

Tím 22

Predicar - Dokumentácia pre kontrolný bod č. 1

Tímový projekt 2019/2020

Vedúci tímu:

Ing. Miroslav Rác

Členovia tímu:

Bc. Baláž Branislav

Bc. Bokor Peter

Bc. Harnúšek Ondrej

Bc. Karel Max Dávid

Bc. Letanec Richard

Bc. Vejčíková Veronika

Bc. Vrtal Tomáš

Dokumentácia k riadeniu	5
R1. Úvod	5
R2. Role členov tímu	5
R3. Aplikácie manažmentov	6
Manažment komunikácie	6
Manažment úloh	6
Manažment verziovania	6
Manažment písania kódu	7
Sumarizácie šprintov	7
Šprint 1 - BENZ:	7
Šprint 2 - LAURIN & KLEMENT:	8
Šprint 3 - FIAT:	9
Globálna retrospektíva ZS	10
Čo bolo dobré?	10
Čo je potrebné zlepšiť?	10
Motivačný dokument	11
Tím 22 - Niečo o nás	11
Motivácia pre tému 6 - Predikcia ceny vozidla na základe inzerátu [CarAd]	12
Prečo chceme túto tému?	12
Prečo sme pre tento projekt ten pravý tím?	12
Motivácia pre tému 12 - Inovatívny portál pre boj s antisociálnym správaním s využitím umelej inteligencie [FireAnt]	13
Prečo chceme túto tému?	13
Prečo sme pre tento projekt ten pravý tím?	13
Motivácia pre tému 18 - Dynamické váženie vozidiel počas premávky [WIM]	14
Prečo chceme túto tému?	14
Prečo sme pre tento projekt ten pravý tím?	14
Príloha k motivačnému dokumentu A - Zoradenie tém podľa priority	14
Príloha k motivačnému dokumentu B - Rozvrh tímu	15

Metodiky	15
Metodika komunikácie	15
1. Dedikácia	15
2. Osobné stretnutia	16
3. Slack	16
4. Email	17
Metodika verziovania kódu	17
1. Dedikácia	17
2. Použité technológie	17
3. Commit	17
4. Vetva	18
5. Pull request	19
Metodika písania kódu v jazyku Python	19
1. Dedikácia	21
2. Zaužívané štandardy	21
3. Používanie nápovedí typov (type hints)	21
4. Dokumentovanie kódu	21
Export evidencie úloh	23
Šprint 1	23
Šprint 2	30
Šprint 3	39
Dokumentácia inžinierskeho diela	50
I1. Úvod	50
I1.1 Hlavný cieľ na zimný semester	50
I2. Základná štruktúra systému	50
I2.1 Architektúra	50
I2.2 Dátový model	51
I2.3 Zoznam priložených elektronických dokumentov	52

I3. Popis modulov	52
I3.1 Data Collection	52
I3.2 Text Processing	53
I3.3 Training Center	53
I3.4 Computing Center	53
I3.5 Web App	54
Príloha k dokumentácii inžinierskeho diela A – GUI Design Document	55
1. Úvodná strana (pred vložení inzerátu)	55
Logo	55
Textové pole	55
Tlačidlo pre výpočet ceny	56
Indikátor pre nápovedu	56
2. Analýzy, štatistiky (po vložení inzerátu)	57
Sekcia 1: Základné analýzy	58
Sekcia 2: Nájdene črty	58
Sekcia 3: Detailné analýzy a grafy	58
3. Detailné analýzy a grafy	59
Graf 1a: os cien	59
Graf 1b: os času	59
Graf 2: graf funkcie medzi cenou a časom	60
Analýza 1: podobné/najpodobnejšie vozidlá	61
Graf 3a, 3b: dostupné/predané podobné vozidlá v čase	61
Graf 4: Zobrazenie podobných áut na predaj na mape	62
Graf 5: Pohyb ceny vzhľadom na najazdené kilometre (alebo iný atribút)	63
Graf 6: Sezónnosť vývoja cien podobných vozidiel (napr. cez rok)	64

Dokumentácia k riadeniu

R1. Úvod

Tento dokument popisuje riadenie tímu Predicar (v zoskupení tak ako je popísané v kapitole Role členov tímu) pri práci na projekte CARad v rámci predmetu Tímový projekt. Taktiež obsahuje všetky dokumenty, ktoré súvisia s organizáciou v rámci školského projektu ako sú motivačný dokument.

R2. Role členov tímu

Väčšinu kľúčových povinností sme si v rámci tímu rozdelili tak aby každá pozícia bola obsadená aspoň dvoma členmi. Potom v prípade neprítomnosti vie projekt bezproblémovo napredovať.

Keďže si viacerí z nás chceli vyskúšať pozíciu Scrum mastera striedame sa po jednom šprinte. Číslo šprintu, v ktorom sa daný člen tímu zhostil úlohy je v zátvorke za menom pozície.

Bc. Baláž Branislav

- Machine learning
- Scrum master (3)
- Frontend developer
- Backend developer

Bc. Bokor Peter

- Machine learning
- Project manager
- Databázy
- Backend developer

Bc. Harnúšek Ondrej

- NLP specialist
- Backend developer
- Tester

Bc. Karel Max Dávid

- Administrácia tímového servera
- Git master
- Scrum master (2)

Bc. Letanec Richard

- Backend developer
- Documentation Manager

Bc. Vejčíková Veronika

- NLP specialist

- Správa tímového webu
- Scrum master (1,4)

Bc. Vrtal Tomáš

- Frontend developer
- Backend developer

R3. Aplikácie manažmentov

Na efektívne riadenie projektu sme napísali nižšie popísané metodiky. Tímy by sa mal riadiť každý člen tímu a ich kompletná forma sa nachádza v kapitole metodiky.

Manažment komunikácie

Komunikácia je neoddeliteľnou súčasťou efektívneho riadenia tímu. Na komunikáciu používame výhradne Slack. Tým zabezpečíme, že všetky informácie sú na jednom mieste a spätne dohľadateľné. V Slacku máme vytvorené kanále pričom, každý má vlastný účel a pravidlá popísané v metodike komunikácie. Okrem toho máme integráciu s tromi Slack botmi: Jira, Slack a Standup Alice. Posledný je na zefektívnenie standup meetingov v dňoch keď sa nevieme stretnúť.

Výmenu dokumentov riešime cez tímový Google Drive. Popis jednotlivých kanálov a pravidiel komunikácie máme presne opísanú v "[metodike komunikácie](#)".

Manažment úloh

Na sledovanie stavu šprintu a jeho organizáciu používame nástroj Jira Software, ktorý máme nasadený na tímovom serveri. Každý člen tímu môže pridávať do backlogu *Tasky* alebo *Story*. Následne je na product ownerovi aby sa staral o backlog a zaradoval jednotlivé položky v backlogu podľa priority.

Ohodnotenie úloh sa deje priebežne metódou *planning poker*. Tým chceme doceliť aby sme pri začínaní nového šprintu neminuli príliš veľa času na hodnotenie úloh ale radšej rozhodli a diskutovali čo má byť cieľom šprintu. Pred samotným hodnotením musí každá úloha obsahovať definition of done pre danú úlohu. Existencia jej podúloh nieje nutná.

Na záver šprintu je každá úloha odprezentovaná a product owner rozhoduje či bude akceptovaná alebo nie.

Manažment verziovania

Pri spolupráci viacerých developerov je nutné efektívne zdieľať kód. Na to používame technológiu git a repozitáre máme hostované na github.com. Rozhodli sme sa pre každú mikroslužbu vytvoriť vlastný repozitár, taktiež máme jeden pre tímovú stránku.

Detaily s špecifikácie ako pracovať s touto technológiou máme v "[metodike verziovania](#)".

Manažment písania kódu

Na písanie kódu zatiaľ používame Python. Riadime sa štandardmi PEP (Python Enhancement Proposal) a používame typový Python. Tým chceme docieľiť efektívnejšie prehliadky kódov a zlepšiť čitateľnosť kódov. Typový Python zas používame na

Detailnejšie použitie spomenutých pravidiel je popísané v [metodike písania kódu](#).

Sumarizácie šprintov

Šprint 1 - BENZ:

V tomto šprinte sme používali angličtinu na označenie úloh. Ukázalo sa, že to spôsobuje komplikácie a zdržania pri písaní definícií úloh. Preto sme neskôr prešli na Jiru v slovenčine.

Cieľ: To implement prototype of the microservice architecture.

Trvanie: 11/Oct/19 11:24 AM - 24/Oct/19 3:20 PM

Názov úlohy	Popis	Zoznam podúloh	Odpracovaný čas	Akceptované
[CAR-6] Integrate Jira	As a team, we want to have integrated Jira software with our team website, in order to work efficiently and track our progress.	CAR-11	4h	Áno
[CAR-4] Create functional system prototype	As a product owner, I want a prototype of the system, to validate the proposed system architecture. no need of graphical UI <ul style="list-style-type: none">• just build REST API (json)• be aware of GET/POST/PATCH methods meaning	CAR-29, CAR-32, CAR-33, CAR-34, CAR-35, CAR-36, CAR-37	2d2h10m	Áno
[CAR-3] Submit TP Cup application	As a team, we want to attend TP Cup, in order to win.	CAR-7	1h30m	Áno
[CAR-2] Finalize team web site design	As a team, we need to have a website, to present our work.	CAR-8, CAR-9, CAR-10, CAR-28, CAR-30, CAR-31	7h15m	Áno
[CAR-1] Create	As a product owner,	CAR-12, CAR-	5h40m	Áno

1st design of system architecture	I want to know the architecture of the system, to identify the components that will be used.	13, CAR-14, CAR-15, CAR-16		
-----------------------------------	--	----------------------------	--	--

Šprint 2 - LAURIN & KLEMENT:

Ciel': Chceme zjednodušiť deployment a preskúmať ďalšie možnosti smerovania našej aplikácie z pohľadu používateľa

Trvanie: 24/Oct/19 5:20 PM - 07/Nov/19 3:33 PM

Názov úlohy	Popis	Zoznam podúloh	Odpracovaný čas	Akceptované
[CAR-44] Low fidelity user interface wireframe	<p>As a user I want to see a low fidelity wireframe of a report, so I do know what to expect from the application.</p> <p>DOD:</p> <ul style="list-style-type: none"> visual wireframe <ul style="list-style-type: none"> possible variants well documented <ul style="list-style-type: none"> each component has it's description why is it there what does it tell what kind of data does it need it provides a base for UX survey questions 	CAR-54, CAR-55, CAR-56, CAR-57	1d35m	Áno
[CAR-43] Dockerise the application	<p>As a programmer, I want to easily run up-to-date application with all components, so I can easily add new fancy features.</p> <p>DOD:</p> <ul style="list-style-type: none"> each member can run the application by typing some simple command each repository gets configuration from some environment file (or compose.yaml) 	CAR-47, CAR-48, CAR-49, CAR-67	6h30m	Áno
[CAR-42] Create and connect with	As a user, I want to be provided with reports based on long-term history,	CAR-50, CAR-51, CAR-52, CAR-53	2h30m	Áno

system database prototype	so reports are as precise as possible. DOD: <ul style="list-style-type: none"> • data is persisted • each service which do need to persist data uses mongoDB 			
[CAR-40] Create writing code in Python methodology	I want to know how write and testing code in Python	CAR-60	4h	Áno
[CAR-39] Create code versioning methodology	I want to know how create and name: commit, branch, pull request		1h	Áno
[CAR-24] Integrate Jira and Slack	As a team we need to have Jira and Slack integrated together to be up to date with changes and progress in Jira.	CAR-27, CAR-45, CAR-46	1h	Áno

Šprint 3 - FIAT:

Ciel': Získavanie, predspracovanie a analýza dát.

Trvanie: 07/Nov/19 5:23 PM - 07/Nov/19 3:33 PM

Názov úlohy	Popis	Zoznam podúloh	Odpracovaný čas	Akceptované
CAR-5	Analyze NLP tools for Slovak Language	CAR-63, CAR-64, CAR-73		Áno
CAR-68	Pridat github bota do slacku	----	15m	Áno
CAR-62	Vytvorenie parsera pre inzeráty od Miroslava Ráca	CAR-76, CAR-82, CAR-99		Áno
CAR-59	Extrakcia údajov z textu inzerátu - word2vec	CAR-81, CAR-83	4h	Áno
CAR-75	Tool pre rychle anotovanie a anotovanie dat	CAR-80, CAR-84		Nie – *
CAR-71	Analyzovať portály inzerátov	CAR-77, CAR-		Áno

		78, CAR-88		
CAR-66	Skompletizovať veci na odovzdanie po 9. týždni	CAR-74, CAR-85, CAR-86, CAR-87		Nie – **
CAR-58	Dátová analýza - dĺžka predaja auta	CAR-79		Áno

* - CAR-59 - nebol ukončený z dôvodu zlého nastavenia programu na anotáciu. Vzniknuté dáta boli nepoužiteľné

** - CAR-66 - Vytvorené subtasky nedovoľovali ukončiť úlohu pred skončením šprintu. Konkrétne CAR-87 na odovzdanie dokumentov musí byť vytvorený report z posledného šprintu.

Globálna retrospektíva ZS

Čo bolo dobré?

V rámci sme dokázali už pred začatím práce na projekte rýchlo pripraviť všetky potrebné komunikačné prostriedky. Súčasťou tohto procesu bolo aj definovanie komunikačných kanálov v aplikácii Slack, aby zostali informácie, ktoré si posielame vždy prehľadné.

O trochu neskôr sme tiež zaviedli kanál určený pre standup, pomocou ktorého máme lepší kontakt s postupom práce ostatných v tíme. Časom sme sa tiež naučili dodržiavať pravidlo o povinnej účasti na standup-och v stredu a nedeľu.

Okrem toho sme rýchlo zostavili metodiky, o ktoré sa môžeme pri práci oprieť.

Čo je potrebné zlepšiť?

Počas prvých troch šprintov sa stalo, že sme dorábali veci tesne pred koncom šprintu, kvôli čomu nám zostávalo málo času na dôkladné review. Čakanie na review bolo tiež vo veľa prípadoch dlhšie, než by bolo vhodné.

Na štvrtkových stretnutiach, hlavne pri plánovaní, nám často zostalo málo času, čoho následkom boli nedostatky v plánovaní a nedorozumenia v popisoch úloh, ktoré sme si neskôr museli vysvetľovať.

Retrospektíva bude doplnená až po skončení semestra.

Motivačný dokument

Tím 22 - Niečo o nás

Mailový kontakt: fiit.tp.22@gmail.com

Sme tím pripravený s chuťou využívať svoje poznatky z doterajšieho štúdia a praxe, a učiť sa novým veciam. Radi pri svojej práci používame moderné technológie.

Všetci v tíme máme skúsenosti s programovacím jazykom Python. Štyria z nás mali bakalársku prácu v oblasti strojového učenia alebo dátovej analýzy s použitím Pythonu. Z toho v troch prácach išlo o hlboké učenie v oblastiach spracovania textu a vizuálnych dát. Využívali sme technológie ako Numpy, Scikit-Learn, Tensorflow + Keras, Pandas, Matplotlib, Seaborn, StanfordNLP, Gensim, či BeautifulSoup a nadobudli sme znalosti o neurónových sieťach pre NLP, konvolučných neurónových sieťach na klasifikáciu, segmentáciu a tiež autoenkóderoch. Ďalší členovia tímu vo svojich prácach tiež využívali Python pre podporné účely. Máme tiež skúsenosti s tradičnými aplikáciami strojového učenia ako používanie rozhodovacích stromov, jednoduchých klasifikátorov a zhlukovania, ale aj s predspracovaním, úpravou a generovaním dát, selekciou a vytváraním črt - hlavne z predmetu IAU, ktorý piati z nás absolvovali na bakalárskom stupni štúdia.

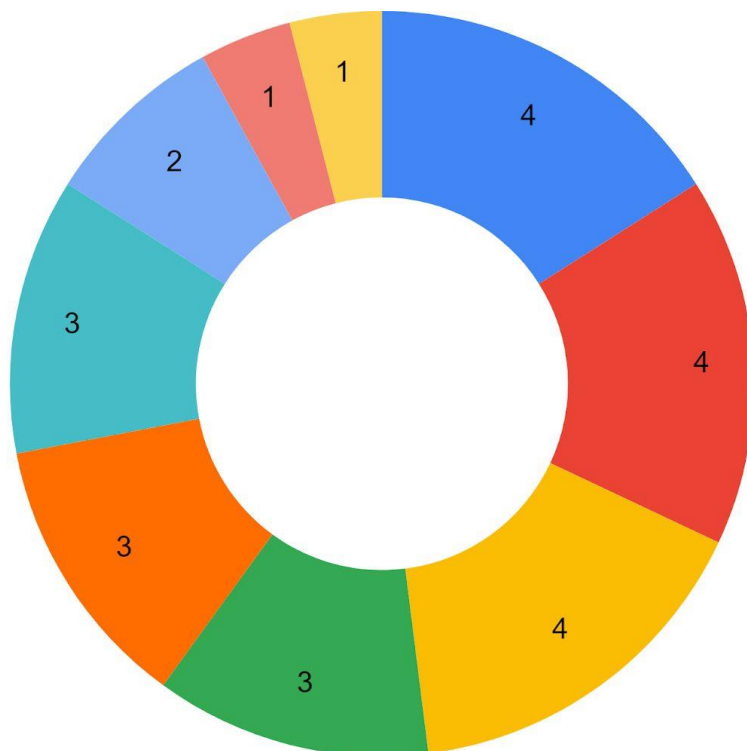
Niektorí z nás sa stretli s jazykom Python v práci, či vo vlastných projektoch, kde sme získali skúsenosti s vývojom backendu webových aplikácií s použitím knižnice Flask alebo grafického používateľského rozhrania s použitím PyQt.

Taktiež sme si všetci prešli základmi z bakalárskeho štúdia: Java, C, C++, SQL, GIT, paralelné a systémové programovanie v C, a niektorí z nás svoje schopnosti využívajú aj v praxi pri práci - MS SQL Server (TSQL).

V praxi sme sa tiež stretli s automatizovaným testovaním, nasadením produktov do prevádzky pomocou technológií Docker, AWS, Jenkins alebo Travis a cloudovými službami.

Skúsenosti máme aj s implementovaním webových aplikácií v prostredí Java s aplikačným serverom Tomcat a využitím webových služieb (REST, SOAP).

- Vyhľadávanie informácií
- Pokročilé databázové technológie
- Štatistické metódy vyhodnocovania experimentov
- Neurónové siete
- Vizualizácia dát
- Objektovo orientovaná analýza a návrh softvéru
- Prírodou inšpirované počítanie
- Aplikačné architektúry softvérových systémov
- Vnorené systémy



Naše vybrané predmety v prvom roku inžinierskeho štúdia

Motivácia pre tému 6 - Predikcia ceny vozidla na základe inzerátu [CarAd]

Prečo chceme túto tému?

Páči sa nám, že projekt má jasne definované použitie - jednoznačne pomenovaný optimalizačný problém: že predaj vozidla chceme docieľiť za minimalizácie času predaja a maximalizácie ceny vozidla, ktoré sú však negatívne korelované. Zároveň v realizácii detailov v rámci tohto cieľa máme voľnú ruku.

Viacerí v tíme máme záujem ďalej sa rozvíjať v oblasti strojového učenia/hlbokého učenia, zberu, spracovania a analýzy dát. Využitie metód spracovania prirodzeného jazyka na slovenčinu považujeme za veľmi zaujímavú výzvu, keďže väčšina dostupných “hotových” riešení pre NLP vhodných pre napr. Angličtinu, pre slovenský jazyk nie je dostupná. Možné nadobudnuté skúsenosti so spracovaním slovenčiny považujeme za veľkú výhodu do budúcnosti pre nás aj pre výskum v slovenskom jazyku.

Taktiež ide o využitie strojového učenia na reálnom použiteľnom “produkte” čo pri mnohých výskumoch napr. v hlbokom učení nie je samozrejmé. Čiže spojenie strojového učenia a reálnej užitočnej aplikácie vnímame ako veľké plus.

Prečo sme pre tento projekt ten pravý tím?

Lebo sú nám blízke a známe spomínané technológie vyžadované pre projekt a Vue.js sa rýchlo naučíme. Viacerých nás zaujíma strojové učenie/hlboké učenie – štyria z

nás mali takto orientovanú BP z toho dvaja z nás spracovanie jazyka – z toho jeden v slovenčine.

To, že táto téma je pre nás naozaj vhodná svedčí aj naša voľba predmetov - až štyria z nás absolvujeme v tomto semestri Vyhľadávanie informácií a traja Neurónové siete. V našom repertoári sa nachádzajú schopnosti pre scrappovanie a crawlovanie webu. Používali sme Docker, máme skúsenosti s Flask Python frameworkom pre backend, REST API a mikroslužbami. V neposlednom rade viacerí z nás mali BP zameranú na neurónové siete a prácu s dátami. Tensorflow a Keras nám nie je cudzí a tiež sme pracovali s knižnicou scikit-learn.

S úlohou vhodnej vizualizácie výsledkov pre podporu informovaného rozhodnutia používateľa si poradíme.

Motivácia pre tému 12 - Inovatívny portál pre boj s antisociálnym správaním s využitím umelej inteligencie [FireAnt]

Prečo chceme túto tému?

Myslíme si, že ide o užitočnú a veľmi aktuálnu tému. Klamstvá šíriace sa na internete nás trápia a naše riešenie by mohlo pomôcť v tejto oblasti. Takto by sme mohli zlepšiť informovanosť širokej spoločnosti, čo by napr. v medicínskej doméne mohlo znamenať lepšiu zdravotnú úroveň.

Viacerí sa zaujímame o oblasť spracovania prirodzeného jazyka.

Páči sa nám sloboda v možnosti implementovať viaceré nástroje pre boj s antisociálnym správaním. Konkrétne nás zaujala myšlienka “sprievodcu” - rozšírenia prehliadača, ako veľmi praktické riešenie pre neskúseného prehliadača internetu, ktorý by takto mohol byť informovaný o kvalite webu a kvalite informácií, s ktorými sa stretáva.

Chceli by sme tento projekt dotiahnuť do podoby, aby bol vhodný na nasadenie do reálnej prevádzky a prístupný verejnosti.

Prečo sme pre tento projekt ten pravý tím?

Lebo viacerí z nás sa zaujímajú o spracovanie prirodzeného jazyka a ešte viac z nás sa zaoberá hlbokým učením. Lebo nás baví získavanie, analyzovanie dát a strojové učenie, o čom svedčí aj náš výber predmetov - Vyhľadávanie informácií a Neurónové siete. Používali sme tiež Docker, máme skúsenosti s Flask Python frameworkom pre backend, REST API a mikroslužbami.

Motivácia pre tému 18 - Dynamické váženie vozidiel počas premávky [WIM]

Prečo chceme túto tému?

Táto téma nás zaujala z dôvodu, že sa jedná o už existujúci projekt, ktorý by sme radi skúsili vylepšiť. Vďaka tomu vieme aj porovnávať nami vytvorené riešenie a taktiež dostať spätnú väzbu od skúsenejších, pôvodných autorov tohto systému. Na projekte sa nám páči, že má reálne využitie a veľkú perspektívu do budúcnosti. Tiež nás láka výzva integrácie nami vytvoreného riešenia do reálneho, už existujúceho systému.

Taktiež je to projekt na ktorom si každý z nás vie nájsť časť v ktorej sa cíti viac doma. Či už je to príprava dát, tréning neurónových alebo vnorené systémy.

Prečo sme pre tento projekt ten pravý tím?

Pri spracovaní tejto témy vieme využiť naše skúsenosti so strojovým učeníím. Viacerí máme skúsenosti so zberom, úpravou a predspracovaním dát a vytváraním číť, či už zo školy z predmetu Inteligentná analýza údajov, kde sme riešili projekt s podobným formátom a postupom ako táto téma, ale aj z bakalárskych prác. Máme skúsenosti s používanými technológiami, hlavne s jazykom Python - z bakalárskych prác a praxe.

S operačným systémom GNU/Linux sme sa všetci stretli na predmete Operačné systémy a niektorí ho používame v praxi/denne. S poslednou technológiou, jazykom C/C++, sme sa v škole stretli všetci, pričom jeden člen tímu mal v C++ bakalársku prácu. Ani kompilovať programy zo zdroja nám nerobí problém.

Rozumieme, že pre mnohé úlohy algoritmické riešenia nepokrývajú všetky prípady, ktoré môžu nastať. Tu sa ponúka potenciál iných metód napríklad využitia hlbokého učenia, s ktorým sme sa viacerí stretli a chceme využiť príležitosť skúsiť ho uplatniť v tomto prípade

Príloha k motivačnému dokumentu A - Zoradenie tém podľa priority

1. Téma č. 6. Predikcia ceny vozidla na základe inzerátu [CarAd]
2. Téma č. 12. Inovatívny portál pre boj s antisociálnym správaním s využitím umelej inteligencie [FireAnt]
3. Téma č. 18. Dynamické váženie vozidiel počas premávky [WIM]
4. Téma č. 8. Vyhľadávanie pomocou obrázkov [ImageSearch]
5. Téma č. 2. Prehľadavanie a vizualizácia textov [TextTool]
6. Téma č. 11. We AR city - Inteligentná komunikácia občana s mestom [city4us]
7. Téma č. 5. Pomoc zvieratám v núdzi [AnimalRescue]
8. Téma č. 19. Automatizácia procesu spracovania a klasifikácie textov [TextProcessing]
9. Téma č. 9. Multibanková mobilná aplikácia pre manažment financií [MultiBank]
10. Téma č. 22. Smart pivovar [SmartBrew]
11. Téma č. 20. Použitelnosť mobilných aplikácií [MobeUX]

12. Téma č. 15. Inteligentný tvorca príbehov z dát [TellStoryAI]
13. Téma č. 1. Event Navigation [EventNav]
14. Téma č. 3. Prostredie na vizualizáciu mikrogridu [GridBox]
15. Téma č. 4. Vizualizácia softvéru vo virtuálnej a rozšírenej realite 2.0 [VizReal]
16. Téma č. 7. Animované architektúry [AnimArch]
17. Téma č. 13. Podpora kvality služieb pre budúci Internet [QoSbySDN]
18. Téma č. 14. Bezpečné manažovanie siete v prostredí Internetu vecí [SecIoT]
19. Téma č. 16. Virtuálna identita [Virtual ID]
20. Téma č. 17. Blockchain platobné brány [BlockPay]
21. Téma č. 21. Kolaboratívne virtuálne prostredie [Coven]
22. Téma č. 10. Databanka otázok a úloh [FIIT-DU]

Príloha k motivačnému dokumentu B - Rozvrh tímu

Deň	8:00 - 8:50	9:00 - 9:50	10:00 - 10:50	11:00 - 11:50	12:00 - 12:50	13:00 - 13:50	14:00 - 14:50	15:00 - 15:50	16:00 - 16:50	17:00 - 17:50	18:00 - 18:50
Pondelok	VINF				NSIETE						
Utorok		VNOS		NSIETE			ASS	VISS	TP1		
Streda	NSIETE	VINF					MTS		MTS		
Štvrtok	VNOS	PPGSO		VINF					SOGAM		
Piatok	PDT			PDT							

Max Dávid Karel	Tomáš Vrhel	Stretnutie s vedúcim
Branislav Baláž	Veronika Vejčíková	Stretnutie tímu
Ondrej Harnúšek	Všetci	
Peter Bokor	Prednáška	
Richard Letanec		

Metodiky

Metodika komunikácie

1. Dedikácia

Metodika definuje princípy a pravidlá komunikácie pre tím Predicar. Opisuje komunikáciu medzi členmi tímu, medzi členmi tímu a vedúcim tímu a komunikáciu s verejnosťou.

2. Osobné stretnutia

Najväčšia časť komunikácie prebieha na spoločných stretnutiach tímu. Formálne stretnutie tímu aj s vedúcim prebieha v štvrtok od 14:00 do 17:00, kde sa hlavne diskutuje o projekte a jeho vývoji.

Okrem toho sa tím stretáva aj v pondelok od 14:00 do 18:00, kde okrem diskusie tím spoločne pracuje na projekte.

3. Slack

Na spoločnú komunikáciu ak nie je možnosť osobného stretnutia využívame nástroj Slack. Každý člen tímu musí pravidelne kontrolovať Slack, či už na mobile alebo počítači - teda každý deň v prípade potreby komunikovať. V tomto nástroji máme viacero komunikačných kanálov:

- **brainstorming** – Kanál, ktorý slúži na zdieľanie akýchkoľvek nápadov či pripomienok.
- **general** – Kanál, ktorý slúži na všeobecnú diskusiu o projekte a úlohách a taktiež aj v prípade, že člen tímu potrebuje radu alebo pomoc.
- **jira** - Kanál, ktorý slúži na notifikácie z nástroja Jira. Nachádzajú sa tu správy o vytvorení, editovaní, priradení a komentovaní príbehu alebo úlohy a ešte správy o tom ak úloha sa dostane do stavu “Ready for Review” a “Done”, čo nám slúži na lepší prehľad o tom čo kto robí.
- **methodics** – Kanál, ktorý slúži na komunikáciu o upravovaní existujúcich a tvorení nových metodík.
- **random** – Kanál, ktorý slúži na komunikáciu netýkajúcu sa tímového projektu.
- **standup** – Kanál, na ktorom každú stredu a nedeľu večer do 20:00 každý člen tímu napíše, na čom aktuálne pracuje a stav úlohy. Na pripomínanie nám slúži Standup Alice bot, ktorý každý deň vyzve každého člena tímu aby vyplnil formulár o jeho činnosti, s tým že v ostatné dni okrem stredy a nedele sa nemusí vyjadriť.
- **tim-web** – Kanál, na ktorom sa riešia všetky veci týkajúceho sa webovej stránky tímu.
- **tp_cup** – Kanál, na ktorom sa rieši všetko súvisiace s TP Cup-om.
- **uzitocne_linky** – Kanál, kde sa pridávajú užitočné odkazy aj s popisom aby sa dali jednoducho kedykoľvek dohľadať.

Okrem týchto kanálov, môže člen využiť priamu komunikáciu s iným členom tímu v súkromnom chate.

4. Email

Náš tím používa spoločný email pre komunikáciu s verejnosťou, odkiaľ sa preposielajú emailové správy všetkým členom tímu - fiit.tp.22@gmail.com. Prístup k tomu mailu majú všetci členovia tímu okrem vedúceho tímu.

Metodika verziovania kódu

1. Dedikácia

Metodika definuje princípy a pravidlá uchovávania a verziovania zdrojového kódu pre tím Predicar. Opisuje použité technológie a spôsob tvorby commitov, vetiev a pull requestov.

2. Použité technológie

Na uchovávanie a verziovanie kódu používame webovú službu Github podporujúcu vývoj softvéru s pomocou verziovacieho nástroja Git.

Na Githube máme vytvorenú organizáciu <https://github.com/predicar>, ktorá obsahuje všetky repozitáre vytvorené našim tímom.

3. Commit

Pomenovaná množina zmien v kóde.

Pravidlá pre vytváranie commitov:

- Commit správa začína vždy slovesom: Add, Fix, Remove, Change, Revert
- Commit správa musí byť pochopiteľná aj bez znalosti daného programovacieho jazyka
- Commit správy sa píšú výlučne v angličtine
- Commitovať treba ucelené, ideálne aj funkčné celky práce
- Jeden commit pokrýva práve jednu zmysluplnú zmenu v kóde
- Zmeny vo formátovaní kódu sú commitované samostatne
- Dĺžka commit správy by nemala presiahnuť viditeľnú dĺžku
- Commit nesmie pokrývať skryté zmeny, ktoré nie sú opísané v správe
- Ak commit súvisí s github issue treba v správe commitu spomenúť vo formáte “#<číslo issue>”
- Ak commit opravuje pripomienky z code review treba v commit správe vo formáte “#<číslo pull requestu>”

- Obsahuje referenciu na Jira task

4. Vetva

Pomenovaná množina commitov určená pre implementáciu konkrétnej funkcionality.

Rozdelenie vetiev:

- master
 - hlavná vetva, ktorá odráža aktuálny stav na produkčnom serveri
 - je zamknutá na priame pushovanie do vetvy, povolené len cez pull requesty z dev vetvy
- dev
 - vývojová vetva, ktorá odráža aktuálny stav na testovacom serveri
 - je zamknutá na priame pushovanie do vetvy, povolené len cez pull requesty z ďalších vetiev
- ďalšie vetvy:
 - vetvy pre implementáciu novej funkcionality:
 - následne sa zlučujú s vetvou dev pomocou pull requestu
 - názov vetvy je Jira issue ID s predponou “feature” (napr. *feature/CAR-1*)
 - pre každú Jira issue je vytvorená maximálne jedna vetva v danom repozitári
 - ak robia viacerí na jednej Jira issue, každý má vytvorenú vlastnú a potom sa mergeje. Názov takejto vetvy je napr *feature/CAR-1/meno*
 - vetvy pre rýchlu opravu chýb:
 - používajú sa len v prípade, že je nevyhnutná okamžitá zmena kódu na produkcii
 - následne sa zlučujú s vetvami master a dev pomocou pull requestu
 - názov vetvy je stručné pomenovanie opravy s predponou “hotfix” (napr. *hotfix/update-db-credentials*)

5. Pull request

Žiadosť o presunutie implementovanej funkcionality z jednej vetvy do inej.

Pravidlá pre vytváranie pull requestov:

- Názov pull requestu je rovnaký ako ID a názov príslušného Jira issue (napr. *CAR-1 Vytvorenie 1. dizajnu systémovej architektúry*)
- Opis pull requestu obsahuje doplňujúce informácie dôležité najmä pri code review
- Názov a opis pull requestu sa píše výlučne v slovenčine
- Každá vetva môže mať maximálne jeden schválený pull request
- Po schválení a zlúčení vetvy sa vetva vymaže (pokiaľ nejde o vetvu dev)

Metodika písania kódu v jazyku Python

Beautiful is better than ugly.

Explicit is better than implicit.

Simple is better than complex.

Complex is better than complicated.

Flat is better than nested.

Sparse is better than dense.

Readability counts.

Special cases aren't special enough to break the rules.

Although practicality beats purity.

Errors should never pass silently.

Unless explicitly silenced.

In the face of ambiguity, refuse the temptation to guess.

There should be one—and preferably only one—obvious way to do it.

Although that way may not be obvious at first unless you're Dutch.

Now is better than never.

Although never is often better than *right* now.

If the implementation is hard to explain, it's a bad idea.

If the implementation is easy to explain, it may be a good idea.

Namespaces are one honking great idea—let's do more of those!

The Zen of Python (PEP 20)

1. Dedikácia

Metodika definuje princípy a pravidlá písania kódu v jazyku Python. Opisuje štýl písania, štandardy ktoré používame a dokumentáciu kódu.

2. Zaužívané štandardy

Pri písaní kódu sa riadime zaužívanými štandardmi [PEP 8](#) (Style Guide for Python Code) pre písanie kódu, [PEP 257](#) (Docstring Conventions) pre dokumentáciu kódu a [PEP 484](#) (Type Hints) pre typovanie kódu.

Kód píšeme v anglickom jazyku, premenným, metódam a triedam dávame výstižný názov.

Názvy premenných a metód píšeme celé malými písmenami, jednotlivé slová oddelené podtržníkom (_).

Názvy tried píšeme štýlom UpperCaseCamelCase.

Funkcie a definície tried sa oddeľujú 2 prázdnymi riadkami. Definície metód v triedach sa oddeľujú 1 prázdny riadok. Na konci súboru musí byť 1 prázdny riadok.

```
class ExampleClass():  
  
    def example_method():  
  
        example_var1 = 1  
        example_var2 = 2  
  
        return example_var1 + example_var2
```

3. Používanie nápovedí typov (type hints)

Pri písaní kódu používame nápovede typov premenných dovážaných do funkcií a typov vracaných funkciami.

```
def example_method(number: int, letter: String) -> String:  
  
    return letter + str(number)
```

4. Dokumentovanie kódu

Komentáre, podobne ako kód, píšeme v **anglickom** jazyku. Riadime sa pri tom štandardom PEP8. Blokové komentáre platia pre kód pod nimi a sú odsadené na rovnakej

úrovni ako kód ktorý komentujú. Každý riadok blokového komentáru začína znakom # a jednou medzerou.

Pri písaní docstringu pre metódy a triedy sa riadime štandardom PEP257 pre docstring konvencie a používame Google štýl písania docstringov. Každý docstring obsahuje krátky a výstižný opis metódy, alebo triedy. Každý docstring opisuje privezené argumenty ako Args, typ návratu ako Returns a vyvolané výnimky ako Raises.

```
def connect_to_next_port(self, minimum):
    """Connects to the next available port.

    Args:
        Minimum(int): A port value greater or equal to 1024.

    Returns:
        int: The new minimum port.

    Raises:
        ConnectionError: If no available port is found.
    """
    if minimum < 1024:
        # Note that this raising of ValueError is not mentioned in the doc
        # string's "Raises:" section because it is not appropriate to
        # guarantee this specific behavioral reaction to API misuse.
        raise ValueError('Minimum port must be at least 1024, not %d.' % (minimum,))
    port = self._find_next_open_port(minimum)
    if not port:
        raise ConnectionError('Could not connect to service on %d or higher.' % (minimum,))
    assert port >= minimum, 'Unexpected port %d when minimum was %d.' % (port, minimum)
    return port
```

Viac o Google štýle písania docstringov:

https://sphinxcontrib-napoleon.readthedocs.io/en/latest/example_google.html

<http://google.github.io/styleguide/pyguide.html>

Export evidencie úloh

Kompletné exporty evidencie úloh z Jira Software sa dajú taktiež nájsť na tímovej stránke.

Šprint 1

[CAR-6] Integrate Jira Created: 10/Oct/19 Updated: 16/Oct/19 Resolved: 16/Oct/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Highest
Reporter:	Peter Bokor	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Management		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	0 minutes
Σ Time Spent:	4 hours	Time Spent:	4 hours
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-11	Set up apache config files	Sub-task	Done	Max Karel
Sprint:	CAR Sprint 1				

Description

As a team, we want to have integrated Jira software with our team website, in order to work efficiently and track our progress.

[CAR-4] Create functional system prototype Created: 10/Oct/19 Updated: 24/Oct/19 Resolved: 24/Oct/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Peter Bokor	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	UserStory		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	2 days, 2 hours, 10 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Issue Links:	Blocks				
	is blocked by CAR-4 Create 1st design of system architecture Done				
Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-29	Create git repositories	Sub-task	Done	Max Karel
	CAR-32	Create crawler microservice prototype	Sub-task	Done	Tomáš Vrtal
	CAR-33	Create text processor microservice pr...	Sub-task	Done	Peter Bokor
	CAR-34	Create predictor microservice prototype	Sub-task	Done	Tomáš Vrtal

	CAR-35	Create manager microservice prototype	Sub-task	Done	Peter Bokor
	CAR-36	Test the prototype as a whole	Sub-task	Done	Peter Bokor
	CAR-37	Create example for a microservice pro...	Sub-task	Done	Ondrej Harnůšek
Sprint:	CAR Sprint 1				

Description

As a product owner,
I want a prototype of the system, to validate
the proposed system architecture.

- no need of graphical UI
 - just build REST API (json)
 - be aware of GET/POST/PATCH methods meaning

[CAR-3] Submit TP Cup application Created: 10/Oct/19 Updated: 16/Oct/19 Resolved: 16/Oct/19

Status: Done
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Highest
Reporter:	Peter Bokor	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Auxiliary		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	1 hour, 30 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-7	Create document	Sub-task	Done	Veronika Vejckikova
Sprint:	CAR Sprint 1				

Description

As a team, we want to attend TP Cup, in order to win.

[CAR-2] Finalize team web site design Created: 10/Oct/19 Updated: 22/Oct/19 Resolved: 22/Oct/19

Status: Done
Project: carAD

Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Highest
Reporter:	Peter Bokor	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Auxiliary		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	7 hours, 15 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-8	Add info about team members	Sub-task	Done	Veronika Vejckikova
	CAR-9	Create documents areas	Sub-task	Done	Tomáš Vrtal
	CAR-10	Review and deploy web site	Sub-task	Done	Max Karel
	CAR-28	Add responsivity for added areas (spr...	Sub-task	Done	Tomáš Vrtal
	CAR-30	Add document for communication method...	Sub-task	Done	Tomáš Vrtal
	CAR-31	Add images and names for sprints to web	Sub-task	Done	Tomáš Vrtal
Sprint:	CAR Sprint 1				

Description As a team, we need to have a website, to present our work.

[CAR-1] Create 1st design of system architecture Created: 10/Oct/19 Updated: 23/Oct/19 Resolved: 23/Oct/19

Status: Done
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Highest
Reporter:	Peter Bokor	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Infrastructure		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	5 hours, 40 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Issue Links:	Blocks				
	blocks	CAR-4	Create functional system prototype	Done	
Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-12	Create DB model	Sub-task	Done	Peter Bokor
	CAR-13	Create architecture overview	Sub-task	Done	Ondrej Harnůšek
	CAR-14	Create diagram for automated data col...	Sub-task	Done	Richard Letanec
	CAR-15	Create diagram for user interaction	Sub-task	Done	Richard Letanec
	CAR-16	Create diagram for automatic model ma...	Sub-task	Done	Branislav Baláž
Sprint:	CAR Sprint 1				

Description As a product owner,

I want to know the architecture of the system, to identify the components that will be used.

Šprint 2

[CAR-44] Low fidelity user interface wireframe Created: 24/Oct/19 Updated: 07/Nov/19 Resolved: 07/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Miroslav Rác	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	UserStory		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	1 day, 35 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-54	Create GUI Wireframe	Sub-task	Done	Branislav Baláž
	CAR-55	Document GUI Design	Sub-task	Done	Veronika Vejckova
	CAR-56	Put together questions for UX research	Sub-task	Done	Branislav Baláž
	CAR-57	Design GUI	Sub-task	Done	Richard Letanec
Sprint:	CAR Sprint 2				

Description

As a **user** I want to see a *low fidelity wireframe of a report*, so I do know what to expect from the application.

DOD:

- visual wireframe
 - possible variants
- well documented
 - each component has it's description
 - why is it there
 - what does it tell
 - what kind of data does it need
- it provides a base for UX survey questions

Comments

Comment by Peter Bokor [07/Nov/19]

prezijem

[CAR-43] Dockerise the application Created: 24/Oct/19 Updated: 07/Nov/19 Resolved: 07/Nov/19

Status: Done
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Medium
Reporter:	Miroslav Rác	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Infrastructure		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	6 hours, 30 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-47	Create Dockerfile for each repository	Sub-task	Done	Tomáš Vrtal
	CAR-48	Link Python configuration with enviro...	Sub-task	Done	Peter Bokor
	CAR-49	Create compose.yaml file	Sub-task	Done	Ondrej Harnůšek
	CAR-67	Mock microservices for manager testing	Sub-task	Done	Ondrej Harnůšek
Sprint:	CAR Sprint 2				

Description

As a **programmer**,

I want to easily run up-to-date application with all components, so I can easily add new fancy features.

DOD:

- each member can run the application by typing some simple command
- each repository gets configuration from some environment file (or compose.yaml)

[CAR-42] Create and connect with system database prototype Created: 24/Oct/19 Updated:

07/Nov/19 Resolved: 07/Nov/19

Status: Done
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Medium
Reporter:	Ondrej Harnůšek	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	UserStory		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	2 hours, 30 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Description

As a user,
I want to be provided with reports based on long-term history, so
reports are as precise as possible.

DOD:

Issue Links:	Blocks
	is blocked by CAR-49 Create compose.yaml file Done

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-50	Create MongoDB instance	Sub-task	Done	Max Karel
	CAR-51	Link all services with MongoDB	Sub-task	Done	Peter Bokor
	CAR-52	Analysis: what services do need to pe...	Sub-task	Done	Richard Letanec
	CAR-53	Load data from persistent volume	Sub-task	Done	Max Karel
Sprint:	CAR Sprint 2				

- data is persisted
- each service which do need to persist data uses mongoDB

[CAR-40] Create writing code in Python methodology Created: 22/Oct/19 Updated: 01/Nov/19 Resolved: 01/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Ondrej Harnúšek	Assignee:	Veronika Vejckikova
Resolution:	Done	Votes:	0
Labels:	Auxiliary		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	4 hours	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-60	Prepisat existujuci kod podla python ...	Sub-task	Done	Peter Bokor
Sprint:	CAR Sprint 2				

Description

I want to know how write and testing code in Python

[CAR-39] Create code versioning methodology Created: 22/Oct/19 Updated: 07/Nov/19 Resolved: 06/Nov/19

Status:	Done
----------------	------

Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Max Karel	Assignee:	Ondrej Harnůšek
Resolution:	Done	Votes:	0
Labels:	Auxiliary		
Remaining Estimate:	0 minutes		
Time Spent:	1 hour		
Original Estimate:	Not Specified		

Sprint:	CAR Sprint 2
Tester:	Max Karel
Reviewer:	Max Karel

Description

I want to know how create and name: commit, branch, pull request

[CAR-24] Integrate Jira and Slack Created: 17/Oct/19 Updated: 29/Oct/19 Resolved: 29/Oct/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Veronika Vejckikova	Assignee:	Max Karel
Resolution:	Done	Votes:	0
Labels:	Management		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	1 hour	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-27	Create HTTPS certificate	Sub-task	Done	Max Karel
	CAR-45	Reconfigure Jira	Sub-task	Done	Max Karel
	CAR-46	Setup Jira/Slack bot	Sub-task	Done	Max Karel
Sprint:	CAR Sprint 2				

Description

As a team we need to have Jira and Slack integrated together to be up to date with changes and progress in Jira.

Šprint 3

[CAR-71] Analyzovať portály inzerátov Created: 07/Nov/19 Updated: 21/Nov/19 Resolved: 21/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Branislav Baláž	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	None		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	4 hours, 20 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Issue Links:	Blocks				
	blocks	CAR-72 Implementovať parsery na vybrané portály			To Do
	blocks	CAR-89 Stiahnutie inzerátov z auto-moto inze...			To Do
Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-77	Analyzovať vhodné portály	Sub-task	Done	Veronika Vejcikova
	CAR-78	Zdokumentovať nájdené portály	Sub-task	Done	Veronika Vejcikova
	CAR-88	Párová dokumentácia k tasku CAR-78	Sub-task	Done	Richard Letanec
Sprint:	CAR Sprint 3				

Description

Ako tím, chceme poznať portály z ktorých môžeme zbierať data.

DoD:

Budeme vedieť, ktoré portály chceme ďalej spracovávať.

[CAR-68] Pridať github bota do slacku Created: 01/Nov/19 Updated: 07/Nov/19 Resolved: 07/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Max Karel	Assignee:	Max Karel
Resolution:	Done	Votes:	0
Labels:	None		
Remaining Estimate:	0 minutes		
Time Spent:	15 minutes		
Original Estimate:	Not Specified		

Sprint: CAR Sprint 3

Description

Ako tím chceme byť informovaný o udalostiach na našich repozitároch

[CAR-62] Vytvorenie parsera pre inzeráty od Miroslava Ráca Created: 31/Oct/19 Updated: 20/Nov/19 Resolved: 20/Nov/19

Status: Done
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Medium
Reporter:	Ondrej Harnúšek	Assignee:	Peter Bokor
Resolution:	Done	Votes:	0
Labels:	None		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	3 hours	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Issue Links:	Blocks				
	blocks	CAR-99	Naplniť databázu na produkčnom serveri		Done
Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-76	Parser pre dáta	Sub-task	Done	Peter Bokor
	CAR-82	Nasadiť aplikáciu na server	Sub-task	Done	Max Karel
	CAR-99	Naplniť databázu na produkčnom serveri	Sub-task	Done	Max Karel
Sprint:	CAR Sprint 3				

Description

Ako vlasník produktu, chcem aby boli

spracované zozbierané inzeráty aby sa dali

d'alej spracovávať ako text

Potrebne uloženie spracovaných dát, paralelizácia, naplnenie do db

Vysledne dáta obsahujú text inzerátu.

[CAR-59] Extrakcia údajov z textu inzerátu - word2vec Created: 31/Oct/19 Updated: 20/Nov/19 Resolved: 20/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Miroslav Rác	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	None		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	4 hours	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Issue Links:	Blocks is blocked by CAR-75 Tool pre rychle anotovanie a anotovan... In Progress				
Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-81	Implementovať word2vec	Sub-task	Done	Ondrej Harnúšek
	CAR-83	Analyzovať možnosti Word2Vec	Sub-task	Done	Ondrej Harnúšek
Sprint:	CAR Sprint 3				

Description

- predspracovanie textu: ◦ tokenizacia ◦ lematizacia ◦ diakritika?
- natréňovanie word2vec
- klasifikácia čísel do tried podľa ich blízkeho okolia ◦ kilometre
◦ ročník ◦ cena
- evaluácia podľa ručne anotovaných dát

DoD:

Extrahovaná čiara z inzerátu pomocou word2vec

[CAR-58] Dátová analýza - dĺžka predaja auta Created: 31/Oct/19 Updated: 21/Nov/19 Resolved: 21/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Miroslav Rác	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	dátová-analýza		
Σ Remaining Estimate:	Not Specified	Remaining Estimate:	Not Specified
Σ Time Spent:	Not Specified	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-79	Analýza pre zistenie dĺžky predaja áut	Sub-task	Done	Peter Bokor
Sprint:	CAR Sprint 3				

Description

Ako používateľ chcem vedieť ako dlho sa zvkynú predávať autá, aby som vedel robiť rozhodnutie pri nacenení.

DoD:

Výsledné čísla z dátovej analýzy.

[CAR-5] Analyze NLP tools for Slovak Language Created: 10/Oct/19 Updated: 21/Nov/19 Resolved: 21/Nov/19

Status:	Done
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Veronika Vejckikova	Assignee:	Unassigned
Resolution:	Done	Votes:	0
Labels:	Spike		
Σ Remaining Estimate:	0 minutes	Remaining Estimate:	Not Specified
Σ Time Spent:	30 minutes	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-63	Stretnúť sa s Ing. Miroslav Blšták, P...	Sub-task	Done	Veronika Vejckikova
	CAR-64	Zostaviť zoznam pravidiel pre slovník...	Sub-task	Done	Branislav Baláž
	CAR-73	Pripraviť sa na stretnutie	Sub-task	Done	
Sprint:	CAR Sprint 3				

Description

Ako vlastník produktu chcem preskúmať možnosti automatického spracovania slovenského jazyka za účelom vytvorenia prototypu spracovania textu DoD:

Dokument opisujúci možnosti NLP a zdokumentované odskúšané prístupy.

Comments

Comment by Miroslav Rác [11/Oct/19]


<http://arl6.library.sk/nlp4sk/api> - kontakt je Miroslav Blstak

Nesplnené stories:

[CAR-75] Tool pre rychle anotovanie a anotovanie dat Created: 07/Nov/19 Updated: 21/Nov/19

Status: In Progress
Project: carAD
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type:	Story	Priority:	Medium
Reporter:	Branislav Baláž	Assignee:	Unassigned
Resolution:	Unresolved	Votes:	0
Labels:	None		
Σ Remaining Estimate:	Not Specified	Remaining Estimate:	Not Specified
Σ Time Spent:	Not Specified	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Attachments:	 image-2019-11-21-12-46-47-881.png				
Issue Links:	Blocks				
	blocks	CAR-59	Extrakcia údajov z textu inzerátu - w...		Done
Sub-Tasks:	Key	Summary	Type	Status	Assignee

	CAR-80	Anotovať dostupné dáta	Sub-task	In Progress	Tomáš Vrtal
	CAR-84	Analyzovať nástroje pre anotovanie dát	Sub-task	Done	Tomáš Vrtal
Sprint:	CAR Sprint 3				
Reviewer:	Branislav Baláž				

Description

DoD:

500 anotovaných inzerátov.

Dôvod: zlou analýzou nástroja bol použitý zle – výstup bol nevyhovujúci - nepoužiteľný

[CAR-66] Skompletizovať veci na odovzdanie po 9. týždni Created: 31/Oct/19 Updated: 21/Nov/19

Status:	In Progress
Project:	carAD
Component/s:	None
Affects Version/s:	None
Fix Version/s:	None

Type:	Story	Priority:	Medium
Reporter:	Peter Bokor	Assignee:	Max Karel
Resolution:	Unresolved	Votes:	0
Labels:	None		
Σ Remaining Estimate:	Not Specified	Remaining Estimate:	Not Specified
Σ Time Spent:	Not Specified	Time Spent:	Not Specified
Σ Original Estimate:	Not Specified	Original Estimate:	Not Specified

Sub-Tasks:	Key	Summary	Type	Status	Assignee
	CAR-74	Zistiť čo sa odovzdáva	Sub-task	Done	Max Karel
	CAR-85	Vytvoriť dokumentáciu inžinierskeho d...	Sub-task	Done	Branislav Baláž
	CAR-86	Vytvoriť dokumentáciu riadenia projek...	Sub-task	Done	Max Karel
	CAR-87	Odovzdať všetko	Sub-task	In Progress	Branislav Baláž
Sprint:	CAR Sprint 3				

Description

DoD:

Pripravené veci na odovzdanie podľa pokynov na stránke

Dôvod: Zle definované splnenie story pre odovzdanie po 9. týždni - pre jeho dokončenie bol potrebný export po skončení šprintu, takže nemohol byť splnený.

Dokumentácia inžinierskeho diela

I1. Úvod

Tento dokument opisuje realizáciu nášho zadania v rámci predmetu Tímový projekt na tému “Predikcia ceny vozidla na základe inzerátu [CarAd]”. Projekt vyvíjame pod názvom Predicar a ide o produkt, ktorý má na základe textu inzerátu odhadnúť cenu vozidla a poskytnúť iné užitočné štatistiky alebo odhady spojené práve s predajom daného vozidla.

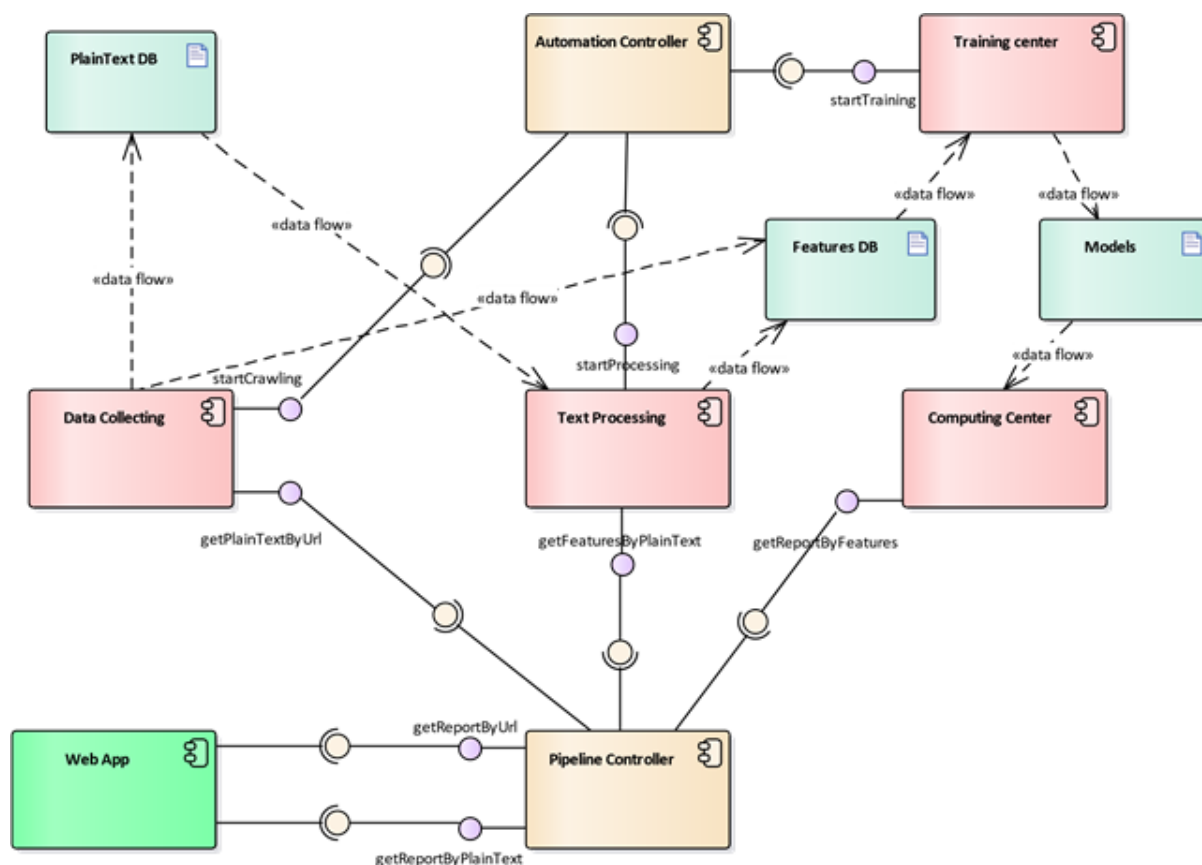
I1.1 Hlavný cieľ na zimný semester

“Chceme vidieť vypočítanú predikciu ako odhad ceny v našom rozhraní pre štandardný vstup.”

I2. Základná štruktúra systému

V tejto kapitole ponúkame náhľad na nami navrhnutú architektúru systému. Okrem toho zhrnieme aj dátový model, podľa ktorého uchováваме v našom systéme dáta o vozidlách a jednotlivých inzerátoch.

I2.1 Architektúra



Obrázok 1 Základný pohľad na architektúru systému Predicar

Náš systém je zameraný na strojové učenie, pomocou čoho dokáže odhadnúť ceny vozidiel. Kvôli tomuto si architektúra vyžaduje časť zameranú na **zber dát**. Tieto dáta potom uchováame v dokumentovej databáze.

Strojové učenie sa v našej architektúre objavuje v dvoch inštanciách: **spracovanie textu** pre extrakciu číť z textu inzerátu a samotný **model pre odhad cien**. Zatiaľ čo spracovanie textu nepotrebujeme sústavne trénovať, ak dokáže správne interpretovať texty inzerátov, odhad cien musíme pravidelne trénovať, aby odhady cien vždy brali do úvahy aktuálne dáta. Preto je súčasťou našej architektúry aj **trénovacie centrum**.

Interakcia so systémom je zabezpečená cez jednoduché **webové rozhranie**. Pomocou tohto rozhrania môže používateľ zadať odkaz na inzerát alebo jeho samotný text a pozrieť sa na odhady a štatistiky od nášho produktu.

Okrem týchto častí musíme pre systém zabezpečiť aj iné **riadiace moduly**, ktoré sú znázornené na obrázku 2.

I2.2 Dátový model

Dátový model nášho systému pozostáva z jednej entity, keďže sme zvolili dokumentovú databázu ako úložisko našich dát. Tieto dáta odrážajú vlastnosti vozidla a kolekciu inzerátov, ktoré sa naň vzťahujú. Údaje, ktoré o vozidlách uchováame sú nasledovné:

- **VIN:** VIN číslo predstavuje unikátny identifikátor vozidla
- **licence_plate:** zoznam ŠPZ značiek pre vozidlo s daným VIN číslom
- **price:** výsledná cena vozidla agregovaná z cien nájdených v inzerátoch pre dané vozidlo
- **features:** vlastnosti vozidla, ktorých presný zoznam určíme pri zbere dát
- **ads:** zoznam objektov s nasledovnou štruktúrou, ktoré reprezentujú nájdené inzeráty pre vozidlo:
 - **url:** unikátny odkaz na inzerát
 - **version:** verzia inzerátu
 - **created_at:** čas vytvorenia inzerátu
 - **ad_price:** cena uvedená v inzeráte
 - **sold:** dátum predaja alebo prázdna hodnota, ak ešte predané nebolo
 - **ad_features:** vlastnosti vozidla získané z daného inzerátu
 - **parsed_raw_text:** znenie inzerátu
 - **raw_html:** html obsah stránky s inzerátom

I2.3 Zoznam priložených elektronických dokumentov

- **Príloha A - GUI Design Document** - Návrh rozhrania webovej stránky pre interakciu so systémom

I3. Popis modulov

I3.1 Data Collection

Analýza	Produkt si vyžaduje, aby sme vždy pracovali s aktuálnymi dátami. Preto potrebujeme zabezpečiť neustály zber dát z viacerých portálov na predaj áut cez inzerát.
Návrh	Časť systému zodpovedná za zber dát sa skladá z viacerých komponentov: <ul style="list-style-type: none"> • Processing queue - udržiava odkazy na inzeráty, ktoré treba spracovať • New URL Discovery - objavuje nové odkazy na inzeráty • Data Updater - aktualizuje stav inzerátov v databáze • Downloader - stiahne obsah stránky • Parser - vytiahne zo stránky znenie inzerátu
Implementácia	Zber dát, teda crawler, je

	implementovaný ako mikroslužba volaná riadiacim modulom. Jazyk: Python
--	--

I3.2 Text Processing

Analýza	Pre spracovanie textu potrebujeme vhodnou metódou implementovaný model pre spracovanie textu v slovenčine. Modul Text processor bude do databázy ukladať nájdené črty, cez ktoré budeme môcť trénovať predikčný model.
Návrh	Pre spracovanie jazyka v slovenčine sú pre nás relevantné 2 možnosti: slovníková metóda a strojové učenie s použitím word2vec.
Implementácia	Text processing je implementovaný ako mikroslužba volaná riadiacim modulom. Jazyk: Python

I3.3 Training Center

Analýza	Model pre odhad cien musí vždy odrážať aktuálny stav trhu, preto musíme zabezpečiť, že tento model sa bude pravidelne učiť na aktuálnych dátach. Toto tréningovanie chceme spúšťať v pravidelných intervaloch.
---------	--

I3.4 Computing Center

Analýza	Pre odhad cien je potrebný regresný model, ktorý pomocou strojového učenia dokáže tieto odhady generovať. Okrem toho potrebujeme iné modely, aby sme získali dáta pre vizualizácie zhrnuté v priloženom elektronickom dokumente.
Návrh	Natrénované modely máme uložené

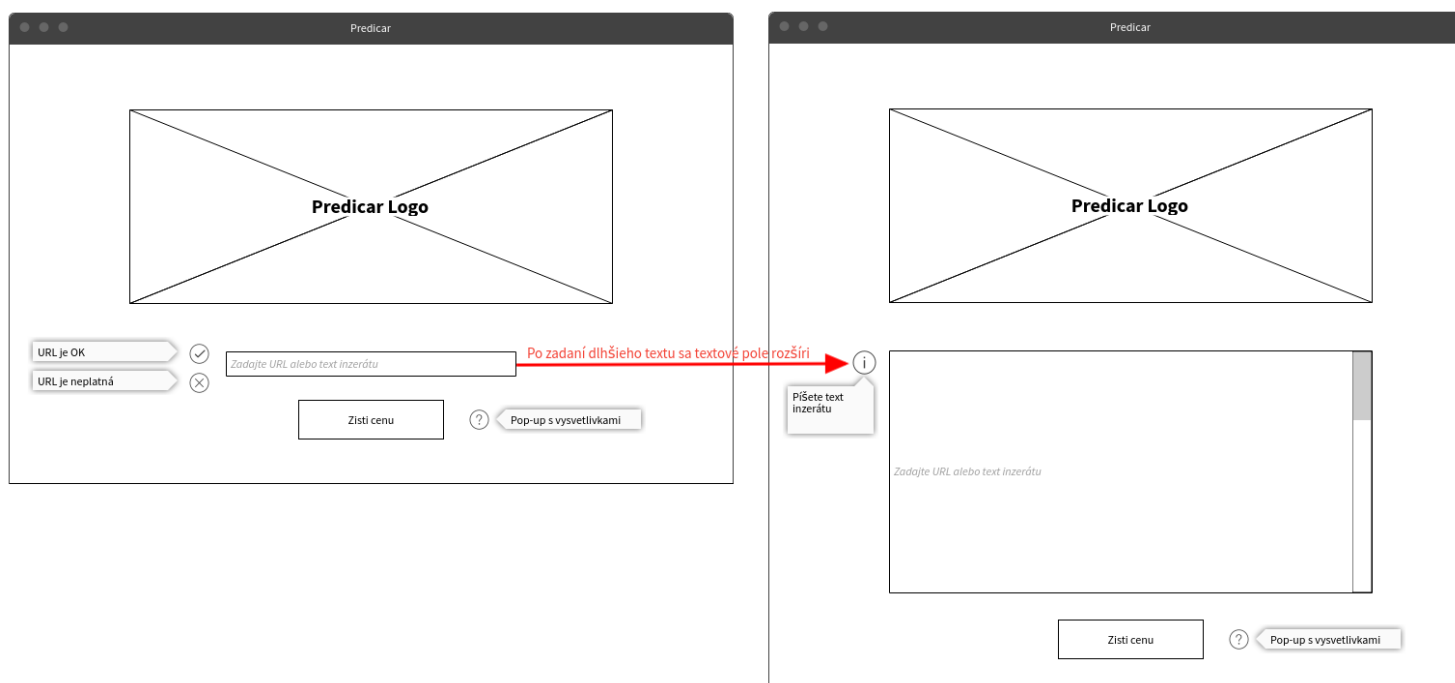
	v pamäti, pričom ich trénovaním získané váhy uchováame v databáze.
Implementácia	Výpočtové centrum je implementované ako mikroslužba volaná riadiacim modulom. Jazyk: Python

I3.5 Web App

Analýza	Pre intuitívnu interakciu so systémom chceme implementovať používateľské rozhranie vo forme webovej stránky, ktorá poskytne prehľadnú časť, kam používateľ môže zadať samotné znenie alebo odkaz na inzerát vozidla. Potom mu taktiež musí prezentovať vypočítané dáta.
Návrh	Návrh rozhrania webovej stránky nájdete medzi priloženými elektronickými dokumentmi.

Príloha k dokumentácii inžinierskeho diela A – GUI Design Document

1. Úvodná strana (pred vložením inzerátu)



Úvodná strana má byť jednoduchá pre intuitívne použitie. Bude obsahovať: logo, textové pole s informačným indikátorom pre obsah, tlačidlo pre výpočet ceny, indikátor pre nápovedu k použitiu.

Logo

Pôjde o logo projektu. Malo by byť výstižné a zapamätateľné.

Textové pole

Textové pole ako vstup môže prijímať buď URL inzerátu alebo text inzerátu. V úvodnej podobe má textové pole veľkosť jedného riadka.

Ak dôjde k detekcii, že je zadávaná URL adresa, overí sa, či ide o nám známy autobazár, ktorý vieme spracovať. Ak áno, informačný indikátor pre obsah bude zelený s popisom, že vložená URL je OK a vieme z nej získať dáta. Ak URL nevieme spracovať, indikátor bude červený s popisom "URL je neplatná".

Ak dôjde k detekcii, že do textového poľa je vkladán text, informačný indikátor pre obsah bude modrý s informáciou, že sa vkladá text. Textové pole sa od určitej dĺžky textu predĺži pre lepší používateľský zážitok pri vkladaní textu. Predĺženie bude konečného počtu riadkov a pre zobrazenie dlhšieho textu bude použitý scroll bar.

Tlačidlo pre výpočet ceny

Toto tlačidlo aktivuje výpočet predikcií. Jeho umiestnenie a popis by mal byť taký aby jeho účel bol jednoznačný.

Indikátor pre nápovedu

Nachádza sa vedľa tlačidla pre vypočítanie ceny. Slúži na zorientovanie sa menej počítačovo zdatného človeka alebo návštevníka, ktorý s našou aplikáciou ešte nemal skúsenosť. Po jeho stlačení sa pod úvodnou sekciou rozbalí stručný text v rozsahu cca 1-2 odseky opisujúci používanie a účel aplikácie. Po opätovnom stlačení sa znova zavrie. Po stlačení tlačidla pre výpočet predikcií sa tiež automaticky zavrie.

2. Analýzy, štatistiky (po vložení inzerátu)

Časť s výpisom štatistík sa zobrazí po stlačení tlačidla pre ich výpočet pod úvodnou stránkou.

?

CENA

ČAS

% ÚROVEŇ SPOĽAHLIVOSTI

Tu budú niekoľkými vetami opísané hlavné body reportu.
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Nunc maximus, nulla ut commodo sagittis, sapien dui mattis dui.
non pulvinar lorem felis nec erat.

^

NÁJDENÉ ČRTY

Upravitel'né polia

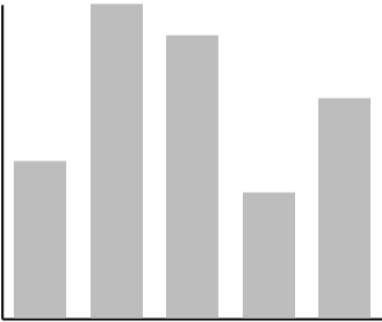
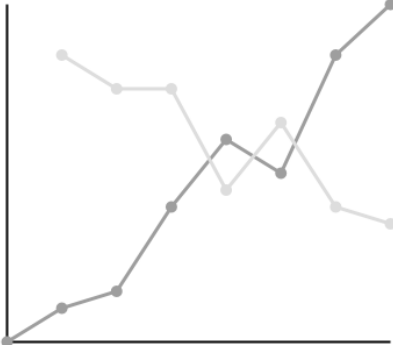
Názov črty	Nájdená hodnota
Model	
Značka	
Najazdené km	
Vek	
...	

Prepočítaj

^

?

PREDIKCIE A ŠTATISTIKY



^

?

Časť s výpisom štatistík sa skladá z 3 sekcií:

1. Základné analýzy
2. Nájsené črty
3. Detailné analýzy a grafy

Sekcia 1: Základné analýzy

Účel tejto sekcie je poskytnúť základnú predikciu. Bude zobrazovať odhadovanú cenu a čas, za ktorý odhadujeme predaj vozidla a spoľahlivosť tohto odhadu.

Cieľom je poskytnúť základný smerodajný odhad aby používateľ nadobudol predstavu akú cenu a čas môže približne očakávať. Tiež táto analýza môže postačovať používateľovi, ktorý nemá záujem o hlbšie analýzy, štatistiky a grafy o predajnosti vozidiel.

Sekcia 2: Nájsené črty

Táto sekcia zobrazí črty, ktoré boli z inzerátu identifikované. Údaje budú vypísané prehľadne, pod sebou.

Účelom je informovať používateľa, na základe akých parametrov boli určené predikcie. Druhým prínosom pre používateľa je možnosť upresniť charakteristiku vozidla ručným vpísaním parametrov, ktoré neboli z textu úspešne identifikované alebo ich text neobsahoval.

Používateľ bude informovaný, že upresnením charakteristiky vozidla môže získať presnejšie predikcie. Bude kladený dôraz na atribúty, ktoré majú najvyššiu váhu na cenu vozidla.

Po vyplnení dodatočných atribútov sa vypočíta nová predikcia pomocou tlačidla prepočítaj.

Sekcia 3: Detailné analýzy a grafy

Posledná sekcia bude obsahovať detailné analýzy, štatistiky, výpisy podobných vozidiel atď. Bude slúžiť v prípade, keď si užívateľ bude chcieť lepšie preskúmať stav a vývoj na trhu vozidiel v rôznej miere podobných.

3. Detailné analýzy a grafy

V tejto kapitole opíšeme analýzy a grafy, ktoré navrhujeme použiť v tretej sekcii. Opisy obsahujú popis, odôvodnenie použitia a aké dáta sú preň potrebné. Poradie, v akom sú zoradené grafy neurčuje ich prioritu a poradie na stránke.

V našom predikčnom systéme bude možné využiť časti z opisov z tohto dokumentu po úpravách aj ako vysvetlenie a interpretáciu pre používateľa.

Graf 1a: os cien

Pre informáciu o škále, v ktorej sa pohybujú podobné vozidlá chceme pre ne zobrazit' os cien a umiestniť na ňu konkrétne vozidlo pre porovnanie s ostatnými podobnými vozidlami.



Účelom grafického zobrazenia je informovať používateľa a dať mu predstavu, v akom cenovom rozmedzí sa pohybuje auto z jeho inzerátu a jemu podobné autá.

Pre toto zobrazenie potrebujeme zistiť podobné vozidlá a získať ich min, avg a max cenu.

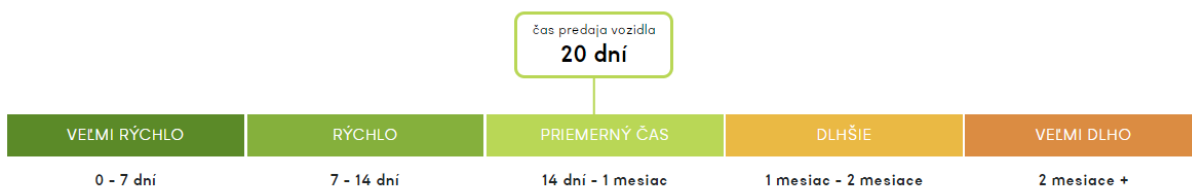
Prichádzajú do úvahy 2 možné varianty/podoby tohto typu grafu:

- A. interval medzi min a max cenou sa rozdelí na intervaly rovnomernej dĺžky
- B. interval medzi min a max cenou sa rozdelí na quartily (alebo n quantilov)

Graf 1b: os času

Zobrazenie obdobnej osi pre čas ako v Grafe 1a.

Informácie o čase predaja vozidla



Účelom grafického zobrazenia je informovať používateľa a dať mu predstavu, v akom časovom horizonte sa odhaduje doba predaja auta z vloženého inzerátu voči časom predaja podobných áut.

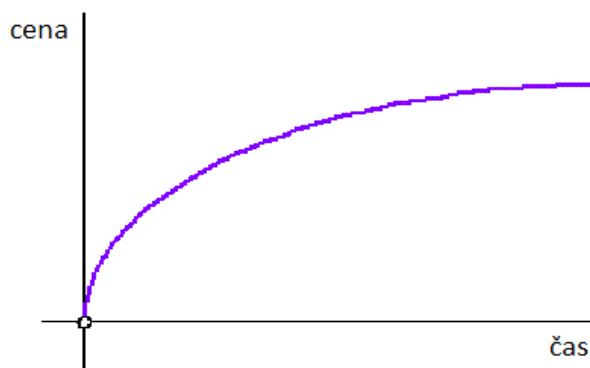
Pre toto zobrazenie potrebujeme zistiť podobné vozidlá a získať ich min, avg a max cenu alebo môže existovať jednotná časová os bez ohľadu na podobné vozidlá pre celý systém pre škálu od 0 do max dĺžku času predaja, ktorú budeme evidovať v systéme.

Taktiež prichádzajú do úvahy 2 možné varianty/podoby tohto typu grafu:

- C. interval medzi min a max sa rozdelí na časové intervaly trvania rovnomernej dĺžky
- D. interval medzi min a max sa rozdelí na quartily (alebo n quantilov)

Graf 2: graf funkcie medzi cenou a časom

Tento graf by mal zobrazovať vzájomný predpovedaný vývoj ceny a času predaja vozidla. Pôjde o rastúcu funkciu - čím rýchlejšie chcem predat' auto tým musí byť cena nižšia. Pri "hoveri" nad krivkou by sa mala vypísať cena a čas v danom bode. Graf nemusí byť spojitý ako zobrazuje obrázok - stačia diskkrétne hodnoty v granularite po dňoch.



Vyjadruje funkciu vzťahu medzi cenou a časom.

Účelom tohto zobrazenia je informovať používateľa a dať celkovú predstavu o vzájomnom vývoji vzťahu medzi cenou a časom predaja daného auta.

Pre toto zobrazenie potrebujeme predikovať cenu za ktorú sa predá vozidlo pre všetky časové body, ktoré chceme vykresliť.

Analýza 1: podobné/najpodobnejšie vozidlá

System zobrazí informácie k stavu predaných a dostupných podobných áut.

Cieľom tejto analýzy je odbremeniť používateľa od interpretácie grafov a abstraktných štatistík a zobrazíť mu konkrétne “hmatateľnejšie” výsledky.

Používateľovi budú zobrazené 3 najpodobnejšie predané autá. Výpis auta bude obsahovať základné parametre, cenu a čas, za ktorý sa predalo a dátum kedy sa predalo (pre prípadné zohľadnenie vývoja cien). Tiež mu budú zobrazené 3 momentálne dostupné podobné autá obdobným spôsobom. V prípade, že používateľ predáva auto aby vedel s akou konkurenciou súperí na momentálnom trhu.

Taktiež bude zobrazená tabuľka celkových podobných dostupných a predaných áut.

	<u>Teraz na predaj</u>	<u>Predané spolu</u>	?
<u>Počet áut</u>	24	251	
<u>Priemerná cena</u>	5860	6320	

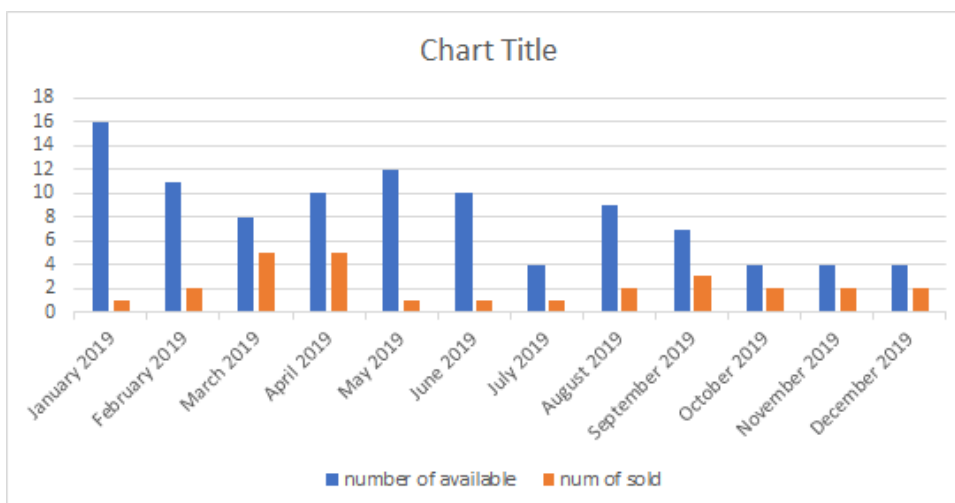
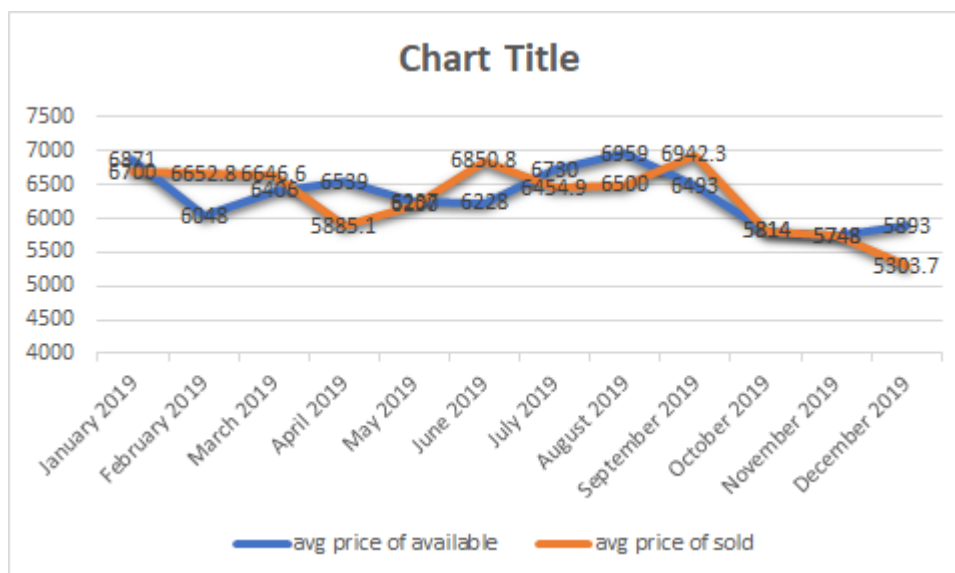
Pre toto zobrazenie budeme potrebovať zoznam predaných a dostupných podobných áut - ich počty a priemer cien a top 3 najpodobnejšie z každej skupiny.

Rozšírením tejto analýzy by mohla byť skupina tabuliek - jedna tabuľka pre región Slovenska (kraj).

Graf 3a, 3b: dostupné/predané podobné vozidlá v čase

Súvisia s Analýzou 1. Narozdiel od momentálneho stavu budú zobrazovať vývoj v čase pre:

- priemernú cenu podobných dostupných (3a)
- priemernú cenu podobných predaných (3a)
- počet podobných dostupných áut (3b)
- počet podobných predaných áut (3b)



Grafmi chceme informovať používateľa o vývoji na trhu, aké sú trendy pre podobné autá - či pribúdajú alebo ubúdajú podobné ponuky, ako sa vyvíja cena, či sa drží v rovnakej hladine alebo výraznejšie klesá a oplatí sa mu nejaký čas počkať.

Pre toto zobrazenie budeme potrebovať záznamy o predaných a dostupných autách - ich cenách a počte pre intervaly zvolenej časovej granularity (napr. pre každý mesiac).

Graf 4: Zobrazenie podobných áut na predaj na mape

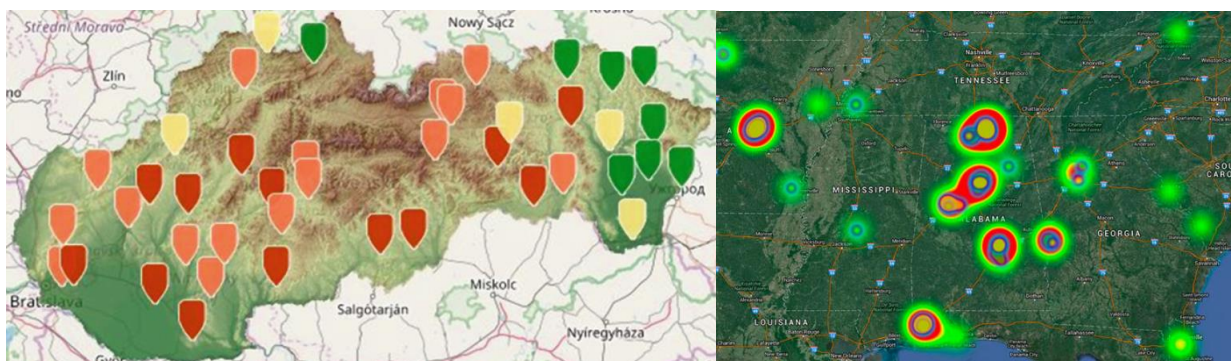
Pôjde o grafické zobrazenie, ktoré bude ukazovať, kde sa predávajú podobné autá na mape slovenska.

V prípade predaja auta bude tak predajca vedieť, či sú predajcovia podobných áut pre neho reálnou konkurenciou - či predávajú v rovnakom regióne alebo nie. V prípade

nákupu auta si bude používateľ môcť pozrieť, v akých regiónoch je ponuka na podobné auto analyzovanému inzerátu.

Vizualizácia môže mať 2 podoby - podľa počtu áut:

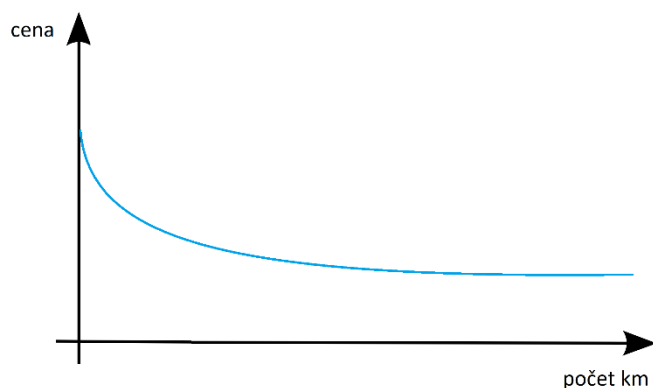
- body na mape - farba identifikátorov môže indikovať cenu, pri "hoverti" nad bodom sa môže zobrazíť presná cena vozidla alebo iné stručné informácie
- heatmapa - ak bude podobných vozidiel veľa - nebude možné zobrazíť jednotlivé body - heatmapa bude symbolizovať lokality, v ktorých regiónoch je aké množstvo podobných áut na predaj



Pre zobrazenie na mape bude potrebný zoznam podobných áut s geografickou lokalitou ich ponuky.

Graf 5: Pohyb ceny vzhľadom na najazdené kilometre (alebo iný atribút)

Graf bude zobrazovať vzťah ako vplýva počet najazdených kilometrov na vývoj ceny (pri štandardnej fixnej dĺžke predaja).



Používateľ tak bude môcť vidieť aký kompromis medzi ojazdenosťou a cenou sa mu oplatí urobiť. Či napríklad od určitého množstva najazdených kilometrov ešte cena výrazne klesá alebo nie a má zmysel uvažovať o takom vozidle.

Pre túto vizualizáciu bude potrebné predikovať vývoj ceny podobných vozidiel vzhľadom na množstvo najazdených km pri fixnej dobe predaja (určenej bežnej dĺžky predaja o čom používateľa informujeme).

Graf rovnakého typu bude možné zostrojiť aj pre iné parametre vozidla, o ktorých zistíme, že majú najvyšší vplyv na vývoj ceny (napr. rok výroby, výkon, objem motora, ...).

Graf 6: Sezónnosť vývoja cien podobných vozidiel (napr. cez rok)

Hypotéza: Existujú trendy v cenách - obdobia v roku, kedy sa autá predávajú vo všeobecnosti drahšie/lacnejšie. (Např. v období zimných mrazov alebo pred letnými dovolenkami sa ľudia môžu rozhodnúť že potrebujú lepšie auto a to môže mať vplyv na ponuku a ceny na trhu) Túto možnosť je potrebné najprv preskúmať či dochádza k javu že ceny áut kolíšu v rámci roka. Ak takýto jav existuje, môžeme ho vizualizovať.

Vizualizácia by zobrazovala priemernú cenu áut pre dané obdobia vo všeobecnosti. Používateľ by tak vedel, kedy je dobré obdobie na kúpu auta - kedy je na trhu najlepšia cena/najväčší výber pre podobné autá.

Pre túto vizualizáciu by bola potrebná štatistika vývoja cien v priebehu niekoľkých rokov.