

Extrakcia dát z webu

[WebExtraction]

Modul Browser extension

Tím:	č. 16, WebX
Vedúci tímu:	Ivan Srba
Členovia tímu:	Ján Brechtľ, Tomáš Juhaniak, Martin Kalužník, Rastislav Krchňavý, Michal Kren, Martin Lacek, Andrej Vaculčíak
Akademický rok:	2016/2017
Autor:	Tomáš Juhaniak, Martin Kalužník
Verzia číslo:	1.1
Dátum poslednej zmeny:	14.12.2016

1 Moduly systému	2
2 Browser extension	2
2.1 Analýza	2
2.2 Návrh	2
2.3 Implementácia	3
2.4 Testovanie	3

1 Moduly systému

Vychádzajúc z dokumentácie k inžinierskemu dielu sa systém dekomponoval na menšie časti (moduly), ktoré sa následnej ešte podrobnejšie rozdeľujú na US (User stories) a tie na úlohy.

Identifikované sú tieto moduly (podrobnejšie info v jednotlivých dokumentoch):

- User management
- Project management
- Browser extension
- Data management
- Script management
- Script scheduler

Jednotlivé moduly postupne prechádzajú 4 základnými procesnými štádiami. Sú dekomponované na jednoduchšie US a tie prechádzajú procesom analýzy a návrhu zväčša priamo na stretnutí, vo fáze prideľovania US do šprintu.

Nasleduje samostatná fáza implementácie v šprinte, počas ktorej sa návrh môže meniť. Aby boli jednotlivé US uznané ako hotové, musí prebehnúť aj úspešné testovanie.

2 Browser extension

2.1 Analýza

Na výber konkrétneho skriptu a anotáciu dát je potrebné rozšírenie do prehliadača. Toto rozšírenie musí poskytovať overenie totožnosti, resp. prihlásenie používateľa, výber konkrétneho projektu a skriptu a možnosť jednoducho zadať skript.

2.2 Návrh

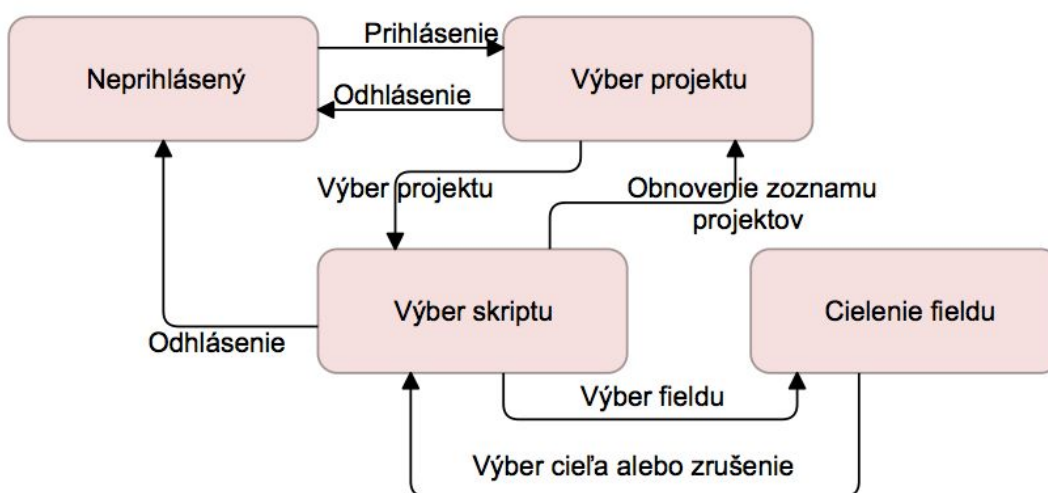
Rozhodli sme sa vytvoriť rozšírenie pre prehliadač Google Chrome. Okrem klasického popup okna vyvolaného z ikony vpravo od adresového panelu, sme sa rozhodli väčšinu logiky implementovať takzvaným DevTools rozšírením. Používateľské rozhranie takéhoto rozšírenia sa nachádza v okne, ktoré sa zobrazí po stlačení F12. Jeho hlavnou výhodou je, že neprekrýva obsah webovej stránky, vďaka čomu sa môžeme v budúcnosti vyhnúť problémom a tiež môže byť integrované so "select element" funkcionalitou (klávesová skratka ctrl+shift+c), ktorá uľahčí definovanie skriptu.

Na získanie informácií o používateľovi sme sa namiesto prihlasovania rozhodli použiť autorizáciu, v budúcnosti to uľahčí tvorbu rozšírení do iných prehliadačov a hlavne to umožní tretím stranám tvorbu aplikácií, ktoré dokážu spolupracovať s naším systémom bez toho, aby sme museli zasahovať do systému.

2.3 Implementácia

Na komunikáciu medzi aplikáciou rozšírením sme vytvorili REST API pomocou gemu ‘grape’. Autentifikáciu používateľa realizujeme pomocou OAuth2 protokolu, metódou implicitného toku, kedy musí používateľ autorizovať našu browser extension na vykonávanie požiadaviek v jeho mene.

Rozšírenie je postavené na použití javascriptového rámca AngularJS. Vďaka nemu je rozšírenie možné implementovať ako SPA. Zobrazenie jednotlivých súčastí funguje na princípe stavov. Rozlišujeme stav prihlásený / neprihlásený. V stave prihlásený rozšírenie rozlišuje stavy podľa naplnenia modelov projektov, skriptov a data schém. Po zvolení nového projektu sa modely skriptov a data schém znovu načítajú a predtým zvolené modely sa resetujú.



Obr. 1 - Stavový diagram súčastí rozšírenia do prehliadača

2.4 Testovanie

Testovanie tohto modulu zatiaľ nebolo vyžadované (výnimka v DoD) a teda nebolo riešené. Zatiaľ je v stave prototypu a v nasledujúcom semestri ho čakajú veľké zmeny.