Data acquisition

Data scraping (keo data tu tren mang ve de keo du lieu ve va xu ly)

1. Truy cập vào trang

Urlopen tra ra mot pipe de chay du lieu tu server den client

Ascii: he bieu dien cac so cua tieng anh

Extend len thanh utf-8, moi con so se anh xa thanh chu cai.

1. Tìm dữ liệu cần copy (ROI: Regioin of Interest):

Vùng quan tâm là ROI. Phân tích

Soup có 2 lệnh cơ bản, find và find\_all()

Prettify chỉ dùng cho soup, nếu là list thì không được

A giống như dictionary, dùng truy cập dựa vào key href

1. Copy
2. Tổ chức lại dữ liệu

HTML

1. Tag
2. Content/ string
3. Contributes?
4. Parent – children
5. Siblings

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="UTF-8">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<meta http-equiv="X-UA-Compatible" content="ie=edge">

<title>HTML intro</title>

</head>

<body>

<a href="https://facebook.com" target="#">Click vào em đi</a>

</body>

</html>

Thẻ div để phân vùng, hiển thị không khác gì bình thường nhưng sau này chỉnh sửa, bê cục div đi chứ không đẩy từng phần :)

<div>

<h6>Heading 6</h6>

<a href="https://facebook.com" target="#">Click vào em đi</a>

</div>