

# Extrakcia dát z webu

[WebExtraction]

*UX testovanie*

<b>Tím:</b>	č. 16, WebX
<b>Vedúci tímu:</b>	Ivan Srba
<b>Členovia tímu:</b>	Ján Brechtl, Tomáš Juhaniak, Martin Kalužník, Rastislav Krchňavý, Michal Kren, Martin Lacek, Andrej Vaculčíak
<b>Akademický rok:</b>	2016/2017
<b>Autor:</b>	Martin Lacek
<b>Verzia číslo:</b>	1.0
<b>Dátum poslednej zmeny:</b>	1.05.2017

# 1 Úvod

Obsahom tohto dokumentu sú výsledky UX testovania nášho systému a extensionu. Každý člen tímu vykonal niekoľko jednoduchých úloh spoločne s jednou zložitejšou úlohou na konci. Okrem zadaných úloh testerí mali možnosť si pozrieť inštruktážne video k práci s extensionom.

## 2 Zadanie

Každý tester si pred začiatkom pozrel päťminútové inštruktážne video, ktoré obsahovalo informácie ako vytvoriť projekt a náležitosti k nemu a následne ako vyklikat' extension pre postprocessing nested.

Prvá časť pozostávala z nasledujúcich úloh:

- Zaregistrovanie sa a následne prihlásenie
- Vytvorenie projektu
- Vytvorenie 2 – 3 data fieldov pre vytvorený projekt (+ data field pre nested)
- Vytvorenie scriptu spoločne s nastavením intervalu spúšťania

V druhej časti sa prešlo do extensionu, kde bolo potrebné:

- Na vybranej stránke vyklikat' určený post procesing typu nested
- Uložit' vybraný script

Následne bolo potrebné narábať s extrakciou:

- Najprv spustiť extrakciu
- Prezrieť detail extrakcie (dáta)
- Stiahnuť dáta danej extrakcie vo formáte CSV

Posledná časť obsahovala prácu s API, kde bolo potrebné za pomoci písomného tutoriálu s príkladom vyplniť API link.

## 3 Testovanie

Testovanie podstúpilo päť kandidátov, pričom boli medzi nimi aj osoby, čo mali absolvovaný minimálne prvý stupeň vysokej školy v odbore IT, alebo len študenti IT.

### 3.1 Prvý kandidát

Prvý kandidát mal na stránke Alza.sk za pomoci nášho nástroja extrahovať všetky stolové počítače na prvej stránke a pre tieto počítače ich meno, cenu a popis.

Prvá časť:

Registrácia prebehla úspešne, ale nepotvrdil si email, pretože aplikácia nevyžaduje pri prvom prihlásení potvrdenie emailu na prihlásenie. Po vytvorení projektu však chvíľku hľadal prekliknutie na profil. Aj keď to bolo vidieť na videu, snažil sa zakliknúť riadok obsahujúci meno profilu a nie presné meno profilu, ktoré obsahuje link. Predtým ako mohol pracovať ďalej bol odhlásený a musel potvrdiť email. Vytváranie data fieldov a scriptu prebehlo v poriadku, ale pri nastavení intervalu si kandidát myslel, že kolónka interval, kde sa zadáva čas, slúži ako meno a interval je fixný na možnosti ako: každú minútu, hodinu, deň, atď.

Extension:

Pri vyklikávaní extensionu nastal problém, že kandidát si musel znovu pozrieť video a stále potreboval trochu pomoc pri vysvetlení ako presne nastaviť post-processing. Následne si uložil skript pri spúšťaní extrakcie manuálne, nevidel tlačidlo „run now“, pretože hľadal niečo v zmysle „run extraction“. Najst' dáta nebol problém a taktiež nebol problém ich stiahnuť vo vybranom formáte.

API:

V API časti kandidát mal určené vylistovať extrakcie. Teda bolo potrebné vyplniť do linku parametre token a id skriptu. Pri vyplňaní vzniklo nedorozumenie, pretože v inštrukciách je napísane token a v linku treba meniť „YourToken“ pričom slovo token predtým sa nesmie zmeniť. Taktiež pri hľadaní skript id to po chvíli vzdal, pretože id skriptu sa nachádza v linke pri prezeraní skriptu. Pri zmene API kľúča vyhodilo permission denied.

Komentár:

Video bolo náročné, ale po prvom vyklikaní už bolo viac menej jasné ako to funguje. Pri API sú id a token mäťúce. Pridať do extension tooltipy.

### ***3.2 Druhý kandidát***

Druhý kandidát mal na stránke Alza.sk za pomoci nášho nástroja extraktovať všetky notebooky na prvej stránke a pre tieto notebooky ich meno, cenu a popis.

Prvá časť:

Pri registrácii vyplnil krátke heslo a očakával chybovú hlášku na dolnej strane stránky. Preto mu chvíľku trvalo, kým si všimol chybovú hlášku na hornej časti stránky. Okrem zdržania s chybovou hláškou kandidát v tejto časti nemal žiadny problém.

Extension:

Pri vyklikávaní kandidáta zmiatlo slovíčko squash pri xpathoch. Tak isto nevedel pri výbere xpathu, či už môže vyberať alebo nie, keďže sa target tlačidlá nevyfarbia a jediný indikátor bolo vyfarbovanie divov pri prechode myškou po stránke. Následne si uložil skript pri spúšťaní extrakcie manuálne. Takisto ako prvý kandidát nevidel tlačidlo „run now“, pretože hľadal niečo v zmysle „run extraction“. Nájst' dáta nebol problém a taktiež nebol problém ich stiahnuť vo vybranom formáte. Pri tomto kroku však zistil, že nevyplnil správne xpathy a musel ich opätovne vyklikáť správne.

API:

V API časti kandidát mal určené vylisovať extrakcie. To znamená, že bolo potrebné vyplniť do linku parametre token a id skriptu. Pri hľadaní skript id to po chvíli vzdal, pretože id skriptu sa nachádza v linke pri prezeraní skriptu.

Komentár:

Post-processing bol zložitejší na prvé pochopenie. Chybové hlášky pri registrácii lepšie zviditeľniť. Skript id lepšie zviditeľniť.

### ***3.2 Tretí kandidát***

Tretí kandidát mal na stránke Alza.sk za pomoci nášho nástroja extraktovať všetky hrácke notebooky na prvej stránke a pre tieto notebooky ich meno, cenu a popis.

Prvá časť:

V tejto časti nebol žiaden problém.

Extension:

Pri vyklikávaní si kandidát musel znovu pozrieť video, pretože to bolo pre neho náročné na zapamätanie. Taktiež mu chýbali tooltipy pri tlačidlách. Po druhom zhliadnutí videa bol schopný vyklikáť xpathy. Následne si uložil skript pri spúšťaní extrakcie manuálne. Tak isto ako prvý kandidát nevidel tlačidlo „run now“, pretože hľadal niečo v zmysle „run extraction“. Nájst' dáta nebol problém a taktiež nebol problém ich stiahnuť vo vybranom formáte.

API:

V API časti kandidát mal určené ukázať dáta z extrakcie za pomoci API. Teda bolo potrebné vyplniť do linku parametre token, id extrakcie a id skriptu. Pri vypĺňaní vzniklo nedorozumenie, pretože v inštrukciách je napísane token a v linku treba meniť „YourToken“ pričom slovo token predtým sa nesmie zmeniť. Okrem toho úlohu dokončil len s pomocou, pretože nielen, že nevedel nájsť skript id, ale ani id extrakcie, ktoré sa nachádzajú v linke pri prezeraní extrakcie.

Komentár:

Post-processing bol zložitejší na prvé pochopenie. API návod mätúci.

### ***3.2 Štvrtý kandidát***

Štvrtý kandidát mal na stránke orange za pomoci nášho nástroja extrahovať všetky telefóny na prvej stránke a pre tieto telefóny ich meno, cenu a popis. Tento kandidát si pozorne pozrel video.

V prvej časti nebol žiaden problém.

Extension:

Pri vyklikávaní nastal jediný problém a to ten, že bol označený vonkajší div na prekliknutie pre nested a nie konkrétny link. Preto po spustení extrakcie kandidát opravil xpath pre link a script fungoval ako mal.

API:

V API časti kandidát mal určené ukázať dáta z extrakcie za pomoci API. Teda bolo potrebné vyplniť do linku parametre token, id extrakcie a id skriptu. Kandidát mal iba problém nájsť lokáciu idčok.

Komentár:

Lokácia idčok je mätúca pre bežného používateľa.

### ***3.2 Piaty kandidát***

Posledný kandidát mal na stránke Alza.sk za pomoci nášho nástroja extrahovať všetky monitory na prvej stránke a pre tieto monitory ich meno, cenu a popis. Kandidát mal predošle skúsenosti s xpathmi.

Prvá časť:

V tejto časti bol jediný problém - skoré odhlásenie bez potvrdenia emailu.

Extension:

Pri vyklikávaní kandidátovi chýbali tooltipy.

API:

V API časti kandidát mal určené ukázať dáta z extrakcie za pomoci API. Teda bolo potrebné vyplniť do linku parametre token, id extrakcie a id skriptu. Keď nevedel nájsť chvíľu idčka skriptu a extrakcie, vybral ich z linku bez pomoci.

Komentár:

Nástroj je určený pre pokročilejších používateľov, ale stále je potrebný tutoriál k prvému vyklikávaniu.

## 4 Zhrnutie

Najčastejšie chybou pri testovaní bolo skoré odhlásenie bez potvrdenia emailovej adresy. Každý očakával, že bude musieť potvrdiť emailovú adresu po jednom dni, alebo najbližšom odhlásení po registrácii. Okrem tohto problému sa v prvej časti, pri práci na serverovej strane, teda na webe WebX jednému kandidátovi podarilo prehliadnúť chybovú hlášku počas registrácie. Chybová hláška nebola veľmi viditeľná oproti obyčajnému textu.

Pri vyklikávaní extension časti, každý kandidát očakával tooltipy na tlačidlách. Ale dokázali úlohu splniť, pretože chýbajúce tooltipy neboli až tak potrebné, keďže vo video tutoriáli bola väčšina tlačidiel využitá. Okrem týchto maličkostí, tlačidlo pre spustenie extrakcie má názov „Run now“. Tento názov nepúta pozornosť a niektorým kandidátom sa stalo, že nevideli na prvý pohľad tlačidlo pre spustenie extrakcie.

V poslednej časti sa pracovalo s API. Každý kandidát označil návod pre API ako mätúci, keďže v popise bolo napísané, že treba zmeniť „token“, ale v skutočnosti bolo potrebné zmeniť „YourToken“. Ďalšiu prekážku tvorila lokácia idčok scriptov a extrakcií. Nie každý očakáva, že je potrebné idčko kopírovať z adresy stránky.



## **5 Zhodnotenie**

Testovanie prebehlo úspešne a odhalilo niekoľko nedostatkov, na ktorých treba popracovať. Všetci kandidáti testovanie ukončili v úspešnej forme, napriek tomu, že sa občas preukázalo, že bolo potrebné isté veci viac objasniť, alebo vymedziť viac času na pochopenie.