Резюме:

1. Представяне

Аз съм ….

Темата на моят дипломен проект е …

1. Въведение

Аз съм любител на спорта , много често ми се налага да влизам в интернет за да проверявам датите и резултатите на определени мачове. След като получих своето задание за дипломен проект се зарадвах защото мага да направя нещото което ми се налага да използвам постоянно. Приложение за извличане на информация от сайтове, на което аз мога да му задам данни от някои сайт, като например Gong.bg. Тoва в моя случаи ще се получи чрез web scraping.

1. Parsing и scraping

Терминът парсинг най-общо означава синтактичен анализ. В програмирането под парсер се разбира програма, която преглежда входните данни, анализира ги и построява някаква изходна структура.

Парсер който извличаме данни от уебсайтове се нарича скрейпер, а самият процес – скрейпинг.

1. WEB сървърите

WEB сървърите се основават на Hyper Text Transfer Protocol и осигуряват достъп до уеб сайтове, уеб приложения и други ресурси. Те работят на принципа заявка/отговор. Kлиентът осъществява връзка със сървъра и изпраща заявка. Сървърът анализира заявката, намира търсената информация и я връща на клиента. Най-често това HTML код на някаква уеб страница.

1. HTML

HTML е основният език за описание и дизайн на уеб страници. Командите в него се наричат тагове. За оформление на външния вид се използват т.н. стилове. Стиловете най-често са разположени в отделен файл – CSS файл. HTML може да включва и кратки програми, написани на скриптов език като JavaScript.

1. Python

Python е универсален програмен език от високо ниво. Той е лесен за научаване и ползване. Създаден е с идеята да бъде удобен и ефективен за разработка на софтуер. Python има много библиотеки с отворен код, които разширяват възможностите му в една или друга област.

1. Принципи на скрейпинга

Пример – gong.bg

Beautiful Soup

1. Заключение

. Не всички сайтове позволяват ботове. Не всеки скрейпинг е легален. Но много често това е необходим или просто полезен процес.

Трудно е да се разработи сериозно приложение, но пример се прави лесно.

1. Край

Благодаря за вниманието!

Ако имате въпроси съм готов да отговоря.

Въпроси:

1. Поясни frontend и backend.