BCCP web scraping course

Day 1

Table of Contents

Day 1

- very short intro to Python
- Intro to Webscraping
- APIs

Day 2

- HTML parsing
- Text pattern matching

Day 3

- Browser automation
- Own script

Day 1

very short intro to Python

Very short intro to Python

Day 1

Intro to Webscraping

Introduction to Webscraping

Day 1

APIs

Application Programming Interface

Twitter API

- "Conduct historical research and search from Twitter's massive archive of publicly-available Tweets posted since March 2006?"
- "Listen in real-time for Tweets of interest?"

Day 2

Table of Contents

Day 1

- very short intro to Python
- Intro to Webscraping
- APIs

Day 2

- HTML parsing
- Text pattern matching

Day 3

- Browser automation
- Own script

Day 2

HTML parsing

Basic principle

1. Load URL and save HTML source code: `requests` or `urllib2`
2. Convert source code to Python object: `beautifulsoup4`
3. Navigate the source code and save needed elements

Day 2

Text pattern matching

Day 3

Table of Contents

Day 1

- very short intro to Python
- Intro to Webscraping
- APIs

Day 2

- HTML parsing
- Text pattern matching

Day 3

- Browser automation
- Own script

Day 3

Browser automation

Day 3

Own script