



**FAKULTA
INFORMAČNÍCH
TECHNOLGIÍ
ČVUT V PRAZE**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Název: Webová aplikace pro online web scraping
Student: Jakub Drahoš
Vedoucí: Mgr. Martin Podloucký
Studijní program: Informatika
Studijní obor: Webové a softwarové inženýrství
Katedra: Katedra softwarového inženýrství
Platnost zadání: Do konce letního semestru 2019/20

Pokyny pro vypracování

Cílem bakalářské práce je vytvořit webovou aplikaci, která bude nabízet online prostředí umožňující scraping webových stránek uživatelům neznalých programování.

Práce nejprve zhodnotí dostupný software pro online web scraping. Na tomto základě vznikne webová aplikace s přívětivým grafickým rozhraním, které nabídne uživateli různé pohledy na data obsažená v HTML struktuře - fulltextové vyhledávání, stažení všech obrázků (audia, videa), texty ze sémantických elementů (<h1>-<h6>, <section>, <a> apod.), výběr elementů podle CSS tříd a další.

Získaná data bude možné exportovat do standardních formátů (např. JSON, CSV apod.). Při vývoji bude použit klasický softwarový vývojový cyklus - analýza, design, implementace, testování.

Seznam odborné literatury

Dodá vedoucí práce.

Ing. Michal Valenta, Ph.D.
vedoucí katedry

doc. RNDr. Ing. Marcel Jiřina, Ph.D.
děkan

V Praze dne 25. listopadu 2018