

Zápisnica pre stretnutie č.2

Téma stretnutia: Plánovanie prvého šprintu

Prítomní: Ing. Róbert Móro, PhD. (vedúci tímu)
Bc. Michal Polák
Bc. Michal Selický
Bc. Michal Števuliak
Bc. Marek Števuliak
Bc. Martin Fukas
Bc. Nghia Pham Van
Bc. Ľubomír Samotný

Neprítomní:

Zapisovateľ: Bc. Nghia Pham Van
Bc. Michal Selický
Ing. Róbert Móro, PhD.

Miesto stretnutia: FIIT STU, miestnosť 4.26

Dátum stretnutia: 3.10.2017

Priebeh stretnutia:

1 Najprv nám Ing. Michal Barla, PhD. prezentoval dôležité informácie o projekte a povedal nám niečo o

Kompozícii projektu – projekt tvoria:

1. Webové rozhranie
2. Získavanie zmlúv
3. Rozhranie pre administrátora - veľmi základné

Používaných technológiách:

- Na serveri sa používa Linux, Ruby 1.9.3, postgresql 9.2, Elastic na vyhľadávanie a facet 0.19
- Elastic bude potrebné premigrovať na novšiu verziu. Taktiež rozdeliť indexy po určitých časových dobách.
- Postgres je hlavná databáza. Je tam rake task na preindexovanie do Elasticu (bude treba prepísať/aktualizovať).
- Redis - resque, použitý ako cache
- Stiahnuté súbory (pdf, obr) su na file systéme.
- Na výrobu obrázkov je Tesseract, GraphicMagick, Headless, nejaké lib z libre/openoffice na extrahovanie.

Dátovom modeli:

- Údaje sa získavajú z 2 zdrojov – CZV, egov ale je to pripravené na rozšírenie
- Document - môže byť Contract alebo Appendix (príloha k zmluve)
- Document má foreign key na Document Detail - tam sú špecifické veci dokumentu
- Komentý ľudí, možnosť sledovať zmluvu, hlasovanie atď. - aktivity ľudí, eventy
- Spaceship_settings tabuľka - perzistentné nastavenia – napr. koľko zmlúv sa ma zobrazovať, od akého prahu sa ma dačo robiť, za aké obdobie sa ma rátať, atď..

Skórovaníu dokumentov:

- Veľmi sa s tým nerobilo, nebolo to public, pripravené ale nie veľmi spoľahlivé

Sťahovaní zmlúv:

- Z Egov sa sťahuje cez API
- CRZ je „scrapovanie“ webu, naplánovaný job prechádza zmluvami ktoré naplánujú ďalšie joby na stiahnutie a spracovanie
- Ďalší job stiahne metadáta, .pdf, extrahujú z .pdf obrázky a skúsia aj texty – cele a teda jeden job, trvá dlho
- Bolo by dobre rozdeliť na menšie časti aby sa dáta získavali iba počas noci, extrahovanie by sa potom mohlo neskôr dokončiť.
- Pre CRZ je väčší limit medzi 8-6 ráno
- CRZ ma teraz nejaké API, ale asi nekompletné. Možno nebude treba parsovať celú stránku
- Možno prerobiť na XML rozdielový súbor – potreba pozrieť, čo ponúka CRZ na stiahnutie
- Resque mal problém so zombie workermi
- Zvážiť vymeniť Resque za Sidekiq
- Je tam nejaké monitorovanie – Sentry alebo niečo iné.
- Regis (register právnických osôb štatistického úradu)
- Deduplikácia, podľa IČO nájsť záznam v RPO alebo štatistický úrad a tak zistiť názov/metadáta.
- Vek dodávateľa je tiež z regisu – na heuristiku
- slovensko.digital ma RPO, asi lepšie

Lokálnom vývoji

- Verejné github repo asi nie je aktuálne
- Spustiť workera na chvíľu
- Vyrobiť dummy
- Stiahnuť dump
- <http://otvorenezmluvy.sk/data/>

2 Zaoberali sme sa aj metodológiou Scrum

Metodika:

- Definition of Done:
 - Programátor naprogramoval 100% funkcionality z jeho pohľadu
 - Testy napísane a zelene
 - Code review spravene - pozrie niekto iný alebo automatické hodnotenie kvality kódu (CodeClimate, Gem v Ruby - RubyCritic, Rails_Best_Practices, AwesomeRuby)
 - Akceptuje to product owner
 - Nasadenie je potom
 - Merge do developu asi po Code Review

3 Získali sme spätnú väzbu od Aliancie Fair Play

Čo robiť:

- Zabezpečiť funkcionality aby sa dala navonok embednúť zmluva
- Párovanie súvisiacich vecí - zmluva a prílohy, dodatok (technicky nová zmluva)
- Dodatok povie číslo zmluvy ku ktorej je.
- Aktuálne je admin rozhranie na manuálne párovanie, nestačí
- Môže pomôcť slovensko.digital, treba sa cez AFP dohodnúť na prístupe a spísať čo potrebujeme

Nedostatky:

- Prezentácia dát
- Spracovanie dát a vzťahov medzi nimi, viac dynamicky, prelinkovane
- Administrácia dát
- Editor metadát
- Sprievodné dáta zadáva uploader
- Môžu tak vzniknúť chyby, preklepy, skratky a iné
- Bolo by dobre ak by zmluvy o verejnom obstarávaní boli priamo prepojene s verejným obstarávaním.
- Dáta zo Slovensko Digital by sa mohli používať - ešte sa da vedieť

4 Šprint Plánovanie

Naplánovali sme si nasledujúce úlohy pre prvý šprint:

Úloha: Webová stránka tímu v piatom týždni

Kategória: Management story

Ohodnotená náročnosť: 2

Pridelená komu: Bc. Martin Fukas

Čo urobiť:

- Naplniť obsahom, fotky, zápisky
- Nasadiť na server

Úloha: Tímový server – setup servera

Kategória: Infrastructure story

Ohodnotená náročnosť: 5

Pridelená komu: Bc. Marek Števuliak

Čo urobiť:

- Vykomunikovať
- Rozbehať prostredie - rvm, ruby, postgres, redis, elastic
- Zdokumentovať postup

Úloha: Nasadenie aktuálnej verzie – setup nasadzovania

Kategória: Infrastructure story

Ohodnotená náročnosť: 8

Pridelená komu: Bc. Nghia Pham Van, Bc. Michal Selický

Čo urobiť:

- Forknut repo - Robo
- Rozbehať nasadzovanie - capistrano nasadenie
- Pozrieť sa na Continuous Integration
- Naplnenie dátami -> pozrieť otvorenezmluvy.sk/data

Úloha: Metodiky

Kategória: Management story

Ohodnotená náročnosť: 1

Pridelená komu: Viacerí členovia tímu

Čo urobiť:

- Zdieľanie súborov - komunikácia v tíme - Michal Polák
- Tvorba zápisníc zo stretnutí - Michal Polák
- Git - sprava verzii - Michal Selický
- Code review - definition of done, stavu úlohy - Kanban
- CI - ak s nim začneme - 2
- Testovanie - RSpec, Cucumber - 2
- Dokumentácia - na akej úrovni sa komentuje kód, kedy sa vytvára dokumentácia user stories, modulov, etc.
- Setup lokálneho vývojového prostredia - vagrant - Michal Selický

Úloha: Zdokumentovanie aktuálneho stavu – vytvoriť diagramy

Kategória: Infrastructure story

Ohodnotená náročnosť: 8

Pridelená komu: Bc. Lubomír Samotný

Čo urobiť:

- Dátový model
- Architektúra
- Triedy – models

Úloha: Zdokumentovanie parsovania

Kategória: Infrastructure story

Ohodnotená náročnosť: 5

Pridelená komu: Bc. Michal Števuliak

Čo urobiť:

- Parsovanie - sekvencia - pozrieť sa na to, dokumentácia aj kód
- sequence alebo activity diagram aby sme vedeli ako to fungujú

Ďalšie možné úlohy:

Úloha: Prototypy na zahodenie

Katégória: Infrastructure story

Ohodnotená náročnosť:

Pridelená komu:

Čo urobiť:

- stiahnuť si zmluvy z CRZ pozrieť - XML
- script v ruby na stiahnutie a zobrazenie

Úloha: Zabezpečenie slovensko.digital - register verejných obstarávaní - API

Katégória: Infrastructure story

Ohodnotená náročnosť: 5

Pridelená komu: Bc. Michal Števuliak

Čo urobiť:

- prototyp/proof of concept script - stiahnutie z ich API, vypísať, pozrieť sa na contract ID - či to matchne v API pre contracty