

Advanced Web Development and Web Scraping

Dr. Garrett Dancik

Why learn “advanced web development”

- You can develop web pages (obviously)
- You can add features to web pages (i.e., develop browser extensions)
 - Examples: <https://chrome.google.com/webstore/category/extensions>
- You can answer questions by extracting and parsing (scraping) web page data
 - Compare weather across multiple cities
 - Find the ages of a list of actors
- You can retrieve information and answer questions by using web page application program interfaces (APIs)
- You can automate web-tasks such as logging onto a web site

How I accidentally became a web developer

- Useful bioinformatics research tools, such as for analyzing genomic data, need to be web-based
 - <https://gdancik.github.io/bioinformatics/software.html>
- What can we learn from web data?
 - What programming languages currently have the largest demand?
 - What possible schedules exist if I need to take CSC 210, MAT 243, etc
 - <https://gdancik.github.io/EasternScheduler/>
 - How does information spread on social media

Web development overview

- **Hypertext Markup Language (HTML)** describes the structure of a web page
 - HTML pages are composed of elements that are specified using tags
 - `<p> This is a paragraph </p>`
 - `<h1> This is a header </h1>`
- **Cascading style sheets (CSS)** describe how HTML elements are displayed
 - `<p style = "background-color:yellow"> This is a paragraph with yellow background </p>`

Web development overview (con't)

- **Javascript** is a programming language for interacting with HTML and CSS

```
// get body of web page and change its background color  
var b = document.getElementsByTagName("body");  
var body = b[0];  
body.style.backgroundColor = "yellow";
```

Examples

- See <http://gdancik.github.io/> for the example.html web page
- Let's use the Web Inspector to look at the structure of various web pages