

Podiel práce

	Hodiny	Podiel práce
Branislav Baláž	8h	11.8%
Peter Bokor	16h	23.6%
Ondrej Harnúšek	8h	11.8%
Max Dávid Karel	4h	6%
Richard Letanec	6h 30m	9.5%
Veronika Vejčíková	14h	20.6%
Tomáš Vrtal	11h 15m	16.7%
SPOLU	67h 45m	100 %

Úspešne dokončené stories

[CAR-187] [Vytvorit prvu verziu technickej dokumentacie](#) Created: 17/Mar/20
Updated: 31/Mar/20 Resolved: 31/Mar/20

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Max Karel](#)
Resolution: Done
Labels: None
Σ Remaining Estimate: 0 minutes
Σ Time Spent: 6 hours, 50 minutes
Σ Original Estimate: Not Specified

Priority: Medium
Assignee: [Richard Letanec](#)
Votes: 0

Remaining Estimate: 0 minutes
Time Spent: 20 minutes
Original Estimate: Not Specified

Sub-Tasks:

Key	Summary	Type	Status	Assignee
CAR-197	Opravit a doplnit docstringy pre vset... Vygenerovat technicku dokumentaciu z ...	Sub-task	Done	Richard Letanec
CAR-198	Z ...	Sub-task	Done	Richard Letanec

Sprint: CAR Sprint 8

Description

Ako tim chceme vygenerovat prvu verziu technickej dokumentacie

Akceptacne kriteria:

- Skontrolovat a doplnit docstringy pre všetky existujúce metódy a classy
 - Vygenerovat technickú dokumentáciu do formátu pdf alebo html
-

[CAR-185] [Parsovanie URL obrazkov z inzeratov](#) Created: 17/Mar/20 Updated: 30/Mar/20 Resolved: 30/Mar/20

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Tomáš Vrtal](#)
Resolution: Done
Labels: None
Remaining Estimate: 0 minutes
Time Spent: 4 hours, 55 minutes
Original Estimate: Not Specified

Priority: Medium
Assignee: [Tomáš Vrtal](#)
Votes: 0

Sprint: CAR Sprint 8

Description

Ako pouzivatel chcem vidiet pre zadany inzerat (a jemu podobne inzeraty) realne fotky auta.

Akceptacne kriteria:

- Pre 4 bazare (okrem auto.bazar.sk) pridane parsovanie URL obrazkov (fotiek auta) do pola (napr. atribut 'ad_images') a su ukladane do db
- Ukladat len URL obrazkov auta

[CAR-184] Kompresia archivnych suborov Created: 17/Mar/20 Updated: 31/Mar/20 Resolved: 30/Mar/20

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Max Karel](#)
Resolution: Done
Labels: None
Remaining Estimate: 0 minutes
Time Spent: 4 hours
Original Estimate: Not Specified

Priority: Medium
Assignee: [Max Karel](#)
Votes: 0

Sprint: CAR Sprint 8

Description

Ako **pouzivatel** nechcem stratit informacie a histori u o novych inzeratoch. Ak su vsak archivne subory prilis velke a niesu skomprimovane zaberaju prilis vela miesta na disku. V takom pripade budu nove data zahodene z dovodu limitacie miesta na produkcnom serveri. To by malo byt vyriesene kompresiou archivnych suborov, ktora bude prebiehat vzdy o jednej rano

Akceptacne kriteria:

- Subor bude skomprimovany
- Povodna verzia bude vymazana

[CAR-183] [BE] Optimalizovať DB Refill Created: 17/Mar/20 Updated: 31/Mar/20 Resolved: 24/Mar/20

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Branislav Baláž](#)
Resolution: Done
Labels: None
Remaining Estimate: 0 minutes
Time Spent: 1 day, 1 hour, 25 minutes
Original Estimate: Not Specified

Priority: Medium
Assignee: [Branislav Baláž](#)
Votes: 0

Sprint: CAR Sprint 8

Description

Ako **používateľ** chcem aby pri prípadnom vypadku alebo potreby pretrenovania prebehlo nactenie dat do databazy co najrychlejsie. Momentalne nactenie je vsak prilis pomale.

Akceptačné kritériá:

- Hľadanie AAA auto záznamov nesmie spomaľovať beh skriptu

[CAR-179] [NLP] Spracovanie freetextu bazosa Created: 10/Mar/20 Updated: 31/Mar/20 Resolved: 31/Mar/20

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Veronika Vejcikova](#)
Resolution: Done
Labels: None
Σ Remaining Estimate: 0 minutes
Σ Time Spent: 2 days, 6 hours, 5 minutes
Σ Original Estimate: Not Specified

Priority: Medium
Assignee: [Veronika Vejcikova](#)
Votes: 0

Remaining Estimate: 0 minutes
Time Spent: 1 hour, 30 minutes
Original Estimate: Not Specified

Sub-Tasks:

Key	Summary	Type	Status	Assignee
CAR-180	Analýza spracovania textu bazosa	Sub-task	Done	Veronika Vejcikova
CAR-181	Spracovanie vstupu z DB SK-specific	Sub-task	Done	Veronika Vejcikova
CAR-194	NLP predspracovanie textu...	Sub-task	Done	Veronika Vejcikova
CAR-195	Natrenovanie embeddingov	Sub-task	Done	Veronika Vejcikova
CAR-196	Analýza natrenovaných embeddingov	Sub-task	Done	Ondrej Harnúšek

Sprint: CAR Sprint 8

Description

Ako **používateľ** chcem aby boli v predikcii zahrnuté aj auta a informácie z bazaru ktorý nieje štruktúrovaný. Na to je potrebné aby vznikol vhodný model s natrenovanými dátami, ktorý bude neskôr použiteľný v extrakcii črt inzerátov z bazos.sk

Preto ako tím musíme poznať možnosti pre spracovanie fulltext inzerátov z bazosu aby sme vedeli nejakú rozumne porovnať auta ktoré nemajú štruktúrované atribúty medzi sebou.

Akceptačné kritéria:

- natrenovať na textoch embeddings pre slova
- spracovanie anotovaných dát
- klasifikácia tokenov na základe anotovaných entít

vystup: model a jednoducha klasifikacia

"Tasky"

- poriesienie vstupnych dat
 - SK NLP predspracovanie (Blstakova vec alebo nieco)
 - vyber vhodneho modelu/trenovanie/embeddings ako vystup
 - klasifikacia (a nejaky zaver z toho)
-

**[CAR-118] [Zefektívnenia práce s databázou pri API requestoch](#) Created: 28/Nov/19
Updated: 28/Mar/20 Resolved: 28/Mar/20**

Status: Done
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Max Karel](#)
Resolution: Done
Labels: None
Remaining Estimate: 0 minutes
Time Spent: 4 hours, 30 minutes
Original Estimate: Not Specified

Priority: Medium
Assignee: [Tomáš Vrtal](#)
Votes: 0

Sprint: CAR Sprint 8

Description

Ako **používateľ** chcem aby boli poskytované API efektívne a volanie netrvalo príliš dlho. V súčasnom stave trvá načítanie 100 najnovších inzerátov približne minútu. Momentálne je celá databáza stiahnutá do pamäte čo nie je najefektívnejší spôsob.

Pri hľadaní 100 najnovších inzerátov treba upraviť spôsob hľadania tak aby zoradzoval inzeráty podľa `date_modified` v najvyššom leveli dokumentu.

Užívateľ vie špecifikovať limit koľko aut API posiela. Číslo môže byť od 1 do 100

Akceptačné kritéria:

- Zrýchlenie API pre nájdenie 100 najnovších inzerátov
- Pre každé auto vrátiť celý mongo dokument (všetky inzeráty)

Neúspešne dokončené stories

[CAR-176] [ML] [Hľadanie podobnosti cez vektory](#) Created: 10/Mar/20 Updated: 19/Apr/20

Status: To Do
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Max Karel](#)
Resolution: Unresolved
Labels: None
Σ Remaining Estimate: 0 minutes
Σ Time Spent: 1 week, 2 days, 6 hours
Σ Original Estimate: Not Specified

Priority: Medium
Assignee: [Peter Bokor](#)
Votes: 0

Remaining Estimate: 0 minutes
Time Spent: 1 week, 2 days, 6 hours
Original Estimate: Not Specified

Sub-Tasks:

Key	Summary	Type	Status	Assignee
CAR-188	Vytvorit z atributov auta feature vector Rozumne	Sub-task	Done	Peter Bokor
CAR-189	zmensit dimenzionalitu vektor... Rozumne	Sub-task	In Progress	
CAR-190	zmensit dimenzionalitu vektor... Vyskusat	Sub-task	To Do	
CAR-191	zistovanie podobnosti na men...	Sub-task	In Progress	
CAR-192	Vygenerovanie vektoru pre vsetky auta... analiza -	Sub-task	To Do	
CAR-193	vytvaranie vektorov pre nov... upravit	Sub-task	To Do	
CAR-227	endpoint na pouzivanie pocita...	Sub-task	To Do	
CAR-228	Natrenovat finalny model na vsetkych ...	Sub-task	To Do	
CAR-229	Vytvorit a naplnit novu kolekciu urce...	Sub-task	To Do	
CAR-	Vytvorenie	Sub-task	To Do	

Description

Ako **zakaznik** chcem aby boli najdene **podobne** inzeraty ku mne vybranemu inzeratu.

Podobne auto: Kategoricke hodnoty - by mali byt rovnake. Spojite - co najblizsie

Endpoint - o polnoci sa vsetkym inzeratom ktore nemaju vypocitany feature vektor tento vektor vypocita a pripoji sa do tabulky v pamati

Akceptacne kriteria:

- Kazde auto ma vypocitany feature vector
- Budeme vediet vypocitat skore podobnost dvoch aut
- Pre dane auto vieme najst 5 najpodobnejsich aut
- Model bude pouzivat triplet loss
- Endpoint pouziva vektory podobnosti vypocitane modelom, ktore su ulozene v tabulke v pamati

[CAR-186] [Opravenie logovania pre multiprocessing](#) Created: 17/Mar/20 Updated: 14/Apr/20

Status: In Progress
Project: [carAD](#)
Component/s: None
Affects Version/s: None
Fix Version/s: None

Type: Story
Reporter: [Max Karel](#)
Resolution: Unresolved
Labels: None
Remaining Estimate: Not Specified
Time Spent: Not Specified
Original Estimate: Not Specified

Priority: Medium
Assignee: [Max Karel](#)
Votes: 0

Sprint: CAR Sprint 8

Description

Ako **pouzivatel** chcem aby boli pripadne chyby co nasjkor odstranene. Momentalny stav logovania nedovoluje spravne zaznamenavat logy pri multithreadovych castiach projektu. Tym padom bude nemozne zistit kde vznikol problem.

Akceptacne kriteria:

- Viacero procesov (python Pools) vie pristupovat k jednemu outputu (stdout)

Comments

Comment by [Max Karel](#) [17/Mar/20]

<https://mattgathu.github.io/multiprocessing-logging-in-python/>

Generated at Mon Apr 20 21:54:43 UTC 2020 by Max Karel using Jira 8.4.1#804002-sha1:94e96d6294939fc41f873484cf96b18ab6f2be0a.