44
Návrh riešenia	45
Ciele	45
Funkčné vlastnosti	45
Scenár prípadu použitia	45
Návrh UI	46
Vizualizácia výstupu NLP analýzy textu	48
Implementácia	49
Obrazovky používateľského rozhrania	49
Extrakcia tripletov	55
Testovanie	57
VANET (tím č. 7): Posudok na prototyp tímu č. 14: OncoAgents	57
1.1. Úvod	57
1.2. Hodnotenie prototypu	57
1.3. Hodnotenie práce tímu	57
1.4. Zhodnotenie	57
Používateľské akceptačné testovanie	58
Inštalačná príručka	59
Backend	59
Frontend	59
Používateľská príručka	60
Prihlásenie a odhlásenie	60
Vyhľadávanie	61
Interakcia s grafom	62
Detail článku	64
Technická dokumentácia	65
Backend	65
Použité technológie	65
Endpointy a opis tried	65
Articles	65
FullArticleFormat	66
Images	66

Triples	66
Wordcloud	66
Frontend	67
Použité technológie	67
Opis tried	68
/views	68
Login	68
Articles	68
/components	68
ArticleDetailLeftDrawer	68
ArticleItem	68
ArticlesList	68
ArticlesSearch	69
ArticlesText	69
DescriptionPanel	69
ImageGrid	69
KnowledgeGraph	69
Navigation	69
Panel	69
PanelTabs	69
ProtectedRoute	70
RadialTree	70
TabSelector	70
TreeSvg	70
WordCloud	70
Export evidencie úloh	71
Šprint 1 - The one where it all began (07.10. 2021 – 18. 10. 2021)	71
Šprint 2 - The one where we had call during holiday (18. 10. 2021 – 1. 11. 2021)	71
Šprint 3 - The one with all the documentation (01. 11. 2021 – 15. 11. 2021)	72
Šprint 4 - The one where we tried NLP (15. 11. 2021 – 29. 11. 2021)	72
Šprint 5 - The one where we created our own path (29. 11. 2021 – 13. 12. 2021)	73
Šprint 6 - The one with new project (28.02.2022 – 07.03.2022)	74
Šprint 7 - The one where we created base project (07.03.2022 – 21.03.2022)	75

Šprint 8 - The one where server fell (21.03.2022 – 04.04.2022)	76
Šprint 9 - The one where we created base project (04.04.2022 – 25.04.2022)	77
Šprint 10 - The one where we finished (25.04.2022 – 09.05.2022)	78
Príloha A - Používateľská príručka	80
OVERENIE FUNKČNOSTI PROTOTYPU	91
Príloha B - Analýza medicínskej domény podľa obrázkov	94
Príloha C - Analýza rôznych štruktúr vedeckých článkov	104
Výskumné články	104
Štruktúra článku	104
Dôležité časti:	104
Dáta zo stránky https://clinicaltrials.gov/	104
Štruktúra dát	104
Dôležité časti	104
Dáta zo stránky https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/	104
Príloha D - Analýza NLP	105
Analýza nástroja vytvoreného Kinit -SlovakBERT	106
Analýza nástrojov pre anglický jazyk	107
Natural Language Toolkit - NLTK	107
Stanford OpenIE	108
Ollie	109
spaCy	110
Apache OpenNLP	112
Analýza nástrojov pre český a slovenský jazyk	112
Nástroje pre slovenský jazyk	112
Bednárík - NLP web	112
STANZA	113
SLOVAK-LEXER	113
dl4dp	113
NLP4SK	113
Nástroje pre český jazyk	114
Ajka, Majka	114
Set	114

BERT (Bidirectional Encoder Representations from Transformers)	115
Príloha E - Porovnanie testovaných NLP nástrojov a analýza výstupov	116
Natural Language Toolkit - NLTK (Python)	116
Ollie (Java)	117
Stanford OpenIE (Java)	117
SpaCy (Python)	118
Zhrnutie	118
Príloha F - Analýza podobných prác	120
Komerčné nástroje	120
Connected Papers	120
Inciteful	120
Ďalšie podobné nástroje	121
Výskumné práce	122
Prínos	122
PubMed knowledge graph (PKG)	122
CONSORT-NLP	123
Ukážky rozhrania aplikácie	124
A Case Study in Enhanced Living Environments	125
Bert	125
Príloha G - Analýza API pre získavanie článkov	127
IEEE Xplore API	127
arXiv API	127
CORE API	127
Springer APIs	128
Elsevier	128
Záver	128
Príloha H - Analýza backend rámce	129
Úvod	129
Porovnanie vybraných frameworkov:	129
Django	129
Laravel	129
ExpressJS	129

Spring Boot	130
Flask	130
Príloha I - Analýza získavania sémantických tripletov	131
Machine learning	131
Tree Bank Parser - Stanford parser	131
The Multi-Liaison Algorithm	133

Riadenie projektu

Úvod

V tomto dokumente je výsledok práce tímu číslo 14 (OncoAgents) pre tému s názvom "Vytvorenie inteligentného model-based agenta (umelá inteligencia na báze Knowledge grafov) pre tvorbu komplexných dátových štruktúr a vzťahov pre aplikovaný výskum v klinickej onkológii" po troch šprintoch v zimnom semestri.

Dokument obsahuje predstavenie tímu, rozdelenie rolí, sumarizácie šprintov a retrospektív, motivačný dokument. Dôležitou časťou sú metodiky komunikácie, dokumentácie a správy úloh, ktoré určujú pravidlá, ktoré pri práci na tomto projekte dodržiavame. V závere je sumarizácia vo forme Big Picture ako aj ciele pre aktuálny semester.

Roly členov tímu a podiel práce

Náš tím sa skladá zo 7 členov, vedúcim tímu je Ing. Juraj Vincúr, ktorý je zároveň aj produktovým vlastníkom. V prvej časti sme spolupracovali s externou firmou BigRing Platform, s.r.o., ktorá je zadávateľom projektu ako aj firmou Lambda Life a. s., neskôr sme však pracovali na vlastnom projekte.

Členovia tímu:

Bc. Ivana Baleková - backend developerka

Bc. Martina Billichová - frontend developerka

Bc. Martin Budinský - frontend developer, špecialista na NLP

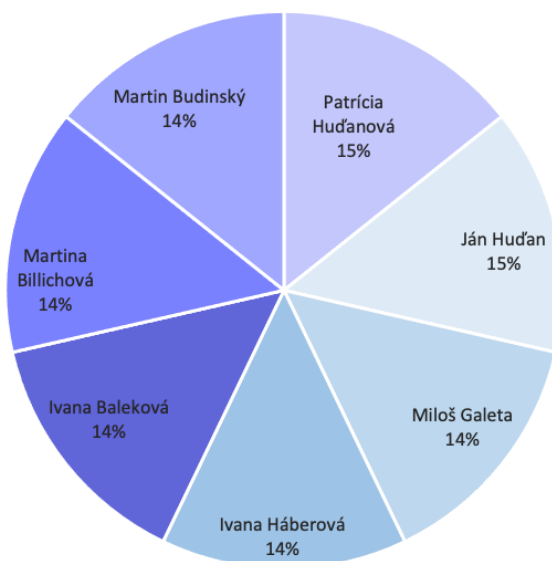
Bc. Miloš Galeta - backend developer

Bc. Ivana Háberová - špecialistka na NLP

Bc. Patrícia Huďanová - scrum master, frontend developerka

Bc. Ján Huďan - manažér testovania, špecialista na nasadzovanie

Podiel práce



Väčšinu času sme pracovali na projekte spoločne v stanovenom čase, preto je podiel práce strávenej na projekte v podstate u každého člena rovnaký - aj čo sa týka zimného, aj letného semestra.

Sumarizácie Šprintov

Tvorbu tímového projektu máme rozdelenú podľa šprintov. Jeden šprint má dĺžku dva týždne a v ojedinelých prípadoch jeden týždeň. V kapitolách zimný semester a letný semester sú zosumarizované všetky vykonané šprinty počas trvania projektu.

Zimný semester

Šprint 1 - The one where it all began (07.10.2021 – 18.10.2021)

Šprint sme začali neskôr, keďže sme prvé stretnutie posunuli až na štvrtok aby sa konalo po prednáške kde nám boli vysvetlené základné princípy scrum a agile. V prvom šprinte sme sa venovali najmä rozbehaniu nástrojov, ktoré budeme používať a prvému stretnutiu s firmami BigRing a Lambda Life.

Do nášho komunikačného kanálu Slack sme pridali Jiru, pri každej zmene nám príde notifikácia. Taktiež sme si v Slacku vytvorili komunikačné kanály pre lepšiu prácu a uľahčenie komunikácie.

K dispozícii sme mali nahrávky a dokumenty z tímového projektu, ktorý minulý rok takisto pracoval s platformou firmy BigRing. Tie sme si v tomto šprinte pozreli a prečítali pre lepšie oboznámenie sa so systémom, s ktorým budeme ďalej na tímovom projekte pracovať. Taktiež sme sa stretli so zástupcami BigRing-u, kde nám námatkovo predstavili ich aplikáciu. Okrem toho sme sa predstavili my im a tiež sme porozprávali, s čím sme už doteraz pracovali.

Ďalej sme si vytvorili webové sídlo tímu. V tomto bode bolo potrebné rozbehať virtuálny stroj. O ten sme požiadali a po pridelení sme tam nastavili potrebné veci pre úspešný beh našej webovej stránky.

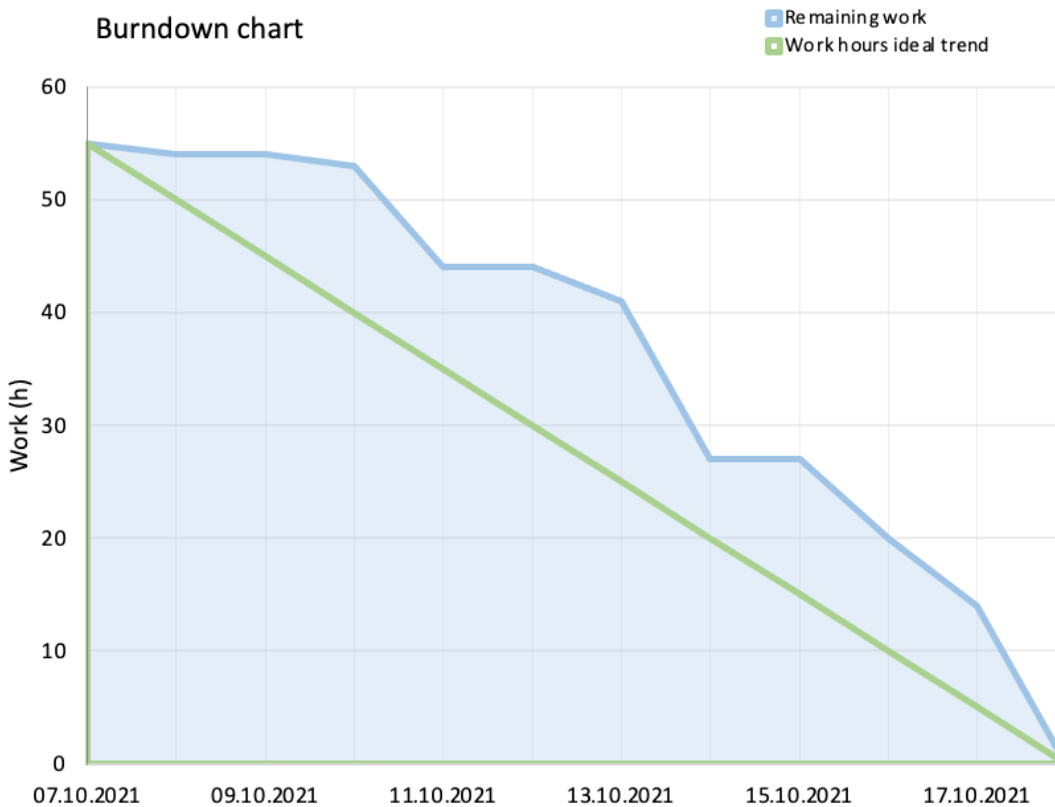
Vytvorili sme taktiež našu webovú stránku, zadefinovali sme si jej obsah a na náš virtuálny stroj sme pridali dokumenty, ktoré si je možné z našej stránky stiahnuť.

Posledný bod obsahoval prípravu šablón. Tu sme si vopred preddefinovali kostru našich šablón pre rôzne dokumenty, ktoré sme používali pri písaní formulárov. Tieto šablóny sme taktiež pridali na našu webovú stránku, skadiaľ si ich je možné stiahnuť.

V tomto šprinte sme splnili všetky úlohy a šprint sme úspešne dokončili.

Splnené stories:

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-4	Oboznámenie sa so systémom	13		Done	Done
Subtask	TEAM-7	Stretnutie s BigRing		Patrícia Huďanová	Done	Done
Subtask	TEAM-6	Prečítanie dokumentov z minulého roka		Martin Budinský	Done	Done
Subtask	TEAM-5	Pozretie nahrávok z minulého roka		Ivana Baleková	Done	Done
Story	TEAM-3	Príprava formulárov	3		Done	Done
Subtask	TEAM-11	Template pre oficiálny doc		Ivana Háberová	Done	Done
Subtask	TEAM-10	Template pre retrospektívu		Martina Billichová	Done	Done
Subtask	TEAM-9	Template pre zápisnice		Ivana Baleková	Done	Done
Story	TEAM-2	Slack- prepojenie s Jirou	3		Done	Done
Subtask	TEAM-8	Integrácia Jiri do Slacku		Miloš Galeta	Done	Done
Story	TEAM-1	Webové sídlo tímu	8		Done	Done
Subtask	TEAM-15	Nahrávanie a sťahovanie súborov		Martin Budinský	Done	Done
Subtask	TEAM-14	Definovanie obsahu stránky		Patrícia Huďanová	Done	Done
Subtask	TEAM-13	Vytvorenie statickej stránky		Ján Huďan	Done	Done
Subtask	TEAM-12	Rozbehovanie virtual machine		Ján Huďan	Done	Done



Šprint 2 - The one where we had call during holiday (18.10.2021 – 01.11.2021)

V druhom šprinte sme sa zamerali na oboznámenie sa s platformou bigring.

Keďže sme mali prístup k základnej dokumentácii k modelovacej platforme, začali sme jej štúdiom. Z dokumentácie sme následne spravili dokument. V dokumente sme uviedli jednotlivé časti spomenuté v dokumentácii s popisom a screenshotom ako ich vieme nájsť.

Následne sme sa venovali tvorbe prvého modelu. Ako referenciu pri tvorbe sme použili minuloročný tímový projekt.

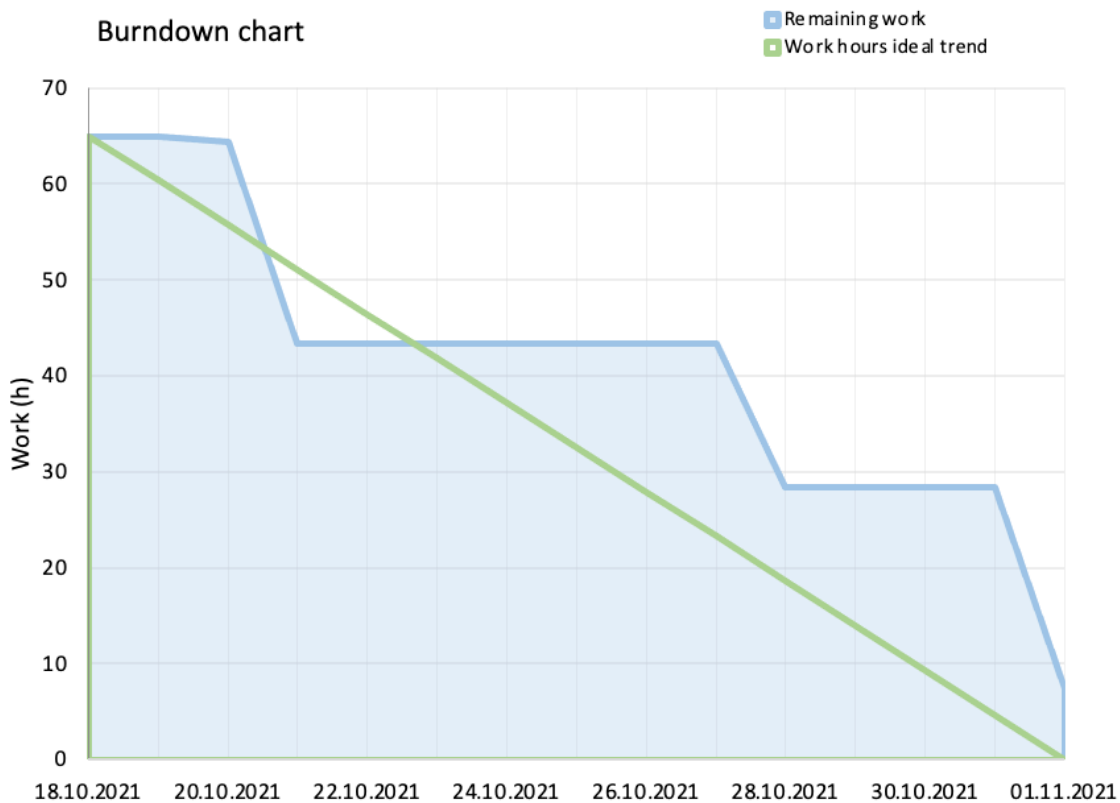
Náš výsledný model sme ešte na stretnutí spoločne s technikom z BigRing mierne upravili aby fungoval správne.

V tomto šprinte sme splnili všetky úlohy a šprint sme úspešne dokončili.

Splnené stories:

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-18	Stretnutie s BigRing-om	10		Done	Done
Subtask	TEAM-20	Stretnutie s BigRing-om		Patrícia Huďanová	Done	Done

Subtask	TEAM-19	Vytvorenie rozvrhu		Martina Billichová	Done	Done
Story	TEAM-17	Štúdium dokumentácie k platforme od BigRing	13		Done	Done
Subtask	TEAM-22	Štúdium dokumentácie k platforme		Ján Huďan	Done	Done
Subtask	TEAM-21	Vytvorenie dokumentu		Miloš Galeta	Done	Done
Story	TEAM-16	Vytvorenie prvého modelu	13		Done	Done
Subtask	TEAM-24	Vytvorenie zadania		Ivana Háberová	Done	Done
Subtask	TEAM-23	Vytvorenie modelu		Martin Budinský	Done	Done



Šprint 3 - The one with all the documentation (01.11.2021 – 15.11.2021)

Tretí šprint sa opäť venoval štúdiu dokumentácie a vytváraniu modelu.

Počas tohto šprintu sme sa venovali aj analýze medicínskej domény, nakoľko je s ňou téma nášho projektu úzko spätá. Analyzovali sme predovšetkým niekoľko kľúčových slov a výstupom našej analýzy je dokument, s opisom niektorých procesov a s linkami na bližšie zdroje.

Navrhli sme a vytvorili druhý funkčný model v platforme od BigRing, ktorý využíva väčšinu poskytovaných entít a vzťahov. Paralelne s tým sme vytvorili dokument s postupom ako sa dostať v platforme od vytvorenia Solutionu až k funkčnému modelu.

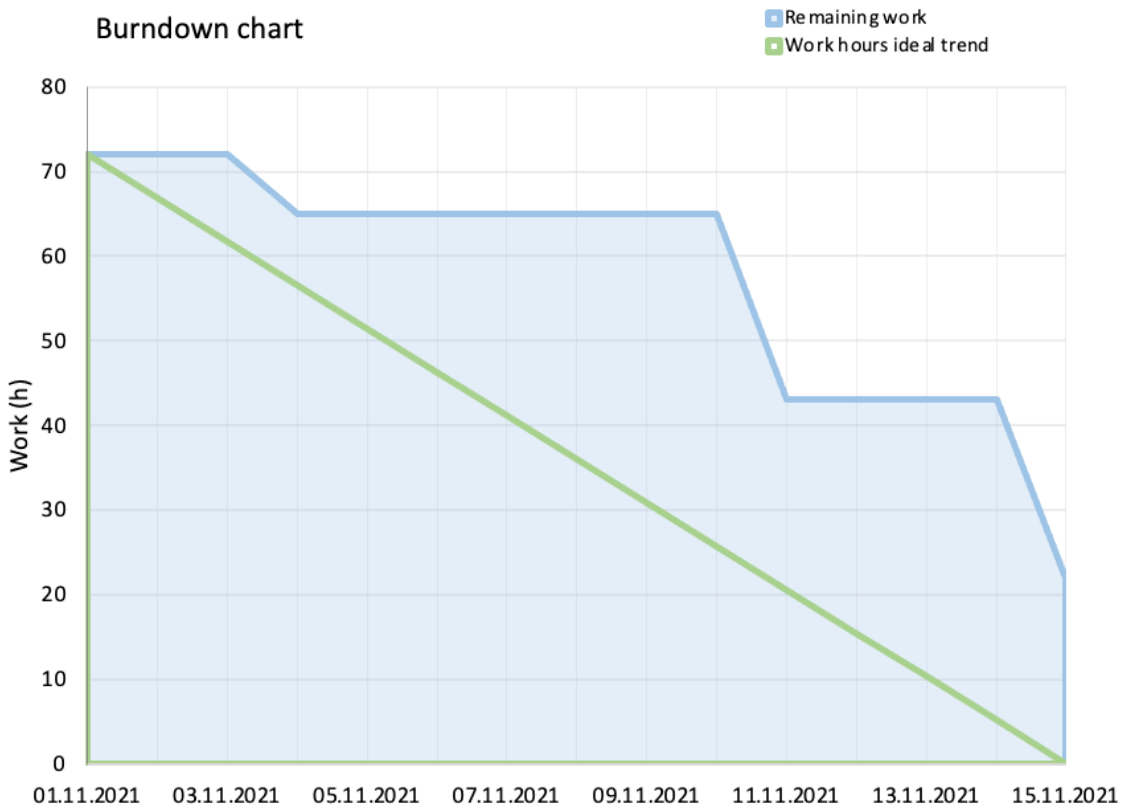
V tomto šprinte sme uzavreli 2 story zo štyroch. Nepodarilo sa nám získať prístup do gitbook dokumentácie a ani zorganizovať stretnutie s BigRingom kvôli vyťažnosti firmy.

Splnené stories:

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-26	Návrh a vytvorenie ďalšieho modelu	13	Ivana Baleková	Done	Done
Subtask	TEAM-32	Vytvorenie funkčného modelu		Ivana Baleková	Done	Done
Subtask	TEAM-31	Vytvorenie zadania		Ivana Háberová	Done	Done
Story	TEAM-25	Analýza podkladov(Lambda Life)	5	Patrícia Huďanová	Done	Done
Subtask	TEAM-30	Analýza podkladov od Lambda Life		Patrícia Huďanová	Done	Done

Nesplnené stories:

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-28	Stretnutie s BigRing-om	13	Martin Budinský	To Do	Unresolved
Subtask	TEAM-34	Stretnutie s BigRing-om		Martin Budinský	To Do	Unresolved
Story	TEAM-27	Štúdium gitbook dokumentácie	20	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-33	Tvorba dokumentu s flow-om		Ján Huďan	Done	Done
Subtask	TEAM-44	Tvorba dokumentu k časti procesy		Patrícia Huďanová	To Do	Unresolved



Šprint 4 - The one where we tried NLP (15.11.2021 – 29.11.2021)

V rámci tohto šprintu sme sa zamerali na analýzu rôznych NLP nástrojov pre rôzne jazyky a vytvorili sme dokument s popisom jednotlivých nástrojov. Popísali sme vstupy aj výstupy jednotlivých nástrojov, ktoré by mohli byť užitočné pre našu analýzu textu.

Ďalej sme sa zamerali na vytvorenie prototypu na spracovanie štúdií/vedeckých článkov. V rámci tohto problému sme si vybrali tri nástroje z predošlej analýzy a otestovali sme ich nad krátkym textom z domény onkológie.

V tomto šprinte sme opäť uzavreli 2 story zo štyroch. Opätovne sa nám nepodarilo získať prístup do gitbook dokumentácie a ani zorganizovať stretnutie s BigRingom, tieto úlohy sme ponechali preto do budúcnosti iba v backlogu a nedávame ich už do ďalšieho šprintu.

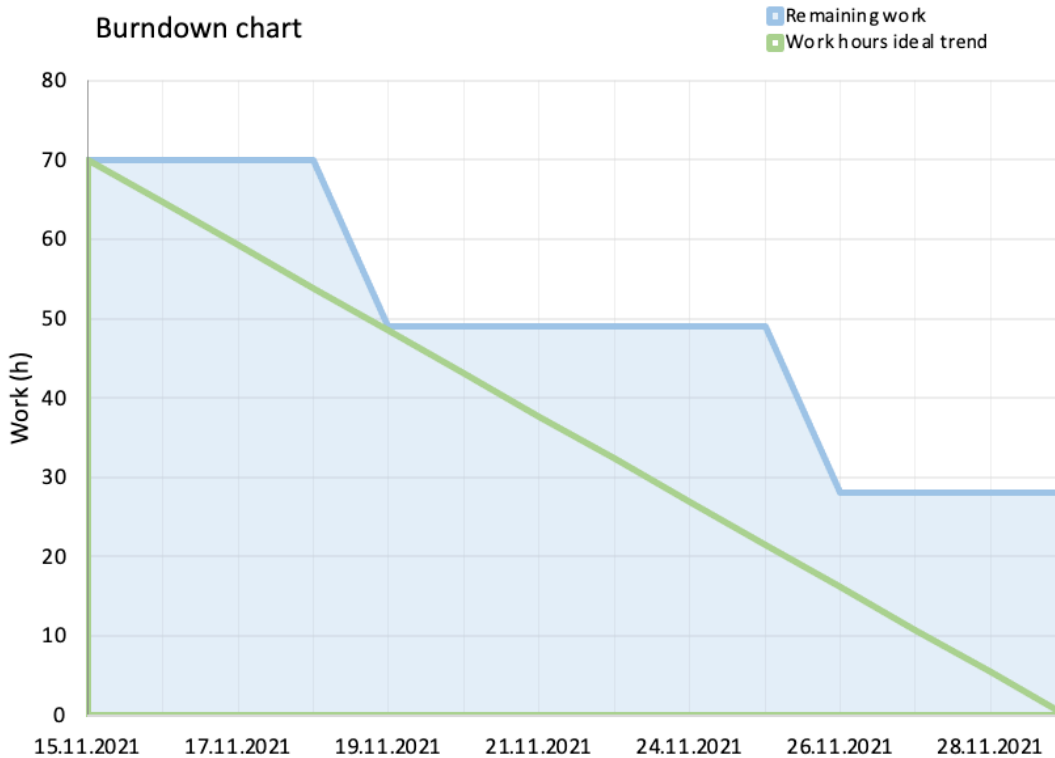
Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-35	Implementácia prototypu na spracovanie štúdií	13		Done	Done
Subtask	TEAM-43	Analýza výsledkov		Ján Huďan	Done	Done
Subtask	TEAM-42	Spustenie a otestovanie NLP knižníc na dátach		Patrícia Huďanová	Done	Done
Subtask	TEAM-41	Predspracovanie datasetu		Martin Budinský	Done	Done

Subtask	TEAM-40	Získanie datasetu klinických štúdií		Ivana Haberova	Done	Done
Story	TEAM-29	Analýza NLP	14		Done	Done
Subtask	TEAM-39	Analýza toolu ktorý vytvoril kinit		Miloš Galeta	Done	Done
Subtask	TEAM-38	Analýza NLP toolov pre český jazyk		Martina Billichová	Done	Done
Subtask	TEAM-37	Analýza NLP toolov pre slovenský jazyk		Ján Huďan	Done	Done
Subtask	TEAM-36	Analýza NLP toolov pre anglický jazyk		Martin Budinský	Done	Done

Nesplnené stories:

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-28	Stretnutie s BigRing-om	13	Martin Budinský	To Do	Unresolved
Subtask	TEAM-34	Stretnutie s BigRing-om		Martin Budinský	To Do	Unresolved
Story	TEAM-27	Štúdium gitbook dokumentácie	20	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-33	Tvorba dokumentu s flow-om		Ján Huďan	Done	Done
Subtask	TEAM-44	Tvorba dokumentu k časti procesy		Patrícia Huďanová	To Do	Unresolved



Šprint 5 - The one where we created our own path (29.11.2021 – 13.12.2021)

V rámci piateho šprintu sme spravili analýzu rôznych komerčných a vedeckých nástrojov kde pomocou NLP robili vyhľadávanie nad textami.

Navrhli a špecifikovali sme si zadanie nášho projektu, ktorý by sme mohli robiť bez konzultácií s BigRing. Ako cieľ sme si dali vytvoriť nástroj pre zjednodušenie a sprehľadnenie vyhľadávania nad vedeckými článkami.

Zanalyzovali sme rôzne API prístupy na získanie vedeckých článkov a požiadali sme o prístup do niekoľkých z nich. Následne sme spravili prototyp pre vizualizáciu vzťahov vo vete.

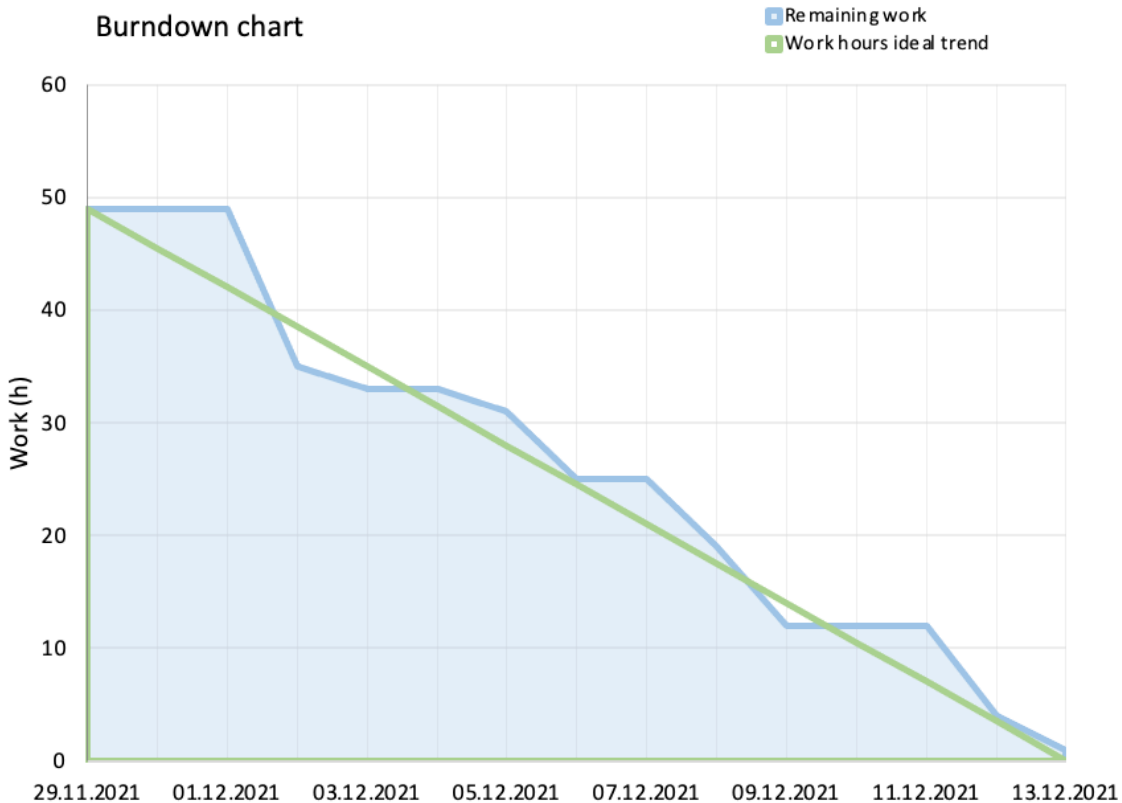
Takisto sme sa v tomto šprinte venovali ukončovaniu dokumentácie pre odovzdanie.

V tomto šprinte sme splnili všetky úlohy a šprint sme úspešne dokončili.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-49	Analýza frameworkov určených pre tvorbu vizualizácií a vytvorenie prototypu	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-60	Vytvoriť prototyp pre vizualizáciu		Patrícia Huďanová	Done	Done
Subtask	TEAM-59	Analyzovať webové frameworky pre vizualizáciu		Ján Huďan	Done	Done

Story	TEAM-47	Existujúce služby a API	8	Ivana Baleková	Done	Done
Subtask	TEAM-58	Analyzovať druhú polovicu API prístupov z libraries.mid.edu		Miloš Galeta	Done	Done
Subtask	TEAM-57	Analyzovať prvú polovicu API prístupov z libraries.mid.edu		Ivana Baleková	Done	Done
Story	TEAM-46	Návrh aplikácie	15	Martina Billichová	Done	Done
Subtask	TEAM-56	Návrh UI		Martina Billichová	Done	Done
Subtask	TEAM-55	Identifikovanie scenárov		Ivana Baleková	Done	Done
Subtask	TEAM-54	Špecifikácia zadanie		Ivana Haberova	Done	Done
Story	TEAM-45	Analýza related work	8	Martin Budinský	Done	Done
Subtask	TEAM-53	Vytvorenie dokumentu pre analýzu výskumných prác		Ivana Haberova	Done	Done
Subtask	TEAM-52	Analýza výskumných prác		Ivana Haberova	Done	Done
Subtask	TEAM-51	Vytvorenie dokumentu pre komerčné nástroje		Martin Budinský	Done	Done
Subtask	TEAM-50	Analýza komerčných nástrojov		Martin Budinský	Done	Done



Letný semester

Šprint 6 - The one with new project (28.02.2022 – 07.03.2022)

Šprint sme začali neskôr, keďže sme prvé stretnutie posunuli kvôli nezhodám v rozvrhu.

V rámci šprintu sme si zadefinovali čo a v akej forme potrebujeme na vizualizáciu jednotlivých článkov (ako by mal vyzerať model).

Následne sme sa venovali implementácii komunikácie s Api serverom od IEEE a zistili aké rôzne možnosti nám táto api ponúka.

Takisto sme vytvorili repozitár pre náš finálny produkt, kde sme po analýze rôznych frameworkov pre BE zvolili Django a pre FE React. Do tohto repozitára sme následne vytvorili základnú kostru stránky - formu pre vyhľadávanie, graf v ktorom sa majú zobrazíť články, zoznam článkov a zobrazenie popupu s detailom článku.

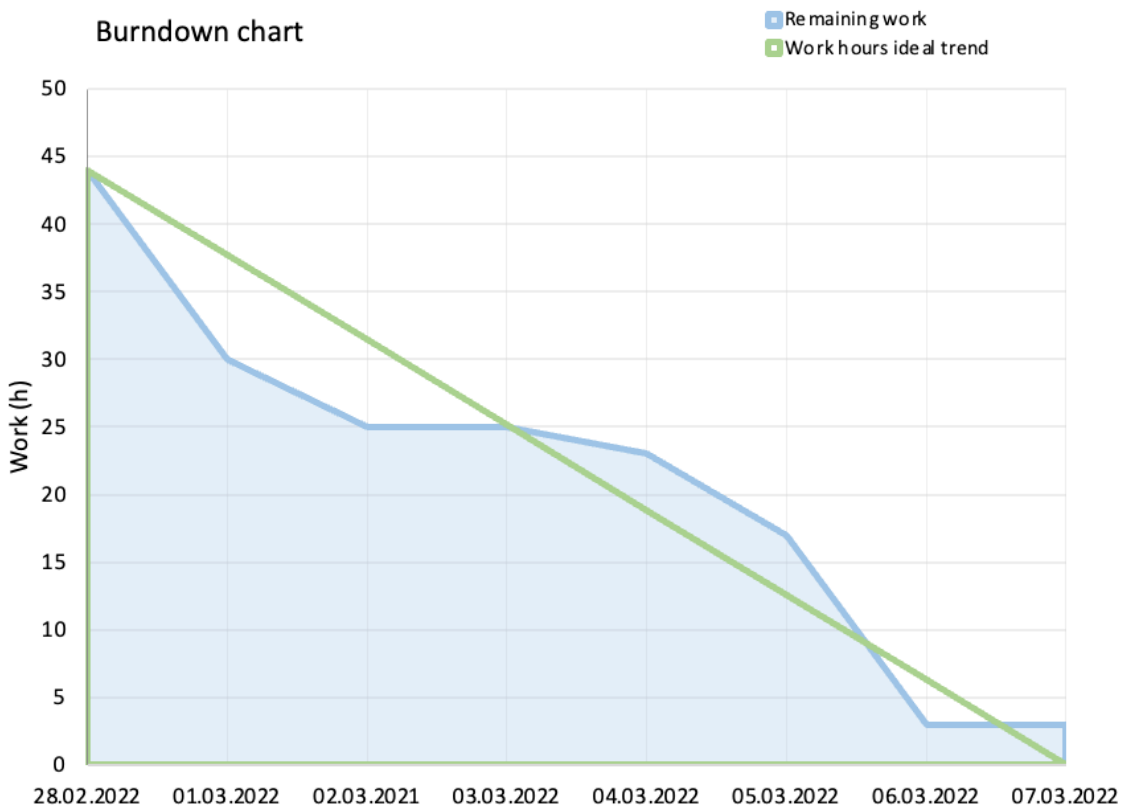
V tomto šprinte sme nespĺnili jednu úlohu, keďže na konci šprintu nespĺňala požiadavky. Šprint sme teda neukončili úspešne.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-62	Prototyp UI	21	Patrícia Huďanová	Done	Done
Subtask	TEAM-69	Implementácia zoznamu článkov		Martin Budinský	Done	Done
Subtask	TEAM-68	Implementácia detailu článku		Martina Billichová	Done	Done
Subtask	TEAM-67	Implementácia prototypu		Patrícia Huďanová	Done	Done
Subtask	TEAM-66	Update mockupov		Ján Huďan	Done	Done
Story	TEAM-63	Analýza backend frameworkov	6	Ivana Haberova	Done	Done
Subtask	TEAM-65	Analyzovanie frameworkov		Ivana Haberova	Done	Done
Subtask	TEAM-64	Vytvorenie dokumentu		Ivana Haberova	Done	Done

Nespĺnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-61	Prototyp využitia API	13	Ivana Baleková	In Progress	Unresolved
Subtask	TEAM-72	Inicializácia repozitára pre BE a FE		Miloš Galeta	Done	Done
Subtask	TEAM-71	Implementácia komunikácie s API serverom		Ivana Baleková	Done	Done
Subtask	TEAM-70	Vytvorenie modelu dát		Ján Huďan	In Progress	Unresolved



Šprint 7 - The one where we created base project (07.03.2022 – 21.03.2022)

V rámci šprintu sme sa venovali implementácii prototypu kde sme prepojili django backend nášho projektu v ktorom sa nachádza pripojenie na IEEE Api s našim frontendom v Reacte.

Náš prototyp po šprinte zvládal základný scenár - Spraviť search IEEE článkov, vrátiť používateľovi výsledky vyhľadávania v grafe (Články s vyparovanými kapitolami).

Vytvorili sme aj prototyp mind mapy z free textu, ktorý chceme neskôr zobrazovať používateľovi v podobe knowledge grafu.

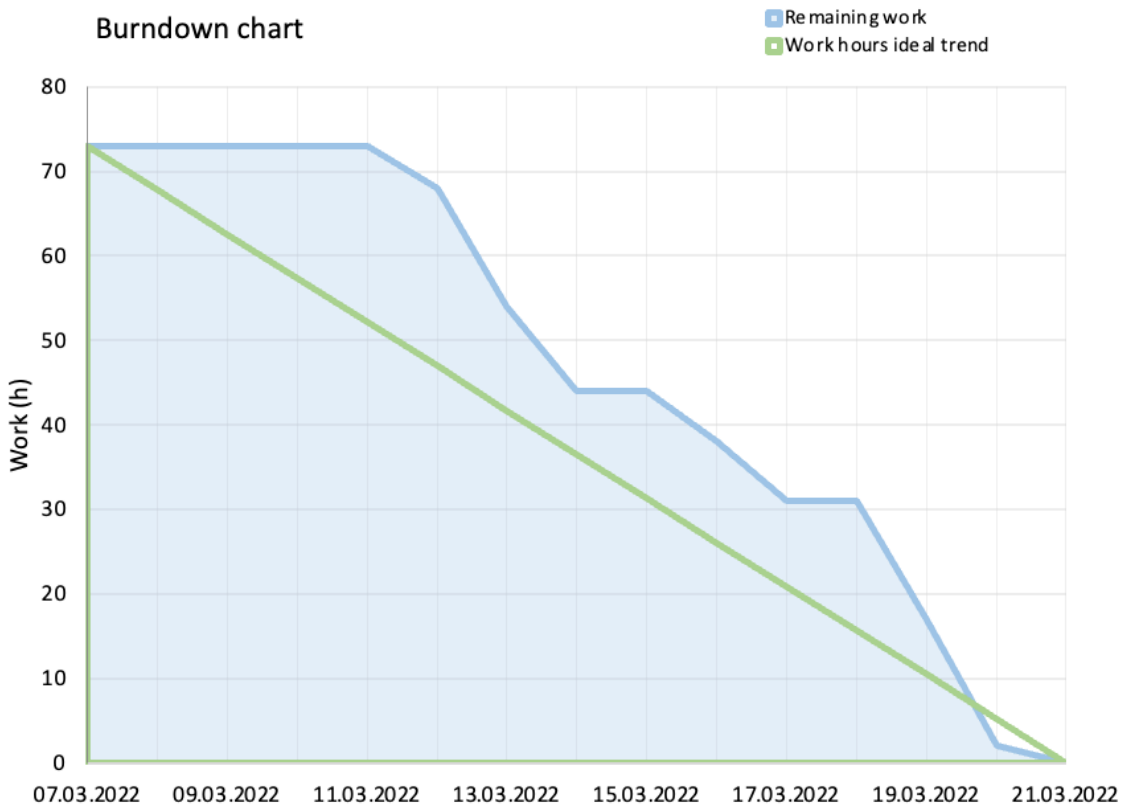
Takisto sme v tomto šprinte analyzovali rôzne možnosti vytvárania štruktúr z free textu a možnosti ako používateľovi zobrazí detail článku.

V šprinte sme dokončili všetky úlohy a úspešne ho ukončili.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-61	Prototyp využitia API	13	Ivana Baleková	Done	Done
Subtask	TEAM-72	Inicializácia repozitára pre BE a FE		Miloš Galeta	Done	Done
Subtask	TEAM-71	Implementácia komunikácie s API serverom		Ivana Baleková	Done	Done
Subtask	TEAM-70	Vytvorenie modelu dát		Ján Huďan	Done	Done
Story	TEAM-76	Analýza možností zobrazenia detailu článku v html	11	Ján Huďan	Done	Done
Subtask	TEAM-85	Implementácia prototypu		Ján Huďan	Done	Done
Subtask	TEAM-84	Vytvoriť dokument s analýzou		Ján Huďan	Done	Done
Story	TEAM-75	Prepojenie BE a FE - prototyp 1	20	Patrícia Huďanová	Done	Done
Subtask	TEAM-83	Prepojenie zoznamu článkov na FE s BE		Martin Budinský	Done	Done
Subtask	TEAM-82	Implementácia searchu na BE		Ivana Baleková	Done	Done
Subtask	TEAM-81	Implementácia searchu na FE		Patrícia Huďanová	Done	Done
Story	TEAM-74	Prototyp mind mapy z free textu	15	Ivana Haberova	Done	Done
Subtask	TEAM-80	Update mockupov		Patrícia Huďanová	Done	Done

Subtask	TEAM-79	Implementácia prototypu mind mapy z free textu		Ivana Haberova	Done	Done
Story	TEAM-73	Analýza možností vytvárania štruktúr z free textu	13	Martina Billichová	Done	Done
Subtask	TEAM-78	Analyzovanie možností vytvárania štruktúr z free textu		Miloš Galeta	Done	Done
Subtask	TEAM-77	Vytvorenie dokumentu		Martina Billichová	Done	Done



Šprint 8 - The one where server fell (21.03.2022 – 04.04.2022)

V rámci ôsmeho šprintu sme prerobili graf s výsledkami vyhľadávania pre lepšiu prehľadnosť do radiálnej formy.

Ďalej sme analyzovali rôzne možnosti zobrazenia detailu článku v pdf/xml a navigáciu v rámci neho.

Takisto sme sa venovali analýze nástrojov na extrakciu semantic triplets z textov, našli sme niekoľko možných knižníc, avšak extrakcia pre zložité texty ako sú vedecké články nefungovala ani v jednom úplne dobre.

Okrem toho sme vytvorili na základe dát zo semantic triplets knowledge graf.

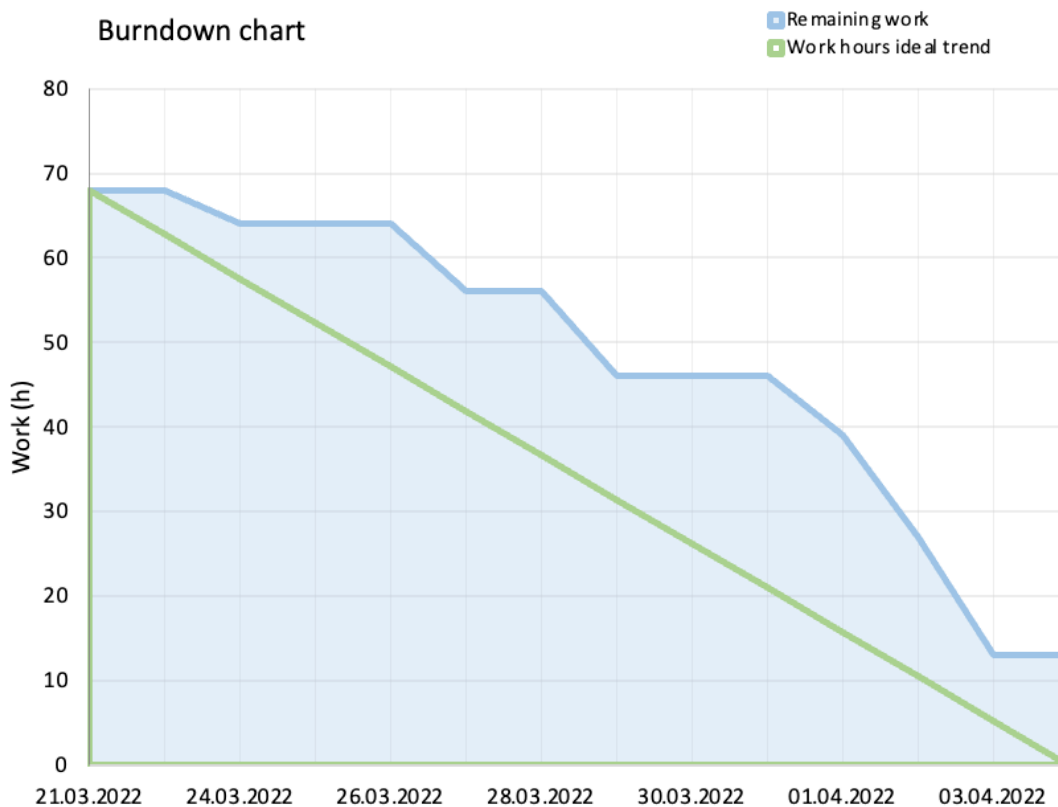
Takisto sme sa pokúsili nasadiť našu aplikáciu na virtuálny server, ale túto úlohu sa nám nepodarilo dokončiť.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-89	Vytvorenie knowledge grafu na základe semantic triplets	5	Ivana Haberova	Done	Done
Subtask	TEAM-97	Prototyp so zobrazením tripletov v knowledge grafe		Ivana Haberova	Done	Done
Story	TEAM-88	Analýza toolov na semantic triplets	21	Ivana Haberova	Done	Done
Subtask	TEAM-96	Otestovanie jednotlivých toolov		Martin Budinský	Done	Done
Subtask	TEAM-95	Analýza toolov		Ivana Haberova	Done	Done
Story	TEAM-87	Analýza možností zobrazenia detailu článku v pdf/xml a navigácia v rámci neho	21	Ivana Baleková	Done	Done
Subtask	TEAM-94	Analýza zobrazenia cez pdf		Miloš Galeta	Done	Done
Subtask	TEAM-93	Analýza zobrazenia s pomocou XML		Ivana Baleková	Done	Done
Story	TEAM-86	Zefektívnenie vizualizácie	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-92	Doplnenie API na BE		Ivana Baleková	Done	Done
Subtask	TEAM-91	Prerobenie stromu na radiálny graf		Patrícia Huďanová	Done	Done

Nesplnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-90	Nasadenie aplikácie	13	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-99	Nasadenie aplikácie na virtuálku		Ján Huďan	In Progress	Unresolved
Subtask	TEAM-98	Úprava dockera		Ján Huďan	In Progress	Unresolved



Šprint 9 - The one with final analysis (04.04.2022 – 25.04.2022)

Spravili sme prototypy pre zobrazovanie detailu článku v rôznych formátoch a upravili sme samotnú vizualizáciu - pridali sme k node-om v radiálnom grafe tlačidlo ktoré otvorí popup s knowledge grafom pre daný článok a detail článku sme zmenili z popup formy na enddrawer.

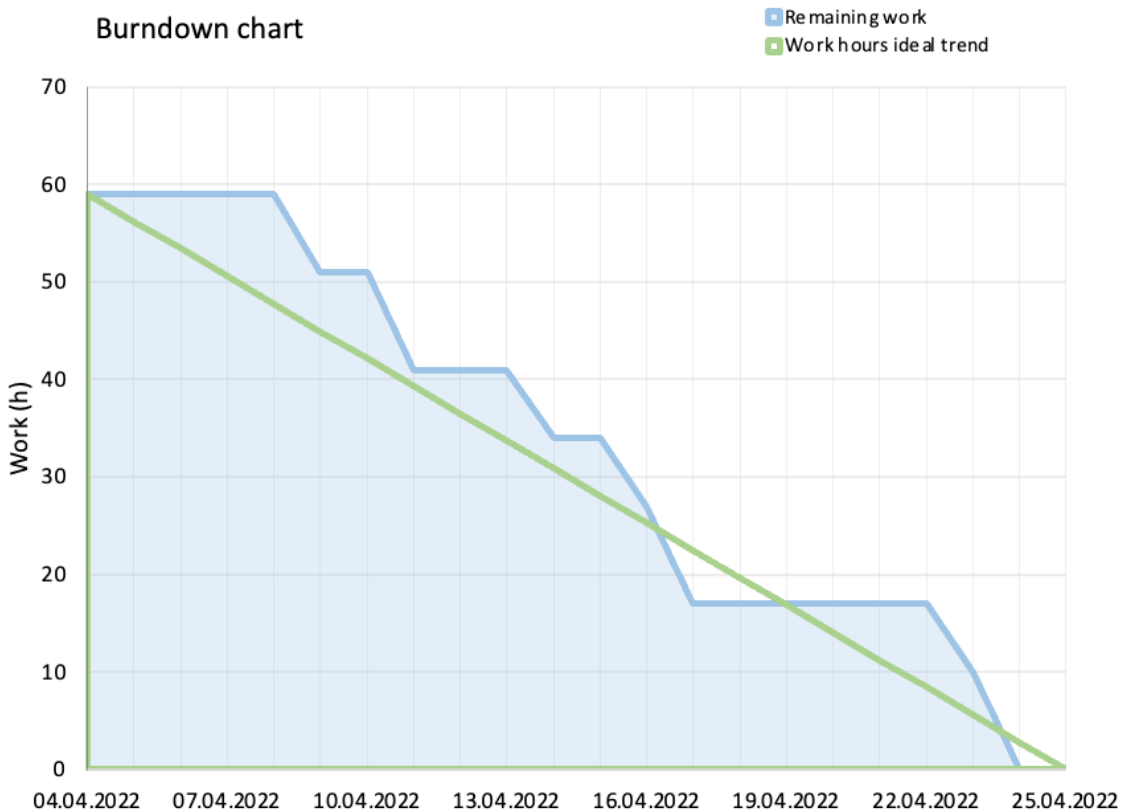
Otestovali sme rôzne knižnice pre tvorbu semantic tripletov z textu. Nakoniec sme sa rozhodli pre implementáciu použiť knižnicu spaCy kde si sami definujeme pravidlá pre extrahovanie tripletov.

Okrem toho sme náš projekt úspešne nasadili na server.

V Šprinte sme dokončili všetky úlohy a úspešne ho ukončili.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-100	Prototypy s rôznymi možnosťami zobrazenia detailu článku	17	Miloš Galeta	Done	Done
Subtask	TEAM-110	Implementovanie prototypu pre html		Ivana Baleková	Done	Done
Subtask	TEAM-109	Implementovanie prototypu pre xml		Ivana Baleková	Done	Done
Subtask	TEAM-108	Implementovanie prototypu pre pdf		Miloš Galeta	Done	Done
Story	TEAM-101	Zefektívnenie vizualizácie	13	Martina Billichová	Done	Done
Subtask	TEAM-111	Prerobenie detailu článku		Martina Billichová	Done	Done
Subtask	TEAM-107	Doplnenie grafu ako listu do radialneho stromu		Patrícia Huďanová	Done	Done
Subtask	TEAM-106	Vytvorenie dokumentu s návrhom		Ivana Haberova	Done	Done
Story	TEAM-102	Testovanie rôznych knižníc na semantic triplets	13	Martin Budinský	Done	Done
Subtask	TEAM-105	Implementovanie prototypu		Martin Budinský	Done	Done
Story	TEAM-90	Nasadenie aplikácie	13	Ján Huďan	Done	Done
Subtask	TEAM-99	Nasadenie aplikácie na virtuálku		Ján Huďan	Done	Done
Subtask	TEAM-98	Úprava dockera		Ján Huďan	Done	Done



Šprint 10 - The one where we finished (25.04.2022 – 09.05.2022)

V tomto šprinte sme spojili všetku funkcionálnosť. Pridali sme knowledge graf a zoznam obrázkov, pridali sme prepojenia na detail článku a pod.

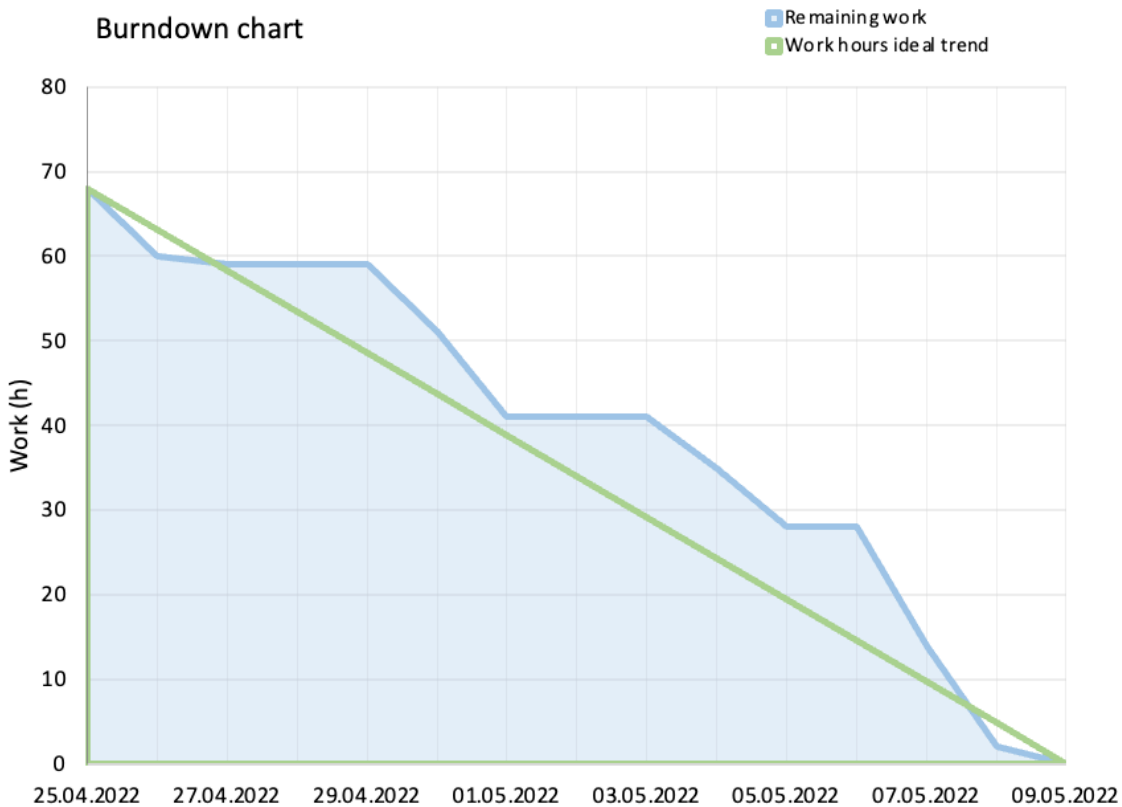
Vytvoril sa systém prihlasovania cez stuba účet a celková autorizácia. Nakoniec sme vytvorili niekoľko testov pre naše komponenty a napísali celkový testovací scenár pre druhý tím, ktorý testuje náš produkt.

V šprinte sme dokončili všetky úlohy a úspešne ho ukončili.

Splnené stories:

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-115	Analýza, návrh a implementácia filtrovania tripletov	20	Martin Budinský	Done	Done
Subtask	TEAM-125	Implementovanie do prototypu		Martin Budinský	Done	Done

Subtask	TEAM-124	Analýza a návrh filtrovania tripletov		Martin Budinský	Done	Done
Story	TEAM-113	Pridať prihlasovanie	13	Miloš Galeta	Done	Done
Subtask	TEAM-123	Autorizácia na BE		Miloš Galeta	Done	Done
Subtask	TEAM-122	Autorizácia na FE		Martina Billichová	Done	Done
Subtask	TEAM-121	Overovanie Ldap		Ivana Baleková	Done	Done
Story	TEAM-112	Pridať taby do popupu	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-120	Prepojenie s BE		Miloš Galeta	Done	Done
Subtask	TEAM-119	Vytvorenie tabu s wordcloudom		Patrícia Huďanová	Done	Done
Subtask	TEAM-118	Vytvorenie tabu s obrázkami		Patrícia Huďanová	Done	Done
Story	TEAM-103	Scenár testovania pre druhý tím + Automatické testy	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-117	Vytvorenie scenárov pre testovanie		Ivana Haberova	Done	Done
Subtask	TEAM-116	Automatické testy		Ján Huďan	Done	Done
Story	TEAM-114	Integrácia prototypu html rest	10	Ivana Baleková	Done	Done



Retrospektíva

Na konci každého šprintu sme sa venovali retrospektíve. Každý člen tímu sa mohol vyjadriť, čo sa mu na danom šprinte páčilo, čo nie, s čím by sme mali pokračovať a podobne.

Zimný semester

Retrospektíva pre šprint 1 - The one where it all began (07.10. 2021 – 18. 10. 2021)

Sprint Review

V rámci šprintu sa vytvorila webová stránka, komunikačné kanály sme prepojili s Jirou, vytvárali sme si šablóny pre jednotlivé dokumenty a prvý krát sme sa stretli so zadávateľmi projektu.

S čím pokračovať ?

- Plnenie taskov načas
- Dobré rozdelenie taskov

S čím začať ?

- Viac komunikácie s bigring – stále nevieme čo ideme robiť = možnosť vytvoriť si s nimi napr. Slack

S čím prestať ?

- Slack – general = prestať používať na všetku komunikáciu (iba hlavné veci)

- Prestať pracovať samostatne

Čo nám prekážalo

- Jira - veľa komplikácií s burndown chartom, nezobraziteľné subtasky, atď.
- Nie je jasné zadanie

Retrospektíva pre Šprint 2 - The one where we had call during holiday (18. 10. 2021 – 1. 11. 2021)

Sprint Review

Vytvorili sme funkčný sample v platforme bigring a preštudovali sme si dokumentáciu.

S čím pokračovať ?

- Spoločné prezenčné stretávania

S čím začať ?

- Story by sme mali rovno posúvať do In progress keď sa začne subtask v rámci storky.
- Stanoviť presný čas nášho spoločného stretnutia.

S čím prestať ?

- Meniť časy stretnutia každý týždeň.

Čo nám prekážalo

- Neaktivita zo strany bigringu.
- Nejasné zadanie z pohľadu developera.

Retrospektíva pre Šprint 3 - The one with all the documentation (01. 11. 2021 – 15. 11. 2021)

Sprint Review

V rámci Šprintu sa nám kvôli vyťažnosti externej firmy nepodarilo zorganizovať stretnutie s BigRingom. V rámci Šprintu sme vytvorili ďalší model v platforme bigring a zanalyzovali sme podklady od Lambda Life. Nebol nám BigRingom pridelený prístup k dokumentácii, takže túto story sa nám podarilo spraviť iba čiastočne – spravili sme dokument s flowom – ako vytvoriť funkčný model v platforme.

S čím pokračovať ?

- Spoločná práca pri tej analýze na intráku bola super - pokračovať v práci spoločne prezenčne

S čím začať ?

- Písanie konkrétnych otázok k modelom z minulého roku a analyzovať ich
- Rozbíjanie story na viac menších taskov
- Efektívnejšie rozdeľovať tasky pri spoločnej práci

S čím prestať ?

- Pre jednu story písať jeden task - vždy rozdeliť na menšie

Čo nám prekážalo

- Neaktivita zo strany firmy BigRing.
- Stále nie je jasné zadanie z pohľadu developera.

Retrospektíva pre Sprint 4 - The one where we tried NLP (15. 11. 2021 – 29. 11. 2021)

Sprint Review

V rámci tohto šprintu sme sa zamerali na analýzu rôznych NLP nástrojov pre rôzne jazyky a vytvorili sme dokument s popisom jednotlivých nástrojov. Ďalej sme sa zamerali na vytvorenie prototypu na spracovanie štúdií/ vedeckých článkov. V rámci tohto problému sme si vybrali tri nástroje z predošlej analýzy a otestovali sme ich nad krátkym textom z domény onkológie. Nepodarilo sa nám spojiť s BigRingom a uskutočniť stretnutie s nimi, a nezískali sme ani prístup do gitbook dokumentácie.

Vzhľadom na minulú retrospektívu:

Podarilo sa nám úspešne rozdeľovať tasky na viac menších issues.

V rámci prezenčných stretávaní sa nám osvedčilo pracovať súčasne avšak online sa nám táto forma nezdá dobrá. Takže navrhujeme rozdeliť si tasky a na spoločnom stretnutí si prejsť spravené tasky, resp. čo sa urobilo.

S čím pokračovať ?

- So všetkým čo robíme doteraz.

S čím začať ?

- Pozrieť sa v rámci retrospektívy vždy na predošlú retrospektívu a vyhodnocovať čo sa podarilo zmeniť vzhľadom na predošlú retrospektívu

Retrospektíva pre Sprint 5 - The one where we created our own path (29. 11. 2021 – 13. 12. 2021)

Sprint Review

V rámci šprintu sme spravili analýzu rôznych komerčných a vedeckých nástrojov pre NLP. Navrhli a špecifikovali sme si zadanie. Ako cieľ sme si dali vytvoriť nástroj pre zjednodušenie a sprehľadnenie vyhľadávania nad vedeckými článkami. Zanalyzovali sme rôzne API prístupy na získanie vedeckých článkov. Následne sme spravili prototyp re vizualizáciu vzťahov vo vete.

S čím pokračovať ?

- Pokračovať v rozdelení práce pre každého

Letný semester

Retrospektíva pre Sprint 6 - The one with new project (28. 02. 2022 – 07. 03. 2022)

Sprint Review

V rámci Šprintu sme si zadefinovali čo a v akej forme potrebujeme na vizualizáciu (ako by mal vyzeráť model). Implementovali sme komunikáciu s Api serverom od IEEE a zistili aké rôzne možnosti nám táto api ponúka. Takisto sme vytvorili repositár pre náš finálny produkt, kde sme po analýze rôznych frameworkov pre BE zvolili django a pre FE react. Do tohto repositára sme následne vytvorili základnú kostru stránky - formu pre vyhľadávanie, graf v ktorom sa majú zobraziť články, zoznam článkov a zobrazenie popupu s detailom článku.

S čím pokračovať ?

- Pokračovať v rozdelení práce pre každého

S čím začať ?

- v jire vždy dávať linky na výsledok

Retrospektíva pre Šprint 7 - The one where we created base project (07. 03. 2022 – 21. 03. 2022)

Sprint Review

V rámci Šprintu sme implementovali prototyp kde sme prepojili django backend nášho projektu v ktorom sa nachádza pripojenie na IEEE Api s našim frontendom v Reacte. Náš prototyp zvládol základný scenár - Spraviť search, vrátiť používateľovi výsledky vyhľadávania v grafe (články s vyparovanými kapitolami).

Okrem toho sme vytvorili prototyp mind mapy z free textu, ktorý chceme neskôr zobrazovať používateľovi v podobe knowledge grafu.

Takisto sme v tomto Šprinte analyzovali rôzne možnosti vytvárania štruktúr z free textu a možnosti ako používateľovi zobraziť detail článku v html.

Vzhľadom na minulú retrospektívu :

V rámci Šprintu sme viac dbali na to aby mal každý task v komentári link na výsledok.

S čím začať ?

- S testovaním FE/BE

S čím prestať ?

- Robiť na tasku v posledný možný deň

Retrospektíva pre Šprint 8 - The one where server fell (21. 03. 2022 – 04. 04. 2022)

Sprint Review

V rámci Šprintu sme prerobili graf s výsledkami vyhľadávania pre lepšiu prehľadnosť do radiálnej formy. Ďalej sme analyzovali rôzne možnosti zobrazenia detailu článku v pdf/xml a navigáciu v rámci neho. Takisto sme sa venovali analýze toolov na extrakciu semantic triplets z textov, našli sme niekoľko možných knižníc, avšak extrakcia pre zložité texty ako sú vedecké články nefungovala ani v jednom úplne dobre.

Okrem toho sme vytvorili na základe dát zo semantic triplets knowledge graf. Pokúsili sme sa nasadiť našu aplikáciu na virtuálny server, ale zatiaľ neúspešne.

S čím pokračovať ?

- pokračovať s obviňovaním Petrika :D

S čím začať ?

- testovanie
- písať metodiky

S čím prestať ?

- strácať prístup na tímový server (za všetko môže Petrík)
- natvrdo definované urlky v kóde

Retrospektíva pre Šprint 9 - The one with final analysis (04. 04. 2022 – 25. 04. 2022)

Sprint Review

V rámci Šprintu sme spravili prototypy pre zobrazovanie detailu článku v rôznych formátoch. Takisto sme upravili samotnú vizualizáciu - pridali sme k node-om v radiálnom grafe tlačidlo ktoré otvorí popup s knowledge grafom pre daný článok a detail článku sme zmenili z popup formy na enddrawer.

Otestovali sme rôzne knižnice na semantic triplets, z ktorých sme sa rozhodli pre implementáciu použiť spacy kde si sami nadefinujeme pravidlá pre extrahovanie tripletov. Okrem toho sme náš projekt nasadili na server.

S čím začať ?

- testovanie
- písať metodiky

Motivačný dokument

Kto sme

Kontakt: teamproject14@googlegroups.com

Členovia: **Ivana Baleková** (ISS), **Martina Billichová** (ISS), **Martin Budinský** (ISS), **Ivana Háberová** (ISS), **Ján Hudčan** (ISS), **Patrícia Hudčanová** (ISS), **Miloš Galeta** (ISS)

O nás

V rámci tímu sa všetci poznáme nielen ako spolužiaci a kolegovia, ale aj ako kamaráti, ktorí spolu trávia svoj voľný čas. Podľa nás je to veľká výhoda, pretože **sme zvyknutí spolu komunikovať a pomáhať si**. Vďaka vzťahom, ktoré medzi sebou máme vybudované je naša skupina súdržná a našou silnou stránkou je práve komunikácia. Už v minulosti sme pracovali na školských zadaniach a projektoch, kde sme spoločne riešili problémy a diskutovali o postupoch, ktoré použijeme. Naša spolupráca sa ukázala ako funkčná a efektívna. A samozrejme náš tím nezaostáva ani v soft skilloch.

Aj keď sú všetci z našej skupiny zapísaní v odbore ISS, každý z nás má **jedinečné skúsenosti s rôznymi technológiami**. Mnohí majú popri škole stáž/prácu, v ktorej získali ďalšie skúsenosti a spoznali prácu na reálnych projektoch, v skutočných tímoch.

Ako tím disponujeme širokou škálou vedomostí a zručností od vývoja frontendu (JavaScript -React, Angular, Vue.js, jquery; HTML; CSS; Bootstrap)/backendu (Python/Django; REST framework/Flask-Restful; Java/Spring; PHP/Laravel; NodeJS/ExpressJS) a práce s Sql/NoSQL databázami, cez tvorbu a deployment (Linux, Docker, AWS) hybridných (Flutter) a natívnych aplikácií (Kotlin) až po data science(Numpy, Pandas, Scikit, Seaborn,..). Väčšina členov tímu má zo štúdia, najmä z bakalárskych prác, nadšenie a skúsenosti s data science a Machine learningom(Pytorch).

Zoradenie tém podľa priority

1. téma 3 - DataHub pre rôzne typy zariadení, ich spracovanie / analýzu / vizualizáciu
2. téma 4 - Adverse Media Screening
3. téma 20 - Aplikačné riešenie pre elektronický volebný systém
4. téma 5 - Vytvorenie inteligentného model-based agenta (umelá inteligencia na báze Knowledge grafov) pre tvorbu komplexných dátových štruktúr a vzťahov pre aplikovaný výskum v klinickej onkológii
5. téma 16 - FIIT WIX
6. téma 19 - Automatizácia procesov KYC (Know your client) a AML (Anti-money laundering)
7. téma 11 - (Q)SAR analýza fototoxických látok
8. téma 17 - Document Wizard
9. téma 9 - Monitorovanie a správa systému pre výrobný areál [LOMON]
10. téma 8 - Educational Content Engineering Hub - Databáza otázok, odpovedí, úloh a riešení [ECEH-DU]
11. téma 14 - IoT platforma na priemyselnú automatizáciu - malý pivovar
12. téma 12 - Spektrometrické rozpoznávanie túh do pera

Rozvrh tímu

	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	20:00 - 20:50
Pondelok			[Cvičenie] VINF_I - Vyhľadavanie informácií: Martina Billichová, Ivana Baleková			Alternatíva 1 - Stretnutie s vedúcim Alternatíva 2 - Stretnutie tímu Alternatíva 3 - Stretnutie tímu Alternatíva 4 - Stretnutie tímu			[Cvičenie] AS_I - Architektúra softvéru: Patrícia Huďanová, Martina Billichová, Ivana Háberová, Ján Huďan		[Cvičenie] AS_I - Architektúra softvéru: Miloš Galeta		
Utorok		[Prednáška] SIPVS_I - Spracovanie informácií v podnikaní a verejnej správe: Ján Huďan		[Cvičenie] VINF_I - Vyhľadavanie informácií: Patrícia Huďanová, Ivana Háberová, Ján Huďan, Martin Budinský			[Prednáška] AS_I - Architektúra softvéru: Celý tím		[Prednáška] VISS_I - Výskum inteligentných softvérových systémov: Celý tím	[Prednáška] TP1_I - Tímový projekt: Celý tím		[Cvičenie] SIPVS_I - Spracovanie informácií v podnikaní a verejnej správe: Ján Huďan	
Streda		[Prednáška] KPAIS_I - Kvalita programových a informačných systémov: Ján Huďan, Miloš Galeta		[Prednáška] VINF_I - Vyhľadavanie informácií: Patrícia Huďanová, Martina Billichová, Ivana Háberová, Ivana Baleková, Ján Huďan, Martin Budinský		[Cvičenie] KPAIS_I - Kvalita programových a informačných systémov: Ján Huďan		[Prednáška] MTS_I - Manažment v tvorbe softvéru: Celý tím		[Cvičenie] MTS_I - Manažment v tvorbe softvéru: Martina Billichová, Ivana Háberová, Ivana Baleková, Martin Budinský	[Cvičenie] MTS_I - Manažment v tvorbe softvéru: Patrícia Huďanová, Ján Huďan, Miloš Galeta		
Štvrtok		[Cvičenie] KPAIS_I - Kvalita programových a informačných systémov: Miloš Galeta			[Prednáška] NMVS_I - Nové médiá v spoločnosti: Martin Budinský, Miloš Galeta		[Prednáška] NMVS_I - Nové médiá v spoločnosti: Martin Budinský, Miloš Galeta		[Cvičenie] AS_I - Architektúra softvéru: Ivana Baleková, Martin Budinský			Alternatíva 2 - Stretnutie s vedúcim	Alternatíva 1 - Stretnutie tímu
Piatok		[Prednáška] PDT_I - Pokročilé databázové technológie: Martina Billichová, Patrícia Huďanová, Ivana Háberová, Ivana Baleková											Alternatíva 4 - Stretnutie s vedúcim

Metodiky

Metodika komunikácie

Metodika komunikácie definuje akým spôsobom jednotlivé strany medzi sebou komunikujú. Jedná sa o strany tím, vedúci a firma Bigring a firma Lambda Life. Jednotlivé strany si medzi sebou dohodli nasledovne komunikačné kanály

Komunikačné kanály

- Facebook messenger

Prostredníctvom tohto komunikačného kanála komunikujú členovia tímu interne, medzi sebou. Dohadujú sa tu stretnutia tímu (bez účasti vedúceho), drobné, nedôležité veci ale i voľnočasové posedenia tímu.

- Slack

Komunikačný kanál Slack slúži na oficiálnu komunikáciu tímu medzi sebou ale i s vedúcim. Slack tímu je rozdelený na viacero kanálov:

- general
 - tu sa riešia veci dôležité pre projekt
- kontakty
 - tu sú sústreďované kontakty na členov tímu ale i vedúceho a firmy
- linky
 - obsahuje odkazy na dokumenty a dôležité weby
- notes
 - tu medzi sebou zdieľame poznámky napr. zo stretnutí
- random
 - aktuálne, nezaraditeľné veci
- zápisnice
 - tu sa riešia zápisnice stretnutí
- webik
 - rieši webové sídlo tímu
- Google Meet

Prostredníctvom tejto platformy sa tím v prípade potreby stretáva offline. Je vytvorený trvalý link pre stretnutia s vedúcim.

- Jitsi Meet

Z prostredia Slack je za pomoci príkazu /jitsi priamo otvorený hovor na tejto platforme. Využívaný pre rýchle stretnutia v menších skupinách.

- E-mail - teamproject14@googlegroups.com

Prostredníctvom emailu tím komunikuje oficiálne veci mimo vyššie spomenutých prípadov.

Metodika dokumentácie

Riešenie projektu plánujeme priebežne dokumentovať, pričom budeme používať nižšie opísanú metodiku. Zásady a pravidlá určujú štýl písania dokumentácie a sú záväzné pre všetkých členov tímu.

Pravidlá pre písanie dokumentácie:

- Pre hlavné kapitoly používať Nadpis 1
- Pre podkapitoly používať Nadpis 2, Nadpis 3, Nadpis 4 v závislosti od hierarchie dokumentu
- V celej dokumentácii používať jeden font pre text – Calibri
- V celej dokumentácii používať jednu veľkosť pre text – 11
- Písať gramaticky a štylisticky správne.

Zverejňovanie dokumentácie

Všetci členovia majú právo prispievať do dokumentácie, čo je zabezpečené jej zdieľaním ako online Word dokument, ktorého link je dostupný v komunikačných kanáloch. Priebežný stav dokumentácie je dostupný aj na stránke tímu, ktorý je aktualizovaný v určitých časových intervaloch. Na stránke sú k dispozícii aj zápisnice zo stretnutí tímu a retrospektív, ktoré sú zverejnené po každom stretnutí.

Revízie dokumentácie

Všetci členovia tímu prispievajú prípadne sa oboznamujú a kontrolujú aktuálnu dokumentáciu v pravidelných intervaloch. V prípade nájdenie gramatických alebo štylistických chýb členovia tímu tieto chyby opravujú. V prípade potenciálnych obsahových nezrovnalostí sú tieto pripomienky prediskutované všetkými členmi tímu, ktorí rozhodnú ako a do akej miery sa daná časť upraví.

Šablóny opakujúcich sa dokumentov

V nasledujúcej časti sú šablóny dokumentov, ktoré vytvárame pravidelne a musia dodržať určitú štruktúru a predpis.

Šablóna zápisnice

OncoAgents - Tím 14

Vypracoval(a): xxx

Zápisnica č.X

Názov Šprintu: xxx

Trvanie Šprintu: dd. mm. YYYY – dd. mm. YYYY (Časové rozmedzie Šprintu zapísané v uvedenom formáte)

Prítomní

zoznam členov tímu, ktorí sa zúčastnili stretnutia

Nepřítomní

zoznam členov tímu, ktorí sa nezúčastnili stretnutia

Obsah stretnutia:

popis obsahu stretnutia

Úlohy:

stav úloh, doplňujúce informácie a poznámky k úlohám

Poznámky k stretnutiu:

prípadné poznámky o stretnutí

Šablóna retrospektívy

OncoAgents - Tím 14

Vypracoval(a): xxx

Retrospektíva Č.X

Názov Šprintu: xxx

Trvanie Šprintu: dd. mm. YYYY – dd. mm. YYYY (časové rozmedzie Šprintu zapísané v uvedenom formáte)

Sprint review

Zhrnutie Šprintu.

Zoznam pripomienok z predošlých retrospektív, ktoré sme zapracovali v aktuálnom Šprinte

S čím pokračovať?

Zoznam pozitívnych vecí, v ktorých chceme pokračovať.

S čím začať?

Zoznam vecí, ktoré ešte nerobíme a chceme s nimi začať.

S čím prestať?

Zoznam negatívnych vecí, ktoré nechceme ďalej robiť.

Čo nám prekážalo?

Zoznam negatívnych vecí, ktoré členom tímu vyslovene vadili.

Metodika pre správu úloh

Úlohy manažujeme v rámci Jira Software-u, kde ako jednu úlohu v našom projekte chápeme Story. Story môže obsahovať viacero subtaskov. Jiru sme takisto integrovali so Slackom aby bol každý člen tímu informovaný o zmenách.

Story

Story má v Jire 4 stavy:

- To Do - novo vytvorená story, ktorá čaká na začatie riešenia
- In Progress - story je v stave riešenia a pracuje sa na nej
- Review and testing - práca na story je ukončená, story musí prejsť kontrolou
- Done - story je uzavretá
- Accepted - story je úspešne ukončená a schválená zadávateľom

Ako členovia tímu môžeme posunúť story iba po štvrtý stav, teda do "Done". Do tohto stavu sa story posunie v prípade, že sú všetky subtasky v rámci story v stave "Done". Story môže ukončiť a posunúť do stavu "Accepted" iba product owner.

Pri vytváraní story si vždy spíšeme DOD (definition of done) a story sa ukončí len ak boli všetky podmienky z DOD splnené.

Subtasky

Subtask má v Jire 4 stavy:

- To Do - novo vytvorený subtask, ktorý čaká na začatie riešenia
- In Progress - subtask je v stave riešenia a pracuje sa na ňom
- Review and testing - práca na subtasku je ukončená, subtask musí prejsť kontrolou
- Done - subtask je uzavretý
- Accepted - subtask je úspešne ukončený a schválený zadávateľom

Subtasky v jire má na starosti človek, ktorý k nim bol pridelený. Pracovať na subtasku, ale môže viacero členov tímu.

Každému subtasku na začiatku pridáme počet hodín, ktoré predpokladáme, že prácou na ňom strávime. Člen tímu, ktorý má story pridelenú musí vždy v Jire zaznamenať, keď sa v rámci subtasku niečo spraví.

Vytváranie Story

Všetky identifikované stories pridáva do backlogu product owner, ktorý im vytvára DOD (definition of done). Pri vytváraní story sa k nej vždy pridá krátky popis, prípadne link alebo odkaz na súvisiace dokumenty.

Pri plánovaní Šprintu vyberáme stories z backlogu a pomocou aplikácie PlanITpoker estimujeme story pointy. Následne podľa potreby story rozdelíme na subtasky.

Vytváranie Subtasku

Subtasky vytvárajú po diskusii spoločne členovia tímu. Každú story sa snažíme rozdeliť do niekoľkých menších subtaskov. Pri vytváraní subtasku mu pridáme počet hodín, ktoré predpokladáme, že prácou na ňom strávime a vždy ho priradíme jednému človeku, ktorý má následne na starosti jeho splnenie. Pracovať na subtasku, ale môže viacero členov tímu.

Metodika verziovania

Metodika verziovania obsahuje základné informácie a pokyny pre prácu s verziovacím systémom. Ako verziovací systém budeme používať Git. Členovia tímu môžu využívať buď webové rozhranie (GitHub), desktopovú aplikáciu GitHub Desktop alebo terminál v kombinácii s git príkazmi.

Repozitár

Tím má na svoju projektovú prácu vytvorený vlastný git repozitár, ktorý je privátny a prístup k nemu majú iba členovia tímu.

Základné pojmy

- Branch - samostatná verzia projektu, ktorá predstavuje líniu vývoja a zároveň umožňuje paralelnú prácu v tíme
- Pull - stiahnutie úprav (aktuálnej verzie) do lokálneho repozitára
- Push - pridanie zmien z lokálneho repozitára do vzdialeného
- Merge - spája dve alebo viac vývojových úprav (vetiev) dokopy
- Commit - slúži na zachytenie zmien v súbore/súboroch
- Pull Request - žiadosť o merge dvoch vetiev

Branch

Pri práci na projekte rozlišujeme spravidla dva druhy vetiev: master a dev.

Vetva master je hlavnou vetvou repozitára, ktorá obsahuje aktuálnu a funkčnú verziu projektu. Do tejto vetvy nie je možné priamo commitovať, ale zmeny je možné aplikovať pomocou merge request z vetvy dev. Vedľajšou, vývojovou, vetvou je vetva dev, ktorá predstavuje rozpracovanú verziu projektu a na ktorej prebiehajú všetky zmeny a vývoj novej funkcionality. Táto vetva je po istých časových úsekoch pridávaná do vetvy master, pričom musí byť vopred dobre otestovaná a musí obsahovať stabilnú a plne funkčnú verziu.

Okrem spomenutých základných vetiev jednotliví členovia tímu vytvárajú vo vetve dev ďalšie menšie vetvy, ktoré spravidla zachytávajú prácu na taskoch.

Pomenovanie vetiev

Názov vetvy pozostáva z čísla a názvu konkrétnej úlohy podľa názvu tasku (nie story) v Jire.

Ukážka: TEAM-[číslo úlohy]_[názov úlohy]

Príklad: TEAM-32_vytvorenie_funkcneho_modelu

Spájanie vetiev

Pri spájaní (mergovaní) jednotlivých vetiev sa riadime nižšie uvedenými pravidlami:

- Člen tímu, ktorý ide vykonať merge vetiev sa najskôr uistí, či sa nachádza na príslušnej vetve
- Člen tímu overí správnosť a funkčnosť verzie danej vetvy

- Člen tímu vytvorí tzv. Pull Request v ktorom uvedie aké zmeny vykonal, ak je to potrebné tak postup ako ich je možné otestovať, prípadne informáciu, ktorý task v jire merge danej vetvy uzavrie
- pred samotným mergovaním sa musí vykonať revízia alebo testovanie iným členom tímu než ten čo vytváral Pull Request
- v prípade nájdania chyby je potrebné členovi tímu, ktorý na danej vetve pracoval, chybu nahlásiť a problém čo najskôr opraviť

Commit

Pri commitovaní zmien je dôležité písať efektívne správy, ktoré majú informačnú hodnotu. Každý člen tímu sa snaží o to, aby správa jeho commitu bola čo najstručnejšia a adekvátne k vykonaným zmenám.

Commit message

Pri písaní commit message sa riadíme nasledujúcimi pravidlami:

- dĺžka správy je maximálne 50 znakov (ak si commit vyžaduje viac ako 50 znakov, je nutné ho rozdeliť)
- commit message začína veľkým písmenom
- na konci commitu nepíšeme bodku
- správa je napísaná v angličtine

Metodika testovania

Metodika testovania definuje typy testov, spôsob ich písania a spúšťania. Testovanie softvéru a jeho častí je dôležitou súčasťou vývoja softvéru. Akákoľvek, hoci aj malá zmena môže vysoko ovplyvniť funkcionality komponentov, preto je potrebné, aby bola táto správnosť funkcionality komponentov vždy otestovaná. Každá nová funkcionality systému, ktorá bola implementovaná, musí prejsť testovaním.

UNIT testy

UNIT testy umožňujú testovanie individuálnych jednotiek kódu nezávisle od ostatných. Za testovateľnú jednotku považujeme samostatne testovateľnú časť programu, v našom prípade komponent. Pre vykonávanie testov v Reacte využívame knižnicu enzyme. V UNIT teste porovnávame vopred špecifikované hodnoty s výstupmi. Všetky testy sú uložené v adresári /src/tests. Pomenovanie jednotlivých testov musí dodržiavať menovaciú konvenciu *.test.js, pričom na mieste hviezdinky musí byť názov komponentu, pre ktorý je test určený. Testy je možné spustiť príkazom `npm run test`.

Nepretržité testovanie (Continuous testing)

Nepretržité testovanie je proces vývoja softvéru, pri ktorom sa aplikácie testujú nepretržite počas celého životného cyklu vývoja softvéru. Cieľom nepretržitého testovania je vyhodnotiť kvalitu softvéru naprieč celého životného cyklu vývoja softvéru, poskytnúť kritickú spätnú väzbu a umožniť kvalitnejšie a rýchlejšie dodávky softvéru. Nepretržité testovanie sa vykonáva cez CI/CD github

actions. To nám umožňuje celé testovanie automatizovať. Testy sa vykonávajú na githube po každom commite zmien na github a github nás o úspešnom/neúspešnom teste upozorní na stránke a mailom.

Inžinierske dielo - Zimný semester

Big picture

Nasledujúca časť dokumentu obsahuje dokumentáciu projektu v rámci predmetu Tímový projekt I. Dokumentácia obsahuje informácie o globálnych cieľoch nášho projektu, ako aj analýzy vyhotovené pre účely vytvorenia jednotlivých častí projektu.

Ciele projektu na zimný semester

Globálnym cieľom projektu pre zimný semester je oboznámiť sa s no-coding platformou BigRing, pomocou ktorej tak budeme vedieť v budúcnosti vytvoriť inteligentného model-based agenta (na báze knowledge grafov) pre tvorbu komplexných dátových štruktúr a vzťahov pre aplikovaný výskum v klinickej onkológii. Jedným z cieľov je preto analyzovať doménu biológie so zameraním na rôzne typy buniek (napr. T-lymfocyty, kmeňové bunky) a ich vplyv na rozličné ochorenia (rakovina, autoimúnne ochorenia). Okrem oboznámenia sa so systémom platformy BigRing bolo cieľom zimného semestra navrhnuť a vytvoriť komplexné modely, ktoré využívajú väčšinu poskytovaných entít a vzťahov. V druhej polovici semestra bolo naším primárnym cieľom analyzovať možnosti spracovania prirodzeného jazyka (NLP - Natural Language Processing) a existujúce nástroje používané na prácu s textom. Okrem iného sme sa zamerali aj na navrhnutie riešenia, opísanie funkčných vlastností spolu so scenárom použitia a navrhnutie používateľského rozhrania našej webovej aplikácie.

Celkový pohľad na vyhotovené časti projektu

V tejto časti dokumentácie je opísaný stav projektu ku dňu 17. 12. 2021, teda po ukončení prvých piatich šprintov. Opisujeme výstupy projektu, na ktorých sme počas týchto šprintov pracovali, a to konkrétne: dokumentácia k no-coding platforme BigRing, modelovanie, analýza medicínskej domény, analýza NLP, možnosti získavania odborných článkov a návrh riešenia.

Dokumentácia k no-coding platforme BigRing

V úvodnej časti projektu sme sa venovali predovšetkým platforme BigRing a učili sa s ňou pracovať. Nakoľko sa jedná o komplexný moderný no-coding softvér so širokou škálou rôznych funkcionalít, kládli sme dôraz na dôkladné štúdium dokumentácie a prácu so systémom BigRing. Výsledkom tejto časti bolo okrem oboznámenia sa so systémom aj vytvorenie podrobnej dokumentácie tejto platformy a flow - postup krokov ako sa dostať od vytvorenia tzv. Solution až k funkčnému modelu.

Prvý dokument obsahuje názvy a opis jednotlivých častí bigRing systému spolu so screenshotmi, kde sú vyznačené dôležité elementy. Druhý dokument opisuje postup vytvárania modelu krok po kroku. V dokumente sú dôkladne opísané všetky konkrétne kroky, ktoré treba vykonať na vytvorenie funkčného modelu. Okrem slovného opisu týchto krokov sú v dokumente tiež screenshoty so zvýraznenými dôležitými časťami. Tento dokument sa nachádza v prílohe A.

Tieto vytvorené dokumenty nám poskytujú potrebné informácie na prácu s platformou BigRing a vytváranie komplexných a funkčných modelov.

Modelovanie

V ďalšej časti sme počas šprintov pracovali na vytváraní komplexných modelov, ktoré využívajú väčšinu poskytovaných entít a vzťahov. Podarilo sa nám vytvoriť nasledujúce dva modely.

Model 1

Cieľ: Oboznámenie sa so systémom platformy BigRing

Úlohy: Vytvorenie komplexného modelu, ktorý využíva väčšinu poskytovaných entít a vzťahov. Preskúmanie dostupnej dokumentácie a záznam o nájdených/využitých funkciách platformy vo forme zoznamu.

Pri vytváraní zadania sme využili informácie a dokumentáciu tímu, ktorý pracoval s platformou spoločnosti BigRing v akademickom roku 2019/2020 a venoval sa téme zameranej na automatizáciu procesu verejného obstarávania.

Pôvodný opis: Nahratie podkladov pre vyhlásenie VO (Verejného obstarávania) V tomto modeli vystupuje ako hlavný používateľ verejný obstarávateľ, ktorý pre vytvorenie verejného obstarávania nahrá podklady, ktoré bližšie špecifikujú samotné verejné obstarávanie. Medzi tieto podklady patrí oznámenie o vyhlásení verejného obstarávania, oznámenie o vyhlásení užšej súťaže a oznámenie o kvalifikačnom systéme.

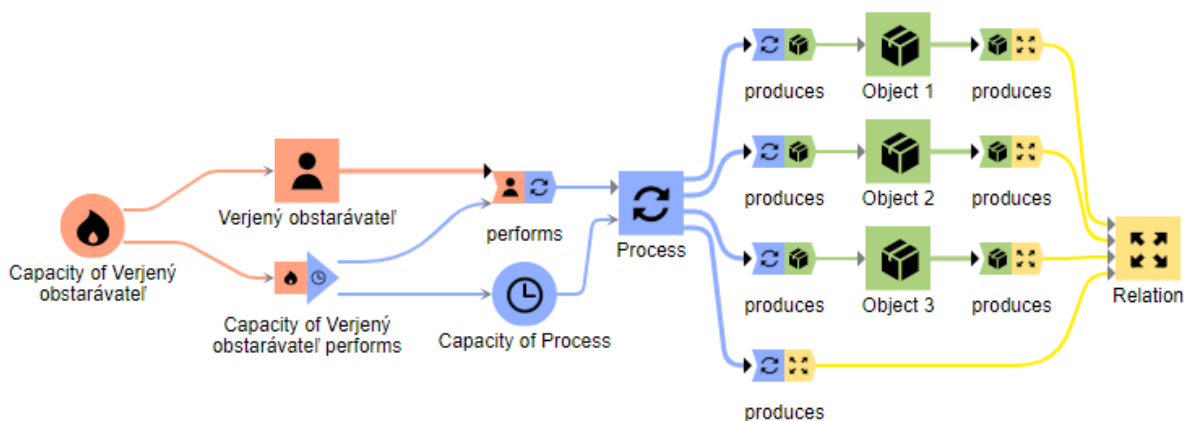
Používateľ (user): verejný obstarávateľ

Objekty (object): oznámenie o vyhlásení užšej súťaže, oznámenie o vyhlásení verejného obstarávania (VO), oznámenie o kvalifikačnom systéme

Procesy (process): nahratie podkladov pre vyhlásenie VO, Výber VO

Vzťahy (relation): verejné obstarávanie

Výsledný model:



Model 2

Cieľ: Vytvorenie druhého funkčného modelu v platforme od BigRing + vytvorenie flow-u

Úlohy: Vytvorenie komplexného modelu, ktorý využíva väčšinu poskytovaných entít a vzťahov. Vytvorenie dokumentu s postupom ako sa dostať v platforme od vytvorenia Solutionu až k funkčnému modelu.

Opis modelu: V modeli vystupuje ako hlavný subjekt pacient, ktorý si vyberie vyšetrenie prostredníctvom formulára. Z výberu vyšetrenia vzniká objekt so žiadosťou o vykonanie vyšetrenia, ktorý sa nakoniec vyhodnotí.

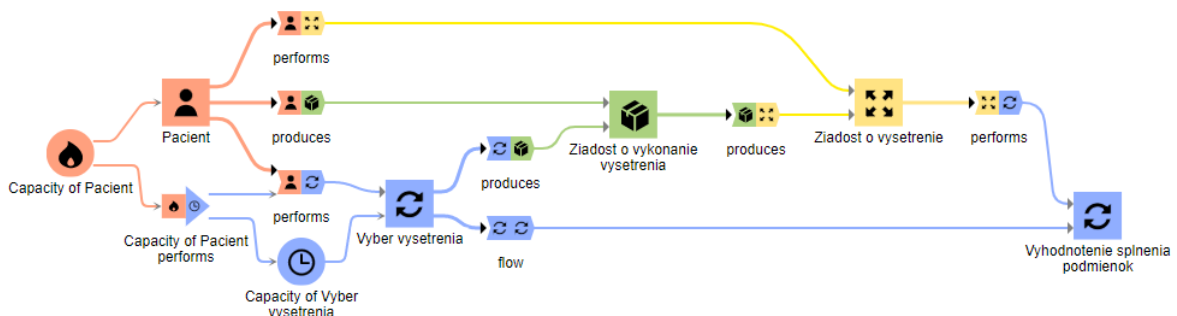
Používateľ (user): Pacient

Objekty (object): Žiadosť o vykonanie vyšetrenia

Procesy (process): Výber vyšetrenia, Vyhodnotenie splnenia podmienok

Vzťahy (relation): Žiadosť o vyšetrenie

Výsledný model:



Dokument s používateľskou príručkou, ktorá nás má odnavigovať od vytvorenia solutionu až po funkčný model, ktorá bola vytvorená počas modelovania sa nachádza v Prílohe A.

Analýza medicínskej domény

V neposlednom rade sme sa počas prvých troch šprintov venovali aj analýze medicínskej domény, nakoľko je s ňou téma nášho projektu úzko spätá. Analyzovali sme predovšetkým rôzne typy buniek a tiež procesy, ktoré prebiehajú na bunkovej úrovni. V rámci analýzy sme skúmali aj rôzne oblasti využitia počítačovej vedy v doméne biológie.

Výsledkom tejto časti bol rozsiahlejší dokument, ktorý obsahuje definíciu dôležitých pojmov z oblasti molekulárnej biológie, medicíny a bioinformatiky a tiež podrobnejší popis vybraných bunkových procesov. Tento dokument sa nachádza v Prílohe B.

Analýza NLP

NLP (Natural Language Processing) - Spracovanie prirodzeného jazyka je disciplína zaoberajúca sa extrakciou informácií z textu. V našom projekte sme sa rozhodli venovať spracovaniu prirodzeného jazyka na získanie potrebných informácií z odborných vedeckých článkov. Tým umožníme používateľovi vyhľadávať nad týmito textami a získavať tak potrebné informácie. Najprv sme analyzovali štruktúru výskumných článkov a identifikovali sme ich dôležité časti ako sú kľúčové slová, abstrakt či úvod. Táto analýza je podrobne opísaná v prílohe C.

Následne sme analyzovali nástroje, ktoré by sme mohli použiť pri spracovaní prirodzeného jazyka. Pozreli sme sa na nástroje spracovanie textov anglického jazyka a následne sme zisťovali možnosti spracovania textov pre slovenský a český jazyk. Vybrali sme si 5 nástrojov pre spracovanie anglického jazyka, s ktorými sa dá pracovať v jazyku Python alebo Java. Analyzovali sme tiež niekoľko nástrojov pre slovenčinu a češtinu - viac informácií v prílohe D.

Vedecké články, ktoré potrebujeme sú písané najmä v angličtine, a preto sme sa zamerali hlavne na analyzované nástroje pre spracovanie anglického jazyka. Vybrali sme si 2 nástroje pre jazyk Java (Ollie, Stanford OpenIE) a 2 nástroje pre jazyk Python (NLTK, SpaCy). Tieto nástroje sme otestovali na jednoduchých príkladoch a porovnali sme ich. Pozreli sme sa aj na grafické zobrazenie týchto dát (hlavne vo forme stromov). Bližšie k porovnaniu je uvedené v prílohe E.

Získavanie odborných článkov

Pre vytvorenie fungujúceho produktu je vhodné sa pozrieť aj na už existujúce práce. Získame tak inšpiráciu či poznanie ako to robiť/nerobiť, z používateľského hľadiska. V prílohe F sme sa bližšie pozreli na dostupné komerčné ale i vedecké riešenia podobné nášmu.

Aby sme mohli spracovávať články, potrebujeme ich najprv získať. Analyzovali sme preto podporu API služieb pre rôzne portály poskytujúce prístupy k článkom. Vyberali sme najmä služby, ktoré poskytovali články súvisiace s odborom, ktorý študujeme. Rozhodli sme sa použiť IEEE Xplore API kvôli dobre zdokumentovanej Python knižnici a možnosti získavania rôznych údajov ako sú metadáta aj celé texty. Analýza rôznych portálov pre podporu API služieb je bližšie opísaná v prílohe G.

Inžinierske dielo - Letný semester (zmena témy)

Big picture

Nasledujúca časť dokumentu obsahuje dokumentáciu projektu v rámci predmetu Tímový projekt II. Dokumentácia obsahuje informácie o globálnych cieľoch nášho projektu, ako aj analýzy vyhotovené pre účely vytvorenia jednotlivých častí projektu.

Ciele projektu na letný semester

Ciele nášho tímového projektu na letný semester sa v porovnaní s tými zo začiatku zimného semestra do značnej miery zmenili, nakoľko sa pri spolupráci s externými firmami vyskytli problémy. Na základe toho sme tak upustili od práce s no-coding platformou BigRing a vytvorenia inteligentného model-based agenta na báze znalostných grafov (knowledge graphs).

Hlavným cieľom letného semestra bolo vytvorenie webovej aplikácie určenej na zobrazovanie vedeckých článkov pomocou radiálneho grafu na základe hľadaného pojmu. Okrem zobrazenia detailu článku bolo cieľom implementovať do aplikácie zobrazovanie základných pojmov článku prostredníctvom znalostného grafu za účelom nielen sprehľadniť, ale aj urýchliť výber vedeckých publikácií. Pri implementácii webovej aplikácie nám pomohli aj výstupy z minulého semestra, predovšetkým analýza NLP a analýza získavania odborných článkov. Následne sme sa tak zamerali na analýzu možností vytvárania štruktúr z free textu a vytvorenie znalostného grafu na základe sémantických tripletov. Okrem toho bolo v úvode semestra cieľom analyzovať frameworky a následne najvhodnejšie z nich sme využili pri implementácii webovej aplikácie.

Celkový pohľad na vyhotovené časti projektu

V tejto časti dokumentácie je opísaný stav projektu ku dňu 18. 05. 2022 - šprinty 6-9 z letného semestra. Opisujeme výstupy projektu, na ktorých sme počas týchto šprintov pracovali, a to konkrétne: návrh riešenia s cieľmi, funkčnými vlastnosťami a scenárom prípadu použitia spolu s návrhom UI, implementáciu a testovanie vytvorenej webovej aplikácie.

Návrh riešenia

Ciele

Cieľom je vytvoriť webovú aplikáciu pre zlepšenie vyhľadávania nad vedeckými článkami. Vyhľadávanie by fungovalo nasledovne: Používateľ zadá dopyt do vyhľadávača a vráti sa mu x stránok, nad týmto zúženým výberom článkov následne spravíme analýzu a indexovanie a budeme vizualizovať kľúčové slová v grafe. Nevyhľadávalo by sa teda iba nad abstraktmi článkov, ale aj nad samotným obsahom článkov.

Funkčné vlastnosti

- analyzovanie a spracovanie výsledkov vyhľadávania z externého API
- analyzovanie celých článkov
- vizualizovanie výsledkov v grafe

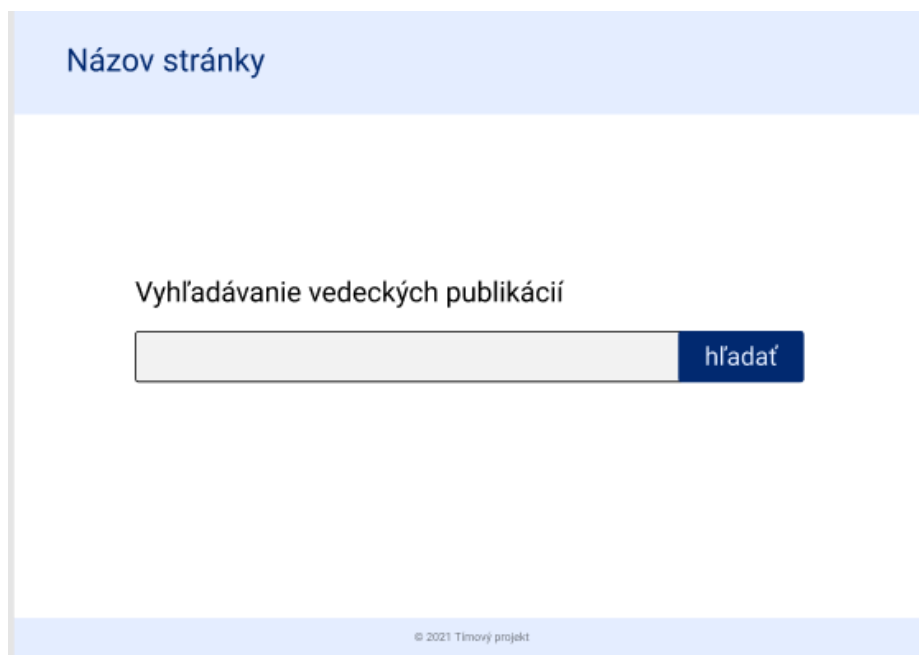
Scenár prípadu použitia

Predpoklad: Otvorená stránka

1. Používateľ zadá oblasti článkov ktoré ho zaujímajú do textového poľa
2. Používateľ stlačí tlačidlo vyhľadávania
3. Aplikácia zobrazí vizualizácie kľúčových pojmov článkov a prepojení medzi nimi
4. Používateľ klikne na jeden z pojmov
5. Aplikácia zobrazí zoznam článkov ku ktorým sa vzťahuje pojem

Dôsledky: Používateľ získa prehľad článkov najviac sa vzťahujúcich k jeho vyhľadávanému dopytu

Návrh UI



Obr. 1: Úvodná stránka



Obr. 2: Vizualizácia výsledkov vyhľadávania



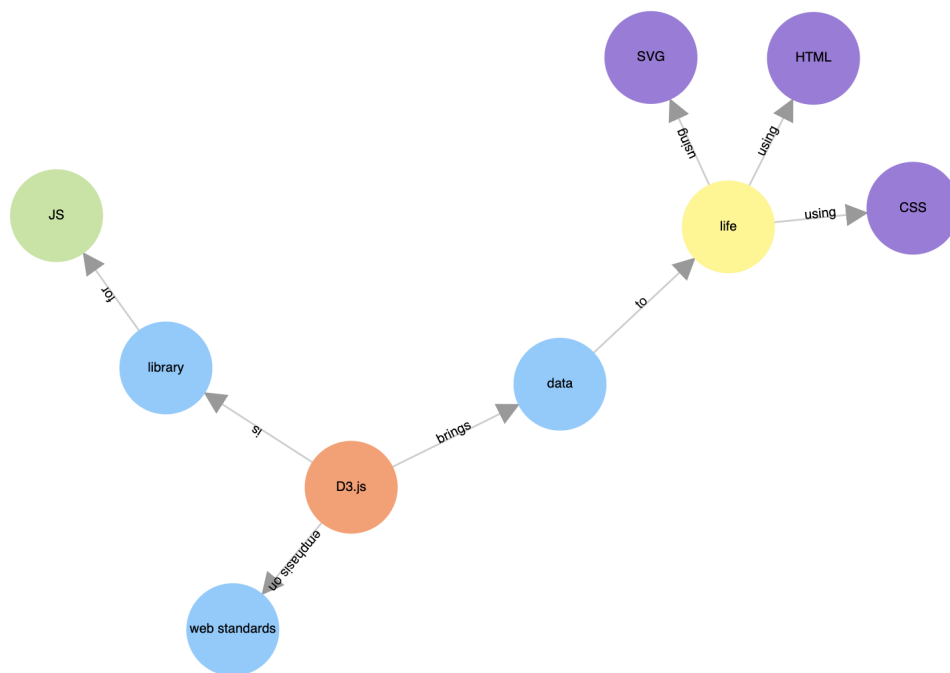
Obr. 3: Detail po zvolení konkrétneho článku- jeho prislúchajúce časti a detail článku



Obr. 3: Detail po zobrazení znalostného grafu pre danú časť článku

Vizualizácia výstupu NLP analýzy textu

Analyzovali sme rôzne možnosti vizualizácie dát na webe pomocou rôznych frameworkov a knižníc ako D3.js, ThreeJS a Sigma.js. V závere sme sa rozhodli naše riešenie implementovať v D3.js, pretože ponúka relatívne jednoduché možnosti úprav grafov pre naše potreby. Výsledkom analýzy bol aj základný prototyp, v ktorom sme zobrazili graf s vizualizáciou vzťahov medzi slovami vo vete.

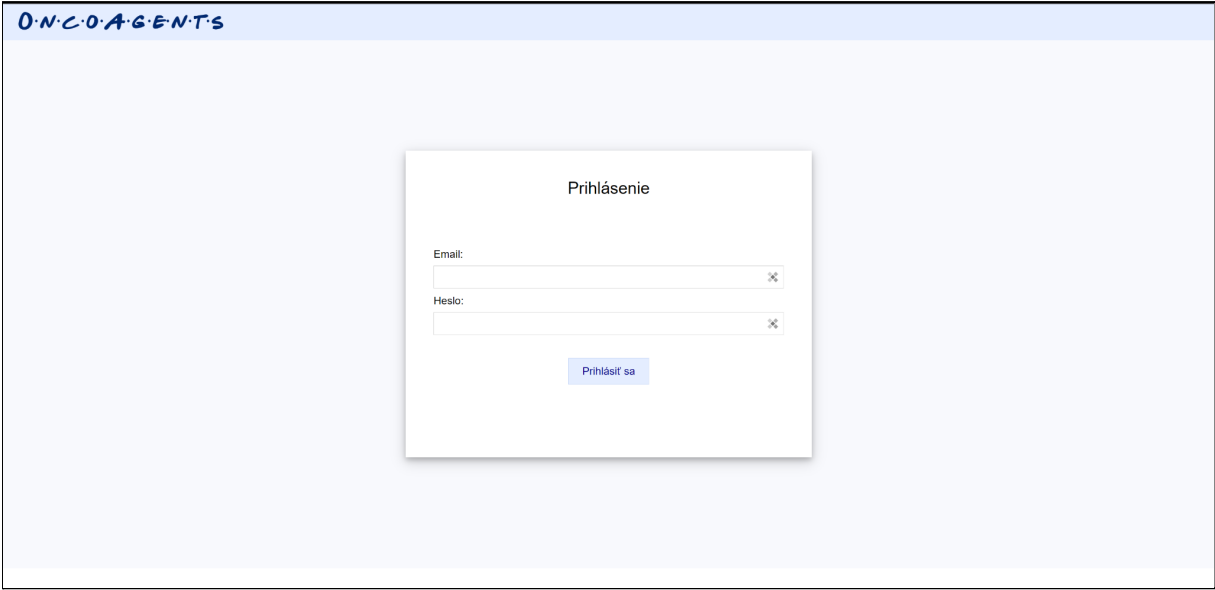


Implementácia

V nasledujúcej časti si popíšeme výslednú implementáciu aplikácie - spôsob realizácie Frontendu a spôsob extrakcie Tripletov z textov Článkov.

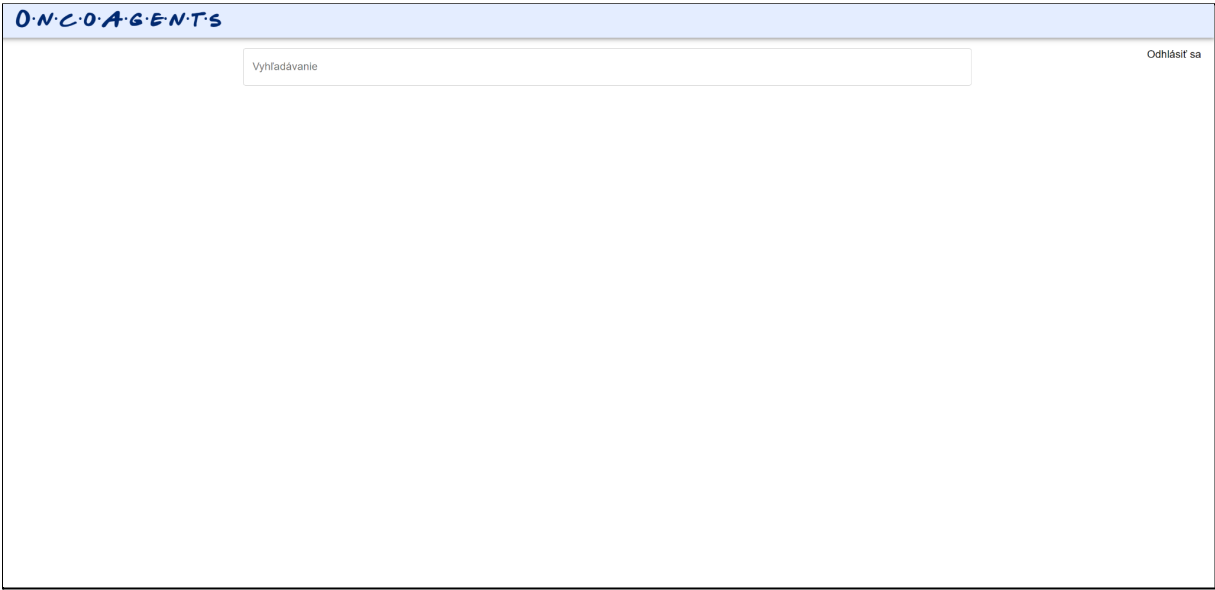
Obrazovky používateľského rozhrania

V tejto časti uvádzame dostupné obrazovky.



The screenshot shows a web browser window with a light blue header containing the text "ONCOAGENTS". The main content area is white and features a centered login form titled "Prihlásenie". The form includes two input fields: "Email:" and "Heslo:". Below the fields is a blue button labeled "Prihlásiť sa".

Prihlasovacia obrazovka - prihlásenie so stuba údajmi



The screenshot shows a web browser window with a light blue header containing the text "ONCOAGENTS". The main content area is white and features a search bar with the placeholder text "Vyhľadávanie". In the top right corner, there is a button labeled "Odhlásiť sa".

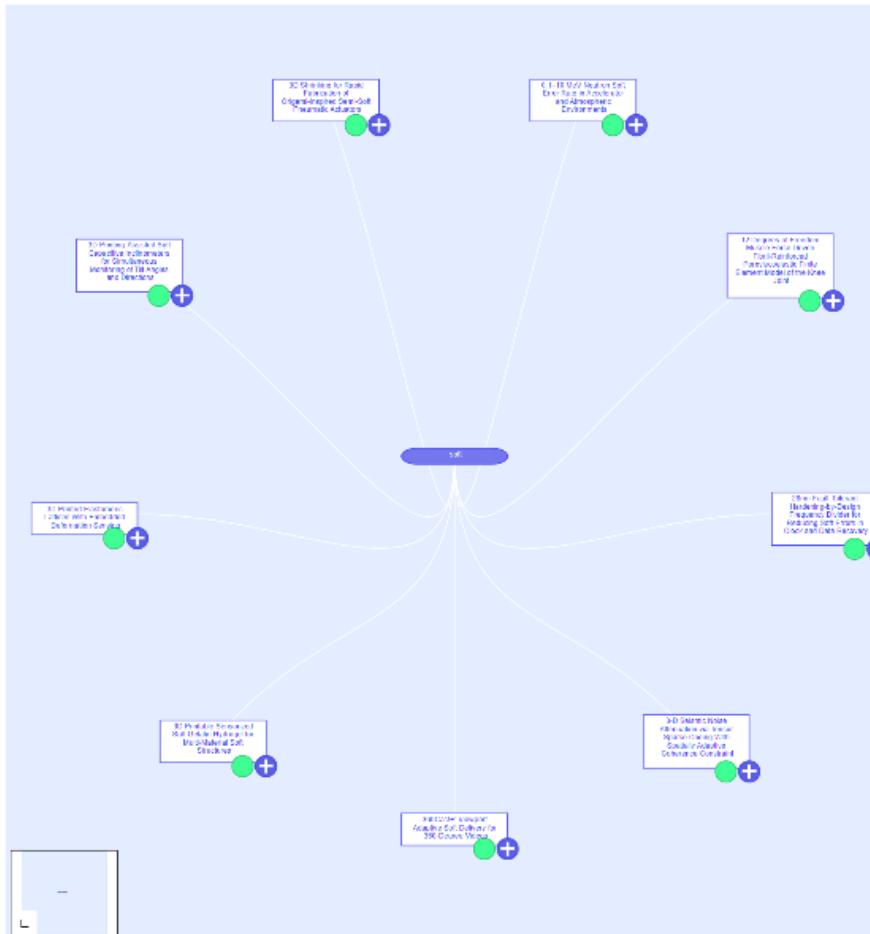
Úvodná obrazovka - vyhľadávanie

O.N.C.O.A.G.E.N.T.S

Odhlásiť sa

Vyhľadavanie

network



Zoznam článkov:

0.1-10 MeV Neutron Soft Error Rate in Accelerator and Atmospheric Environments

Autori: Matteo Cecchetti, Rubén Garcia Alía, Frédéric Wrobel, Andrea Coronetti, Kacper Bilko, David Lucsanyi, Salvatore Fiore, Giulia Bazzano, Elisa Pirovano, Ralf Nolte

Rok: 2021

URL: <https://ieeexplore.ieee.org/document/9373367/>

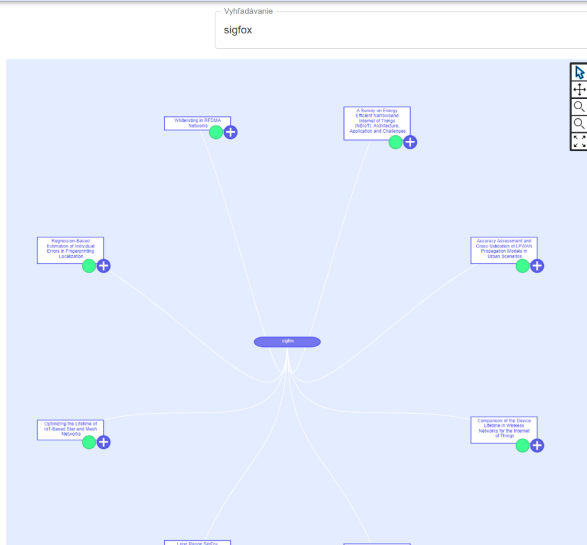
12 Degrees of Freedom Muscle Force Driven Fibril-Reinforced Poroviscoelastic Finite Element Model of the Knee Joint

Autori: A. Esrafilian, L. Stenroth, M. E. Mononen, P. Tanska, S. Van Rossum, D. G. Lloyd, I. Jonkers, R. K. Korhonen

Rok: 2021

URL: <https://ieeexplore.ieee.org/document/9256286/>

Po vyhľadani sa zobrazí graf a zoznam s obrázkami



Detail článku

Accuracy Assessment and Cross-Validation of LPWAN Propagation Models in Urban Scenarios

Authors: Martin Stusek, Dmitri Molchanov, Pavel Masek, Konstantin Mikhaylov, Otto Zeman, Martin Roubicek, Yevgeni Koucheryavy, Jiri Hosek

Year: 2020

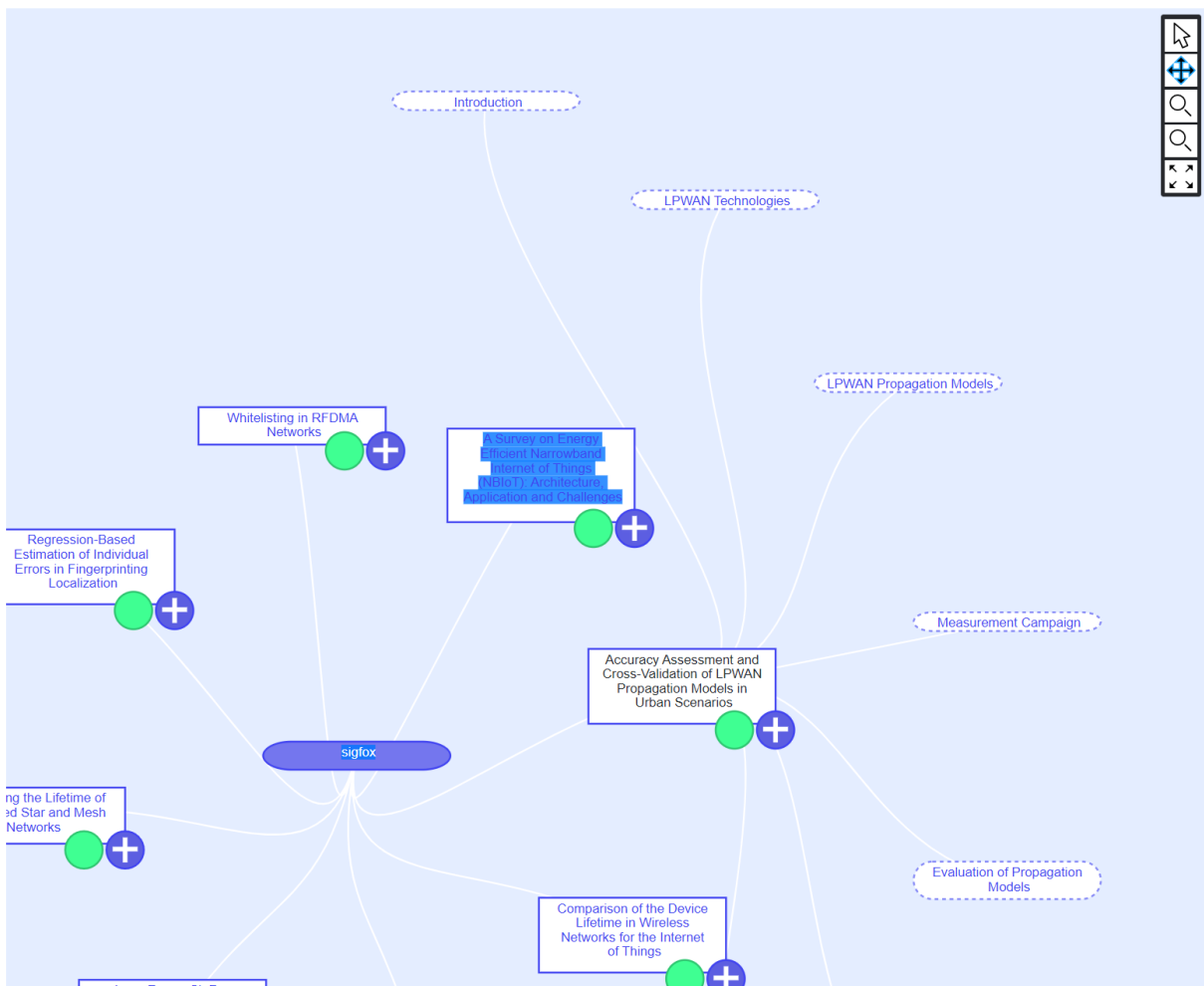
URL: <https://ieeexplore.ieee.org/document/9165731/>

Abstract: With the proliferation of machine-to-machine (M2M) communication in the course of the last decade, the importance of low-power wide-area network (LPWAN) technologies intensifies. However, the abundance of accurate propagation models proposed for these systems by standardization bodies, vendors, and research community hampers the deployment planning. In this paper, we question the selection of accurate propagation models for Narrowband IoT (NB-IoT), LoRaWAN, and Sigfox LPWAN technologies, based on extensive measurement campaign in two mid-size European cities. Our results demonstrate that none of the state-of-the-art models can accurately describe the propagation of LPWAN radio signals in an urban environment. For this reason, we propose enhancements to the selected models based on our experimental measurements. Performing the fine-tuning of the propagation models for one of the cities, we select Ericsson Urban (NB-IoT, LoRaWAN) and 3GPP (Sigfox) models as the ones providing the closest match. Finally, we proceed to perform cross-validation of the propagation models using the data set for another city. The tuned models demonstrate an excellent match with the real data in the cross-validation phase. They outperform their competitors by at least 20-80% in terms of relative deviation from the measured signal levels presenting the accurate option for NB-IoT, LoRaWAN, and Sigfox deployments planning in mid-size cities.

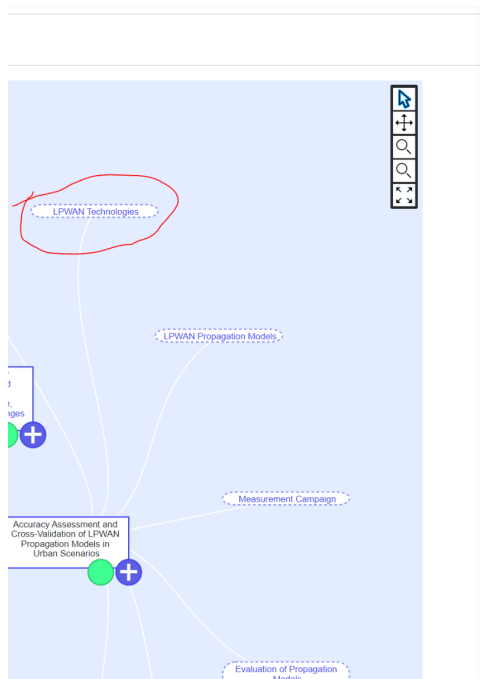
CCBY - IEEE is not the copyright holder of this material. Please follow the instructions via <https://creativecommons.org/licenses/by/4.0/> to obtain full-text articles and stipulations in the API documentation.

SECTION I.
Introduction

Detail článku - zobrazený je celý článok



Po kliknutí na '+' sa zobrazia klúčové slová v grafe



Deploying the LPWAN systems is a challenging task as they have to satisfy not only capacity requirements but also provide ubiquitous coverage for various sets of indoor and outdoor applications. Therefore, the propagation models are vital tools used to plan the network (i.e., the locations of both base stations and end devices) and to estimate/predict the quality of service and communication performance. However, there are many propagation models that differ in their structure and factors hampering clear conclusions about their choice for a particular technology. This is especially important for complex city-scale urban deployments of LPWAN systems [7]. Therefore, the intended models cover the whole spectrum of primary sources, i.e., standardization, vendors/operators, and academia. From each group, we selected the most commonly referenced models in the literature, which are supposed to provide an accurate prediction in the whole operating spectrum of the selected **LPWAN technologies** (mainly sub-GHz band) [8]–[9][10].

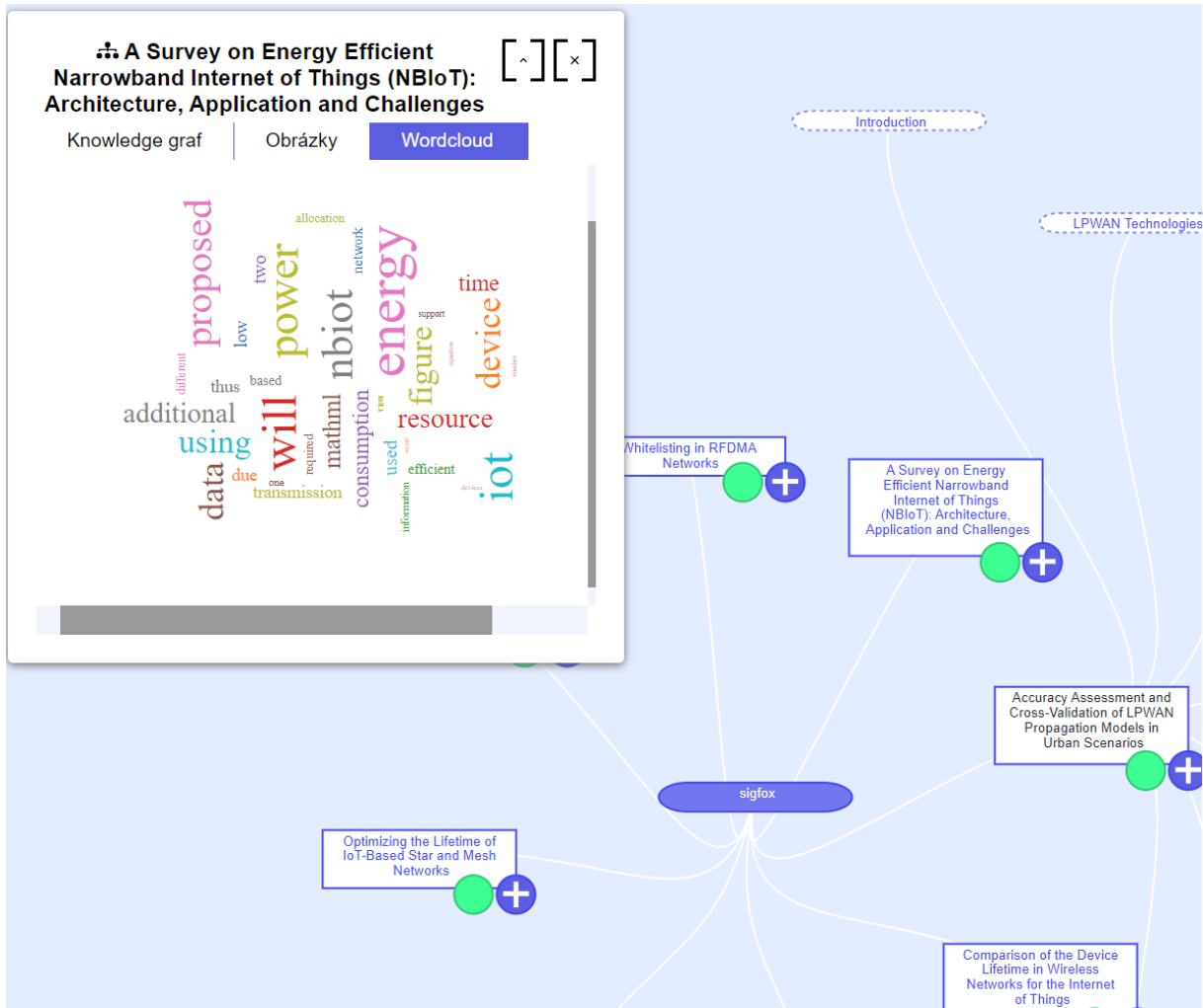
A. Main Contributions

In this paper, aiming to improve the accuracy of propagation models, which can be employed, e.g., for LPWAN deployments planning in an urban city environment, we have evaluated and improved the accuracy of standardized propagation models. To this aim, we start by over-viewing the standardized propagation models suitable for **LPWAN technologies**. Then we carry out an extensive measurement campaigns in two midsize European cities in the Czech Republic, i.e., Brno and Ostrava for three dominant **LPWAN technologies**: (i) NB-IoT, (ii) Sigfox, and (iii) LoRaWAN. Utilizing the results of our measurements, we propose and apply the two-steps refinement procedure based on the fine-tuning of models' parameters and cross-validation of the proposed propagation models.

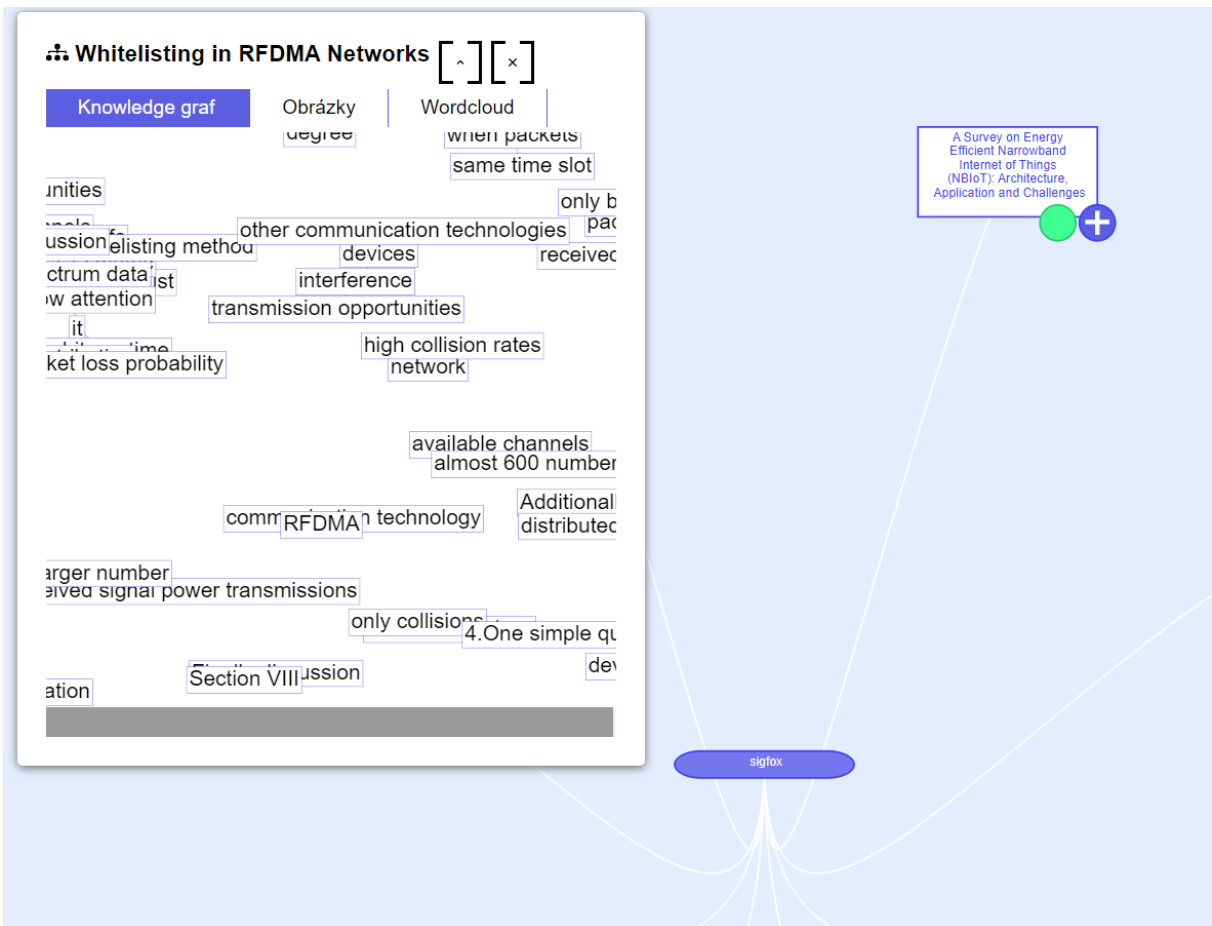
The key contributions are:

- We show that none of the standardized models provide accurate approximation for considered **LPWAN technologies** and needs to be fine-tuned to match the specifics of urban environment.
- We deliver a methodology for fine-tuning of the propagation models for the **LPWAN technologies** basing on the experimental results. The utilization of proposed methodology is demonstrated and cross-validated for the two cities in the Czech Republic – Brno and Ostrava.
- The proposed and reported propagation models tuned using the real-life measurements outperform

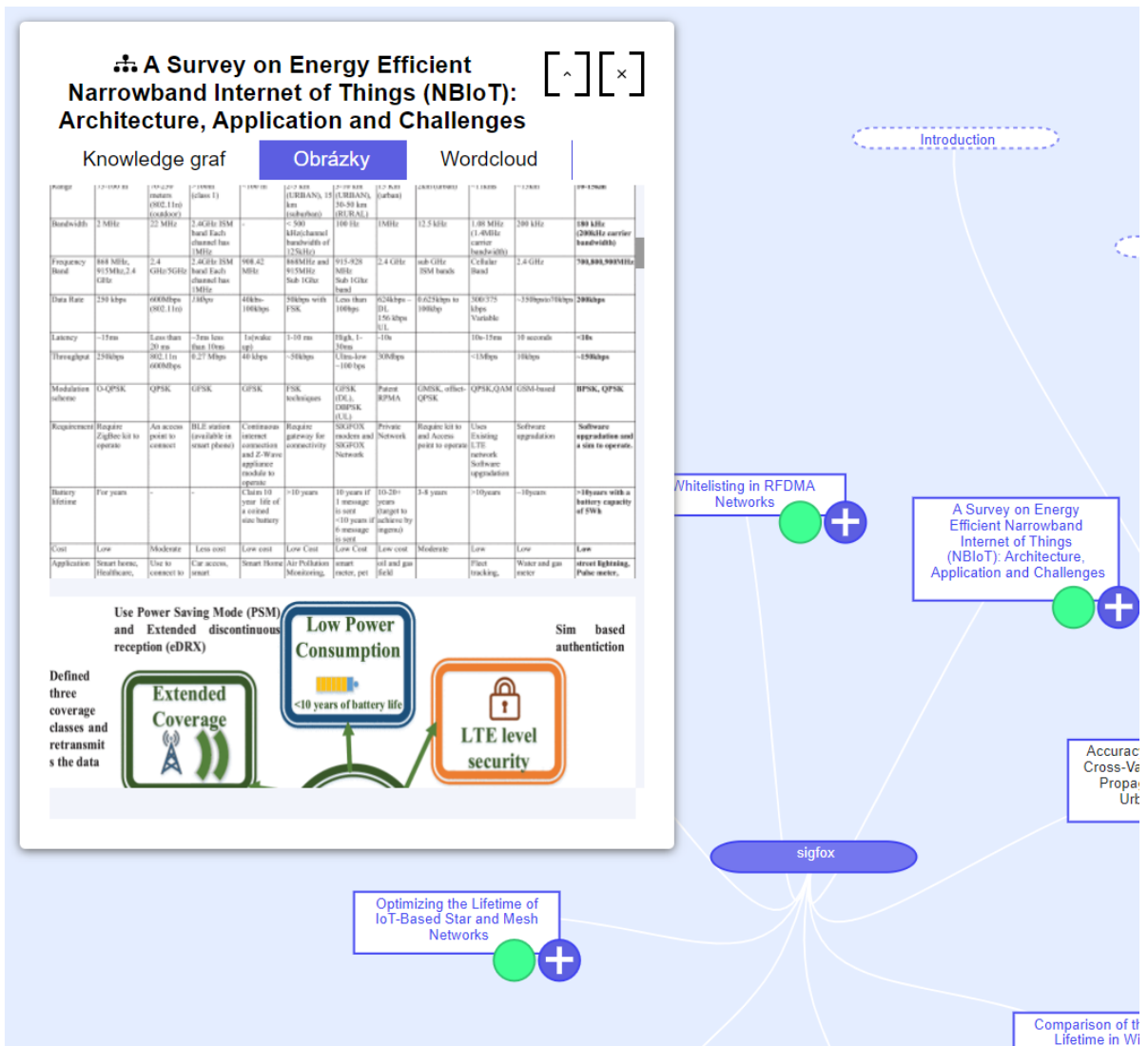
Po kliknutí na klúčové slovo v grafe sa dané slovo zvýrazní v texte



Po kliknutí na zelený kruh v grafe sa zobrazia informácie o článku - Wordcloud



Po kliknutí na zelený kruh v grafe sa zobrazia informácie o článku - Knowledge graf



Po kliknutí na zelený kruh v grafe sa zobrazia informácie o článku - obrázky

Extrakcia tripletov

Funkcionalitu extrakcie tripletov z odborného článku sme implementovali pomocou manuálne vytvorených pravidiel. Pravidlá extrakcie sú zostrojené najmä na základe POS a dependency tagov, na ktorých získanie sme použili knižnicu SpaCy.

Celý proces získania tripletov z voľného textu pozostáva z nasledujúcich krokov:

1. získanie voľného textu z HTML článku
2. extrakcia tripletov z voľného textu
3. filtrovanie tripletov
4. prevedenie tripletov do formátu vhodného pre knowledge graf

Implementácia každého z týchto krokov sa nachádza v samostatnom python súbore.

Na získanie voľného textu článku bolo potrebné vyčistiť HTML, ktoré vrátilo IEEE API. Na vyčistenie HTML sme použili triedu *HTMLParser* z knižnice *html.parser* a regulárne výrazy.

Nasledujúci algoritmus extrakcie tripletov pozostáva z týchto krokov:

1. rozdeľ voľný text na vety
2. pre každú vetu:
 - a. získaj entity (subjekty a objekty)
 - b. získaj predikáty
 - c. získaj relácie
 - d. sformátuj výstup

Testovanie

VANET (tím Č. 7): Posudok na prototyp tímu Č. 14: OncoAgents

1.1. Úvod

V tomto dokumente členovia tímu Vanet hodnotia a posudzujú doterajšiu prácu tímu číslo 14-OncoAgents. Hodnotený je ich zvolený postup, kvalita odvedenej práce a dosiahnuté výsledky implementácie prototypu, použiteľnosť vytvorenej webovej aplikácie.

1.2. Hodnotenie prototypu

Tím OncoAgents sa zaoberal projektom zameraným najmä na implementáciu webovej aplikácie, ktorá dokáže efektívnou formou zobraziť vedecké články týkajúce sa daného hľadaného pojmu. Články sú zobrazené formou radiálneho grafu, čo zvyšuje prehľadnosť, pričom po kliknutí na uzol sa zobrazí detail daného článku – čo považujeme za veľmi nápomocné. Rovnako vidíme prínos v zobrazení znalostného grafu, ktorý zachytáva základné pojmy článku a urýchľuje rozhodovanie pri výbere literatúry napr. pre diplomovú prácu.

Vytvorený testovací scenár zachytáva celkovú funkčnosť aplikácie pomerne dobre. Pri testovaní nenastali žiadne neočakávané situácie či problémy.

1.3. Hodnotenie práce tímu

Tím používal manažovací systém JIRA, kde bolo vidieť, že na úlohách pracovali priebežne v logických nadväznostiach. Zdrojový kód bol verziovaný nástrojom git, pričom boli dodržiavané konvencie písania kódu či verziovania uvedené v dokumentácii. Výsledná dokumentácia je prehľadná a zachytáva celý priebeh vytvárania výsledného produktu so všetkými potrebnými časťami ako napr. špecifikácia zadania, potrebné metodiky, opis inžinierskeho diela či rôzne analýzy.

1.4. Zhodnotenie

Výsledky tímu OncoAgents sú veľmi dobré a vytvorená webová aplikácia je na dobrej úrovni. Vzhľadom na problémy, s ktorými sa musel tím vysporiadať v zimnom semestri oceňujeme zanietenie a snahu, ktorú tím venoval spracovaniu novej témy.

Webové sídlo tímu je prehľadné a obsahuje všetky potrebné dokumenty ako zápisnice či retrospektívy a finálnu dokumentáciu.

Používateľské akceptačné testovanie

Vstupné podmienky: prístup na internet, funkčná webová aplikácia, prihlasovacie údaje do akademického informačného systému

Krok	Očakávaný výstup	Status	Poznámky
Používateľ otvorí http://team14-21.studenti.fiit.stuba.sk/onco-agents/	Aplikácia zobrazí prihlasovací formulár.	OK	
Používateľ zadá prihlasovacie údaje a klikne na tlačidlo Prihlásiť sa.	Aplikácia overí prihlasovacie údaje. Ak sú správne, aplikácia zobrazí hlavnú stránku. Ak nie sú správne, používateľ je informovaný o zlých prihlasovacích údajoch.	OK	
Používateľ napíše do vyhľadávacieho poľa slovo: Machine learning a stlačí ENTER.	Aplikácia zobrazí zoznam článkov týkajúcich sa vyhľadávaného slova a vyhľadávané slovo ako klikateľné tlačidlo.	OK	
Používateľ klikne na tlačidlo Machine learning.	Aplikácia zobrazí radiálny graf vychádzajúci z daného slova zobrazujúci nájdené články ako primárne vrcholy grafu.	OK	
Používateľ klikne na ľubovoľný vrchol v grafe.	Aplikácia zobrazí detail článku ako okno vyrolované sprava.	OK	
Používateľ klikne na zelený kruh.	Aplikácia zobrazí modálne okno s tromi kartami - Knowledge graf, Obrázky, Word Cloud.	OK	
Používateľ klikne na tab Knowledge graf.	Aplikácia zobrazí v modálnom okne znalostný graf vychádzajúci z daného článku.	OK	
Používateľ klikne na tab Obrázky.	Aplikácia zobrazí v modálnom okne obrázky z daného článku.	OK	
Používateľ klikne na tab Word Cloud.	Aplikácia zobrazí v modálnom okne word cloud vychádzajúci z daného článku.	OK	
Používateľ klikne na X v modálnom okne.	Aplikácia zatvorí modálne okno.	OK	
Používateľ klikne na tlačidlo + pri ľubovoľnom vrchole.	Aplikácia zobrazí podgraf vychádzajúci z daného vrcholu s názvami častí článku ako vrcholmi.	OK	
Používateľ zadá klávesovú skratku CTRL+F.	Aplikácia umožní používateľovi vyhľadať ľubovoľné slovo na danej stránke - v detaile článku aj zozname článkov.	OK	
Používateľ klikne na tlačidlo X v detaile článku.	Aplikácia zatvorí detail článku.	OK	
Používateľ klikne na tlačidlo odhlásiť sa.	Aplikácia zobrazí prihlasovací formulár.	OK	

Inštalčná príručka

V tejto prílohe si opíšeme jednotlivé kroky potrebné k spusteniu našej aplikácie.

Backend

Backend aplikácie vieme spustiť v dockeri pomocou nasledovných príkazov:

```
docker-compose up --build
docker-compose run backend python manage.py migrate
```

Backend je následne dostupný na:

```
Dev - localhost:8000
Debug - localhost:3000
```

Frontend

Následne sa presunieme do priečinku s frontendom aplikácie.

```
$ cd ./tp_fe
```

V tomto priečinku nainštalujeme pomocou príkazu nižšie všetky knižnice potrebné pre spustenie prototypu.

```
$ npm install
```

Príkazom nižšie spustíme frontend:

```
$ npm start
```

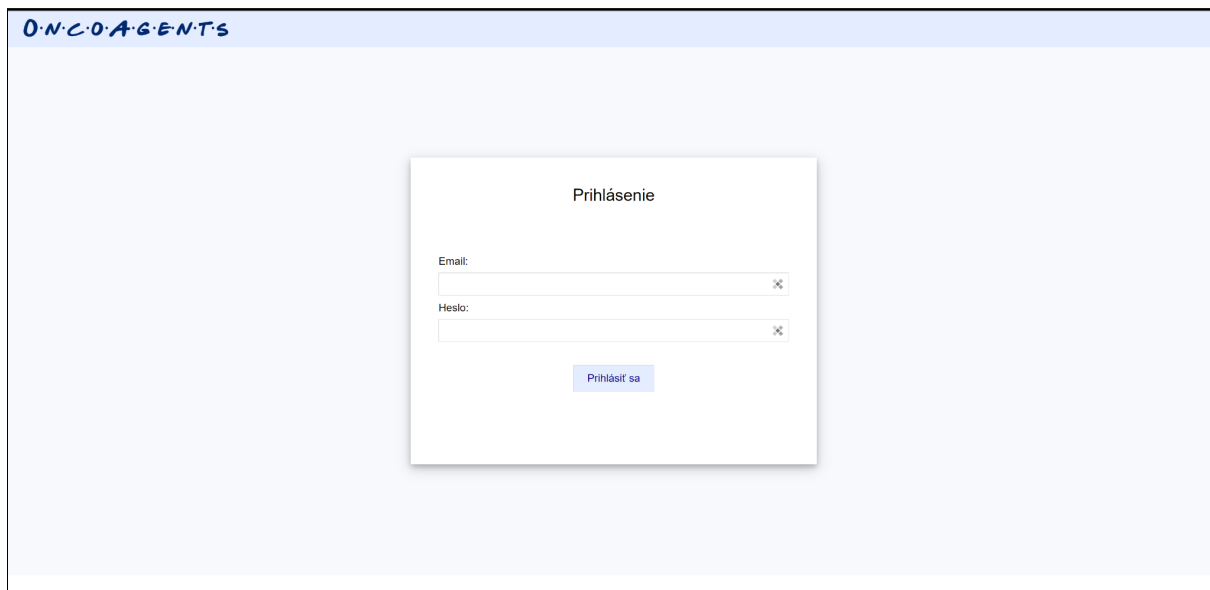
Následne, pokiaľ je port 3001 dostupný, zadaním url *localhost:3001/onco-agents/* do prehliadača spustíme aplikáciu.

Používateľská príručka

V tejto kapitole uvádzame používateľskú príručku pre nami vytvorenú webovú aplikáciu.

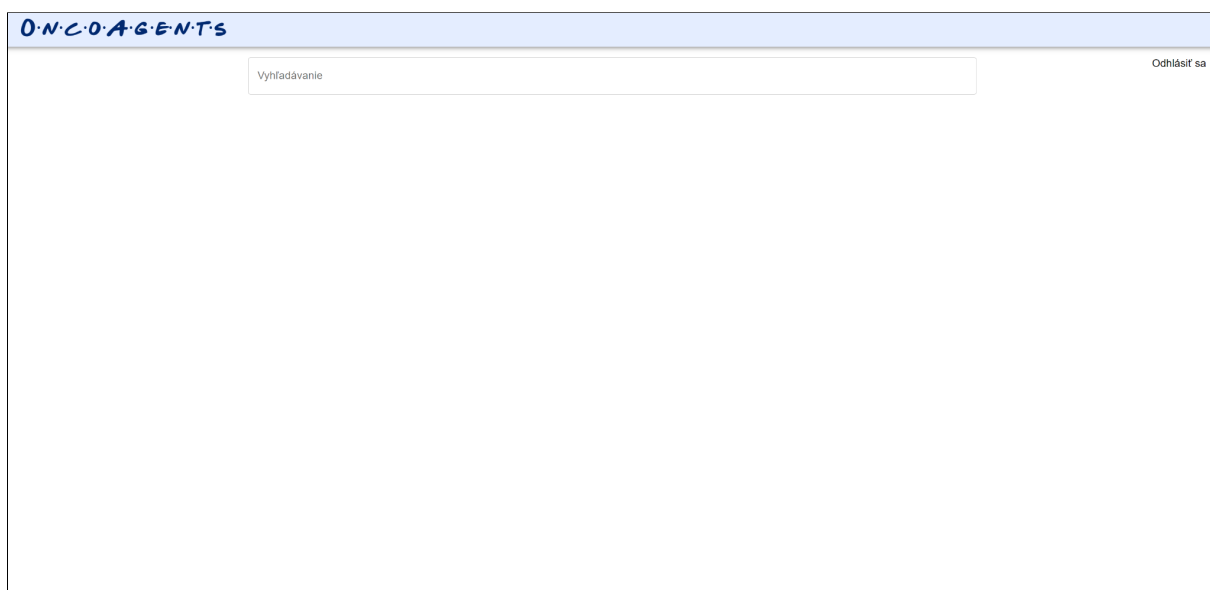
Prihlásenie a odhlásenie

Pred začatím používania aplikácie je nutné aby sa používateľ prihlásil. Preto pri prvom spustení v prehliadači je používateľ presmerovaný na prihlasovaciu obrazovku. Tu musí zadať prihlasovacie údaje do stuba.



Prihlasovacia obrazovka

Po prihlásení sa používateľ dostane na úvodnú stránku s vyhľadávačom. Kedykoľvek sa môže pomocou tlačidla v pravom rohu odhlásiť. Tento krok by ho vrátil opäť na prihlasovaciu obrazovku.

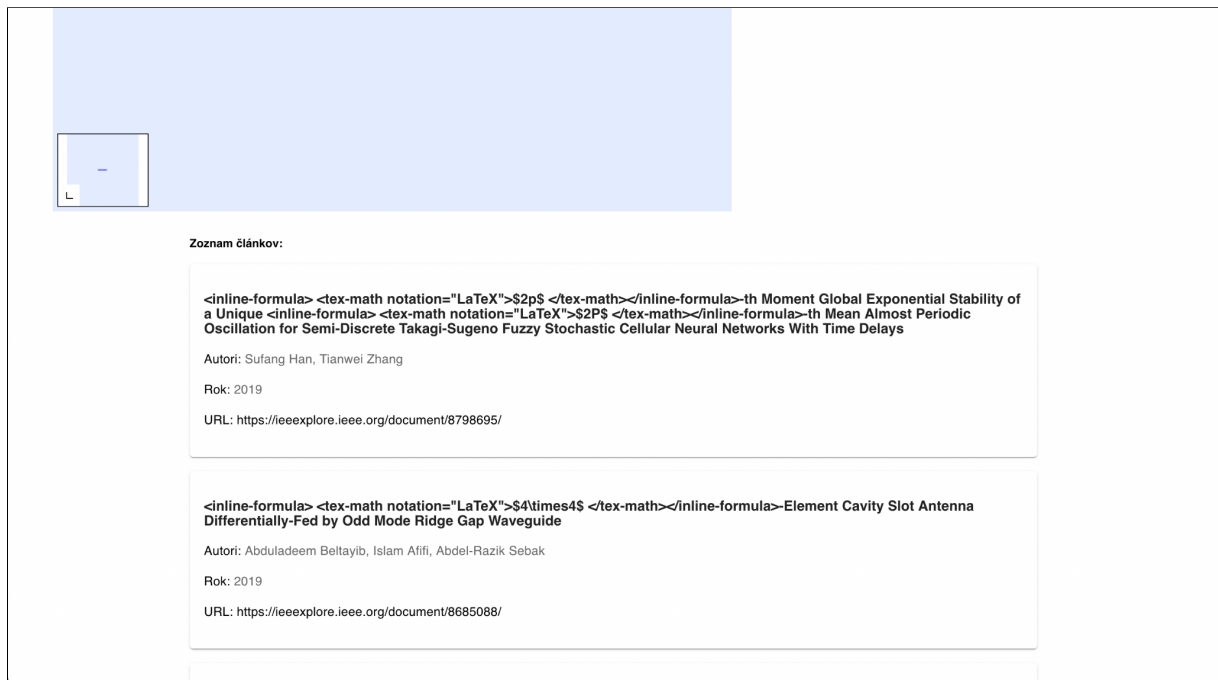
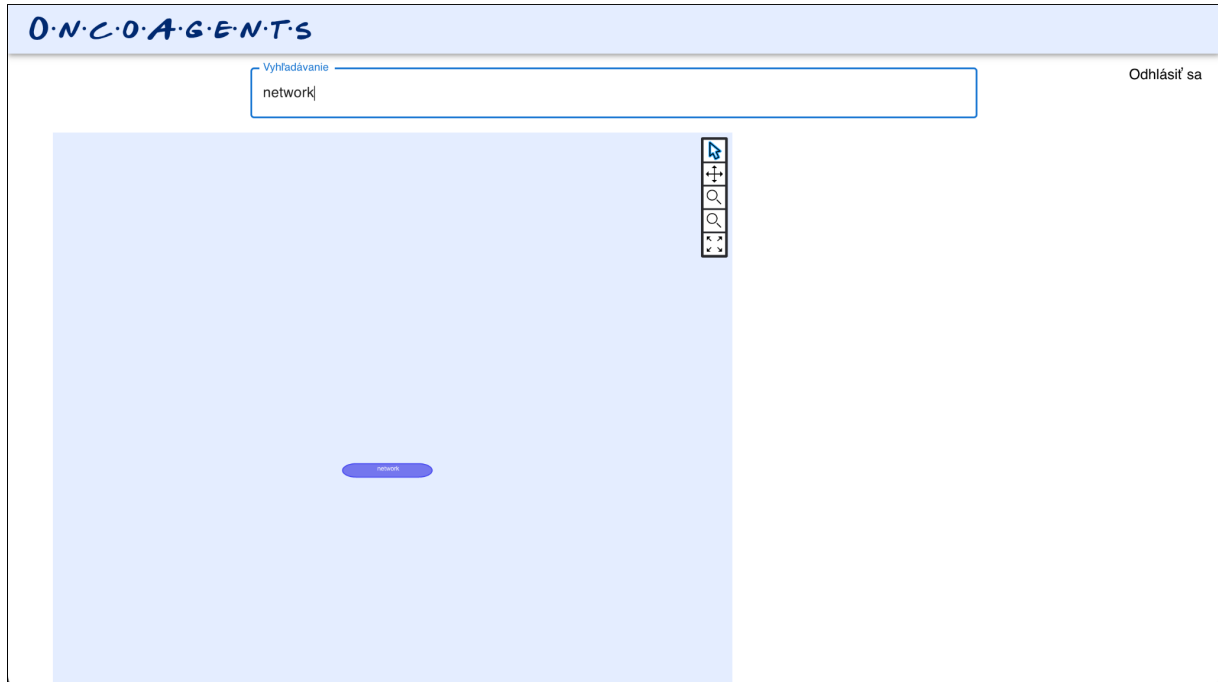


Úvodná obrazovka

Vyhľadavanie

Na úvodnej stránke môže používateľ vidieť vyhľadávaciu lištu. Do tejto lišty môže napísať text ktorý chce vyhľadať a následne stlačiť klávesu ENTER. Po tomto kroku sa mu zobrazí graf s jedným uzlom, ktorý predstavuje text zadaný do vyhľadávača a zoznam článkov.

Po kliknutí na článok zo zoznamu sa zobrazí detail daného článku.



Graf s jedným uzlom, ktorý predstavuje text zadaný do vyhľadávača a zoznam článkov

Interakcia s grafom

Používateľ môže graf ovládať pomocou lišty v pravom rohu:

Prvá ikona - môže interagovať s grafom - klikať na tlačidlá a jednotlivé uzly grafu

Druhá ikona - môže graf ťahaním posúvať

Tretia ikona - môže graf priblížiť

Štvrtá ikona - môže graf oddialiť

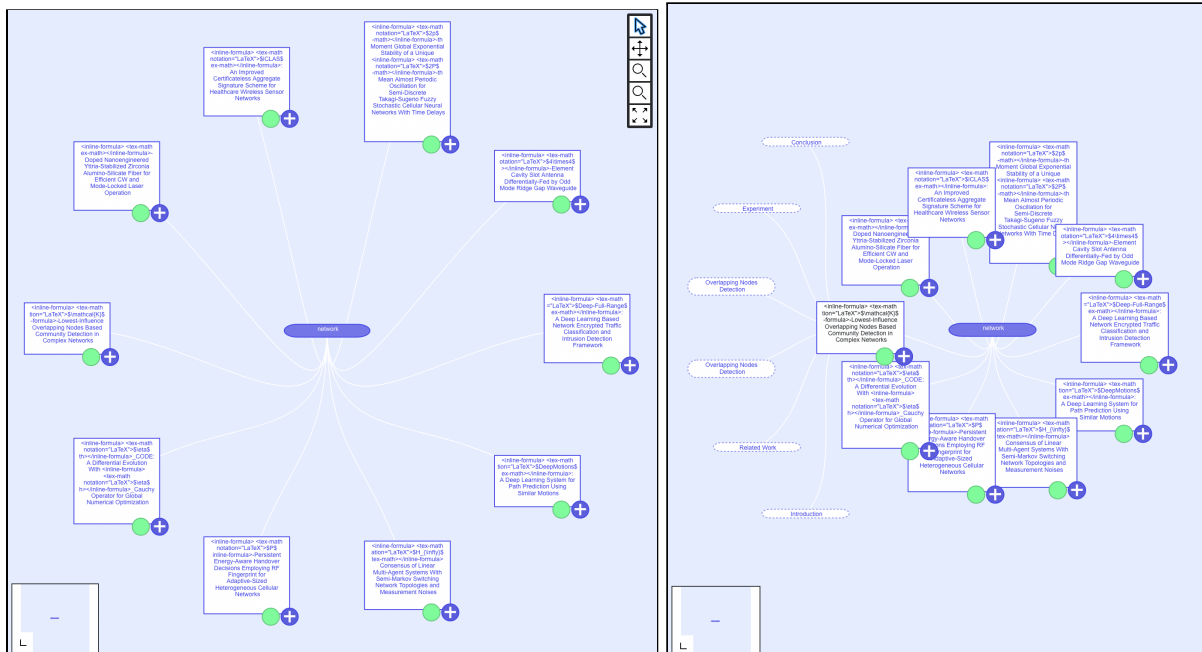
Piata ikona - vycentruje graf



Ovládacia lišta grafu

Ako sme už spomínali v prípade, že má používateľ kliknuté na prvú ikonu v ovládacej lište, môže využívať ďalšiu funkcionlitu, ktorú si ďalej opíšeme.

V prípade ak používateľ klikne na hlavný uzol v grafe, zobrazia sa mu ďalšie uzly, ktoré predstavujú názvy článkov. Ak následne klikne na tlačidlo + zobrazia sa mu ďalšie uzly grafu = tieto predstavujú jednotlivé kapitoly daného článku.

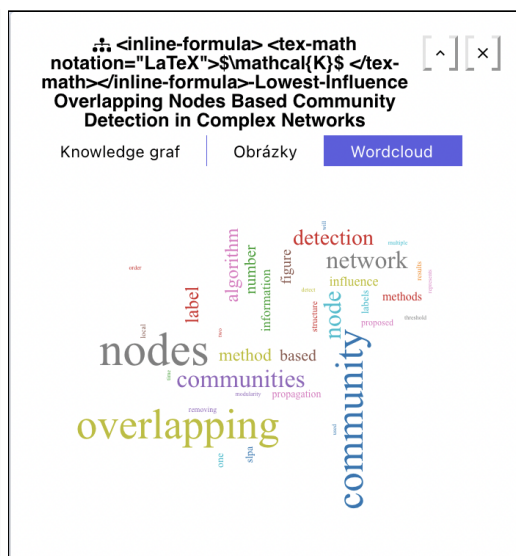
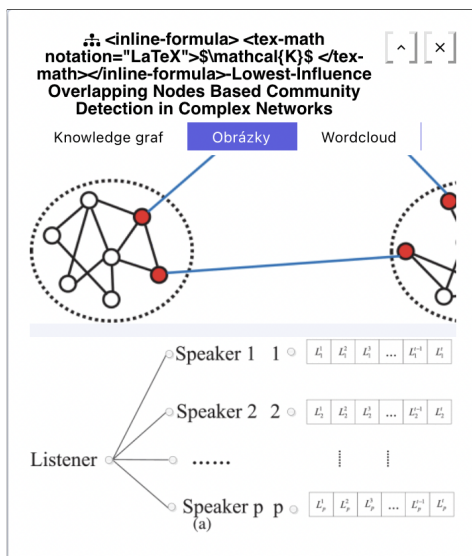
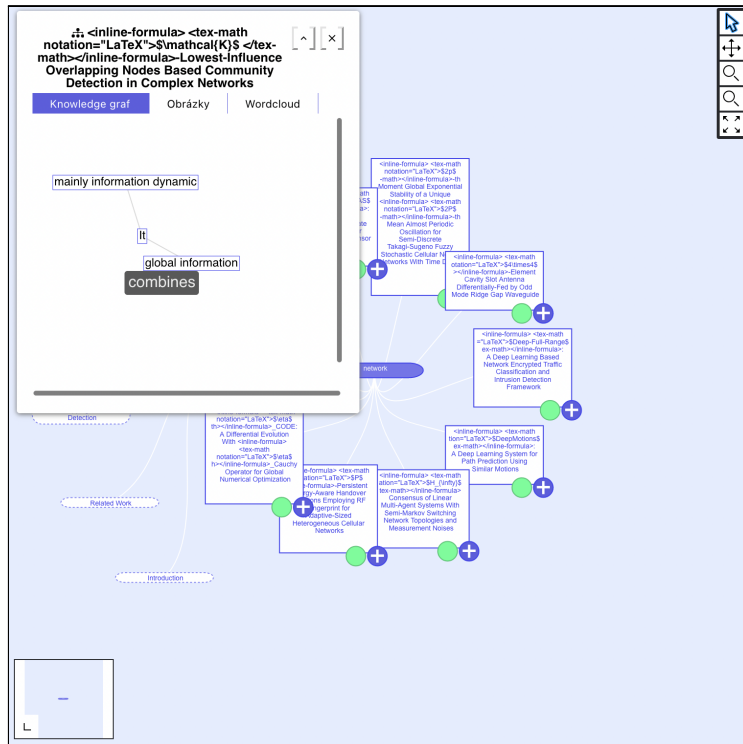


Zoznam článkov a kapitoly

Po kliknutí na ktorýkoľvek okrem hlavného uzla sa zobrazí detail daného článku, ktorému patrí daná vetva grafu.

Ak používateľ klikne na zelené tlačidlo pri uzloch zobrazí sa mu malý panel v ľavom rohu grafu. tento panel vie zavrieť kliknutím na tlačidlo x alebo zrolovať kliknutím na tlačidlo ^. Tento panel má tri taby:

- V prvom tabe môže používateľ vidieť tzv. knowledge graf, ktorý pozostáva z tripletov extrahovaných z článku. Po kliknutí na ľubovoľné slovo knowledge grafu sa mu otvorí detail článku so zvýrazneným slovom na ktoré klikol.
- V druhom tabe môže používateľ vidieť zoznam všetkých obrázkov z daného článku.
- V treťom tabe môže používateľ vidieť tzv. wordcloud, ktorý pozostáva z najčastejšie sa vyskytujúcich slov extrahovaných z článku. Po kliknutí na ľubovoľné slovo wordcloudu sa mu otvorí detail článku so zvýrazneným slovom na ktoré klikol.



Panel s tabmi

Detail Článku

Detail článku je vysúvací panel, ktorý sa otvára už vo viacerých spomínaných prípadoch (zo zoznamu článkov alebo z grafu a panelu). V detaile môže používateľ vidieť názov článku, jeho autorov, rok vydania a url cestu k nemu na IEEE. Okrem toho si tu môže prečítať Abstrakt článku a celý článok.

Ak sa na detail článku preklikol z knowledge grafu alebo z wordcloudu, či po kliknutí na uzly kapitol tak sa mu tieto slová v texte zvýraznia.

The image shows a knowledge graph interface. On the left is a word cloud with terms like 'overlapping nodes', 'community', 'detection', 'algorithm', 'method', 'number', 'figure', 'label', 'influence', 'node', 'information', 'method', 'network', 'nodes', 'community', 'algorithm', 'method', 'number', 'figure', 'label', 'influence', 'node', 'information'. On the right is a network graph with nodes and edges, some highlighted in blue and green. A search bar at the top right contains the text 'x'.

Detail článku

Overlapping Nodes Based Community Detection in Complex Networks

Authors: Meili Tang, Qin Liu, Tinghui Ma, Jie Cao, Yuan Tian, Abdullah Al-Dhelaan, Mohammed Al-Dhelaan

Year: 2019

URL: <https://ieeexplore.ieee.org/document/8769926/>

Abstract: Community detection is a classic and essential task in complex network analysis which aims at understanding their structural properties and functional organization. Community detection based on overlapping nodes has become one of the most popular methods in recent years. But, how to define the overlapping nodes in the real network is an important job. In this paper, we use the speaker-listener label propagation algorithm to find overlapping nodes. Second, we proposed a new metric which is based on local and global attributes to measure the node influence so as to evaluate the overlapping nodes' importance. Community detection is realized with the k-lowest-influence overlapping nodes deleting. Then, we assign the removed nodes into specific communities by voting to find the final community structure. The voting strategy is based on choosing the specific communities containing most of their neighbors. Finally, the extensive experiments on real-world networks demonstrate that our proposed method improves the quality of community detection methods and shows both the effectiveness and

Detail článku

The image shows the same knowledge graph interface as above, but with a search bar at the top containing the text 'network'. A red box highlights the word 'network' in the word cloud, and a red arrow points from this box to the word 'network' in the text of the article preview on the right.

efficiency of the method.

CCBY - IEEE is not the copyright holder of this material. Please follow the instructions via <https://creativecommons.org/licenses/by/4.0/> to obtain full-text articles and stipulations in the API documentation.

SECTION I.

Introduction

In [1], cellular neural networks (CNNs), which have been widely applied in psychophysics, parallel computing, perception, robotics, adaptive pattern recognition, associative memory, image processing pattern recognition and combinatorial optimization. All of these applications heavily depend on the (almost) periodicity and global exponential stability. Specifically, many scholars had focused on the issues of the existence and global exponential stability of the equilibrium point, periodic and almost periodic solutions of CNNs with time delays in literatures [2]–[3][4][5][6][7]. For instance, Xu [7] considered the following CNNs with time delays:

$$\begin{aligned} \dot{x}_i(t) &= -a_i x_i(t) + \sum_{j=1}^n w_{ij} x_j(t) + \sum_{j=1}^n v_{ij} x_j(t-\tau_{ij}) + I_i \\ \dot{y}_i(t) &= -b_i y_i(t) + \sum_{j=1}^n u_{ij} y_j(t) + \sum_{j=1}^n v_{ij} y_j(t-\tau_{ij}) + J_i \end{aligned}$$

where n denotes the number of units in a neural network, $x_i(t)$ corresponds to the state of the i th unit at time t , $y_i(t)$ represents the passive decay rate at time t , a_i and b_i are the neuronal output signal functions, w_{ij} and $v_{ij}(t)$ denote the strength of the j th unit on the i th unit at time t , $v_{ij}(t)$ denotes the external input at time t , the continuous function $\tau_{ij}(t)$ corresponds to the information transmission delay at time t , $i, j = 1, 2, \dots, n$. In [7], the author studied the existence and exponential stability of anti-periodic solutions of system (1.1).

Uncertain models described by stochastic differential equations have

Zvýraznenie slova v texte článku

Technická dokumentácia

Backend

Nižšie popisujeme niektoré z vývojových technológií s ktorými sme sa stretli pri vývoji aplikácie a presné verzie, ktoré sme využili.

Použité technológie

django rest framework

stránka: <https://www.django-rest-framework.org/>

verzia: 3.13.1

REST rámec slúžiaci na vytváranie webových API rozhraní v jazyku Python

lxml

stránka: <https://lxml.de/>

verzia: 4.8.0

Python knižnica na prácu s XML formátmi

ldap3

stránka: <https://ldap3.readthedocs.io/en/latest/>

Python knižnica na prístup k LDAP serveru

spacy

stránka: <https://spacy.io/>

Python knižnica pre spracovanie prirodzeného jazyka

Endpointy a opis tried

V časti backendu používame štandardné rozdelenie, ktoré poskytuje framework Django. Hlavná logika projektu sa nachádza v priečinku *ieee*.

V súbore *views.py* tohto priečinka sú implementované triedy komunikujúce s IEEE API rozhraním a pomocné funkcie potrebné na spracovanie získaných výstupov tohto API:

Articles

trieda *Articles* zabezpečuje funkčnosť pre url adresu *ieee/articles/*. Obsahuje metódu typu POST, ktorá na vstupe (z frontendu) v tele dopytu dostáva vyhľadávaný reťazec zadaný používateľom. V tejto metóde je reťazec odoslaný do IEEE API, ktorá nám poskytne prvých 10 výsledkov pre

vyhľadávaný reťazec. Na komunikáciu s API rozhraním IEEE je potrebné s každým dopytom poslať licenčný kľúč, o ktorý sme ako tím požiadali na oficiálnej stránke IEEE a bol nám pridelený. API ako výsledky nevráti celé texty 10 článkov naraz, no vráti metadáta vo forme JSONU o týchto článkoch, z ktorých vyberieme potrebné informácie. Ďalšími dopytmi do IEEE API pomocou čísel článkov postupne získavame celé texty článkov v XML formáte. Načítaním týchto XML súborov do stromu pomocou lxml knižnice získame prístup k jednotlivým elementom XML súboru. Takýmto spôsobom vieme spracovať článok a zistiť jeho štruktúru - hlavné nadpisy, ktoré sú spolu s vybranými meta informáciami o článku odoslané späť na frontend.

Na komunikáciu s API rozhraním sme použili existujúcu Python knižnicu dostupnú na oficiálnej stránke IEEE: https://developer.ieee.org/Python_Software_Development_Kit. Knižnicu bolo potrebné mierne modifikovať, pretože bola vytvorená pre inú verziu jazyka Python. Túto knižnicu sme stiahli a pridali do projektu ako priečinok s názvom *explore* (tp_be/ieee/xplore).

FullArticleFormat

trieda *FullArticleFormat* zabezpečuje funkcionality pre url adresu *ieee/full_article_format/<str:article_number>*, pričom *article_number* je číslo článku, ktoré poskytuje IEEE API ako jedno pole z metadát. Úlohou metódy GET tejto triedy je odoslať späť na frontend celý článok spolu s tabuľkami a obrázkami vo formáte HTML.

Images

trieda *Images* obsahuje metódu typu GET, ktorá načíta celý text požadovaného článku do objektu stromu a pomocou XML elementov vyhľadá zdroje všetkých obrázkov. Na frontend je potom odoslaná odpoveď vo forme poľa všetkých zdrojov obrázkov. Funkcionalita je dostupná pomocou GET požiadavky na adresu *ieee/images/<str:article_number>*.

Triples

trieda *Triples* obsahuje metódu typu GET, ktorá slúži na získanie tripletov z textu zadaného článku. Metóda najprv načíta celý text požadovaného článku. Následne sa na získaný text aplikuje funkcionality extrakcie tripletov, ktorá sa nachádza v module *nlp*. Na frontend je na záver odoslaná odpoveď vo formáte, priamo využiteľnom v knowledge grafe. Funkcionalita je dostupná pomocou GET požiadavky na adresu *ieee/articles/<str:article_number>/triples*.

Wordcloud

trieda *WordClouds* obsahuje metódu typu GET, ktorá slúži na získanie údajov pre wordcloud vizualizáciu. Metóda najprv načíta celý text požadovaného článku. Následne je získaný text vyčistený a pre jednotlivé slová textu sú vypočítané ich váhy. Na frontend je na záver odoslaná odpoveď vo formáte, priamo využiteľnom vo wordcloud vizualizácii. Funkcionalita je dostupná pomocou GET požiadavky na adresu *ieee/articles/<str:article_number>/wordcloud*.

V priečinku *xauth* sa nachádzajú funkcie a logika potrebné na zabezpečenie funkcionality registrácie a prihlasovania používateľov, ktorá je dostupná pomocou POST požiadaviek na url adrese */xauth/login*. Vytvorená aplikácia je prístupom obmedzená na ľudí, ktorí majú k dispozícii platné prihlasovacie údaje do Akademického informačného systému (tj. študenti, učitelia a ďalší zamestnanci pôsobiaci na STU). Pri registrácii a prihlasovaní do tejto aplikácie je potrebné prihlasovacie meno a heslo, ktoré sú následne overené voči LDAP Akademického informačného systému. Táto funkcionality si vyžaduje pripojenie k školskej sieti - produkčná verzia je nasadená na školskom serveri a v prípade lokálneho spúšťania projektu mimo školy je potrebné VPN pripojenie.

Po registrácii sú údaje používateľa uložené do PostgreSQL databázy (heslo ukladané v hešovanom tvare). Zároveň, po registrácii je vytvorený a vrátený token, prostredníctvom ktorého sa používateľ autentifikuje v každom ďalšom volaní, čím je zabezpečený prístup len ľuďom z STU.

Frontend

Použité technológie

ReactJS

stránka: <https://reactjs.org/>

verzia: 17.0.2

JavaScript knižnica na vytváranie používateľských rozhraní.

d3.js

stránka: <https://d3js.org>

verzia: 7.3.0

D3 umožňuje naviazať ľubovoľné údaje na DOM a následne na dokument aplikovať transformácie riadené údajmi. V našej aplikácii sme napríklad použili knižnicu D3 na vytvorenie interaktívneho grafu SVG s interakciou.

axios

stránka: <https://axios-http.com/>

verzia: 0.26.1

Axios je jednoduchý HTTP klient pre prehliadač a node.js.

MUI

stránka: <https://mui.com/>

verzia: 5.4.4

Material UI- MUI ponúka sadu nástrojov (knižnicu komponentov) používateľského rozhrania pre React.

npm

stránka: <https://www.npmjs.com/>

verzia: 6.14.16

Má viacero použití, v našom prípade sme npm využili na sťahovanie a manažovanie package-ov v našom projekte.

Opis tried

V tejto časti sa nachádza dokumentácia k vytvoreným React triedam. Pri triedach je uvedený ich stručný popis a zoznam ich detí (iba nami vytvorené triedy).

/views

V priečinku */views* sa nachádzajú layout-y aplikácie.

Login

Layout predstavujúci prihlasujúcu obrazovku.

Children:

Articles

Layout tvoriaci základ hlavnej obrazovky aplikácie.

Children: ArticlesSearch, TreeSvg, ArticlesList

/components

V priečinku */components* sa nachádzajú komponenty, z ktorých je zložená aplikácia.

ArticleDetailLeftDrawer

Children: ArticlesText

ArticleItem

Komponent zodpovedný za zobrazenie jednej položky zo zoznamu článkov.

Children:

ArticlesList

Komponent zodpovedný za zobrazenie zoznamu článkov.

Children: ArticleItem, ArticleDetailLeftDrawer

ArticlesSearch

Komponent zodpovedný za zobrazenie a funkcionality poľa pre vyhľadávanie článkov.

Children:

ArticlesText

Children:

DescriptionPanel

Children: Panel, PanelTabs

ImageGrid

Komponent zodpovedný za získanie a zobrazenie zoznamu obrázkov vybraného článku.

Children:

KnowledgeGraph

Komponent zodpovedný za získanie a zobrazenie tripletov vyextrahovaných z vybraného článku.

Children:

Navigation

Komponent predstavujúci horný panel v layoute celej aplikácie.

Children:

Panel

Children:

PanelTabs

Komponent zodpovedný za zobrazenie okna, ktoré sa otvorí po kliknutí na uzol (článok) v grafe.

Children: TabSelector, KnowledgeGraph, ImageGrid, WordCloud

ProtectedRoute

Children:

RadialTree

Komponent zodpovedajúci za zobrazenie a funkcionality radiálneho grafu.

Children:

TabSelector

Komponent zodpovedný za zobrazenie a funkcionality tlačidiel slúžiacich na prepínanie medzi tab-mi v komponente *PanelTabs*.

Children:

TreeSvg

Children: RadialTree, ArticleDetailLeftDrawer, DescriptionPanel

WordCloud

Komponent zodpovedný za zobrazenie wordcloud-u vybraného článku.

Children:

Export evidencie úloh

Export úloh zo systému Jira obsahujúci všetky doposiaľ absolvované Šprinty.

Šprint 1 - The one where it all began (07.10. 2021 – 18. 10. 2021)

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-4	Oboznámenie sa so systémom	13		Done	Done
Subtask	TEAM-7	Stretnutie s BigRing		Patrícia Huďanová	Done	Done
Subtask	TEAM-6	Prečítanie dokumentov z minulého roka		Martin Budinský	Done	Done
Subtask	TEAM-5	Pozretie nahrávok z minulého roka		Ivana Baleková	Done	Done
Story	TEAM-3	Príprava formulárov	3		Done	Done
Subtask	TEAM-11	Template pre oficiálny doc		Ivana Háberová	Done	Done
Subtask	TEAM-10	Template pre retrospektívu		Martina Billichová	Done	Done
Subtask	TEAM-9	Template pre zápisnice		Ivana Baleková	Done	Done
Story	TEAM-2	Slack- prepojenie s Jirou	3		Done	Done
Subtask	TEAM-8	Integrácia Jiri do Slacku		Miloš Galeta	Done	Done
Story	TEAM-1	Webové sídlo tímu	8		Done	Done
Subtask	TEAM-15	Nahrávanie a sťahovanie súborov		Martin Budinský	Done	Done
Subtask	TEAM-14	Definovanie obsahu stránky		Patrícia Huďanová	Done	Done
Subtask	TEAM-13	Vytvorenie statickej stránky		Ján Huďan	Done	Done
Subtask	TEAM-12	Rozbehovanie virtual machine		Ján Huďan	Done	Done

Šprint 2 - The one where we had call during holiday (18. 10. 2021 – 1. 11. 2021)

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-18	Stretnutie s BigRing-om	10		Done	Done
Subtask	TEAM-20	Stretnutie s BigRing-om		Patrícia Huďanová	Done	Done

Subtask	TEAM-19	Vytvorenie rozvrhu		Martina Billichová	Done	Done
Story	TEAM-17	Štúdium dokumentácie k platforme od BigRing	13		Done	Done
Subtask	TEAM-22	Štúdium dokumentácie k platforme		Ján Huďan	Done	Done
Subtask	TEAM-21	Vytvorenie dokumentu		Miloš Galeta	Done	Done
Story	TEAM-16	Vytvorenie prvého modelu	13		Done	Done
Subtask	TEAM-24	Vytvorenie zadania		Ivana Háberová	Done	Done
Subtask	TEAM-23	Vytvorenie modelu		Martin Budinský	Done	Done

Šprint 3 - The one with all the documentation (01. 11. 2021 – 15. 11. 2021)

Issue Type	Key	Summary	Story points	Assignee	Status	Resolution
Story	TEAM-26	Návrh a vytvorenie ďalšieho modelu	13	Ivana Baleková	Done	Done
Subtask	TEAM-32	Vytvorenie funkčného modelu		Ivana Baleková	Done	Done
Subtask	TEAM-31	Vytvorenie zadania		Ivana Háberová	Done	Done
Story	TEAM-25	Analýza podkladov(Lambda Life)	5	Patrícia Huďanová	Done	Done
Subtask	TEAM-30	Analýza podkladov od Lambda Life		Patrícia Huďanová	Done	Done
Story	TEAM-28	Stretnutie s BigRing-om	13	Martin Budinský	To Do	Unresolved
Subtask	TEAM-34	Stretnutie s BigRing-om		Martin Budinský	To Do	Unresolved
Story	TEAM-27	Štúdium gitbook dokumentácie	20	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-33	Tvorba dokumentu s flow-om		Ján Huďan	Done	Done
Subtask	TEAM-44	Tvorba dokumentu k časti procesy		Patrícia Huďanová	To Do	Unresolved

Šprint 4 - The one where we tried NLP (15. 11. 2021 – 29. 11. 2021)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-35	Implementácia prototypu na spracovanie štúdií	13		Done	Done

Subtask	TEAM-43	Analýza výsledkov Spustenie a otestovanie		Ján Huďan	Done	Done
Subtask	TEAM-42	NLP knižnic na dátach		Patricia Huďanová	Done	Done
Subtask	TEAM-41	Predspracovanie datasetu Získanie datasetu		Martin Budinský	Done	Done
Subtask	TEAM-40	klinických štúdií		Ivana Haberova	Done	Done
Story	TEAM-29	Analýza NLP	14		Done	Done
Subtask	TEAM-39	Analýza toolu ktorý vytvoril kinit		Miloš Galeta	Done	Done
Subtask	TEAM-38	Analýza NLP toolov pre český jazyk		Martina Bilichová	Done	Done
Subtask	TEAM-37	Analýza NLP toolov pre slovenský jazyk		Ján Huďan	Done	Done
Subtask	TEAM-36	Analýza NLP toolov pre anglický jazyk		Martin Budinský	Done	Done
Story	TEAM-28	Stretnutie s BigRing-om	13	Martin Budinský	To Do	Unresolved
Subtask	TEAM-34	Stretnutie s BigRing-om		Martin Budinský	To Do	Unresolved
Story	TEAM-27	Štúdium gitbook dokumentácie	20	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-33	Tvorba dokumentu s flow-om		Ján Huďan	Done	Done
Subtask	TEAM-44	Tvorba dokumentu k časti procesy		Patricia Huďanová	To Do	Unresolved

Šprint 5 - The one where we created our own path (29. 11. 2021 – 13. 12. 2021)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-49	Analýza frameworkov určených pre tvorbu vizualizácií a vytvorenie prototypu	13	Patricia Huďanová	Done	Done
Subtask	TEAM-60	Vytvoriť prototyp pre vizualizáciu		Patricia Huďanová	Done	Done
Subtask	TEAM-59	Analyzovať webové frameworky pre vizualizáciu		Ján Huďan	Done	Done
Story	TEAM-47	Existujúce služby a API	8	Ivana Baleková	Done	Done

Subtask	TEAM-58	Analyzovať druhú polovicu API prístupov z libraries.mid.edu		Miloš Galeta	Done	Done
Subtask	TEAM-57	Analyzovať prvú polovicu API prístupov z libraries.mid.edu		Ivana Baleková	Done	Done
Story	TEAM-46	Návrh aplikácie	15	Martina Billichová	Done	Done
Subtask	TEAM-56	Návrh UI		Martina Billichová	Done	Done
Subtask	TEAM-55	Identifikovanie scenárov		Ivana Baleková	Done	Done
Subtask	TEAM-54	Špecifikácia zadanie		Ivana Haberova	Done	Done
Story	TEAM-45	Analýza related work	8	Martin Budinský	Done	Done
Subtask	TEAM-53	Vytvorenie dokumentu pre analýzu výskumných prác		Ivana Haberova	Done	Done
Subtask	TEAM-52	Analýza výskumných prác		Ivana Haberova	Done	Done
Subtask	TEAM-51	Vytvorenie dokumentu pre komerčné nástroje		Martin Budinský	Done	Done
Subtask	TEAM-50	Analýza komerčných nástrojov		Martin Budinský	Done	Done

Šprint 6 - The one with new project (28.02.2022 – 07.03.2022)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-62	Prototyp UI	21	Patrícia Huďanová	Done	Done
Subtask	TEAM-69	Implementácia zoznamu článkov		Martin Budinský	Done	Done
Subtask	TEAM-68	Implementácia detailu článku		Martina Billichová	Done	Done
Subtask	TEAM-67	Implementácia prototypu		Patrícia Huďanová	Done	Done
Subtask	TEAM-66	Update mockupov		Ján Huďan	Done	Done
Story	TEAM-63	Analýza backend frameworkov	6	Ivana Haberova	Done	Done
Subtask	TEAM-65	Analýzovanie frameworkov		Ivana Haberova	Done	Done
Subtask	TEAM-64	Vytvorenie dokumentu		Ivana	Done	Done

				Haberova		
Story	TEAM-61	Prototyp využitia API	13	Ivana Baleková	In Progress	Unresolved
Subtask	TEAM-72	Inicializácia repozitára pre BE a FE		Miloš Galeta	Done	Done
Subtask	TEAM-71	Implementácia komunikácie s API serverom		Ivana Baleková	Done	Done
Subtask	TEAM-70	Vytvorenie modelu dát		Ján Huďan	In Progress	Unresolved

Šprint 7 - The one where we created base project (07.03.2022 – 21.03.2022)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-61	Prototyp využitia API	13	Ivana Baleková	Done	Done
Subtask	TEAM-72	Inicializácia repozitára pre BE a FE		Miloš Galeta	Done	Done
Subtask	TEAM-71	Implementácia komunikácie s API serverom		Ivana Baleková	Done	Done
Subtask	TEAM-70	Vytvorenie modelu dát		Ján Huďan	Done	Done
Story	TEAM-76	Analýza možností zobrazenia detailu článku v html	11	Ján Huďan	Done	Done
Subtask	TEAM-85	Implementácia prototypu		Ján Huďan	Done	Done
Subtask	TEAM-84	Vytvoriť dokument s analýzou		Ján Huďan	Done	Done
Story	TEAM-75	Prepojenie BE a FE - prototyp 1	20	Patrícia Huďanová	Done	Done
Subtask	TEAM-83	Prepojenie zoznamu článkov na FE s BE		Martin Budinský	Done	Done
Subtask	TEAM-82	Implementácia searchu na		Ivana	Done	Done

		BE			Baleková		
Subtask	TEAM-81	Implementácia searchu na FE			Patrícia Huďanová	Done	Done
Story	TEAM-74	Prototyp mind mapy z free textu	15		Ivana Haberova	Done	Done
Subtask	TEAM-80	Update mockupov			Patrícia Huďanová	Done	Done
Subtask	TEAM-79	Implementácia prototypu mind mapy z free textu			Ivana Haberova	Done	Done
Story	TEAM-73	Analýza možností vytvárania štruktúr z free textu	13		Martina Billichová	Done	Done
Subtask	TEAM-78	Analyzovanie možností vytvárania štruktúr z free textu			Miloš Galeta	Done	Done
Subtask	TEAM-77	Vytvorenie dokumentu			Martina Billichová	Done	Done

Šprint 8 - The one where server fell (21.03.2022 – 04.04.2022)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-89	Vytvorenie knowledge grafu na základe semantic triplets	5	Ivana Haberova	Done	Done
Subtask	TEAM-97	Prototyp so zobrazením tripletov v knowledge grafe		Ivana Haberova	Done	Done
Story	TEAM-88	Analýza toolov na semantic triplets	21	Ivana Haberova	Done	Done
Subtask	TEAM-96	Otestovanie jednotlivých toolov		Martin Budinský	Done	Done
Subtask	TEAM-95	Analýza toolov		Ivana Haberova	Done	Done
Story	TEAM-87	Analýza možností	21	Ivana	Done	Done

		zobrazenia detailu Článku v pdf/xml a navigácia vrámci neho		Baleková		
Subtask	TEAM-94	Analýza zobrazenia cez pdf		Miloš Galeta	Done	Done
Subtask	TEAM-93	Analýza zobrazenia s pomocou XML		Ivana Baleková	Done	Done
Story	TEAM-86	Zefektívnenie vizualizácie	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-92	Doplnenie API na BE		Ivana Baleková	Done	Done
Subtask	TEAM-91	Prerobenie stromu na radiálny graf		Patrícia Huďanová	Done	Done
Story	TEAM-90	Nasadenie aplikácie	13	Ján Huďan	In Progress	Unresolved
Subtask	TEAM-99	Nasadenie aplikácie na virtuálku		Ján Huďan	In Progress	Unresolved
Subtask	TEAM-98	Úprava dockera		Ján Huďan	In Progress	Unresolved

Šprint 9 - The one with final analysis (04.04.2022 – 25.04.2022)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-100	Prototypy s rôznymi možnosťami zobrazenia detailu článku	17	Miloš Galeta	Done	Done
Subtask	TEAM-110	Implementovanie prototypu pre html		Ivana Baleková	Done	Done
Subtask	TEAM-109	Implementovanie prototypu pre xml		Ivana Baleková	Done	Done
Subtask	TEAM-108	Implementovanie prototypu pre pdf		Miloš Galeta	Done	Done
Story	TEAM-101	Zefektívnenie vizualizácie	13	Martina Billichová	Done	Done

Subtask	TEAM-111	Prerobenie detailu článku		Martina Billichová	Done	Done
Subtask	TEAM-107	Doplnenie grafu ako listu do radialneho stromu		Patrícia Huďanová	Done	Done
Subtask	TEAM-106	Vytvorenie dokumentu s návrhom		Ivana Haberova	Done	Done
Story	TEAM-102	Testovanie rôznych knižníc na semantic triplets	13	Martin Budinský	Done	Done
Subtask	TEAM-105	Implementovanie prototypu		Martin Budinský	Done	Done
Story	TEAM-90	Nasadenie aplikácie	13	Ján Huďan	Done	Done
Subtask	TEAM-99	Nasadenie aplikácie na virtuálku		Ján Huďan	Done	Done
Subtask	TEAM-98	Úprava dockera		Ján Huďan	Done	Done

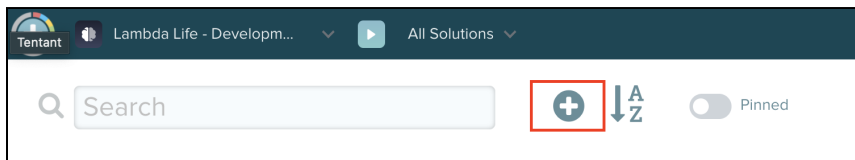
Šprint 10 - The one where we finished (25.04.2022 – 09.05.2022)

Issue Type	Key	Summary	Story point estimate	Assignee	Status	Resolution
Story	TEAM-115	Analýza, návrh a implementácia filtrovania tripletov	20	Martin Budinský	Done	Done
Subtask	TEAM-125	Implementovanie do prototypu		Martin Budinský	Done	Done
Subtask	TEAM-124	Analýza a návrh filtrovania tripletov		Martin Budinský	Done	Done
Story	TEAM-113	Pridať prihlasovanie	13	Miloš Galeta	Done	Done
Subtask	TEAM-123	Autorizácia na BE		Miloš Galeta	Done	Done
Subtask	TEAM-122	Autorizácia na FE		Martina Billichová	Done	Done
Subtask	TEAM-121	Overovanie Ldap		Ivana Baleková	Done	Done

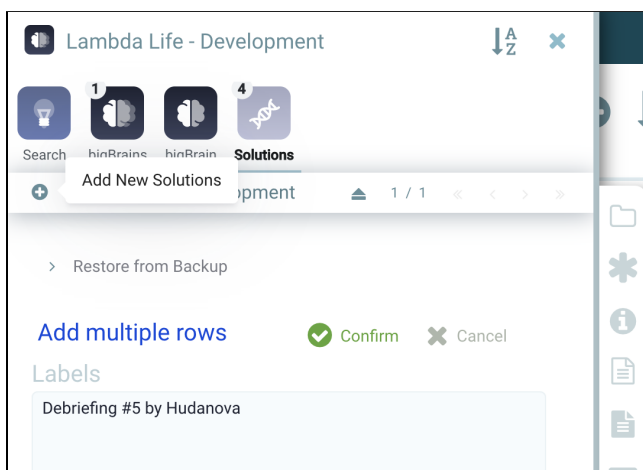
Story	TEAM-112	Pridať taby do popupu	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-120	Prepojenie s BE		Miloš Galeta	Done	Done
Subtask	TEAM-119	Vytvorenie tabu s wordcloudom		Patrícia Huďanová	Done	Done
Subtask	TEAM-118	Vytvorenie tabu s obrázkami		Patrícia Huďanová	Done	Done
Story	TEAM-103	Scenár testovania pre druhý tím + Automatické testy	13	Patrícia Huďanová	Done	Done
Subtask	TEAM-117	Vytvorenie scenárov pre testovanie		Ivana Haberova	Done	Done
Subtask	TEAM-116	Automatické testy		Ján Huďan	Done	Done
Story	TEAM-114	Integrácia prototypu html rest	10	Ivana Baleková	Done	Done

Príloha A - Používateľská príručka pre platformu BigRing

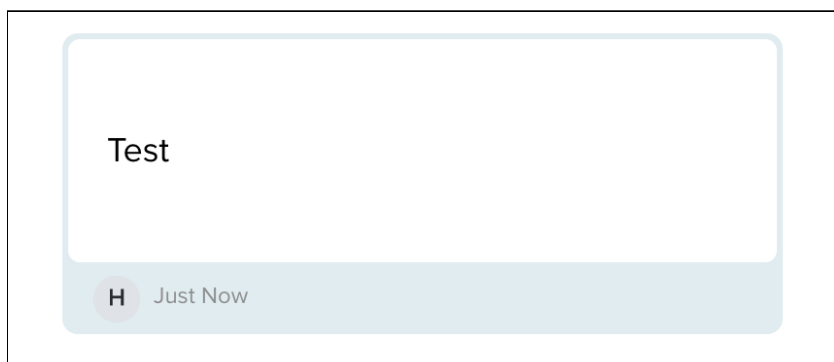
Po prihlásení do aplikácie je potrebné vytvoriť nový solution:



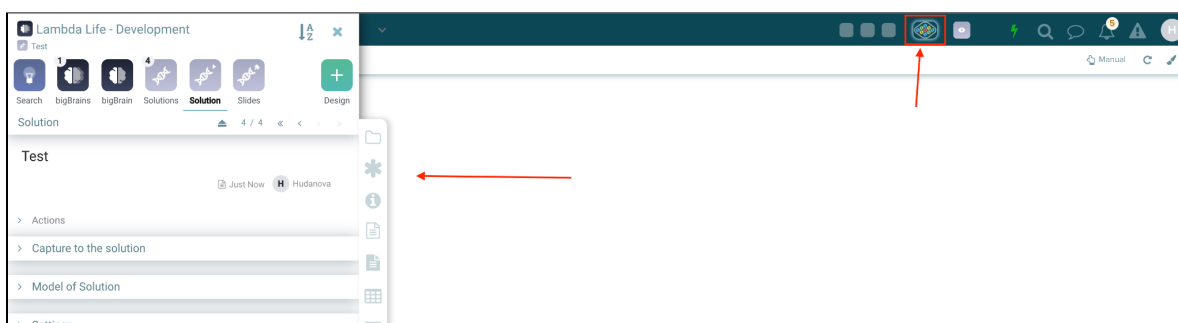
Solution sme nazvali "Test" - názov sa zadáva do textového poľa "Labels":



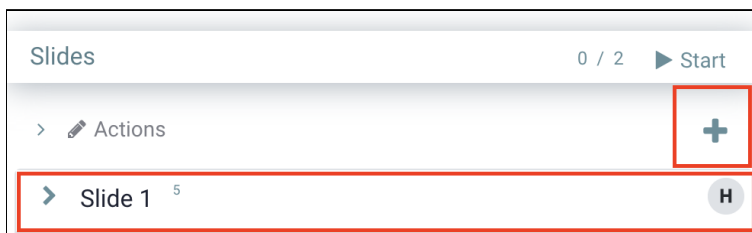
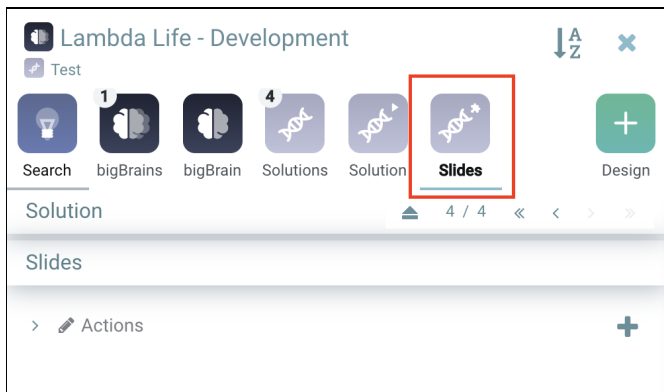
Po vytvorení nového solution sa zobrazil na ploche:



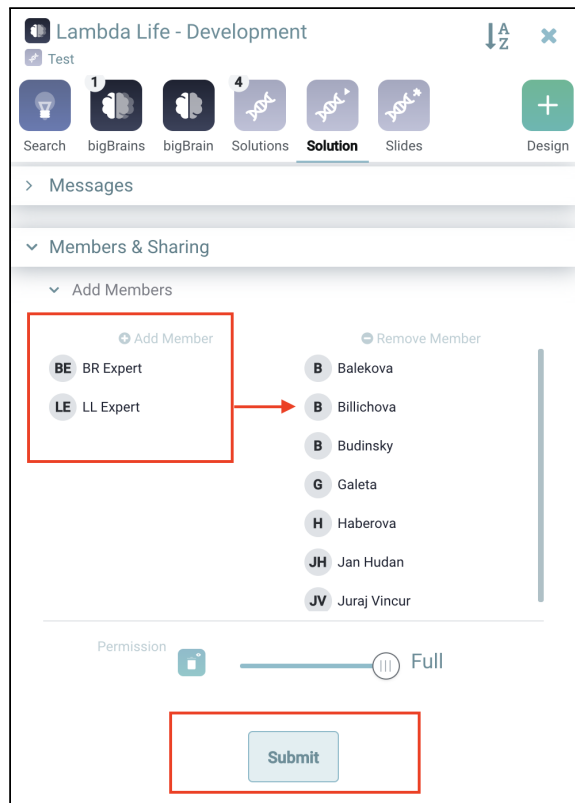
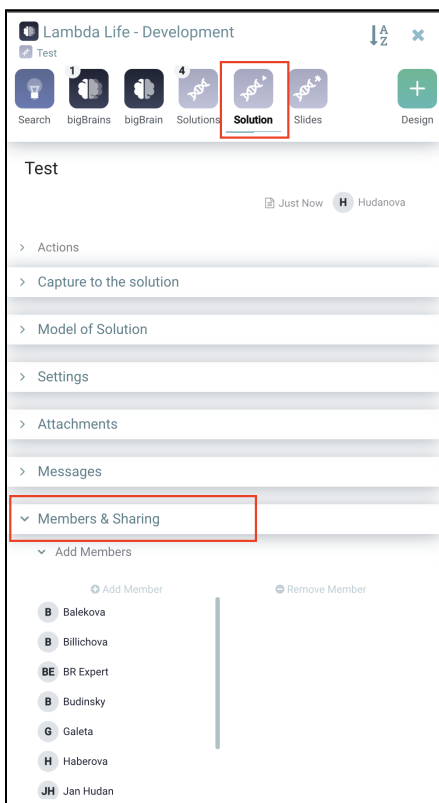
Cez "Toggle" tlačidlo sme sa dostali k editoru.



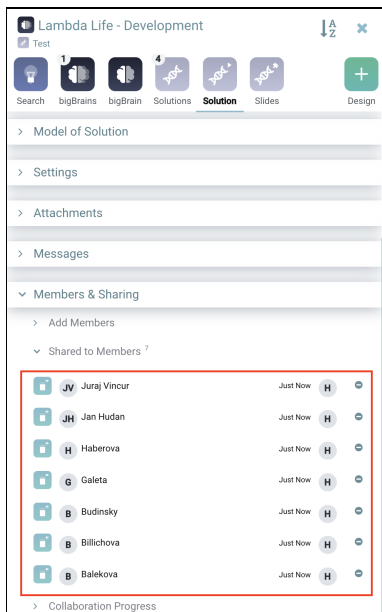
V tabe "Slides" si vytvoríme kliknutím na tlačidlo s ikonou plus nový slide.



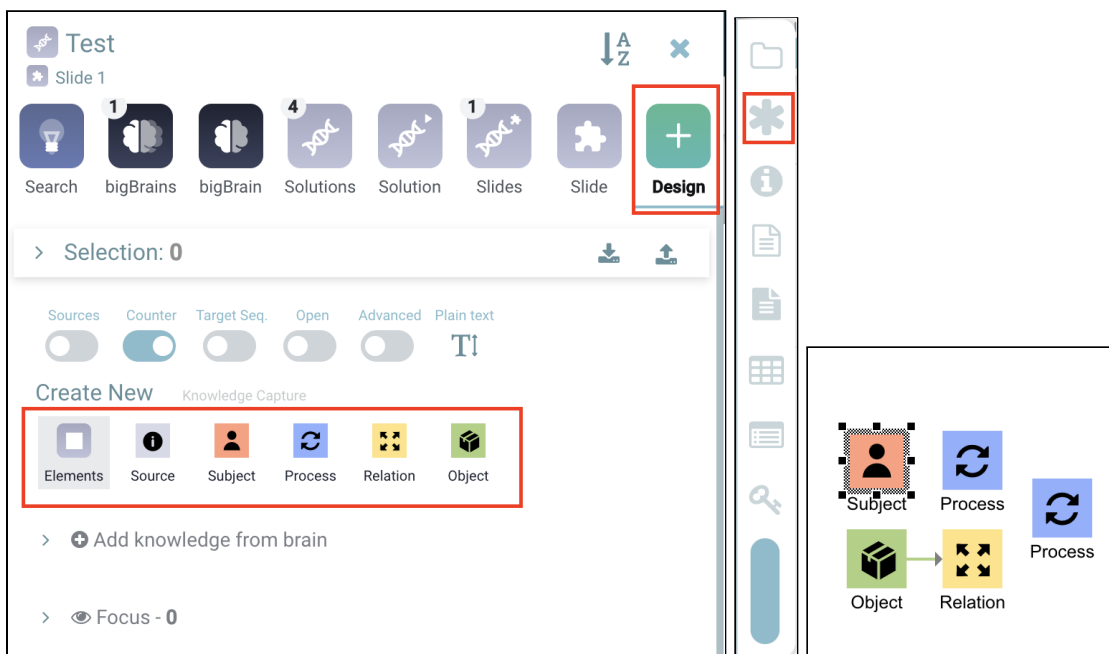
Aby mohli solution upravovať aj ostatní členovia tímu, je potrebné ho s nimi zdieľať:



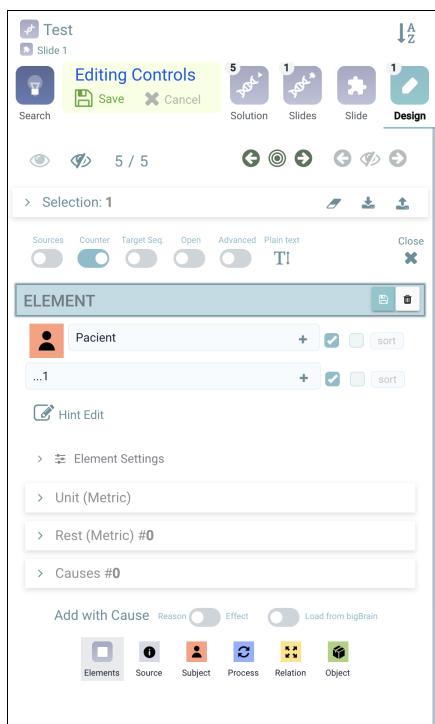
Po pridaní sa členovia zobrazia v tabuľke:



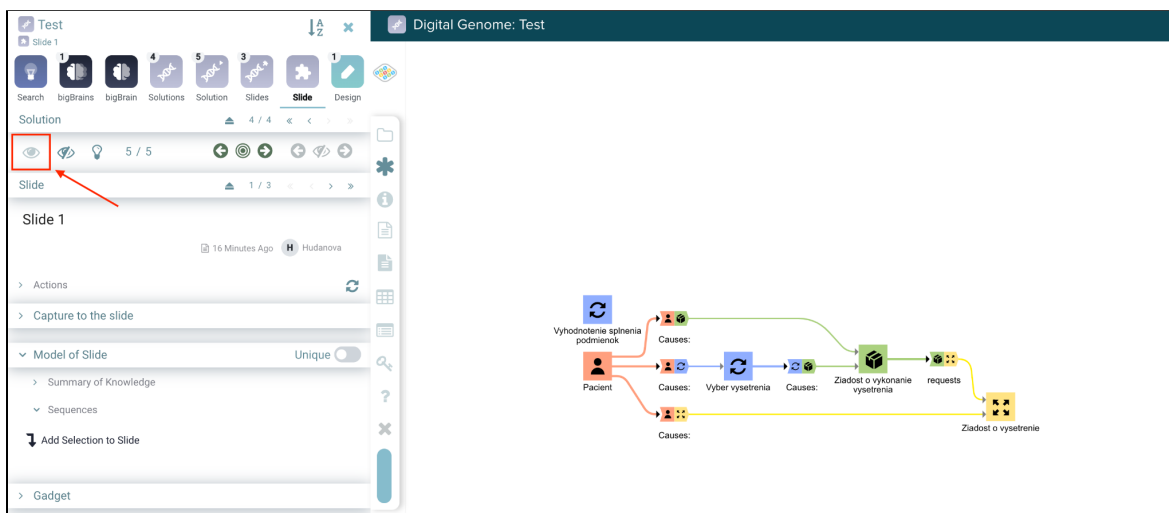
Po prekliknutí sa do tabu "Design" vieme pridávať elementy do nášho grafu- drag and drop (graf zobrazíme kliknutím na hviezdičku)



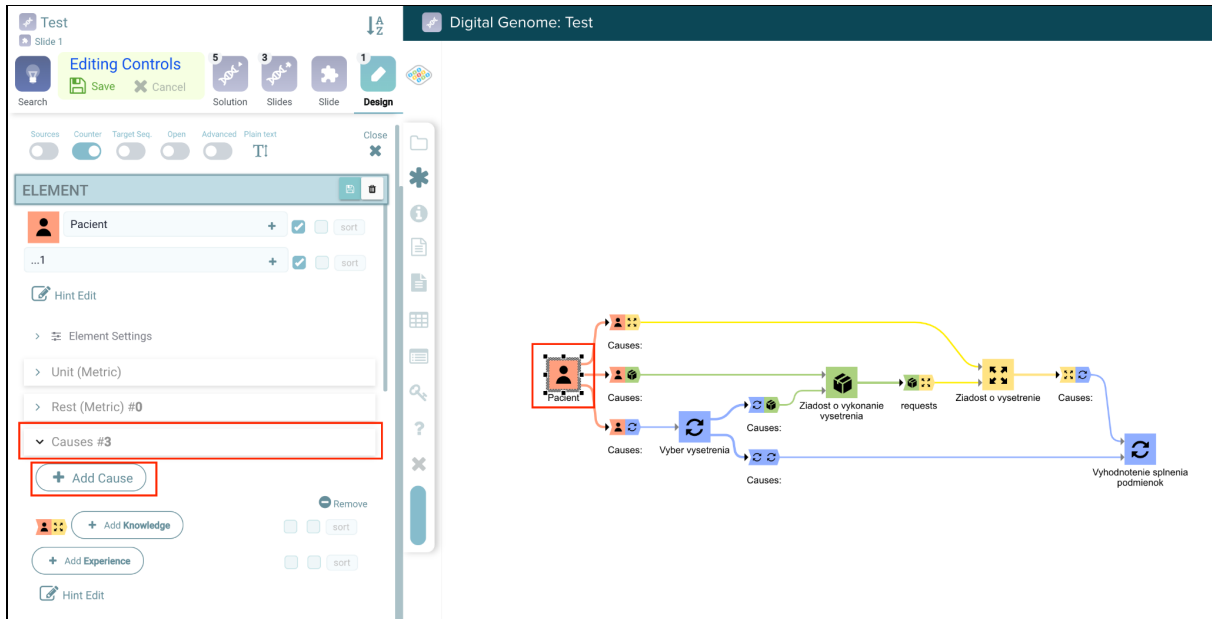
Jednotlivé elementy pomenujeme.



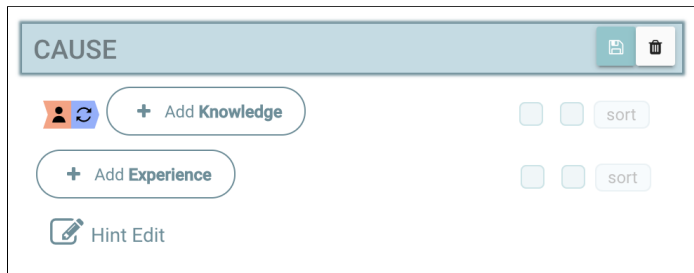
Aby sme videli vytvorený model, v "Slide" je potrebné zakliknúť symbol oka:



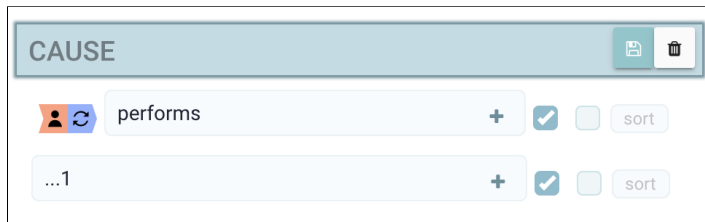
Vzťahy medzi elementami/causes (šípky) pridáme kliknutím na element, cez "+ add Causes", tam vzťahy pomenujeme a v časti "Effects" pridáme elementy predstavujúce začiatok a koniec hrany



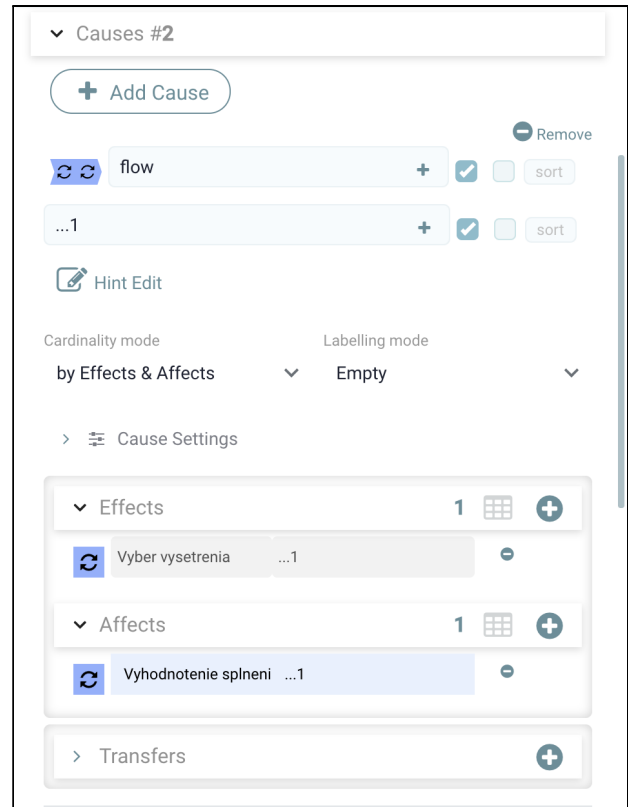
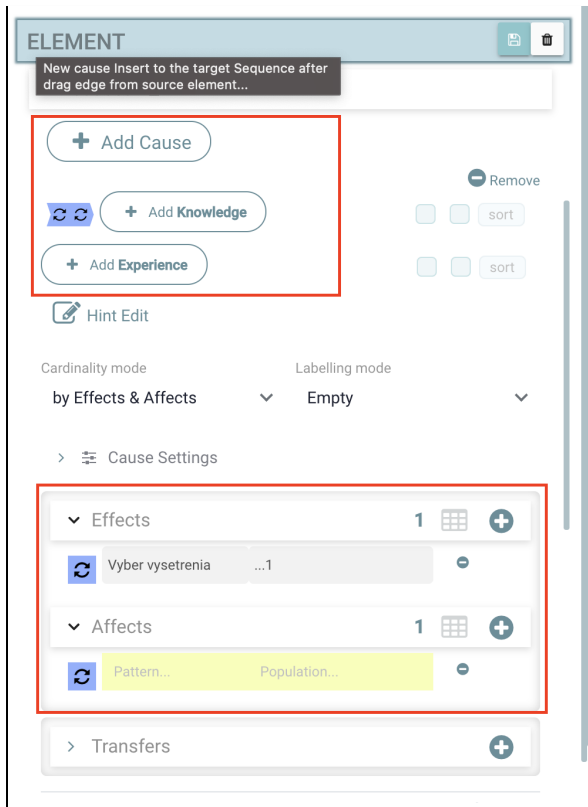
Klikneme na “+ Add Knowledge”:



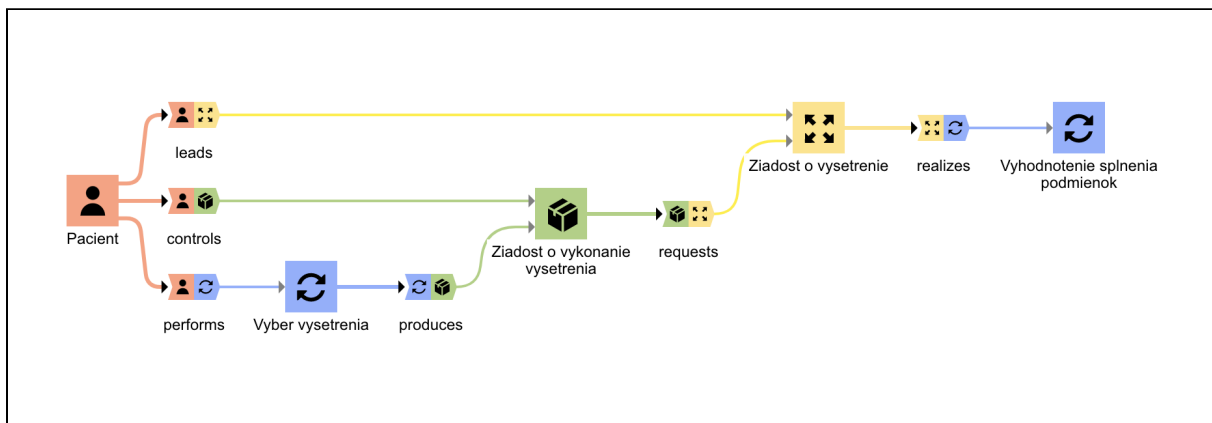
Zadáme názov vzťahu:



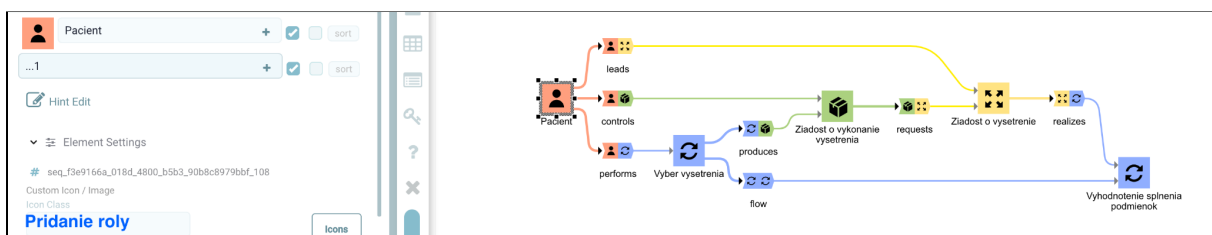
V časti “Effects” pridáme názvy elementov medzi ktorými je pridávaný vzťah a stlačíme tlačidlo uloženia v modrom riadku:



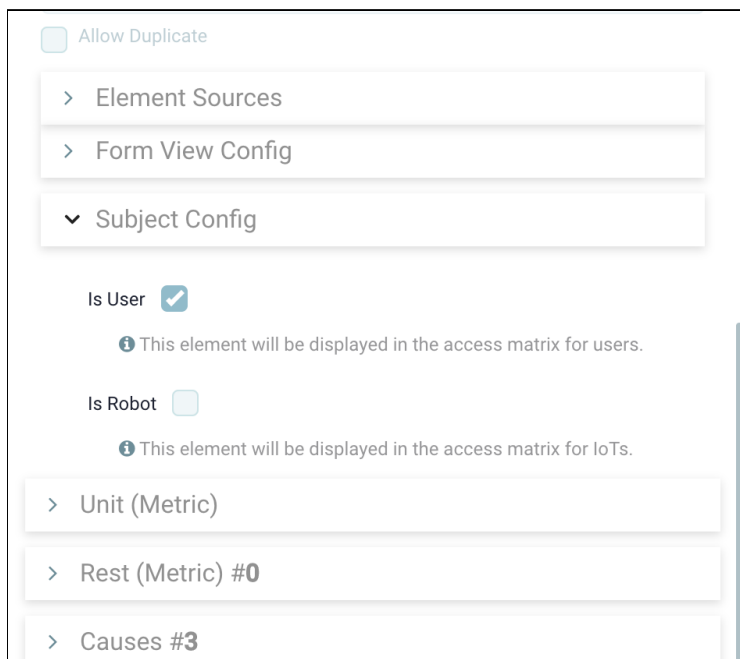
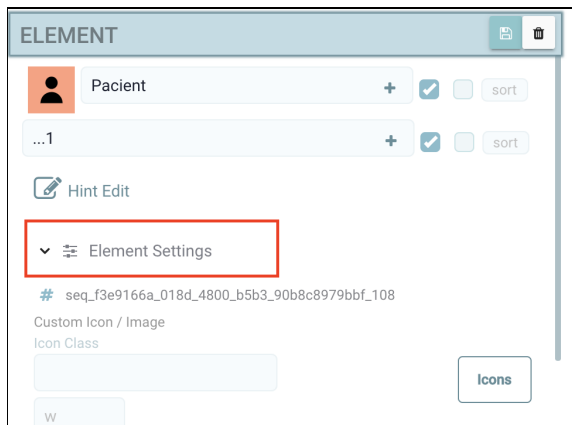
Vzťahy v modeli sú pomenované.



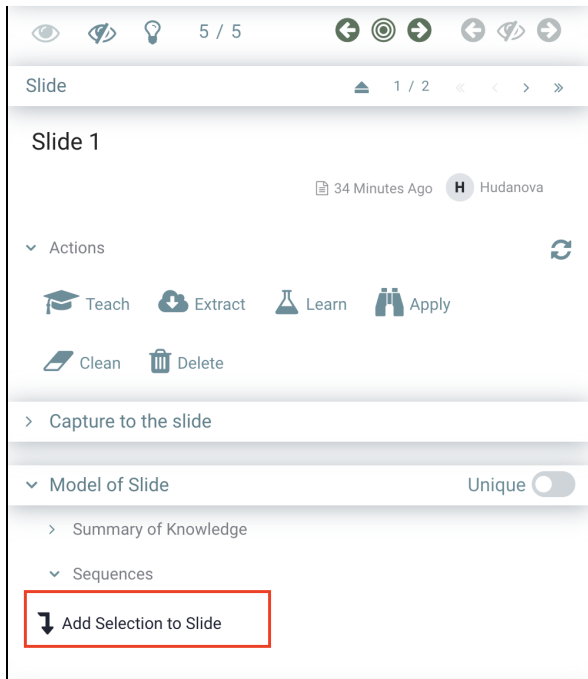
Na pridanie roly pacienta je potrebné najprv v design editore označiť tento element



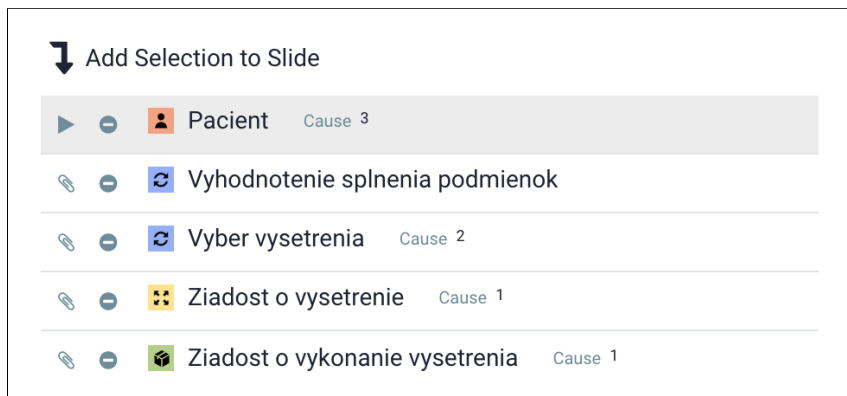
Po rozkliknutí možnosti „Element Settings” a “Subject Config” je potrebné označiť možnosť „Is User”



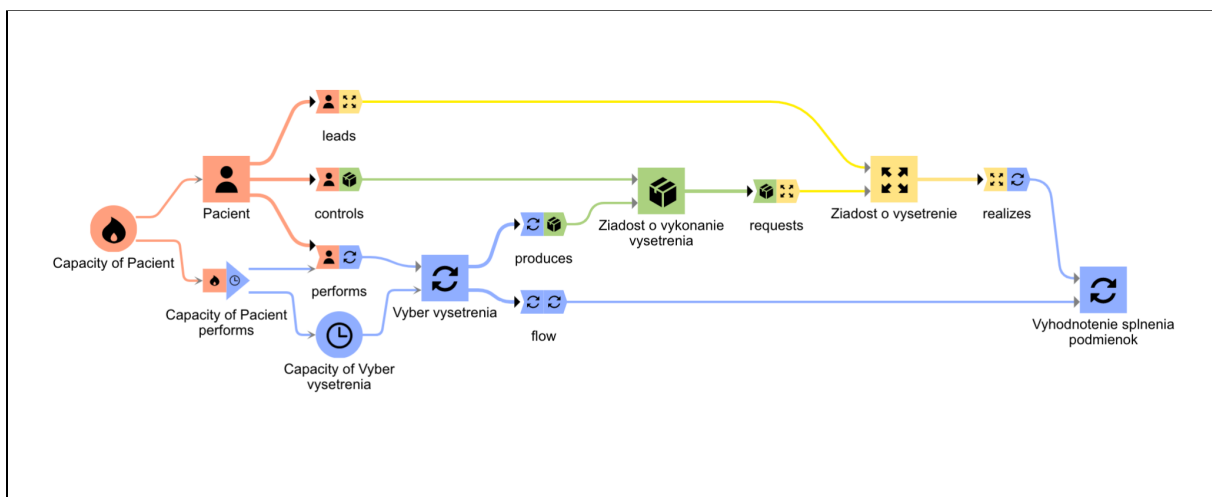
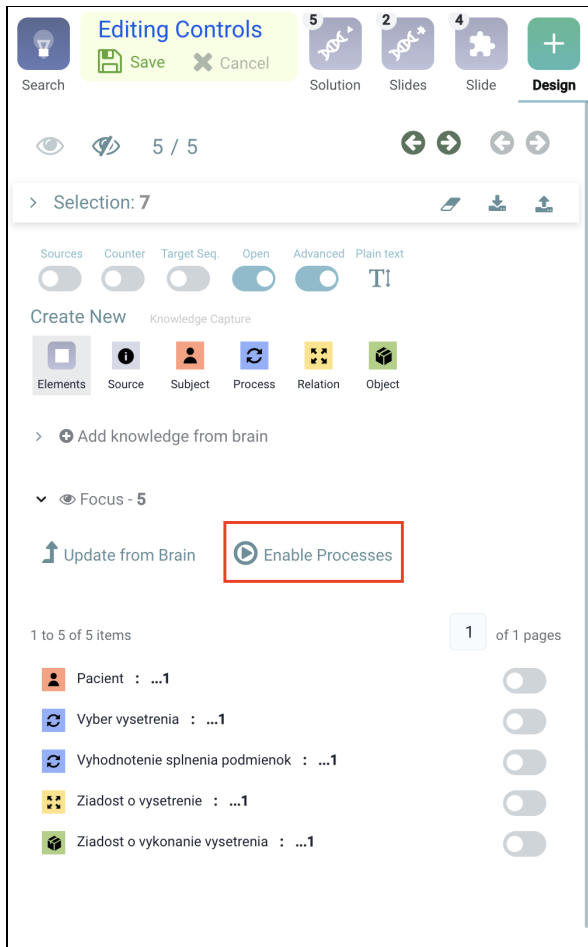
Následne v tabe “Slides”, po rozkliknutí nášho Slide-u vyberieme “Model of Slide”, označíme si celý náš graf a klikneme “Add Selection to Slide”.



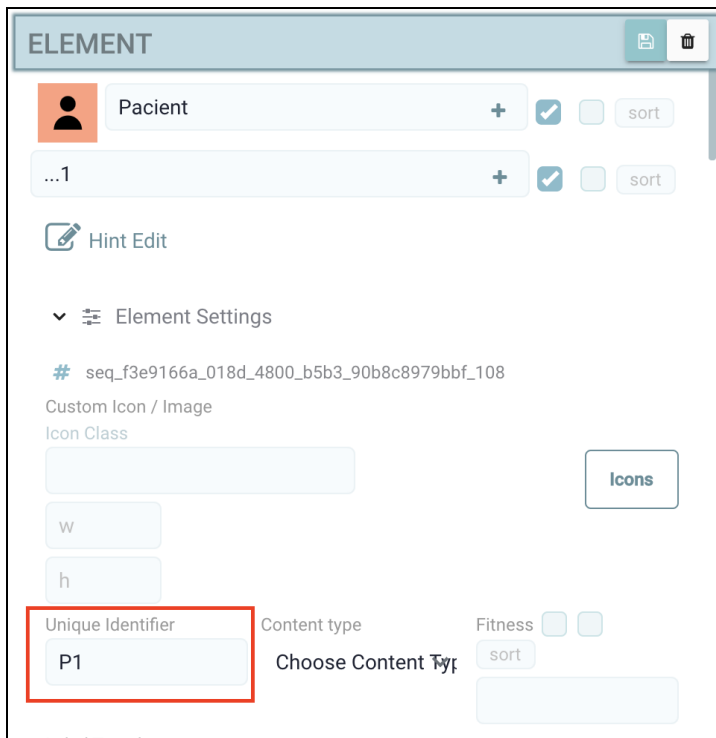
To nám pridá všetky naše elementy do Slide-u. Pri našom "Subjecte" následne klikneme na ikonu, ktorá z neho spraví Primárny element (ikona sa zmení na ikonu play).



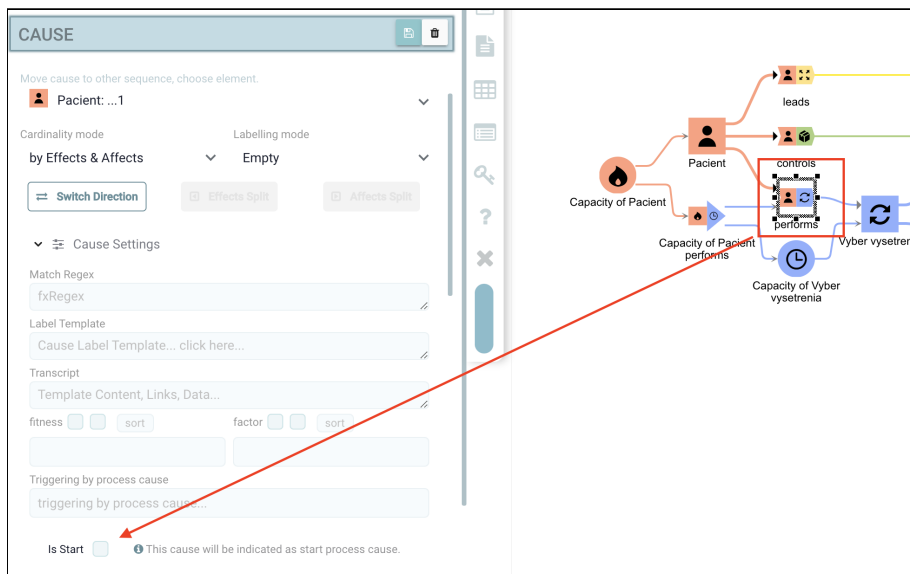
Následne v tabe "Design" rozklikneme "Focus". Klikneme na " Enable Processes". Toto nám do modelu vygeneruje ďalšie elementy.



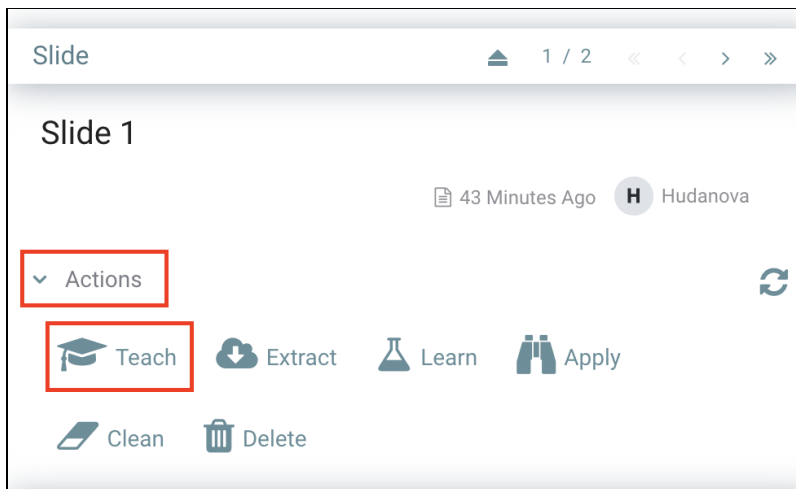
Nášmu hlavnému subjectu pridáme identifikátor - v tabe "Design" keď si v modeli rozklikneme element.



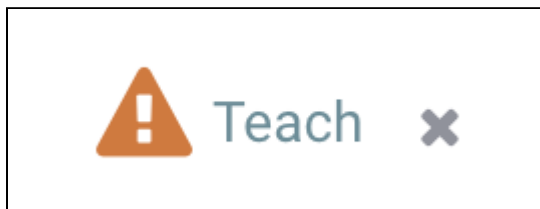
Na určenie začiatočného procesu takisto musíme nastaviť vzťah Is start.



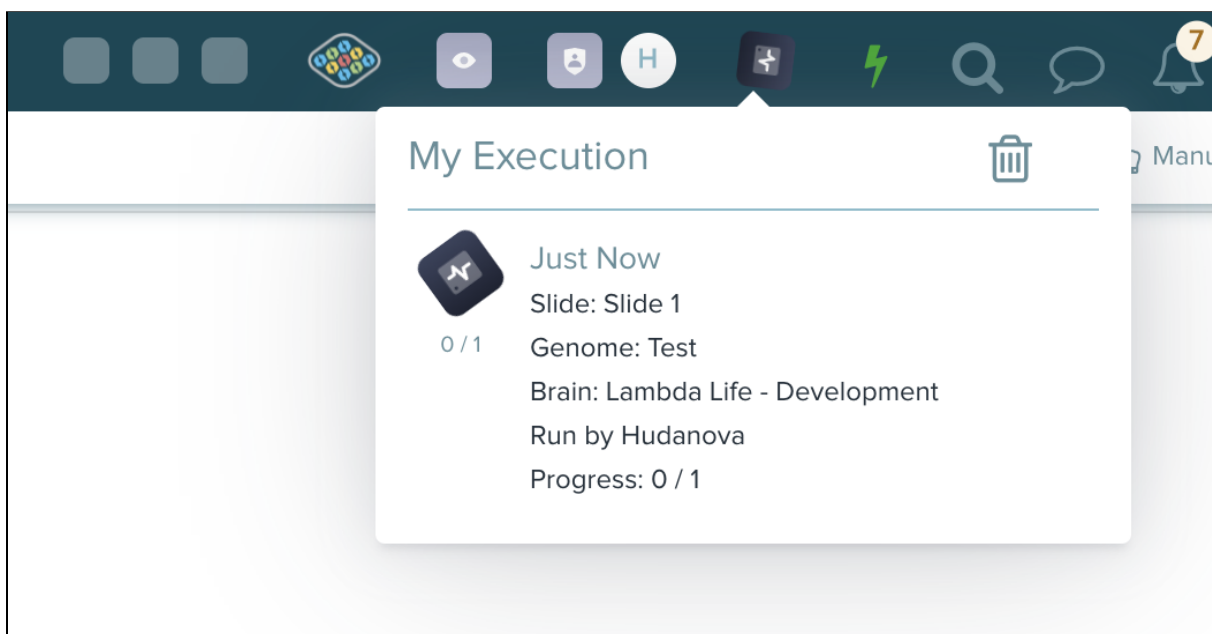
Na naučenie brainu je potrebné v Slide tabe rozkliknúť možnosť „Actions” a následne zvoliť možnosť “Teach”.



Po stlačení tlačidla “Teach” je potrebné ho stlačiť znovu (potvrdenie):



Po zavretí ľavého editovacieho panelu sa v hornej lište objaví ikona s prebiehajúcimi procesmi.

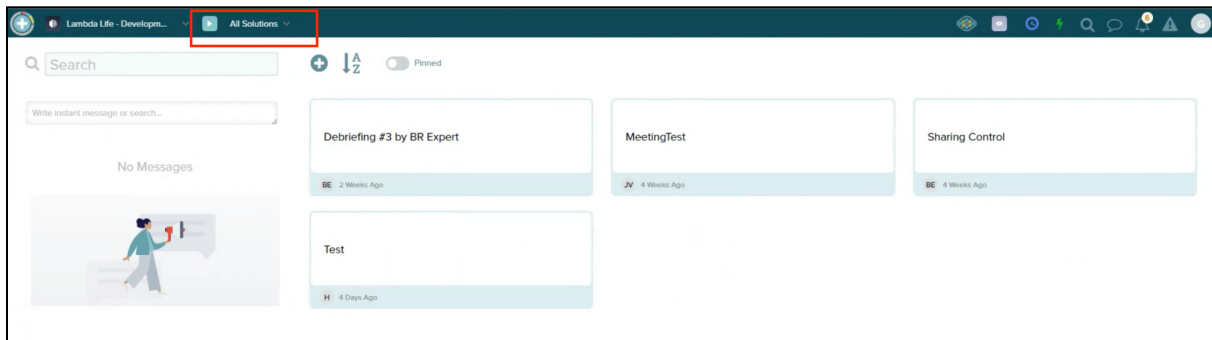


Kliknutím na ikonu kľúča sa po Teachnutí zobrazí zoznam rolí používateľov:

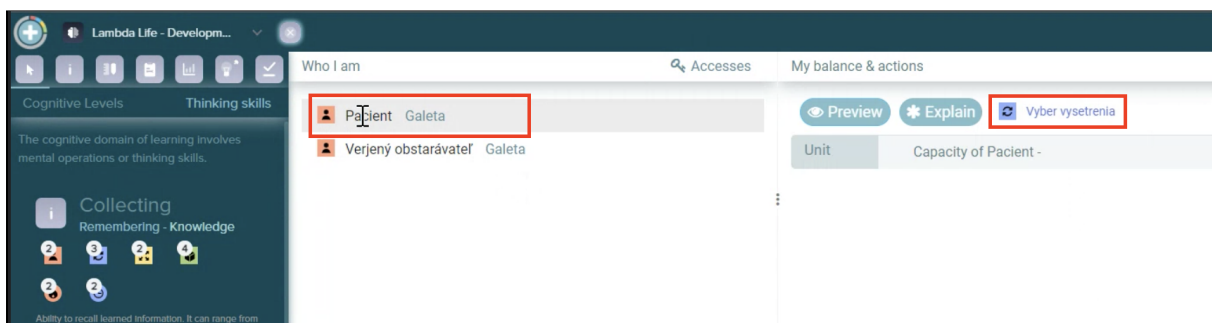
Access Matrix		Users ID	Pacient	Verjený obstará...
Users / Roles				
Balekova	B	109	✓	✓
Billichova	B	112	✓	✓
BR Expert	BE	115	✓	✓
Budinsky	B	113	✓	✓
Galeta	G	110	✓	✓
Haberova	H	111	✓	✓
Hudanova	H	108	✓	✓
Jan Hudan	JH	114	✓	✓
Juraj Vincur	JV	107	✓	✓

OVERENIE FUNKČNOSTI PROTOTYPU

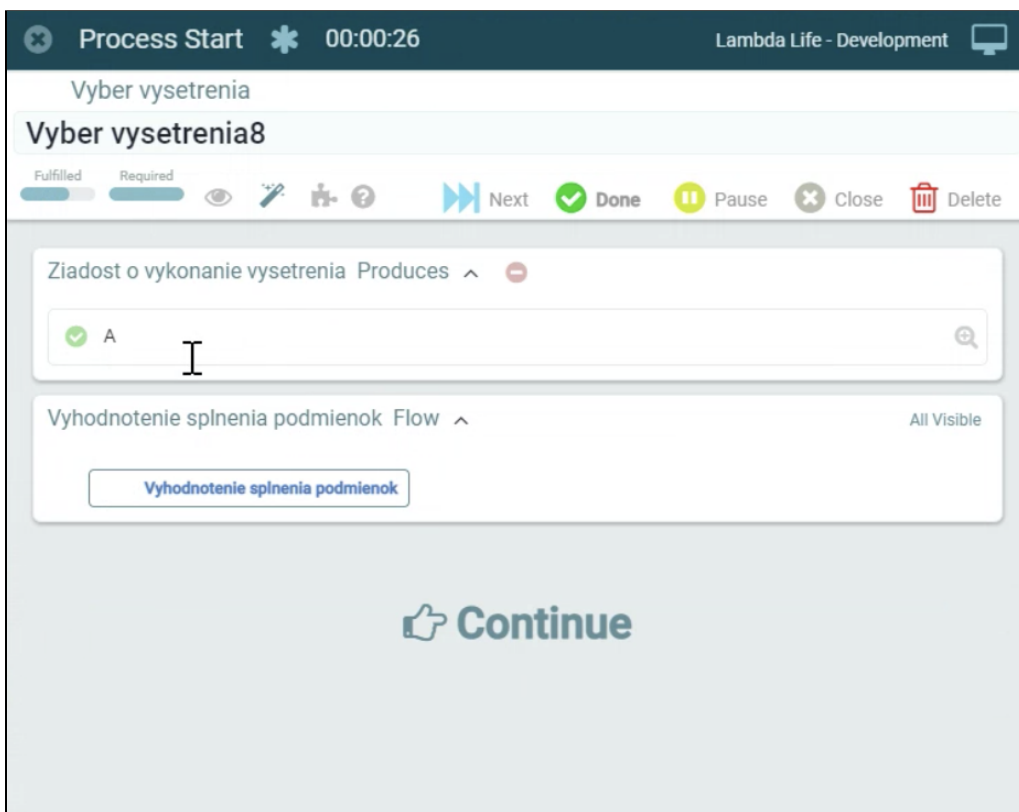
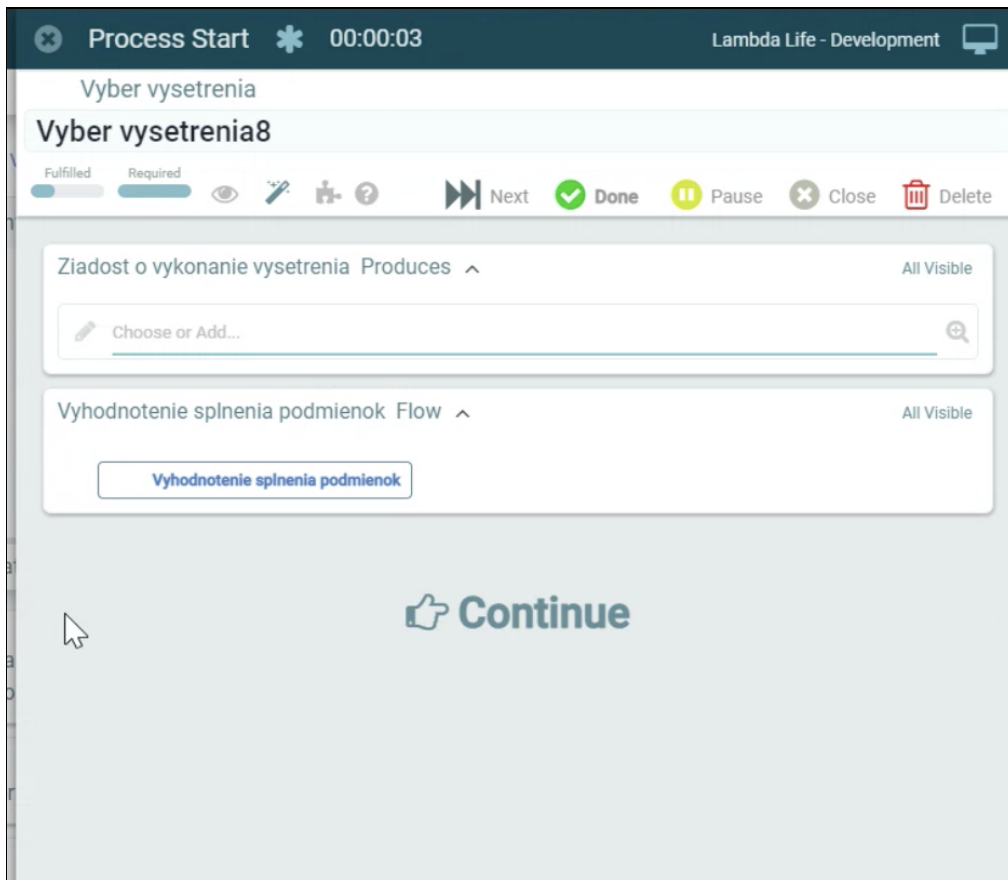
V hornom menu v dropdowne pre výber solutions vyberieme All solutions a klikneme na button s ikonkou play po ľavej strane- Toto nás presmeruje do okna s procesmi.



Následne klikneme na našu rolu- u nás Pacient a potom na proces ktorý sme vytvárali a chceme overiť.



Zobrazí sa okno v ktorom vyplníme polia a následne klikneme na continue.



Keď proces zbehne správne zobrazí sa nám okienko s Done v pravom hornom rohu.

Preview & Browse Lambda Life - Development

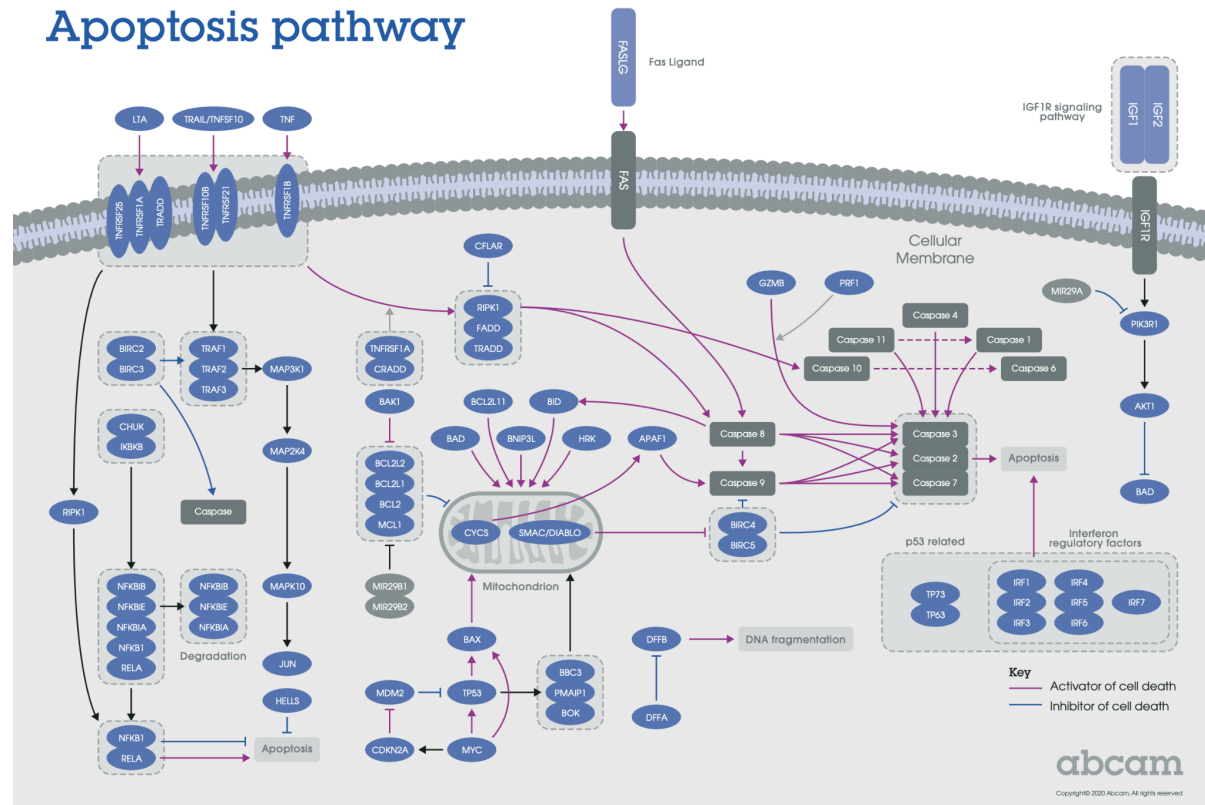
Vyber vyšetrenia ✔ Done

Vyber vyšetrenia9

Completeness Optimum 👁️ 🔧 ? ▶ Run

Unit	Capacity of Vyber vyšetrenia 15/11/2021 1...	0.00
Pacient Performs		
> Galeta		
Capacity of Pacient From 15/11/2021 03:29:49 To 15/11/2021 03:29:54 dur. 0 h.		
Ziadost o vykonanie vyšetrenia Produces		
> A		
Vyhodnotenie splnenia podmienok Flow		
> Messenger		

Príloha B - Analýza medicínskej domény podľa obrázkov



<https://www.cusabio.com/pathway/Apoptosis.html>

Apoptóza je mechanizmus na selektívne odstránenie starnúcich, poškodených alebo inak nežiaducich buniek = programovaná bunková smrť.

<https://en.wikipedia.org/wiki/Apoptosis>

Apoptóza

- nemôže sa zastaviť, keď už začala => vysoko regulovaný proces
- možno spustiť jednou z dvoch ciest
 - vnútorná dráha bunky sa zabíja, pretože cíti bunkový stres,
 - vonkajšia cesta bunka sa zabíja kvôli signálom z iných buniek.

Nedostatočná apoptóza vedie k nekontrolovanej bunkovej proliferácii, ako je rakovina.

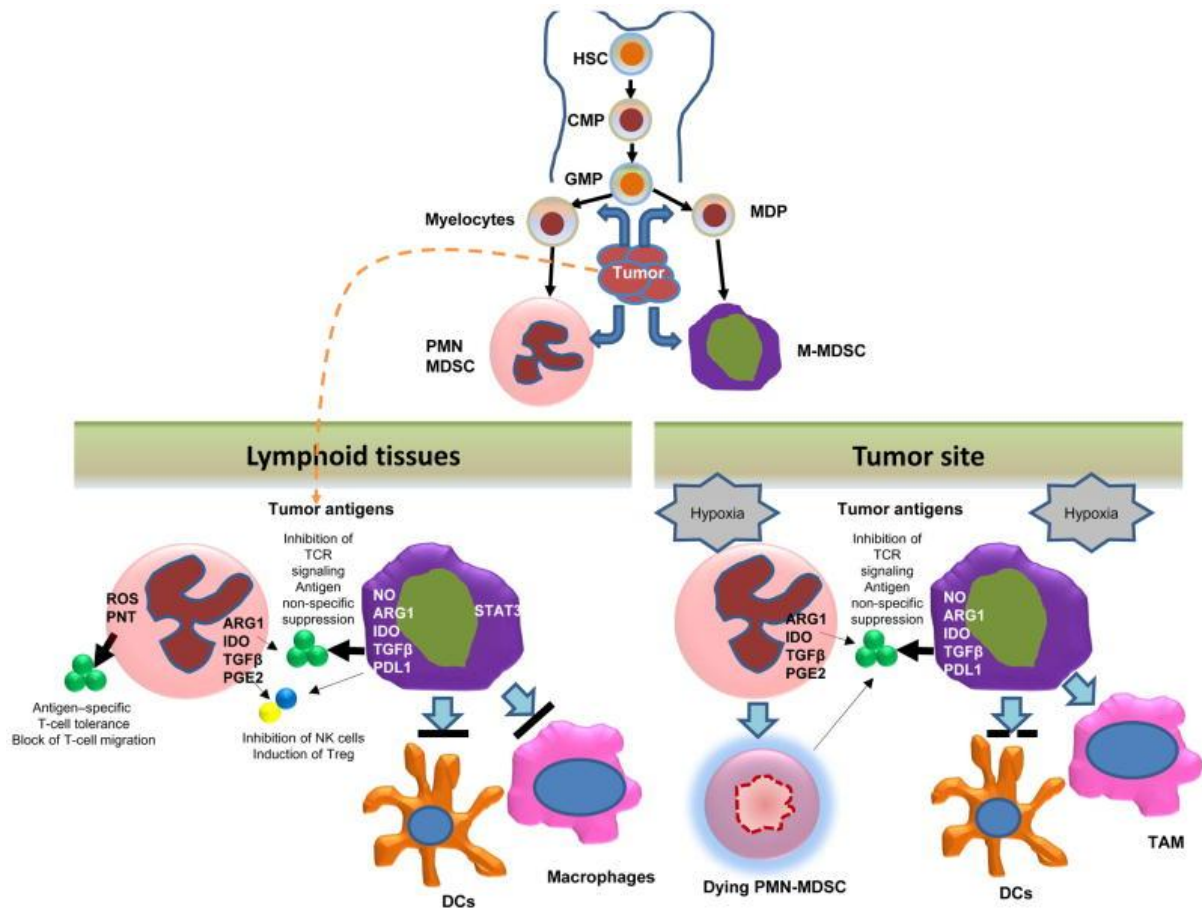
https://www.youtube.com/watch?v=SyvOPXeg4ig&ab_channel=Genentech

-krátky popis toho, ako je tento bunkový proces regulovaný prostredníctvom vnútorných a vonkajších dráh buniek.

Apoptóza sa zásadne podieľa na regulácii tvorby nádorov a tiež kriticky určuje odpoveď aj na jej liečbu.

Články k apoptóze pri protirakovinovej chemoterapii:

<https://www.nature.com/articles/1209608>



Lymphoid tissues - organized structures that support immune responses

<https://www.nature.com/subjects/lymphoid-tissues#:~:text=Definition,examples%20of%20secondary%20lymphoid%20tissue.>

Tumor site - The site refers to the location of the cancer within the body

<https://training.seer.cancer.gov/disease/diagnosis/>

Tumor antigens - Tumor antigen is an antigenic substance produced in tumor cells, i.e., it triggers an immune response in the host.

https://en.wikipedia.org/wiki/Tumor_antigen

Hypoxia - Hypoxia is a state in which oxygen is not available in sufficient amounts at the tissue level

<https://www.ncbi.nlm.nih.gov/books/NBK482316/>

Macrophage - Macrophages are specialised cells involved in the detection, phagocytosis and destruction of bacteria and other harmful organisms.

<https://www.immunology.org/public-information/bitesized-immunology/cells/macrophages>

DCs (Dendritic cells) - Dendritic cells are a type of antigen-presenting cell (APC) that form an important role in the adaptive immune system. The main function of dendritic cells is to present [antigens](#) and the cells are therefore sometimes referred to as “professional” APCs.

<https://www.news-medical.net/health/What-are-Dendritic-Cells.aspx>

TAM - Tumor-associated macrophages (TAMs) are a class of immune cells present in high numbers in the microenvironment of solid tumors

https://en.wikipedia.org/wiki/Tumor-associated_macrophage

TCR signalling - A TCR signal causes global cellular changes within the T cell...

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5801066/>

Antigen - An antigen is a molecule that stimulates an immune response by activating leukocytes (white blood cells) that fight disease. Antigens may be present on invaders, such as bacteria, viruses, parasites, fungi, and transplanted organs, or on abnormal cells, such as cancer cells.

<https://www.verywellhealth.com/what-is-an-antigen-5083801>

T-cell tolerance - T cell-mediated immune tolerance is a state of unresponsiveness of T cells towards specific self or non-self antigens

<https://www.frontiersin.org/articles/10.3389/fimmu.2020.576261/full>

Block of T-cell migration - T cell migration is essential for T cell responses; it allows for the detection of cognate antigen at the surface of antigen-presenting cells and for interactions with other cells involved in the immune response.

<https://www.nature.com/articles/nri.2015.16>

NK cells - Natural killer (NK) cells are effector lymphocytes of the innate immune system that control several types of tumors and microbial infections by limiting their spread and subsequent tissue damage.

<https://www.nature.com/articles/ni1582>

Treg - Regulatory T cells (Tregs) are a specialized subpopulation of T cells that act to suppress immune response, thereby maintaining homeostasis and self-tolerance.

[https://pubmed.ncbi.nlm.nih.gov/20672742/#:~:text=Regulatory%20T%20cells%20\(Tregs\)%20are,critical%20role%20in%20preventing%20autoimmunity.](https://pubmed.ncbi.nlm.nih.gov/20672742/#:~:text=Regulatory%20T%20cells%20(Tregs)%20are,critical%20role%20in%20preventing%20autoimmunity.)

PMN-MDSC - PMN-MDSC emerged as critical negative regulator of immune responses under many pathologic conditions and major partner of mesenchymal cells in promotion of tumor metastases.

<https://www.sciencedirect.com/science/article/pii/S1044532317300076>

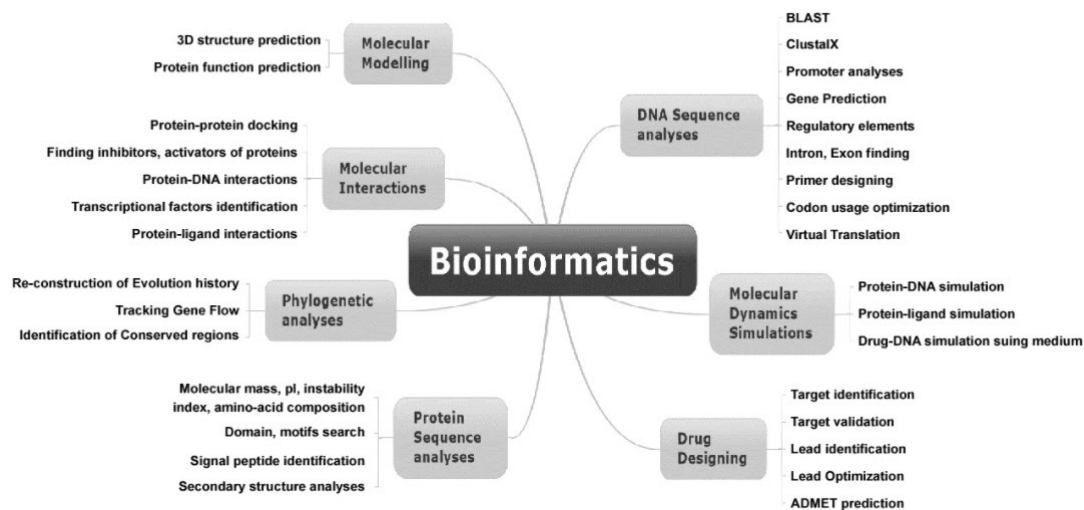


Figure 1: Application of bioinformatics tools in various areas of biological sciences.

Figure 1: Application of bioinformatics tools in various areas of biological sciences.

Published in 2014

Use of Bioinformatics Tools in Different Spheres of Life Sciences

M. Mehmood, Ujala Sehar, N. Ahmad



Molecular modeling

- <https://www.sciencedirect.com/topics/chemistry/molecular-modeling>
- veda o numerickej reprezentácii molekulárnych štruktúr a simulácii ich správania pomocou rovníc kvantovej a klasickej fyziky
- môže byť jedinou metódou, ktorá dokáže odhaliť povahu materiálov priamo na molekulárnej úrovni
- súčasný cenovo dostupný nástroj pri navrhovaní nových molekúl a chemických entít. Ako sa výpočtové zdroje stávajú dostupnejšie a s vývojom užívateľsky prívetivejších balíkov molekulárneho modelovania, výpočtové správy vo forme čisto teoretických štúdií alebo ako doplnky experimentálnej práce sú čoraz bežnejšie

Molecular interactions

- <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/molecular-interaction>
- sú medzi molekulami alebo medzi atómami, ktoré nie sú spojené väzbami
- zahŕňajú kohéziu (príťažlivosť medzi podobnými), adhéziu (príťažlivosť medzi odlišnými) a odpudivé sily medzi molekulami
- boli zavedené do výpočtovej a lekárskej chémie koncom sedemdesiatych rokov minulého storočia, aby opísali trojrozmerné interakcie medzi molekulou a jej prostredím
- medzi opsínovým proteínom a retinálnym chromoforom definujú základné spektrálne vlastnosti konkrétneho vizuálneho pigmentu
- na rozhraniach tekutín sú zodpovedné za mnohé známe fyzikálne procesy, od mazania a tvorby bublín až po zmáčanie pevných látok a kapilárne stúpanie kvapalín v úzkych rúrach
- Molekulárne interakcie uhlíkových nanotrubic s bunkovým povrchom mozgového nádoru, ako aj iných fagocytárnych buniek boli skúmané, ale nie úplne objasnené, a neexistuje všeobecný konsenzus o ich celkovom osude

- medzi lipidmi a proteínmi nedodržiavajú pravidlá, ktoré popisujú väzbu dvoch vo vode rozpustných zlúčenín vo vodnom prostredí

Phylogenetic analyses

- <https://www.news-medical.net/health/What-is-Phylogenetic-Analysis.aspx>
- je štúdium evolučného vývoja druhu alebo skupiny organizmov alebo konkrétnej charakteristiky organizmu
- je prostriedkom na odhadnutie evolučných vzťahov. V molekulárnej fylogenetickej analýze sa sekvencia spoločného génu alebo proteínu môže použiť na posúdenie evolučného vzťahu druhov
- vytvárajú sa diagramy vetvenia, ktoré predstavujú evolučnú históriu alebo vzťah medzi rôznymi druhmi, organizmami alebo charakteristikami organizmu (gény, proteíny, orgány atď.), ktoré sa vyvinuli zo spoločného predka
- Diagram je známy ako fylogenetický strom. Fylogenetická analýza je dôležitá pre zhromažďovanie informácií o biologickej diverzite, genetických klasifikáciách, ako aj pri učení sa vývojových udalostí, ktoré sa vyskytujú počas evolúcie

Protein sequence analysis

- <https://www.creative-proteomics.com/services/protein-sequence-analysis-service.htm>
- je proces podrobenia proteínovej alebo peptidovej sekvencie jednej zo širokého spektra analytických metód na štúdium ich vlastností, funkcie, štruktúry alebo vývoja. Použité metodológie zahŕňajú zoradenie sekvencií, vyhľadávanie v biologických databázach a iné metódy
- môže byť použitá pre veľmi široký rozsah relevantných tém
 - Porovnanie sekvencií s cieľom nájsť podobnosť, často na odvolenie, či sú homológne
 - Identifikácia sekvenčných rozdielov a variácií
 - Identifikácia molekulárnej štruktúry zo samotnej sekvencie
 - Identifikácia vnútorných znakov sekvencie, ako sú aktívne miesta, miesta PTM, génové štruktúry a regulačné prvky
 - Odhalenie evolúcie a proteínovej diverzity sekvencií a organizmov

Drug designing

- https://sk.wikipedia.org/wiki/N%C3%A1vrh_lie%C4%8Diva
- Návrh liečiv (angl. Drug design), často označovaný ako racionálny návrh liečiv, je proces hľadania nových liečiv pomocou znalosti biologického cieľa. Liečivo je často malá organická molekula, ktorá sa používa na ochranu alebo obranu zdravia, či na zmiernenie chorobných príznakov. Presnejším termínom je návrh ligandu (tz. molekuly, ktorá sa bude pevne viazať na svoj cieľ).[2] Konštrukčné techniky na predpovedanie väzobnej afinity sú primárne úspešné avšak existuje mnoho iných vlastností, ktoré sa musia najskôr optimalizovať, aby sa ligand mohol stať efektívnym a hlavne bezpečným liekom.
- V najzákladnejšom slova zmysle zahŕňa návrh liečiva návrh molekúl, ktoré majú komplementárny tvar a náboj ako biomolekulárny cieľ, s ktorým interagujú. Pri navrhovaní liečiv sa často, ale nie vždy, spolieha na techniky molekulového modelovania. Takýto typ modelovania sa často označuje aj počítačový návrh liečiv.

Molecular dynamics simulations

- https://sk.wikipedia.org/wiki/Molekulov%C3%A1_dynamika

- Molekulová dynamika (MD) je počítačová simulačná metóda, ktorá simuluje klasický pohyb atómov v danom silovom poli.
- Pôvodne sa molekulová dynamika používala na výskum relaxačných javov a transportných vlastností kvapalín. Neskôr sa zistilo, že MD simulácie môžu byť účinnou metódou výpočtov štruktúrnych a termodynamických vlastností komplexných kvapalín, tekutých solí, kryštálov, polymérov a proteínov v roztoku. Vývoj molekulodynamických simulácií urýchlil rozvoj počítačovej techniky a MD metód. V súčasnosti sa bežne uskutočňujú MD simulácie systémov veľkosti enzýmov, ktoré sú obklopené tisíckami molekúl vody.

DNA sequence analysis

- https://cs.wikipedia.org/wiki/Sekvenov%C3%A1n%C3%AD_DNA
- Sekvenovanie DNA (tiež sekvenácia či sekvencovanie, často aj „čítanie“ DNA) je súhrnný termín pre biochemické metódy, ktorými sa zisťuje poradie nukleových báz (A, C, G, T) v sekvenciách DNA. Tieto sekvencie sú súčasťou dedičnej informácie v jadre (pri prokaryote skôr tzv. nukleoid), ďalej však aj v plazmidoch, mitochondriách a plastidoch.
- Sekvenovanie DNA má množstvo aplikácií vo vede, medicíne aj napríklad v poľnohospodárstve. Sekvenujú sa celé genómy, prípadne iba niektoré časti, ktoré majú pre daný odbor význam. Napríklad pre štúdium evolučného vývoja organizmov sa spravidla sekvenuje DNA, ktorá kóduje ribozomálnu RNA. V kriminalistike zase majú význam repetitívne oblasti DNA zvanej mikrosatelity. Niekedy sa sekvenujú len tie časti genómu, ktoré kódujú gény: táto filozofia sa nazýva EST (expressed sequence tag).

V lekárstve sľubuje sekvenovanie DNA napríklad revolúciu v diagnostike chorôb a skorému zisteniu náchylnosti jedinca k určitým chorobám (rakovina, kardiovaskulárne ochorenie). Uľahčuje sa aj výroba liekov – vedci môžu zacieliť liek proti konkrétnemu produktu konkrétneho génu. Podľa genotypu jedinca môže byť vybraný správny liečebný postup. Pokiaľ bude zistená istá genetická porucha, ponúka sa v budúcnosti tiež možnosť génovej terapie.

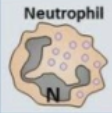
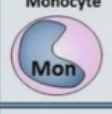

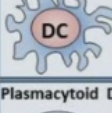
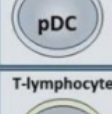
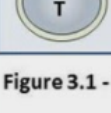
 <p>Neutrophil</p>	<p><i>Key functions:</i> Phagocytosis and killing of bacteria and fungi <i>CD markers:</i> CD66b, CD177 <i>Physical features:</i> Characteristic multi-lobed nucleus, abundant intracellular granules, 12-15 µm <i>Blood frequency:</i> 50-70 %</p>
 <p>Monocyte</p>	<p><i>Key functions:</i> Source of some tissue macrophages, circulate in blood to be recruited to sites of inflammation, detection of PAMPs, secretion of inflammatory cytokines <i>CD markers:</i> CD14. <i>Major subtypes:</i> inflammatory (CD14^{lo} CD16⁺), reparative (CD14^{hi} CD16⁻). <i>Blood frequency:</i> 5-10 %</p>
 <p>Macrophage</p>	<p><i>Key functions:</i> Phagocytosis of microbes, debris and dead host cells, PAMP detection, secretion of cytokines and chemokines to modulate immune responses and tissue repair. <i>CD markers:</i> CD68. <i>Major subtypes:</i> M1 (pro-inflammatory), M2 (involved in tissue repair & remodelling)</p>
 <p>Conventional DC</p>	<p><i>Key functions:</i> Capture of antigen and migration to lymph nodes, presentation of antigen to naïve T-cells, detection of PAMPs, secretion of cytokines to direct immune responses. <i>CD markers:</i> CD11c, CD123. <i>Physical features:</i> Numerous long projections of the cell membrane, motile on activation.</p>
 <p>Plasmacytoid DC</p>	<p><i>Key functions:</i> Detection of viral PAMPs via endosomal TLRs, secretion of large quantities of type I interferon which promotes anti-viral defence, presentation of antigen to T-cells. <i>CD markers:</i> CD123, CD303, CD304. <i>Blood frequency:</i> 0.2-0.4%</p>
 <p>T-lymphocyte</p>	<p><i>Key functions:</i> Discrimination of highly specific antigens presented on MHC, clonal expansion, assisting B-cell and macrophage activation (CD4⁺), or killing virally infected cells (CD8⁺). <i>CD markers:</i> CD3. <i>Major subtypes:</i> CD8⁺ (CTLs), CD4⁺ (inc T-reg, Th1, Th2, Th17). <i>Blood frequency:</i> 5-25%</p>

Figure 3.1 - Key features of major cells of the immune system

Neutrofil - kľúčové funkcie:

1. fagocytóza - proces využívaný bunkami k pohlcovaniu veľkých pevných častíc ($\geq 0.5 \mu\text{m}$) z okolitého prostredia. Bunky so schopnosťou fagocytózy obalia svojou cytoplazmatickou membránou prijímanú časticu a vtiahnu ju do vnútorného prostredia bunky. Vznikne tak nová intracelulárna organela zvaná fagozóm, v ktorej sa pohltý materiál strávi. (zdroj: <https://sk.wikipedia.org/wiki/Fagocyt%C3%B3za>)
2. ničenie baktérií a húb CD markerov - CD markery, tiež známe ako CD antigény, sú špecifické typy molekúl nájdených na povrchu buniek, ktoré pomáhajú diferencovať jeden typ bunky od druhého. CD antigény môžu byť použité na detekciu abnormálneho rastu buniek známych ako novotvar . Novotvary môžu byť benigne (nerakovinové), zhubné (rakovinové) alebo prekancerózne, ale rovnako ako akákoľvek iná bunka majú značky CD, ktoré môžu vedci použiť na ich identifikáciu. Ukazovatele CD sú dôležité nielen pri diagnostike rakoviny, ale môžu pomôcť určiť, ktoré typy liečby môžu byť najúspešnejšie, a merať efektívnosť liečby monitorovaním zmien príslušných markerov CD. (zdroj: <https://sk.approb.com/cd-markery-v-diagnostike-a-liecbe-rakoviny/>)

fyzické vlastnosti: charakteristické mnoholaločnaté jadro, bohaté intracelulárne (vnútro-bunkové) granuly.

Monocyt - kľúčové funkcie:

1. zdroj niektorých tkanivových makrofágov (makrofág je typom bielej krvinky imunitného systému, ktorá je charakteristická svojou schopnosťou fagocytózy,) (zdroj: <https://sk.wikipedia.org/wiki/Makrof%C3%A1g>)
2. cirkulujú v krvi, aby sa dostali do miest zápalu
3. detekcia PAMP (Pathogen-associated molecular patterns - Molekulárne vzory spojené s patogénom)
4. sekrécia (vyučovanie) zápalových cytokínov (cytokíny sú široká kategória malých proteínov, ktoré sú dôležité v bunkovej signalizácii (zdroj: <https://sk.wikipedia.org/wiki/Cytok%C3%ADn>)

Makrofág - kľúčové funkcie:

1. fagocytóza mikróbov, zvyškov a mŕtvych buniek hostiteľa
2. detekcia pamp
3. sekrécia cytokínov
4. oprava tkaniva

(Dendritová bunka je antigén-prezentujúca bunka imunitného systému. Produkuje cytokíny, je schopná fagocytózy a zároveň dokáže spracovať a T lymfocytom prezentovať antigén. Tým prepája vrodené (nešpecifické) a adaptívne (antigénne špecifické) mechanizmy imunitného systému. zdroj: https://sk.wikipedia.org/wiki/Dendritov%C3%A1_bunka)

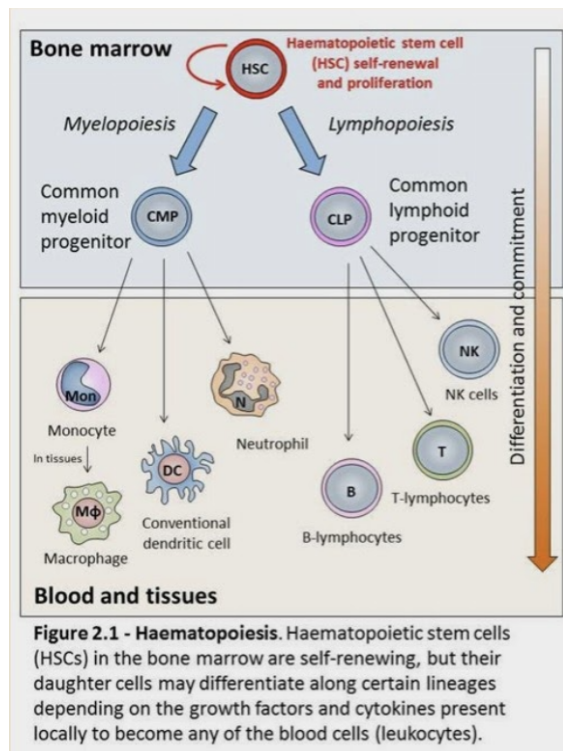
Konvenčné dendritové bunky - kľúčové funkcie:

1. zachytenie antigénu (antigén alebo imunogén je látka, ktorá vyvoláva tvorbu protilátok a môže vyvolať imunitnú odpoveď zdroj: <https://sk.wikipedia.org/wiki/Antig%C3%A9n>) a migrácia do lymfatických uzlín
2. prezentácia antigénu naivným T-bunkám
3. detekcia PAMP
4. sekrécia cytokínov na priame imunitné reakcie

Plasmacytoid DC

Plasmacytoidné dendritické bunky (pDC) predstavujú odlišnú populáciu dendritických buniek (DC) od „klasických“ myeloidných DC (mDC). Na rozdiel od mDC, pDC nefungujú primárne ako antigén prezentujúci bunky, ale ako bunky produkujúce zápalové cytokíny: interferóny I typu (IFN I). Produkciou týchto cytokínov aktivujú „natural-killer“ (NK) bunky (tzv. prirodzené zabijáče) a spúšťajú tým vrodenú imunitnú odpoveď či zahajujú zápalovú reakciu.

https://cs.wikipedia.org/wiki/Plasmacytoidn%C3%AD_dendritick%C3%A1_bu%C5%88ka

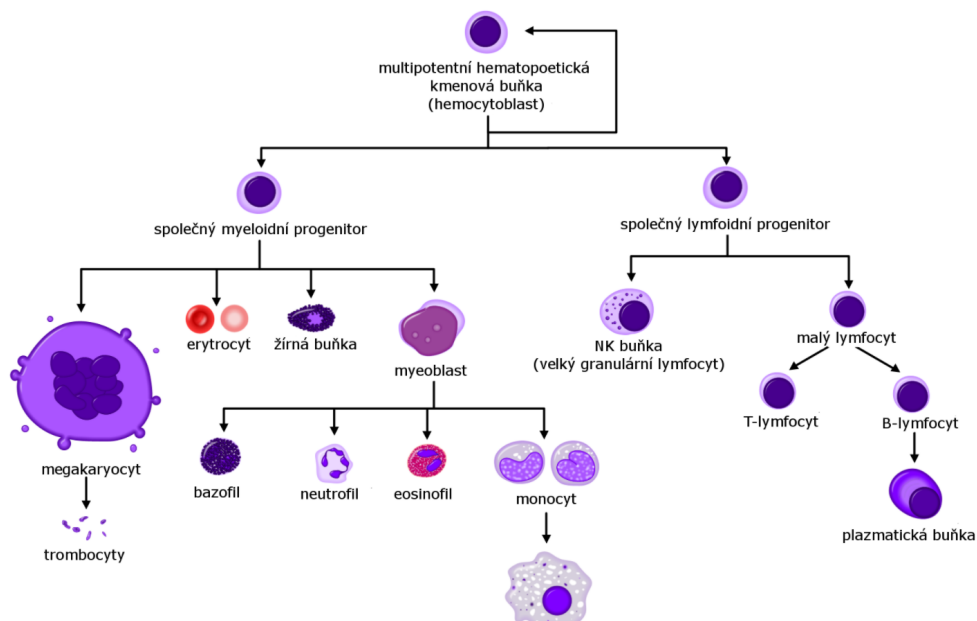


Články týkajúce sa liečby kmeňovými bunkami v súvislosti s rakovinou .

<https://febs.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/febs.16030>

<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2141.2005.05755.x>

<https://www.nature.com/articles/s41467-021-24858-3>



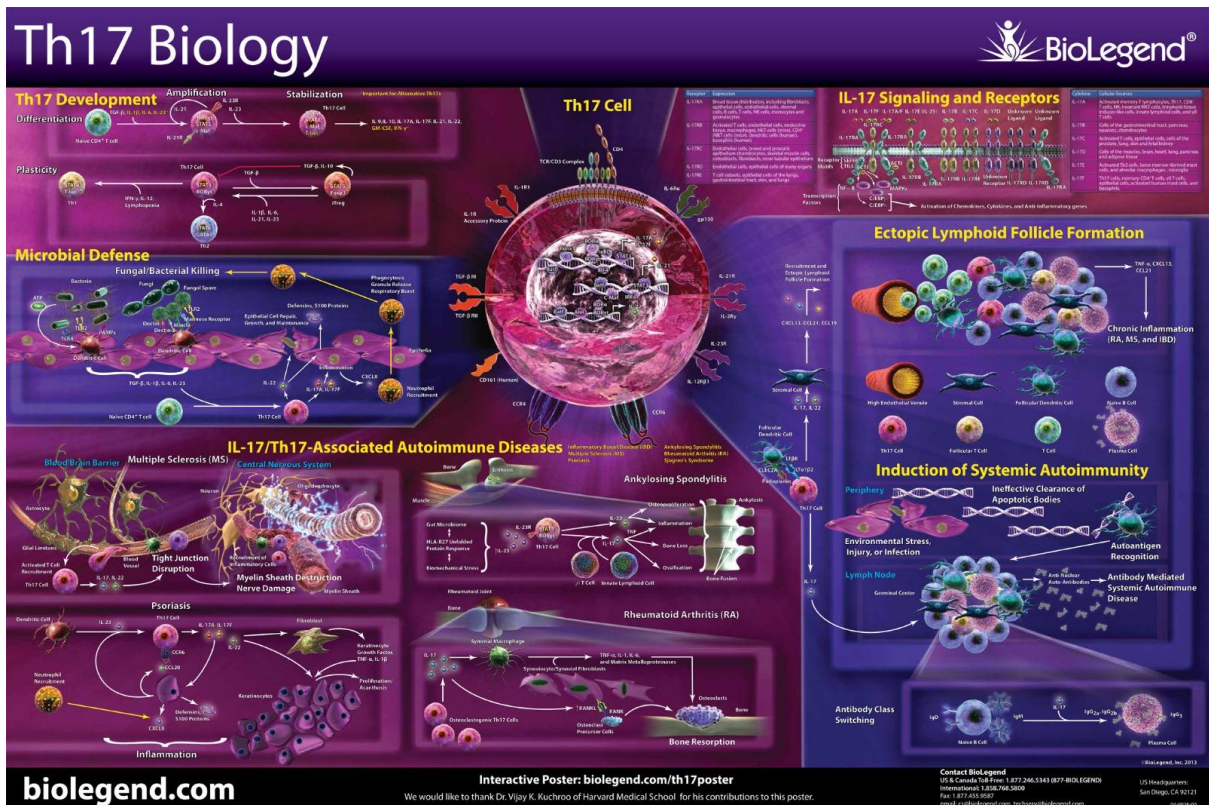
Krvotvorba -

Hematopoetické kmeňové bunky (HSC) sú v (kostnej dreni) a majú jedinečnú schopnosť umožniť vznik všetkým rôznym typom zrelých krviniek a tkanivám.

HSC sú samoobnovujúce sa bunky: keď sa diferencujú, aspoň niektoré z ich dcérskych buniek zostanú ako HSC, takže zásoba kmeňových buniek sa nevyčerpá. Tento jav sa nazýva asymetrické delenie.

Ostatné dcérske bunky - HSC (myeloidné a lymfoidné progenitorové bunky) môžu ísť ktorúkoľvek z iných diferenciačných dráh, ktoré vedú k vytvoreniu jedného alebo viacerých špecifických typov krviniek, ale nemôžu sa obnovovať. Je to jeden z hlavných životne dôležitých procesov v tele.

HSC sú veľmi dôležité pri liečbe/zvládaní onkologických ochorení, ktoré súvisia do veľkej miery so zmenami buniek (mutácie, nekontrolovaný rast, nenávratné poškodenia).



T-lymfocyt: <https://sk.wikipedia.org/wiki/T-lymfocyt>

T-lymfocyt je typ leukocytu ([bielej krvinky](#)), konkrétnejšie [lymfocytu](#), ktorý je hlavnou regulačnou bunkou špecifickej [imunity](#). T-lymfocyty sa vyvíjajú v [kostnej dreni](#) a následne dozrievajú v [týmuse](#). Po aktivácii pomocou tzv. [antigén prezentujúcich buniek](#) sa pod vplyvom množstva faktorov (cytokíny, interleukíny) diferencujú na konečné efektorové T-lymfocyty. Tie môžu aktivovať iné bunky imunitného systému alebo sa priamo podieľať na zneškodňovaní [patogénov](#).

Th-17 lymfocit: https://cs.wikipedia.org/wiki/Th17_lymfocyt

Hlavní funkcí Th17 lymfocytů je regulace zánětu a obrana proti některým extracelulárním patogenům. Th17 lymfocyty specifické pro vlastní antigeny jsou velmi patogenní a jsou spojovány s autoimunitními onemocněními např. [roztroušenou sklerózou](#) a [revmatoidní artritidou](#).

Th17 bunky zabraňujú závažnej infekcii HIV udržiavaním črevnej epiteliálnej bariéry počas infekcie HIV v čreve. Kvôli vysokej hladine CCR5 koreceptora pre HIV sú prednostne infikované. K mikrobiálnej translokácii teda dochádza prostredníctvom deplécie buniek Th17.

Th17 cytokíny IL-17A a IL-17F spúšťajú produkciu prozápalových cytokínov v cieľových tkanivách, ktoré nielen sprostredkovávajú zápal prostredníctvom nábory vrodenných imunitných buniek, ako sú neutrofil, ale tiež podporujú ďalšiu aktiváciu Th17 spôsobom pozitívnej spätnej väzby.

<https://pubmed.ncbi.nlm.nih.gov/22640807/>

<https://www.sciencedirect.com/science/article/pii/S1568997214001633>

Príloha C - Analýza rôznych štruktúr vedeckých článkov

Výskumné články

- články sú dostupné vo formáte .pdf, pre niektoré sú potrebné prihlasovacie údaje
- články je možné získať crawlovaním dát

Štruktúra článku

- prevažná časť text rozdelený vo viacerých sekciách -> vhodné ako vstup pre nástroj automatického spracovania textu
- niekoľko dovysvetľujúcich obrázkov -> skôr vhodné pre následnú kontrolu používateľom nástrojov
- tabuľky so vstupnými dátami resp. výsledkami -> overenie výsledkov

Dôležité časti:

- kľúčové slová -> základné slová definujúce oblasť vedeckého výskumu, dôležité pre danú oblasť
- Abstract -> stručné zhrnutie zámeru článku, využitý prístup bez dosiahnutých výsledkov
- Introduction a Related Work -> úvod do oblasti, všeobecné poznanie potrebné pre porozumenie článku, prehľad podobne orientovaných výskumných prác
- Methods -> popis vykonaného experimentu, využitých metód, skúmaných dát
- Results a Discussion + Conclusion -> Zhodnotenie vykonaných experimentov a dosiahnutých výsledkov, načrtnutie ďalšieho možného smerovania v oblasti
 - text doplnený tabuľkami s číselným vyhodnotením alebo snímkami

Dáta zo stránky <https://clinicaltrials.gov/>

- dáta môžeme získať crawlovaním stránky

Štruktúra dát

- dáta nie sú vo forme článkov, jedná sa skôr o tabuľkové dáta
- rozčlenené do viacerých sekcií
- iba textové dáta
 - číselné
 - dátumy

Dôležité časti

- Description - stručný opis štúdie, na čo je zameraná, opis medicínskeho problému
- Study Design - informácie o štúdiu - trvanie, počet participantov, primárny účel/cieľ
- Arms and Interventions - informácie o participanoch a danej liečbe podľa protokolu štúdie
- Outcome Measures - zoznam výsledkov - primárnych, sekundárnych ale aj iných

Dáta zo stránky <https://www.ncbi.nlm.nih.gov/pmc/tools/ftp/>

- Umožňuje sťahovať pdf-ká článkov individuálne - je ich tam veľa, niektoré sú ale aj staré.
- Cez FTP, sú tu aj iné spôsoby.
- Niektoré pdf sa nedajú strojovo čítať - sú to len scany..
- Download cez windows explorer z tohoto linku: ftp.ncbi.nlm.nih.gov/pub/pmc/oa_pdf/
 - otvoriť to akoby folder ale z netu a stlačí ctrl+c, ctrl+v

Príloha D - Analýza NLP

NLP (Natural Language Processing) - Spracovanie prirodzeného jazyka je disciplína zaoberajúca sa extrakciou informácií z textu. V našom projekte použijeme NLP metódy na spracovanie klinických štúdií do formy štruktúrovaných dát, ktorým bude rozumieť počítač.

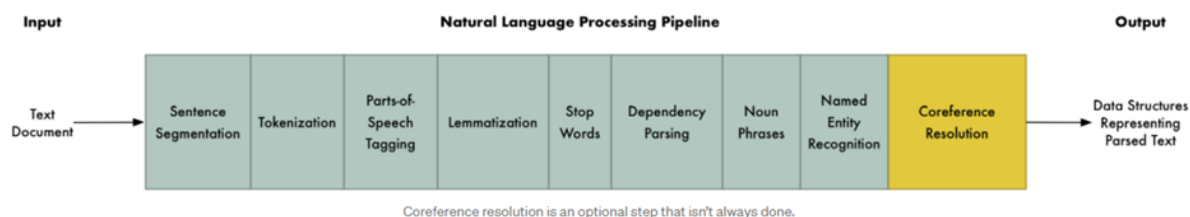
Pri vytváraní NLP modelu budeme postupovať nasledovne:

1. Získanie textov, ktoré budeme analyzovať – najmä klinické štúdie
2. Spracovanie/analýza textov – Spracovanie textu do podoby, ktorej rozumie počítač. Metódy predspracovania v krokoch:
 - a. Rozdelenie textu na segmenty - vety
 - b. Tokenizácia – rozdelenie viet podľa slov na slová. To nám umožní pracovať s menšími časťami textu.
 - c. Odstránenie „stop-slov“ – stop slová sú slová, ktoré chceme ignorovať – sú veľmi bežné, no nemajú význam (napríklad predložky)
 - d. Stemming – upravovanie slov na korene (napríklad „helping“ alebo „helper“ upravíme na „help“)
 - e. POS tagging - gramatické označovanie je proces označovania slova v texte ako zodpovedajúceho konkrétnemu slovnému druhu (podstatné mená...)
 - f. Lemmatizácia – redukovanie slov na ich základný význam – tak aby slovo samo o sebe dávalo zmysel
 - g. Dependency parsing – určovanie ako spolu slová z vety súvisia – výsledkom je strom
 - h. Hľadanie fráz – zoskupenie slov, ktoré predstavujú jednu myšlienku/vec.
 - i. Rozpoznávanie pomenovaných entít – mená ľudí, organizácií, miest
 - j. Coreference Resolution – v texte sa namiesto písania mien sa v každej vete používajú zámená (on, ona..). Pomocou tejto metódy zabezpečíme, že počítač vie na čo sa zámená odkazujú.

Zdroje:

<https://realpython.com/nltk-nlp-python/#getting-text-to-analyze>,

<https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e>



(zdroj obrázka: <https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e>)

Analýza nástroja vytvoreného Kinit -SlovakBERT

GitHub: <https://github.com/gerulata/slovakbert>

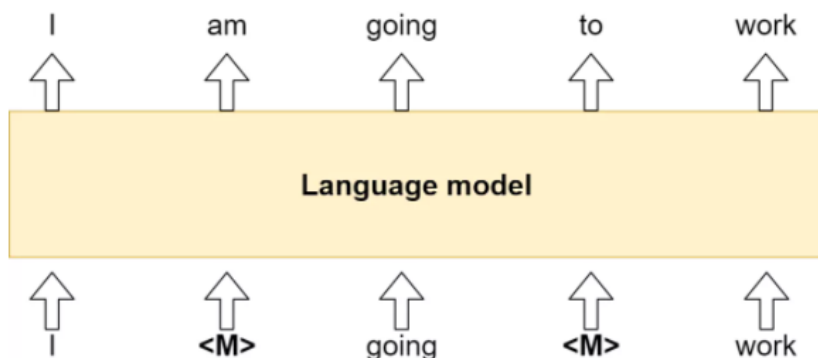
Článok: <https://arxiv.org/abs/2109.15254>

Návod: <https://huggingface.co/gerulata/slovakbert>

- Spustiteľné cez PyTorch / TensorFlow
- Stiahnuteľne z gitu alebo nejakou mágiou cez python - v Readme je kód
- Vie dopĺňať slová do viet
- Píše sa aj o analýze sentimentu - k tomuto nie je návod..
- Je case sensitive
- Je to tzv masked language modeling (MLM) objective. Tu je vysvetlenie:

Chris **Staff** answered 9 months ago

In a **Masked Language Modeling** task, language models don't have access to the full input – but rather to a masked input, where some (10-20 percent) of the input tokens are masked. This simply means replacing the tokens (or some token spans) with a special token representing <mask>. The goal for the MLM task becomes reconstructing the original sequence, i.e. to reveal what is hidden under the mask. This adds complexity on top of regular language modeling tasks, and some works argue that it can help boost performance.



Source:

MachineCurve. (2021, March 2). *Easy masked language modeling with machine learning and HuggingFace transformers*. <https://www.machinecurve.com/index.php/2021/03/02/easy-masked-language-modeling-with-machine-learning-and-huggingface-transformers/>

Napr príklad je takýto readme:

```

from transformers import pipeline
unmasker = pipeline('fill-mask', model='gerulata/slovakbert')
unmasker("Deti sa <mask> na ihrisku.")

[{'sequence': 'Deti sa hrali na ihrisku.',
'score': 0.6355380415916443,
'token': 5949,
'token_str': ' hrali'},
{'sequence': 'Deti sa hrajú na ihrisku.',
'score': 0.14731724560260773,
'token': 9081,
'token_str': ' hrajú'},
{'sequence': 'Deti sa zahrali na ihrisku.',
'score': 0.05016357824206352,
'token': 32553,
'token_str': ' zahrali'},
{'sequence': 'Deti sa stretli na ihrisku.',
'score': 0.041727423667907715,
'token': 5964,
'token_str': ' stretli'},
{'sequence': 'Deti sa učia na ihrisku.',
'score': 0.01886524073779583,
'token': 18099,
'token_str': ' učia'}]

```

Analýza nástrojov pre anglický jazyk

Natural Language Toolkit - NLTK

Hlavná stránka: <https://www.nltk.org/>

Návod a obsluha: https://www.nltk.org/book_1ed/

Github: <https://github.com/nltk/nltk>

Komunita: <https://groups.google.com/g/nltk-users>

- popredný nástroj na vytváranie programov v jazyku Python na prácu s textami
- bezplatný nástroj, dostupné zdrojové kódy, aktívna komunita ľudí používajúcich nástroj
- pre NLP poskytuje množstvo komponentov ako tokenizácia, analýza, parsovanie, klasifikácia, získanie základu slova, vytváranie ngramov alebo skipgramov, taktiež vytváranie stromov
- na nainštalovanie nástroja potrebujeme mať nainštalovaný python a pip, následne ho vieme nainštalovať príkazom “pip install nltk” a ďalej v aplikácii kde chceme nltk použiť “import nltk”
- získanie základu slov podporuje nltk len pre arabský, dánsky, holandský, anglický, fínsky, francúzsky, nemecký, maďarský, taliansky, nórsky, portugalský, rumunský, ruský, španielsky a švédsky jazyk. Pre slovenský jazyk a podobne by sme to museli riešiť inou knižnicou
- je možné kombinovať s knižnicou scikit-learn - strojové učenie
- vstup knižnice je text, s ktorým plánujeme pracovať
- výstup je už podľa toho, čo si naprogramujeme. Buď tokenizovaný text, parsovaný, text len so základmi slov, existujúce ngramy alebo skipgramy a ich počet v texte. Vďaka možnosti strojového učenia môžeme napríklad určovať, s akou pravdepodobnosť sa jedná o vedecké články a podobne

In [3]:

```
from nltk.util import ngrams

n = 3
sentence = 'Whoever is happy will make others happy too'
unigrams = ngrams(sentence.split(), n)

for item in unigrams:
    print(item)
```

[Out] :

```
('Whoever', 'is', 'happy')
('is', 'happy', 'will')
('happy', 'will', 'make')
('will', 'make', 'others')
('make', 'others', 'happy')
('others', 'happy', 'too')
```

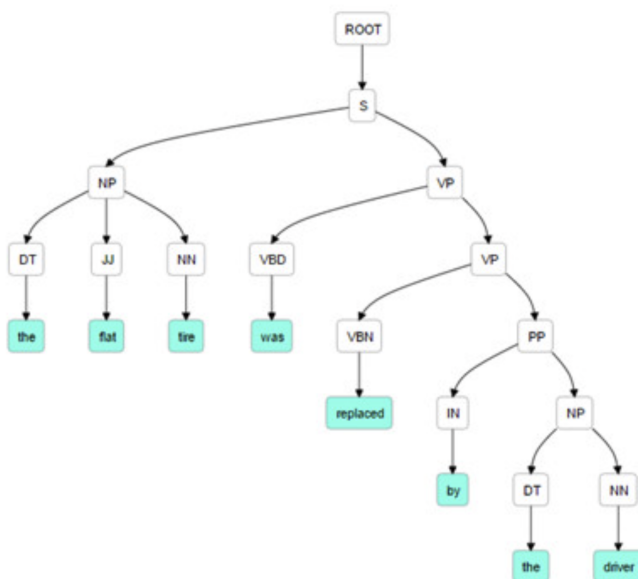
```
>>> tree.pretty_print()
```

```
      s
    -----|-----
    |         vp
    |         -----|----
    |         |         dp
    dp        |         dp
    ---|---    |         ---|---
    d    np    v    d    np
    |     |    |    |    |
    the  dog chased the  cat
```

Stanford OpenIE

Hlavná stránka: <https://nlp.stanford.edu/software/openie.html>

- extrakcia binárnych relácií (tripletovej) z neštruktúrovaného textu



Napr. extrahuje triplet: driver replaced tire

- súčasť Stanford CoreNLP knižnice (<https://stanfordnlp.github.io/CoreNLP/>)
- Java 8+
- vstup
 - text zo štandardného vstupu
 - súbory obsahujúce voľný text
- výstup
 - TSV formát obsahujúci 4 stĺpce (confidence, subject, relation, object)
 - na štandardný výstup
 - do súboru
- použitie
 - analýza textu zo štandardného vstupu:
- `java -mx1g -cp "*" edu.stanford.nlp.naturalli.OpenIE`
 - analýza textu zo súborov:
- `java -mx1g -cp "*" edu.stanford.nlp.naturalli.OpenIE /path/to/file1 /path/to/file2`

Ollie

Hlavná stránka: <https://knowitall.github.io/ollie/>

Github: <https://github.com/schmmd/ollie>

- extrakcia binárnych relácií (tripletov) z anglických viet
- Scala
- pokročilá funkcionálnosť:
 - ošetrovanie podmienok vo vete

```
sentence: If I slept past noon, I'd be late for work.  
extraction: (I; 'd be late for; work)[enabler=If I slept past noon]
```

- detekcia autora výroku

```
sentence: Some people say Barack Obama was not born in the United States.  
extraction: (Barack Obama; was not born in; the United States)[attrib=Some people say]
```

- identifikácia vzťahov, ktoré nie sú vo vete vyjadrené slovesom

```
sentence: Microsoft co-founder Bill Gates spoke at a conference on Monday.  
extraction: (Bill Gates; be co-founder of; Microsoft)
```

- N-árne extrakcie - relácia získaná na základe niekoľkých iných relácií

```
sentence: I learned that the 2012 Sasquatch music festival is scheduled for May 25th until May 28th.  
extraction: (the 2012 Sasquatch music festival; is scheduled for; May 25th)  
extraction: (the 2012 Sasquatch music festival; is scheduled until; May 28th)  
nary: (the 2012 Sasquatch music festival; is scheduled; [for May 25th; to May 28th])
```

- vstup:
 - súbor obsahujúci voľný text

- výstup:
 - viacero možností, napr. TSV súbor obsahujúci vyextrahované vzťahy (nenašiel som presný formát)
- použitie:
 - GUI
 - Maven dependency
 - CLI

```
java -Xmx512m -jar ollie-app-VERSION.jar yourfile.txt --split --outputformat tabbed --output outfile.txt
```

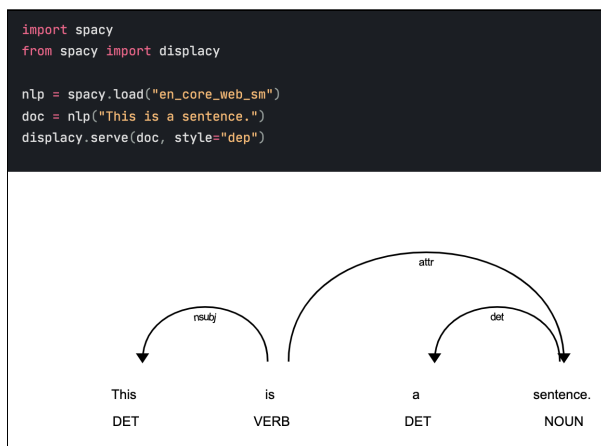
- bez prepínača --split považuje za vetu jeden riadok
- prepínač --outputformat - možnosť zdefinovať formát výstupu, napr. TSV formát cez hodnotu tabbed

spaCy

Hlavná stránka: <https://spacy.io/>

Github: <https://github.com/explosion/spaCy>

- do určitej miery podporuje aj slovenčinu a češtinu
- označovanie slovných druhov, analýzu závislostí, kategorizácia textu, rozpoznávanie pomenovaných entít...
- Python
- podpora PyTorch a TensorFlow
- inštalácia a použitie: <https://spacy.io/usage>
- vstup: text alebo súbory obsahujúce voľný text
- má peknú vizualizáciu výstupov:
 - vizualizácia dependency



- vizualizácia rozpoznávania entít

```
import spacy
from spacy import displacy

text = "When Sebastian Thrun started working on self-driving cars at Google in 2007, few people outside of the company took him seriously."

nlp = spacy.load("en_core_web_sm")
doc = nlp(text)
displacy.serve(doc, style="ent")
```

When Sebastian Thrun PERSON started working on self-driving cars at Google ORG in 2007 DATE , few people outside of the company took him seriously.

- výstup a použitie:
 - Unigramy, bigramy, trigramy... N-gramov

N-gramy = N po sebe idúcich tokenov v texte.

Užitočné pri rozhodovaní o tom, ktoré slová sa vyskytujú spolu, aby sme ich mohli rozdeliť do jednej entity. Je to užitočné aj pri predikcii ďalšieho slova.

```
import spacy

def n_grams(tokens, n):
    return [tokens[i:i+n] for i in range(len(tokens)-n+1)]

English = spacy.load('en_core_web_sm')

text = "Hi how are you? I am fine. So, what's up?"
text = text.lower()

tokens = [str(token) for token in English(text)]

print(n_grams(tokens, 4))

[['hi', 'how', 'are', 'you'], ['how', 'are', 'you', '?'], ['are', 'you', '?', 'i'], ['you', '?', 'i', 'am'], ['?', 'i', 'am', 'fine', '.'], ['am', 'fine', '.', 'so'], ['fine', '.', 'so', ','], [',', 'so', ',', 'what'], ['so', ',', 'what', "'s"], ['up'], ['what', "'s", 'up', '?']]
```

- POS-tagging

Druh klasifikačnej úlohy. Používa sa pri vytváraní stromov. Pre získanie významu slova alebo vytvorenie akéhokoľvek vzťahu medzi slovami je dôležitým krokom.

```
English = spacy.load('en_core_web_sm')

text = "Hi how are you? I am fine. So, what's up?"

tokens = [token for token in English(text)]

for token in tokens:
    print("{} -> {}".format(token, token.pos_))

Hi -> INTJ
how -> ADV
are -> AUX
you -> PRON
? -> PUNCT
I -> PRON
am -> AUX
fine -> ADJ
. -> PUNCT
So -> ADV
, -> PUNCT
what -> PRON
's -> AUX
up -> ADP
? -> PUNCT
```

- Rozpoznávanie pomenovanej entity

Identifikovanie pomenovaných entít, ako je osoba, miesto, organizácia atď. Užitočný na extrahovanie kľúčových prvkov v neštruktúrovaných údajoch, aby sme ich mohli triediť na základe súvisiacich kľúčových jednotiek v texte.

```

English = spacy.load('en_core_web_sm')

text = "European authorities fined Google a record $5.1 billion on Wednesday"

doc = English(text)

for token in doc.ents:
    print("{} -> {}".format(token, token.label_))

```

European -> NORP
 Google -> ORG
 \$5.1 billion -> MONEY
 Wednesday -> DATE

Apache OpenNLP

Stránka: <https://opennlp.apache.org/>

Tutoriál: <https://www.tutorialspoint.com/opennlp/index.htm>

Vzor na použitie: <https://www.baeldung.com/apache-open-nlp>

Knižnica Apache OpenNLP je sada nástrojov založená na strojovom učení na spracovanie textu v prirodzenom jazyku. Podporuje najbežnejšie úlohy NLP, ako je detekcia jazyka, tokenizácia, segmentácia viet, označovanie časti reči, extrakcia pomenovanej entity, blokovanie, analýza a rozlíšenie koreferencií.

Je nad Javou.

Dostupné pre angličtinu a nemčinu a pár ďalších jazykov - pre slovenčinu by sa musel natrénovať model

Analýza nástrojov pre Český a slovenský jazyk

Vyvíjajú sa rôzne nástroje na spracovanie prirodzeného jazyka, pričom existuje predovšetkým veľké množstvo nástrojov a netrénovaných modelov pre anglický jazyk.

Čo sa týka nástrojov na spracovanie slovenského a českého jazyka, tak tie zaostávajú za tými anglickými, a máme ich k dispozícii oveľa menej. Je to zapríčinené aj tým, že slovenčina a čeština sú veľmi náročné jazyky s voľným slovosledom a náročným/nepravidelným skloňovaním.

Nástroje pre slovenský jazyk

Bednárík - NLP web

Jedným z nástrojov na spracovanie prirodzeného slovenského jazyka je NLP web. Je to webová aplikácia vytvorená študentom FIIT ako DP. (<http://nlp.bednarik.top/>)

Táto aplikácia umožňuje širokú škálu textových operácií ako napríklad

- tokenizácia: zadá sa text - výstupom sú tokeny oddelené medzerou

- segmentácia: zadá sa text a konfiguračné nastavenie - výstupom je zoznam viet oddelených novým riadkom
- lematizácia: zadá sa text - výstupom sú lemy (základné tvary slova)
- stemovanie: zadá sa text - výstupom je sekvencia stemov (kmeň slova)
- rozpoznávanie pomenovaných entít

STANZA

Ďalším nástrojom je nástroj Stanza. Jedná sa o zbierku nástrojov na lingvistickú analýzu mnohých jazykov, medzi ktorými je aj slovenčina.

Stanza je pythonovsky package pre analýzu prirodzeného jazyka. Slúži napríklad na konvertovanie reťazca slov do viet, na generovanie základných tvarov slov, ich slovných druhov a morfológických vlastností, na analýzu závislosti syntaktickej štruktúry alebo aj na rozpoznávanie pomenovaných entít.

Tento nástroj je postavený na vysoko presných komponentoch neurónových sietí,

ktoré taktiež umožňujú efektívne tréningovanie a vyhodnocovanie nad vlastnými anotovanými dátami. Moduly sú implementované nade knižnicou PyTorch.

(<https://github.com/stanfordnlp/stanza>)

SLOVAK-LEXER

Je nástroj vytvorený pre lexikálne spracovanie slovenského jazyka. Podporuje textové operácie ako sú tokenizácia a segmentácia. Tento nástroj je implementovaný v jazyku C++.

(<https://github.com/hladek/slovak-lexer>)

dl4dp

Jedná sa o NLP knižnicu Pythonu, ktorá poskytuje nástroje na morfológické značkovanie,

lematizáciu a analýzu závislostí. Hlavnou motiváciou tejto knižnice je poskytovať najmodernejšie nástroje pre slovenský jazyk. Táto knižnica poskytuje rozhranie príkazového riadku, ktoré umožňuje natréningovať vlastný model alebo parsovať dáta.

(<https://github.com/peterbednar/dl4dp>)

NLP4SK

Projekt NLP4SK poskytuje NLP nástroje pre spracovanie slovenského jazyka. Pokrýva viacero služieb od identifikácie slov až po určenie informácií o jednotlivých slovách.

Medzi poskytovanými službami na nachádza napríklad:

- predspracovanie textu: čistenie textu od netextových častí (napr. referencií).
- segmentácia textu na vety a slová: identifikácia slov a začiatku/konca vety
- lexikálna analýza: identifikácia slov v slovníku
- lematizácia a stemovanie
- syntaktická analýza : určenie významu slova vo vete (napr. slovný druh)
- identifikácia sentimentu: určovanie sentimentu slov/viet (pozitívny alebo negatívny)

(http://arl6.library.sk/nlp4sk/#chap_modules)

Nástroje pre Český jazyk

Spracovaniu češtiny venujú veľkú pozornosť v Centru zpracování přirozeného jazyka (CZPJ) na Masarykovej univerzite v Brne, kde sa zameriavajú najmä na tokenizáciu (unitok), segmentáciu na vety a na syntaktickú (syn, set), sémantickú a morfológickú analýzu jazyka (Ajka a Majka).

Ajka, Majka

Analýzátor ajka je program napísaný v jazyku C a je výsledkom diplomovej práce s názvom "Morfológický analyzátor češtiny". Bol vyvinutý v roku 1999, avšak jeho vývoj stále pokračuje.

Všetky zdrojové kódy sú napísané v jazyku C (ANSI). Aktuálna verzia je pripravená na kompiláciu na linuxových strojoch s procesormi Intel a predpokladá znakovú sadu ISO 8859-2 pre češtinu. Kompilácia na strojoch s inými typmi procesorov a inými operačnými systémami nebola testovaná a tak sa neodporúča. Aktuálny vývoj sa zameriava na čo najviac platformovo nezávislú verziu.

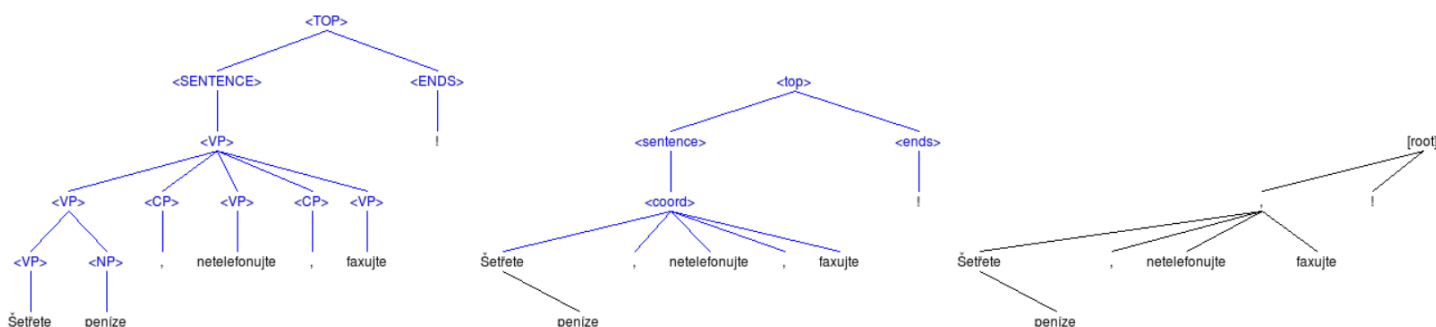
Autor zvolil slovníkový prístup, čo znamená, že všetky dáta potrebné k správne fungovaniu sú uložené v strojovom slovníku češtiny a v definičnom súbore koncovkových množín a vzorov. Taktiež využili dátovú štruktúru trie implementovanú vo forme minimálneho konečného automatu, z dôvodu minimalizácie pamäťových nárokov. Pri morfológickej analýze vybraného slova algoritmus segmentuje slovo na jednotlivé morfémy a na ich základe určí gramatické kategórie. Ajkasa však ukázala v praxi ako príliš pomalá. Tento problém sa snaží odstrániť analyzátor Majka, ktorý je využíva minimálny acyklický deterministický automat, čím je vyhľadávanie rýchlejšie a efektívnejšie.

Set

<https://nlp.fi.muni.cz/trac/set/wiki/documentation>

Projekt SET (Syntactic Engineering Tool) má za cieľ vyvinúť nový prístup k syntaktickej analýze českého jazyka. Nová metóda je založená na pravidlách spájania zhody vzorov. SET možno považovať za nástroj na automatickú syntaktickú analýzu češtiny a na vývoj formalizmov, ktoré sú založené na pravidlách syntaktickej analýzy.

Zahrnuté formalizmy sú založené na detekcii dôležitých položiek vo vstupnej vete a inkrementálnej segmentácii vety.



Porovnanie zložkových (vľavo), hybridných (v strede) a závislých (vpravo) výstupných stromov pre vstupnú vetu: Šetřete peníze, netelefonujte, faxujte!.

BERT (Bidirectional Encoder Representations from Transformers)

<http://docs.deeppavlov.ai/en/master/features/models/bert.html>

BERT je metóda predbežného tréovania jazykových reprezentácií, čo znamená, že trénujeme všeobecný model „pochopenia jazyka“ na veľkom textovom korpuse (ako je Wikipedia) a potom tento model používame pre následné úlohy NLP, na ktorých nám záleží (napríklad otázka odpoveď). Prístup je implementovaný v Python 3.6.

Príloha E - Porovnanie testovaných NLP nástrojov a analýza výstupov

Všetky nástroje dostali na vstupe rovnaký odsek textu - abstrakt článku The role of immune checkpoint inhibitors in prostate cancer (zdroj:

<http://www.aem.pl/The-role-of-checkpoint-Inhibitors-in-Prostate-cancer,96331,0,2.html>)

	NLTK	Ollie	Stanford OpenIE	SpaCy
Jazyk	python	java	java	python
Vstup	raw text	voľný text (Štandardný vstup alebo súbor)	voľný text (Štandardný vstup alebo súbor)	voľný text ale musíme ho predspracovať

Natural Language Toolkit - NLTK (Python)

Nástrojom NLTK vieme ľahko vypočítať štatistickú mieru TF-IDF textu, ktorá nám napríklad môže vrátiť N najdôležitejších slov pre text. Text je ale potrebné predprípraviť do formy, ktorú budeme chcieť na vyhodnotenie použiť. Pre tento test sa z článku odstránili všetky číslice, znaky okrem “:”, “?”, “!” a taktiež všetky slová kratšie ako 4 znaky. Taktiež sa v tomto teste nevyhodnocovali “stop_words” použité z knižnice “nltk.corpus”.

Pri vstupe článku “The role of immune checkpoint” sme získali tento výstup:

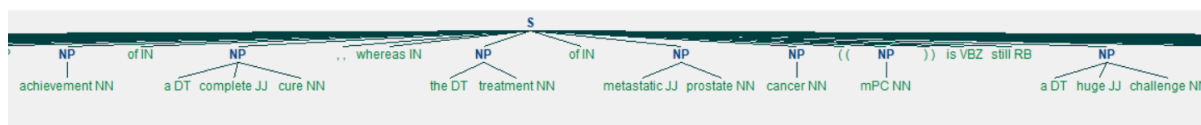
1. expression -> 0.02216235319901451
2. checkpoint -> 0.02073781622330965
3. cancer -> 0.017800140939809042
4. patients -> 0.0166947318613726
5. prostate -> 0.015661274270125387
6. immune -> 0.012130404435087028
7. inhibitors -> 0.011029456685097647
8. therapy -> 0.010070644930151527
9. treatment -> 0.009620896885799647
10. enzalutamide -> 0.00798701302682247

Vo výstupe môžeme vidieť 10 najdôležitejších slov článku, medzi ktorými sa umiestnili prevažne slová, ktoré by sme vo vedeckom článku s témou rakoviny prostaty očakávali. Vďaka nim môžeme neskôr článok zaradiť do správnej skupiny článkov.

NLTK tiež umožňuje vytvorenie chunkov - fráz/menších informácií z textu. Najprv sú slová otagované - je určený slovný druh:

```
[('Prostate', 'NNP'), ('cancer', 'NN'),
```

V tomto prípade NNP znamená singular proper noun a NN singular noun. Tieto skratky sú obširnejšie opísané na stránke <https://www.guru99.com/pos-tagging-chunking-nltk.html> . Pomocou Regex parsera, ktorý si vopred zdefinujeme, vieme vytvoriť na základe otagovaných slov strom chunkov. Príklad:



Ollie (Java)

- výstup - triplety, spolu s ich confidence
- vyextrahovaných vzťahov nie je veľa ale sú zväčša relevantné a podstatné

Pre vetu:

„The key to achieve the maximum anti-tumour response is to choose the best candidates for this therapy and determine the optimal sequence and combination of drugs.”

- vyextrahoval nástroj nasledujúce vzťahy s uvedeným confidence:

0,807: (The key to achieve the maximum anti-tumour response; is to choose; the best candidates)

0,729: (The key to achieve the maximum anti-tumour response; is to determine; the optimal sequence and combination of drugs)

Stanford OpenIE (Java)

- výstup - triplety, spolu s ich confidence
- veľké množstvo vyextrahovaných vzťahov, veľakrát aj nepodstatných a nepresných

Pre vetu:

„The key to achieve the maximum anti-tumour response is to choose the best candidates for this therapy and determine the optimal sequence and combination of drugs.”

vyextrahoval nástroj napr. nasledujúce vzťahy s uvedeným confidence:

1.0	best candidates determine	sequence
1.0	best candidates determine	optimal sequence of drugs
1.0	best candidates determine	optimal sequence
1.0	best candidates determine	sequence of drugs
1.0	candidates determine	optimal sequence
1.0	candidates determine	optimal sequence of drugs

SpaCy (Python)

- výstupom je strom, resp. sekvencia tokenov s anotáciou mnohých vzťahov
- má všetky vlastnosti stromu. Tento strom obsahuje informácie o štruktúre viet a gramatike a možno ho prechádzať rôznymi spôsobmi na extrahovanie vzťahov.

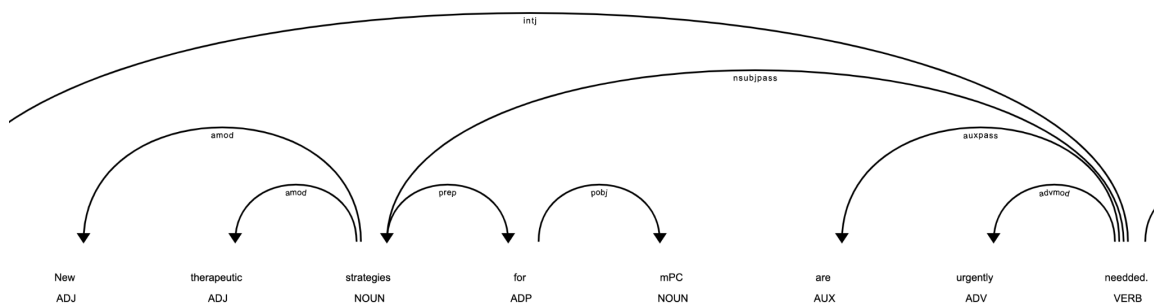
```

TEXT -> LEMMA -> POS -> TAG -> DEP -> SHAPE -> ALPHA -> STOP

Prostate -> prostate -> NOUN -> NN -> amod -> Xxxxx -> True -> False
cancer -> cancer -> NOUN -> NN -> nsubj -> xxxx -> True -> False
( -> ( -> PUNCT -> -LRB- -> punct -> ( -> False -> False
PC -> pc -> NOUN -> NN -> appos -> XX -> True -> False
) -> ) -> PUNCT -> -RRB- -> punct -> ) -> False -> False
is -> be -> AUX -> VBZ -> ROOT -> xx -> True -> True
the -> the -> DET -> DT -> det -> xxx -> True -> True
most -> most -> ADV -> RBS -> advmod -> xxxx -> True -> True
commonly -> commonly -> ADV -> RB -> advmod -> xxxx -> True -> False
diagnosed -> diagnose -> VERB -> VBN -> amod -> xxxx -> True -> False
malignant -> malignant -> ADJ -> JJ -> amod -> xxxx -> True -> False
tumour -> tumour -> NOUN -> NN -> attr -> xxxx -> True -> False
and -> and -> CCONJ -> CC -> cc -> xxx -> True -> True
the -> the -> DET -> DT -> det -> xxx -> True -> True
third -> third -> ADJ -> JJ -> amod -> xxxx -> True -> True
cause -> cause -> NOUN -> NN -> conj -> xxxx -> True -> False
of -> of -> ADP -> IN -> prep -> xx -> True -> True
cancer -> cancer -> NOUN -> NN -> compound -> xxxx -> True -> False
deaths -> death -> NOUN -> NNS -> pobj -> xxxx -> True -> False
among -> among -> ADP -> IN -> prep -> xxxx -> True -> True
men -> man -> NOUN -> NNS -> pobj -> xxx -> True -> False
in -> in -> ADP -> IN -> prep -> xx -> True -> True
Europe -> Europe -> PROPN -> NNP -> pobj -> Xxxxx -> True -> False

```

- anotácie vyzerali prevažne správne, ale niektoré vzťahy neboli dobre určené (strategies a are by malo mať medzi sebou vzťah)



- Zgrupovanie podstatných mien do fráz:

Výsledky vyzerali dobre aj pre odborný text.

```
Noun phrases: ['Prostate cancer', '(PC', 'the most commonly diagnosed malignant tumour', 'the third cause', 'cancer deaths', 'men', 'Europe']
Noun phrases: ['The treatment', 'early-stage PC', 'many cases', 'achievement', 'a complete cure', 'the treatment', 'metastatic prostate cancer', '(mPC', 'a huge challenge', 'clinicians']
Noun phrases: ['New therapeutic strategies', 'mPC']
Noun phrases: ['the most promising methods', 'treatment', 'anticancer immunotherapy', 'the monoclonal', 'immune checkpoint inhibitors']
```

Zhrnutie

PYTHON:

NLTK funkcie sa dajú veľmi prispôbiť a naše riešenie síce neukazovalo najpresnejšie výsledky, ale existuje veľa spôsobov ako ho zlepšiť -> napr. pri zgrupovaní do fráz zdefinovať gramatiku.

SpaCy nám poskytuje už teraz celkom presné výsledky aj pre odborný text vo forme stromov s množstvom informácií o štruktúre viet avšak pri použití tohto nástroja treba myslieť na to že je určený skôr do produkcie než na výskumné účely, čomu zodpovedá aj trochu iný toolset než v NLTK.

JAVA:

Ollie dokázal identifikovať podstatné vzťahy zatiaľčo Stanford OpenIE pri rovnakom tasku extrahoval veľa viac menej pre nás zbytočných a často nepresne určených vzťahov. Ollie tiež zvládol extrahovať pomerne úspešne aj komplexnejšie vzťahy.

Príloha F - Analýza podobných prác

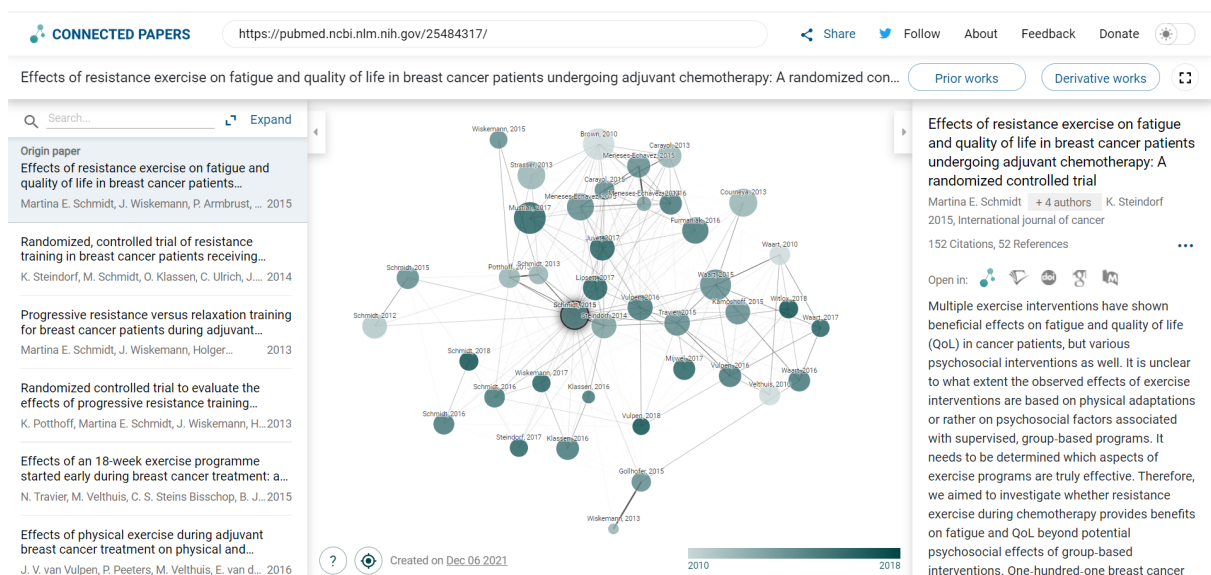
Komerčné nástroje

<https://aarontay.medium.com/3-new-tools-to-try-for-literature-mapping-connected-papers-inciteful-and-litmaps-a399f27622a>

Connected Papers

<https://www.connectedpapers.com/>

- možnosť zadať jeden hlavný, základný článok
- služba vyhledá a zobrazí podobné články
- vizualizácia vzťahov medzi článkami - graf
- veľkosť vrcholu reprezentuje počet citácií daného článku, farba vrcholu predstavuje rok vydania
- algoritmus na výpočet podobnosti je založený na ko-citácii (dve práce sú často citované spolu v iných článkoch) a tiež na „bibliographic coupling“ (dve práce citujú rovnaký iný článok) -> funguje aj pre nové články, ktoré ešte nemajú veľa citácií
- algoritmus tiež berie do úvahy dátum vydania (uprednostňuje články z bližšieho obdobia)
- služba tiež ponúka zoznam „Prior works“ (práce, ktoré sú často citované nájdenými článkami) a „Derivative works“ (práce, ktoré často citujú nájdené články)

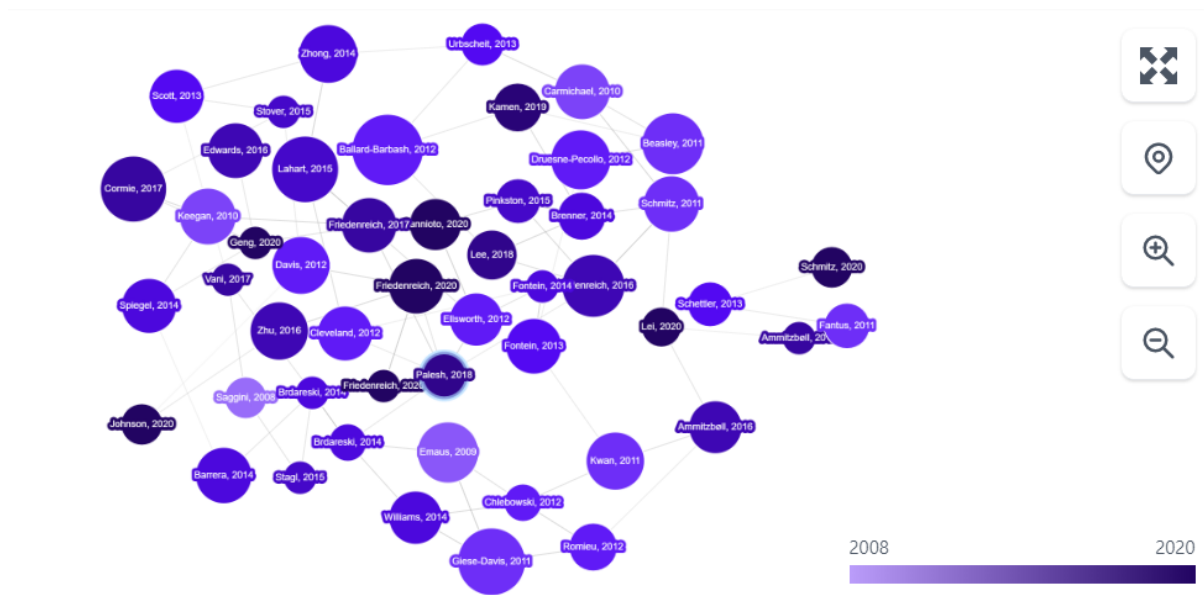


Inciteful

<https://inciteful.xyz/>

- možnosť zadať viacero vstupných článkov
- služba vráti :
 - zoznam podobných článkov zoradené podľa miery podobnosti

- podobné články zoradené podľa PageRank (najdôležitejšie články grafu)
 - nedávne články od top 100 autorov z danej oblasti
 - najdôležitejšie nedávne články
 - top autori, top žurnály z oblasti
- možnosť filtrovať na základe kľúčových slov v nadpisoch
 - založené na citáciach, PageRank, „Link Prediction“ algoritmus, ...



Ďalšie podobné nástroje

- Citation Gecko - <https://www.citationgecko.com/>
- Springer Nature SciGraph - <https://scigraph.springernature.com/explorer>
- ResearchRabbit (potrebná registrácia) - <https://researchrabbitapp.com/>
- Litmaps (potrebná registrácia) - <https://www.litmaps.co/>
- Open Knowledge Map - <https://openknowledgemaps.org/index.php>

Výskumné práce

Keywords: Spracovanie textových dát, extrakcia informácií, vytvorenie znalostných grafov (knowledge graph)

Všeobecný úvod do témy:

https://www.sciencedirect.com/science/article/pii/S1474034619302642?casa_token=CKf4R0IYnQ4AAAA:fWSTVKOHhCtXmrEugUBN4AdI2MdaujgDtYyY_JrsybXFcgOZJ_Vq2RkJfK0mpZLeDfaepc43Fw

Prínos

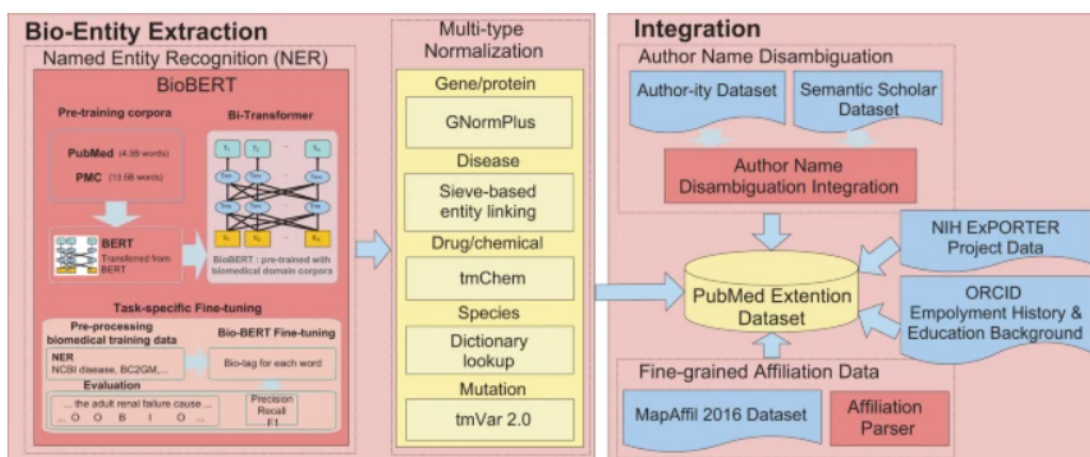
- Zvýšenie dostupnosti informácií z textu
- Rýchle spracovanie veľkého množstva dát
- Zjednodušenie štruktúry textu
- Možnosť spracovania textu automatizovane (počítačom resp. strojovým učením)
- Rýchla kontrola správnosti štruktúry a obsahu textu
- Urýchlenie výberu podobných článkov z danej domény
- Identifikácia vhodných článkov pre meta-analýzy

PubMed knowledge graph (PKG)

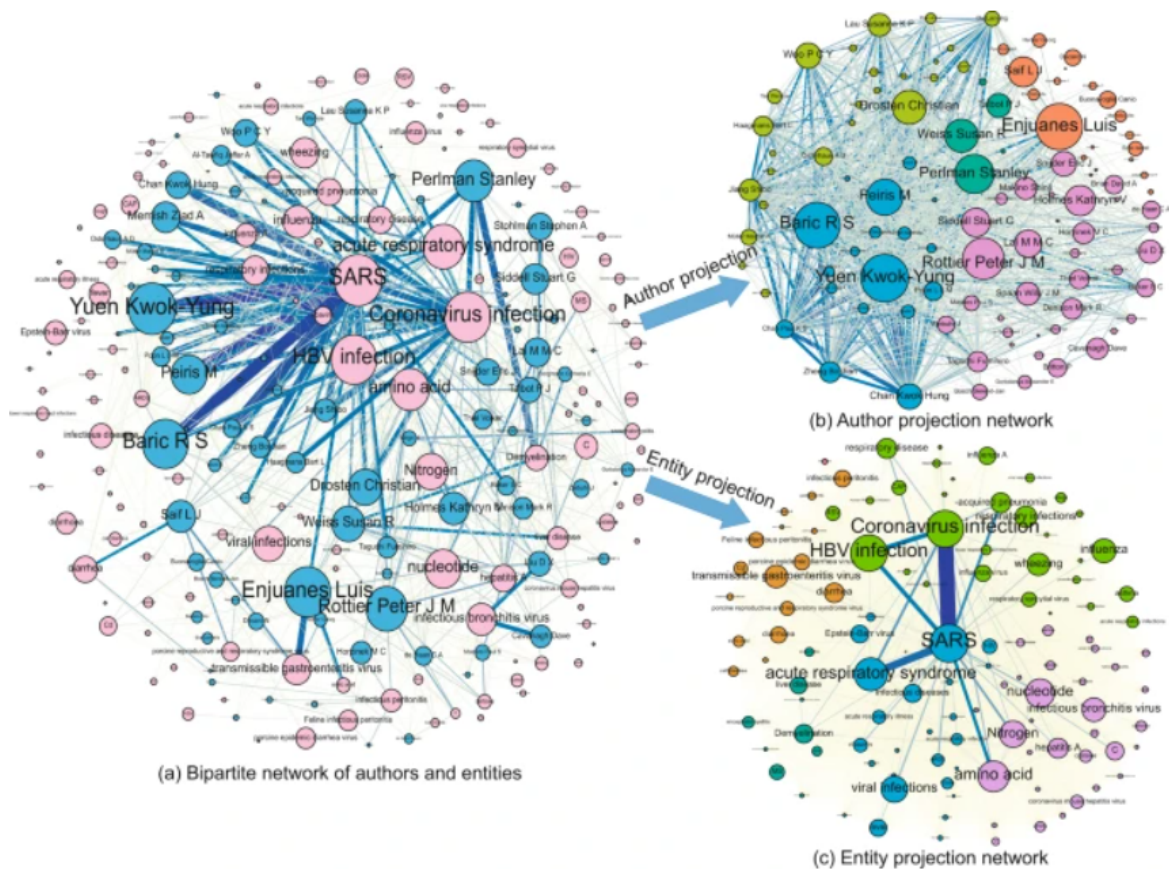
<https://www.nature.com/articles/s41597-020-0543-2>

- Zjednodušenie náročnej a často nejednoznačnej extrakcie informácií
- Potrebné kroky: Extrakcia biologických entít z abstraktov článkov, zistenie mien autorov, ich vzdelania a iných prác, integrácia údajov o financovaní,
- Cieľ: vytvorenie prepojení medzi biologickými entitami, autormi, článkami, pridruženiami a financovaním -> merať vedeckého dopadu, využitie znalostí a prenos znalostí, profilovanie autorov a organizácií

Základný princíp funkcionality nástroja



Ukážka možného výsledku - bipartitná analýza koronavírusu



Bipartite network analysis of coronavirus.

CONSORT-NLP

<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2771515>

- nástroj umelej inteligencie na automatické generovanie checklist hlásenia (reportu) pre paralelné skupiny randomizovaných štúdií.
- Report má mať určitú štruktúru, obsahovať konkrétne informácie vo vopred definovaných častiach podľa CONSORT-statement
- Vstupom je dokument (report), ku ktorému je v aplikácii vytvorený checklist 25 položiek, ktoré má správne napísaný report obsahovať. Farebne je indikovaný stav daného bodu.

Ukážky rozhrania aplikácie

The image displays three screenshots of the CONSORT-NLP application interface, illustrating different views and features.

Article View: The top-left screenshot shows the 'Article View' of a document. The title is 'Hydroxychloroquine Effectiveness in Reducing Symptoms of Hand Osteoarthritis: A Randomized Trial'. The authors listed include Sarah R. Kingsbury, Puvan Tharmanathan, Ada Keding, Sarah J. Ronaldson, Andrew Grainger, Richard J. Wakefield, Catherine Arundel, Fraser Birrell, Michael Doherty, Tonia Vincent, Fiona E. Watt, Krystia Dziedzic, Terence W. O'Neill, Nigel K. Arden, David L. Scott, John Dickson, Toby Garrood, Michael Green, Ajit Menon, Tom Sheeran, David Torgerson, and Philip G. Conaghan. The abstract and background sections are visible.

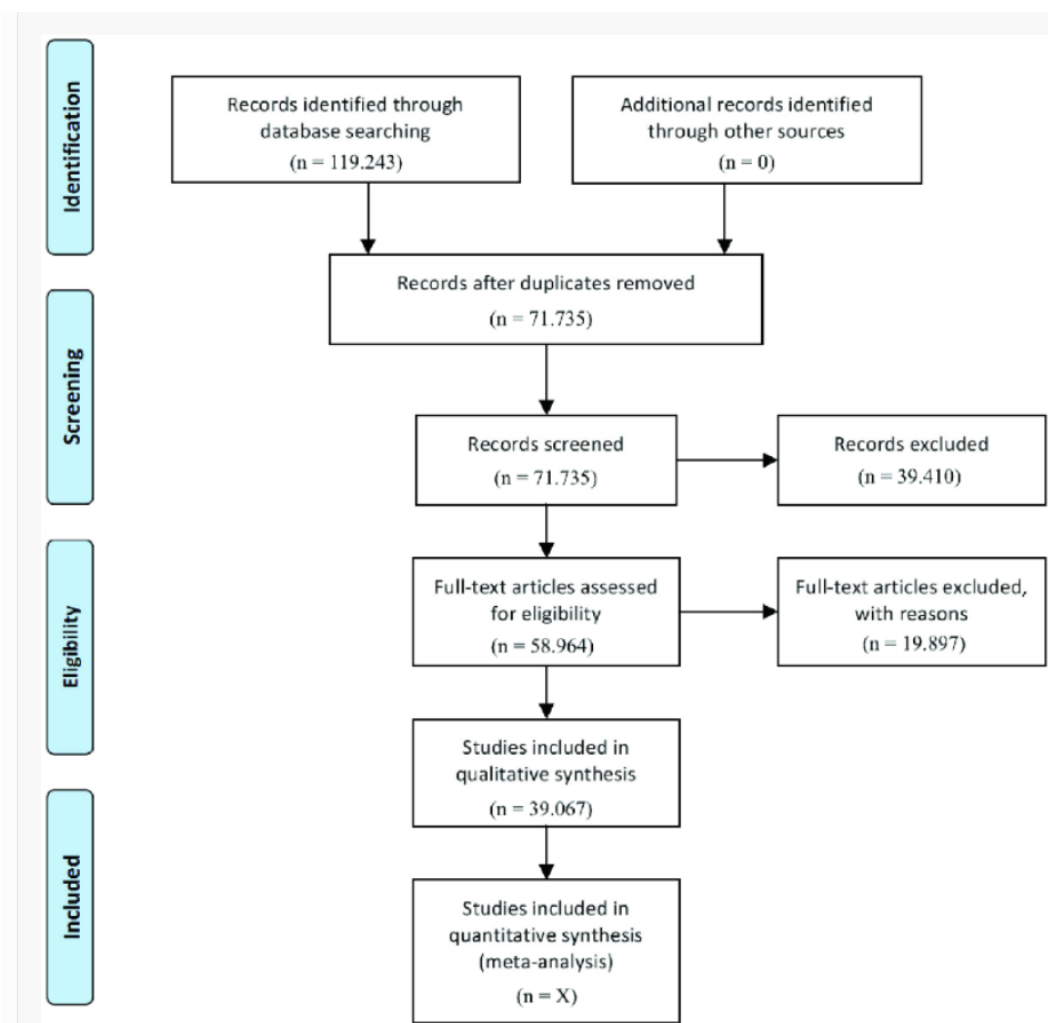
Checklist items (Clickable): The top-right screenshot shows the 'Checklist' panel. It lists 17 items for review, such as '1a. Title', '1b. Abstract', '2a. Background', '2b. Objectives', '3a. Trial Design', '3b. Changes to trial design', '4a. Participants', and '4b. Study settings'. A 'Color Legend' window is open to the right, defining the colors used for checklist items: white for 'Item Not Found', blue for 'Matched Item', cyan for 'Selected Matched Item', pink for 'Notification Item', magenta for 'Selected Notification Item', and grey for 'Item Not Available'.

Matching Sentence View: The bottom screenshot shows the 'Matching Sentence View'. It displays the original text with highlighted segments corresponding to checklist items. For example, '13b. Losses and exclusions' is highlighted in pink, and '14a. Recruitment' is highlighted in magenta. A red arrow points to the '14a. Recruitment' item in the checklist panel on the right, which also has a yellow warning icon. Another red arrow points to the corresponding highlighted text in the original text.

A Case Study in Enhanced Living Environments

https://link.springer.com/chapter/10.1007/978-3-030-10752-9_1#Fig4

- nástroj vytvorený pre spracovanie veľkého množstva textových dát v doméne medicíny
- zameranie sa na výskumné témy v oblasti systémov a služieb ALE/ELE (Enhanced and Assisted living environments) aplikovaných na zdravotnú starostlivosť a pohodu.
- identifikácia potenciálne relevantných článkov pomocou kľúčových slov
- skrátenie a zjednodušenie získavania základného prehľadu a informácií o danej doméne a na urýchlenie procesu prieskumu vedeckých článkov a meta štúdií či analýzy trendov.
- spracovanie prirodzeného jazyka (NLP) a ďalšie metódy softvérového inžinierstva na automatizáciu analýzy vhodnosti a relevantnosti článkov, identifikáciu relevantných článkov, generovanie vizualizácií trendov a vzťahov..
- doménový cieľ: nákladovo efektívne, personalizované monitorovanie, detekcia a odporúčania v reálnom čase pre koncových používateľov aj poskytovateľov zdravotnej starostlivosti

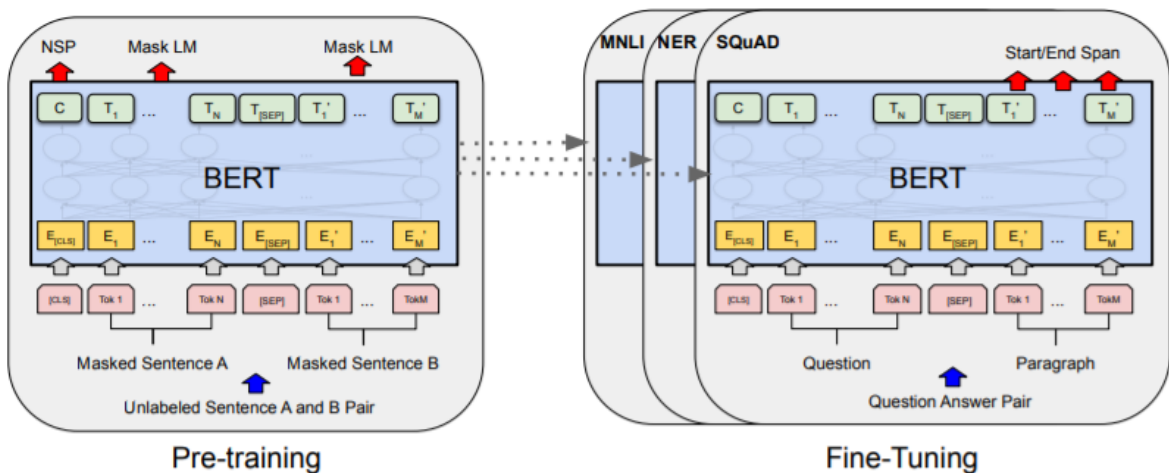


Bert

<https://aclanthology.org/N19-1423.pdf>

BERT -> Bidirectional Encoder Representations from Transformers.

- technika strojového učenia založená pre NLP vytvorená Googlem (využitie v ich search engine)
- architektúra modelu je založená na multi-layer bidirectional Transformer encoder
- dva hlavné kroky:
 - pretrénovanie
 - dve úlohy s neoznačenými dátami:
 - MLM (masked language model) - zakrytie niektorých vstupných tokenov a následná predikcia chýbajúcich slov
 - next sentence prediction - model dostane dvojice viet ako vstup a naučí sa predpovedať, či druhá veta z dvojice je nasledujúcou vetou v pôvodnom dokumente
 - finetuning



<https://arxiv.org/pdf/1905.05950.pdf>

- Cieľ: pochopenie reprezentácie jazyka modelom BERT ako výsledok pretrénovania
- kvantitatívna analýza so zameraním sa na konkrétny encoder

Príloha G - Analýza API pre získavanie článkov

IEEE Xplore API

<https://developer.ieee.org/>

- Poskytuje flexibilné vyhľadávanie a získavanie metadátových záznamov pre viac ako 5 milióny dokumentov.
- V jednom dopyte je možné získať maximálne 200 výsledkov. Dopytový výraz môže obsahovať max 10 slov.
- Aby sme mohli používať toto API, je potrebné sa zaregistrovať a tým získať API kľúč. Tento kľúč musí byť priložený ku každému dopytu.
- Poskytnutá odpoveď je v JSON alebo XML formáte.

Sú k dispozícii 3 typy API:

1. Metadata Search API - záznamy metadát vrátane abstraktov pre viac ako 5 miliónov dokumentov v IEEE Xplore® vrátane časopisov, zborníkov z konferencií, kníh, kurzov a noriem.
2. IEEE Open Access API - fulltextové články označené ako Open Access
3. Digital Object Identifier (DOI) API - Dotazujte sa až na 25 čísel DOI (Digital Object Identifier) a získajte záznamy metadát vrátane abstraktov.

Link na dokumentáciu:

https://developer.ieee.org/docs/read/Searching_the_IEEE_Xplore_Metadata_API

Link na python software development kit:

https://developer.ieee.org/Python_Software_Development_Kit

arXiv API

<https://arxiv.org/>

- arXiv je bezplatná distribučná služba a otvorený archív pre 1 985 422 odborných článkov z oblasti fyziky, matematiky, informatiky, kvantitatívnej biológie, kvantitatívneho financovania, štatistiky, elektrotechniky a systémovej vedy a ekonómie.
- Poskytujú API bez potreby registrácie a API kľúča

<https://arxiv.org/help/api/basics>

- Formát odpovede - Atom - formát založený na xml, ktorý sa bežne používa v informačných kanáloch webových stránok. Je čitateľný človekom a výsledky možno jednoducho prečítať v mnohých webových prehliadačoch.
- Link na python knižnicu: <https://pypi.org/project/arxiv/>

CORE API

- Core <https://core.ac.uk/> je portál, ktorý umožňuje vyhľadávať medzi viac ako 200 miliónmi článkov zhromaždených od 10 397 poskytovateľov údajov z celého sveta.
- Poskytujú API <https://core.ac.uk/services/api>. Na jeho použitie je potrebná registrácia - API kľúč.
- Okrem API je možné údaje stiahnuť ako hromadne, čo umožní spracovať ich na vlastnom počítači alebo v rámci našej infraštruktúry. (Dataset 2020-03-18, veľkosť ~400GB, 2.1TB Extracted)
- Link na python knižnicu: <https://core-api.github.io/python-client/api-guide/document/>

Springer APIs

<https://dev.springernature.com/>

<https://dev.springernature.com/adding-constraints>

- Dostupných cca 650k článkov. Treba sa zaregistrovať ale malo by to byť zadarmo. Vyhľadávanie asi podľa všetkého čo sa dá, titulka, doi, rok, isbn, volný text... aj sortovanie.
- Springer Nature Open Access API by malo vracať aj metadáta aj full text.
- Tu je live dokumentácia na endpointy: <https://dev.springernature.com/documentation>

Majú aj podoobnú prácu ako ideme robiť my:

<https://www.springernature.com/gp/researchers/scigraph>

Elsevier

<https://dev.elsevier.com/>

- vie ťahať články z <https://www.sciencedirect.com/> - pre dopyt "Sigfox" mi našlo dvojnásobok výsledkov oproti IEEE
- Api je zadarmo, len sa treba zaregistrovať
- Dajú sa ťahať abstrakty ale aj celé články
- Majú python knižnicu ktorá rieši implementáciu dopytov - <https://github.com/ElsevierDev/elsapy>

Záver

Analyzovali sme existujúce API služby, ktoré poskytujú dokumenty a články súvisiace s odborom, ktorý študujeme. Po analýze rôznych služieb sme vybrali IEEE Xplore API kvôli dobre zdokumentovanej Python knižnici - https://developer.ieee.org/Python_Software_Development_Kit. Podmienkou bola možnosť využitia jazyka python kvôli prepracovanej NLTK knižnici, ktorú ponúka a ktorú budeme používať pri následnom spracovávaní nájdených článkov.

Príloha H - Analýza backend rámce

Úvod

Backend rámce sa považujú za dôležitú súčasť vývoja interaktívnych webových aplikácií. Kým frontend sa stará o používateľskú skúsenosť, backend je to, čo uľahčuje fungovanie aplikácie. Jedná sa o knižnice jazykov na strane servera, ktoré pomáhajú pri vytváraní serverovej konfigurácie akejkoľvek webovej stránky. Výber správnej technológie pri vývoji webových stránok môže vážne ovplyvniť používanie aplikácie. Rámce pre backend sa zvyčajne zameriavajú na automatizáciu výstupu spojeného s funkciami vývoja softvéru.

Medzi hlavné výhody používania patria:

- Bezpečnosť
- Úspora času
- Integrácie
- Jedinečnosť
- Škálovateľnosť

Medzi najznámejšie rámce patria: Django, Flask, Laravel, ExpressJs, Spring Boot. CakePHP, ASP .NET Core. Výber frameworku závisí do veľkej miery od predchádzajúcich skúsenosti a od komplexnosti a funkčnosti vyvíjanej aplikácie.

Porovnanie vybraných frameworkov:

Django

Jazyk: Python

Známe prípady použitia: Instagram, Mozilla, Spotify

Základné informácie: open-source, rámec na vysokej úrovni, považovaný za najznámejší rámec na backend, MVC architektúra

Výhody: rýchly framework, mnohé funkcie (angl. features), vysoká škálovateľnosť, všestrannosť, ľahko použiteľný admin-panel, jednoduchá správa databáz

Laravel

Jazyk: PHP

Známe prípady použitia: 9GAG, Deltanet Travel

Základné informácie: open-source, MVC - architektúra, integrovaná správa databáz na vysokej úrovni

Výhody: jednoduchá implementácia autentifikácie, predkonfigurované a jednoduché logovanie a správ chýb, PHPUnit pre testovanie, simulácia správania používateľov (klikanie linkov, vyplnenie formulárov, ...), ORM (Eloquent object relational Mapping), jednoduché api (drivers pre Mandril, SMTP, Amazon SES, posielanie mailov), cache backends, vysoká bezpečnosť, dokumentácia a stránka s tutoriálmi, middleware

ExpressJS

Jazyk: NodeJS

Známe prípady použitia: MySpace, GeekList

Základné informácie: využívané ako mean stack spolu s AngularJS pre frontend a NoSQL pre databázy, má MIT licenciu

Výhody: Využíva funkcie Full-stack JS – Node.js ako backend je full-stack JavaScript pre aplikácie na strane servera aj klienta, vysoký výkon (V8 JavaScript engine z Google pre interpretáciu Node.js), komplexné routovanie (aj dynamické URL), využíva middleware, škálovateľnosť, jednoduchá konfigurácia a prispôsobivosť

Spring Boot

Jazyk: Java

Známe prípady použitia: Trivago, Intuit

Základné informácie: open-source,

Výhody: lenivá inicializácia (lazy initialization - vytvorenie beanov na základe požiadaviek), Fluent builder API (SpringApplicationBuilder for building ApplicationContext), vysoká bezpečnosť, dokumentácia a stránka s tutoriálmi, jednoduché testovania (unit a integračné), jednoduchá správa databáz, admin support

Flask

Jazyk: Python

Známe prípady použitia: LinkedIn, Pinterest

Základné informácie: micro web-framework, nepotrebuje špecifické knižnice alebo nástroje, nemá validáciu formulárov, či abstraktnú databázovú vrstvu

Výhody: jednoduchý a ľahký na používanie (najmä po predchádzajúcich skúsenostiach s pythonom), flexibilita, optimálny výkon (pre menej abstrakcie), dostupné rozšírenia pre lepšiu funkčnosť

Príloha I - Analýza získavania sémantických tripletov

Machine learning

Prístup

- využitie algoritmov strojového učenia ako napr. SVM pre získanie tripletov
- potrebný predspracovaný vstup ako aj rozsiahla dátová množina
- pri tréovaní je priradené skóre jednotlivým slovám, ktoré majú byť extrahované ako časť tripletov
- vytvorenie tripletov zo slov s vysokým pozitívnym skóre s prihliadnutím na priradené skóre pre subjekt-predikát(?)-objekt
- zlúčenie tripletov s rovnakými slovami do komplexnejších štruktúr

Naše použitie

- problém nedostupnosti anotovaných dát tripletov
- nutnosť tréovania - časová náročnosť, dáta
- nutné predspracovanie - na tokeny
- graf si vytvárame samy - posledný krok nie je potrebný
- možná implementácia v rôznych jazykoch - aj (náš) python

Tree Bank Parser - Stanford parser

Prístup

- každá veta má priradenú svoju syntaktickú štruktúru
- napr. Stanford parser - <https://nlp.stanford.edu/software/lex-parser.shtml>
- implementácia v Jave -> pravdepodobnostné analyzátory prirodzeného jazyka
- dostupné aj pre ďalšie jazyky
- dostupná vizualizácia
- vstupná veta je rozdelená na tokeny, pričom sú tokeny rozdelené na noun phrase (NP) a verbal phrase (VP) a full stop (ukončenie vety).

Naše použitie

- nie je nutné predspracovanie (stále môže pomôcť)
- možnosť využitia v pythone - napr. pomocou nltk
- ukážka aj v jupyter notebooku
- online ukážky

Ukážka

Vstup:

Histology and histopathology images are very important for diagnosis purposes; these images are a fundamental resource to determine the state of a particular biological structure, to support diagnosis of diseases like cancer, or to analyze anatomy of cells and tissues.

Výstup - priradené zaradenie

Tagging

Histology/NNP and/CC histopathology/NN images/NNS are/VBP very/RB important/JJ for/IN diagnosis/NN purposes/NNS ;/: these/DT images/NNS are/VBP a/DT fundamental/JJ resource/NN to/TO determine/VB the/DT state/NN of/IN a/DT particular/JJ biological/JJ structure/NN ,/, to/TO support/VB diagnosis/NN of/IN diseases/NNS like/IN cancer/NN ,/, or/CC to/TO analyze/VB anatomy/NN of/IN cells/NNS and/CC tissues/NNS ./.

Výstup - stromová štruktúra

Parse

```
(ROOT
  (S
    (S
      (NP
        (NP (NNP Histology))
        (CC and)
        (NP (NN histopathology) (NNS images)))
      (VP (VBP are)
        (ADJP (RB very) (JJ important)
          (PP (IN for)
            (NP (NN diagnosis) (NNS purposes))))))
    (: ;)
    (S
      (NP (DT these) (NNS images))
      (VP (VBP are)
        (NP (DT a) (JJ fundamental) (NN resource)
          (S
            (VP
              (VP (TO to)
                (VP (VB determine)
                  (NP
                    (NP (DT the) (NN state))
                    (PP (IN of)
                      (NP (DT a) (JJ particular) (JJ biological) (NN structure))))))
              (, ,)
              (VP (TO to)
                (VP (VB support)
                  (NP
                    (NP (NN diagnosis))
                    (PP (IN of)
                      (NP
                        (NP (NNS diseases))
                        (PP (IN like)
                          (NP (NN cancer))))))))
              (, ,)
              (CC or)
              (VP (TO to)
                (VP (VB analyze)
                  (NP
                    (NP (NN anatomy))
                    (PP (IN of)
                      (NP (NNS cells)
                        (CC and)
                        (NNS tissues))))))))))
    (. .)))
```

The Multi-Liaison Algorithm

Prístup

- extrahovanie viacerých spojení alebo väzieb medzi subjektom a objektom zo vstupu prirodzeného jazyka (najmä angličtina), ktorý môže mať jeden alebo viac subjektov, predikátov a objektov.
- pomocou stromu analýzy a závislosti, ktoré sú generované napr. Stanford parser
- výstup definuje prepojenie jednotlivých subjektov a objektov ako aj predikátov, ktoré definujú daný vzťah

Naše použitie

- vhodné ako krok v extrakcii tripletov
- dokáže spracovať komplexné vety (vedecké články) a získať triplety
- inšpirácia :
<https://github.com/kj-lai/SentenceTriplet/blob/master/Multi-Liaison%20Algorithm.ipynb>
- pri našom otestovaní bola extrakcia pomerne slabá, dokonca aj na vetách slúžiacich ako príklad, preto bude potrebné skontrolovať prípadné bugy v implementácii

Na základe tejto analýzy ako aj

`TEAM-73_Analyza_moznosti_vytvarania_struktur_z_free_textu` odporúčame pre náš projekt použiť kombináciu prístupov
predspracovanie -> stanford parser -> multi-liaison algoritmus (popr. skúsiť nájsť niečo spoľahlivejšie) -> vizualizácia tripletov.