

5. Stretnutie tímu

Dátum: 25. 10. 2021

Miesto: Discord

Prítomní: všetci

Poznámky:

- implementáciu scrapera je potrebné spustiť na počítači každého člena tímu
- predvedenie spustenia kódu
- rozdelenie zoznamu zločinov
- po spustení máme analýzu čo nám ako dlho trvalo a koľko miesta to zabralo

5. Stretnutie s vedúcim

Dátum: 26. 10. 2021

Miesto: Miestnosť 5.27

Prítomní: všetci

Poznámky:

- diskusia k úlohám
- review k videu, ktoré sme si mali minuly tyzden pozriet
- data scraper diskusia:
 - odmietnutie prístupu - VPN, anonymne proxy
 - odkladat nabok DB liniek, ktoré nepresli
 - neskôr to poriesit
 - veľkosť uložiska - v terabytoch
 - kompresia dát - odstrániť hlavne zbytočné tagy
 - indexovanie mien? vlastný index
 - nechať bežať vyhľadávanie cez noc?
 - napr. transakcie banky tiež prebehnu až na druhý deň

- alternativa - nerobit archiv
 - extrahovat vsetky mena, ulozit vsetky zlociny, ktore nasiel
 - budeme vratat len odkaz
- ponechanie linkov v textoch
- problem s neuzavretymi tagmi - uvidime, ked to pride
- urcite je lepsie, ked data budeme spracovavat hned pri scrapovani
- extrakcia tagov, ktore neobsahuju text - zistime, kolko miesta usetrime
 - vieme z toho urobit hned report na clanok - archivovany
 - ale archiv v podobe html nebudeme mat
 - nieco pekne zo stranky budeme generovat az vtedy, ked to niekto bude chciet - az vtedy pojdeme na stranku
- mongo DB ma automaticku kompresiu

veduci ma za ulohu zistit moznosti zvacsenia uloziska dat

Šprint č. 1 review

Čo hodnotíme pozitívne?

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

Čo chceme zlepšiť?

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

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Stránka tímu	Jakub Müller	6	áno	6
Otvorenie portov na serveri	Dávid Silady	6	áno	6

Vyriešiť "sudo su"	Táňa Poláková	2	áno	2
User DB	Táňa Poláková	6	nie	6
Prieskum DB	Jakub Müller	1	áno	1
Prihláška na TP CUP	Dávid Silady	2	áno	2
Set up Elasticsearch v docker container	Jakub Hlavačka	4	áno	4
Získanie dát pre prototyp	Jakub Hlavačka	4	áno	4
Získanie dát pre prototyp	Adam Šípka	8	áno	8
Získanie dát pre prototyp	Dominik Horváth	8	áno	8

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

Plán šprintu: zaciatok 26.10.2021, dlzka 2 tyzdne, koniec 9.11.2021

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