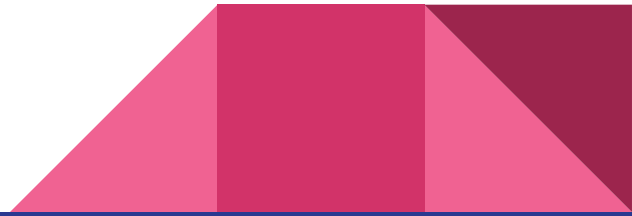
PokeCrawler

Ahmed Mohamad Bakri
Matheus Miranda

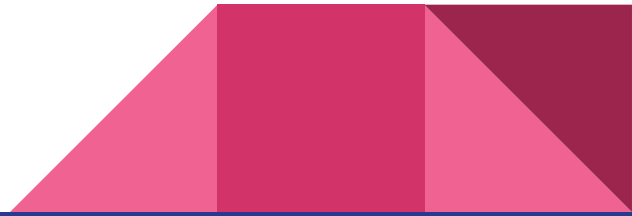
PokeCrawler

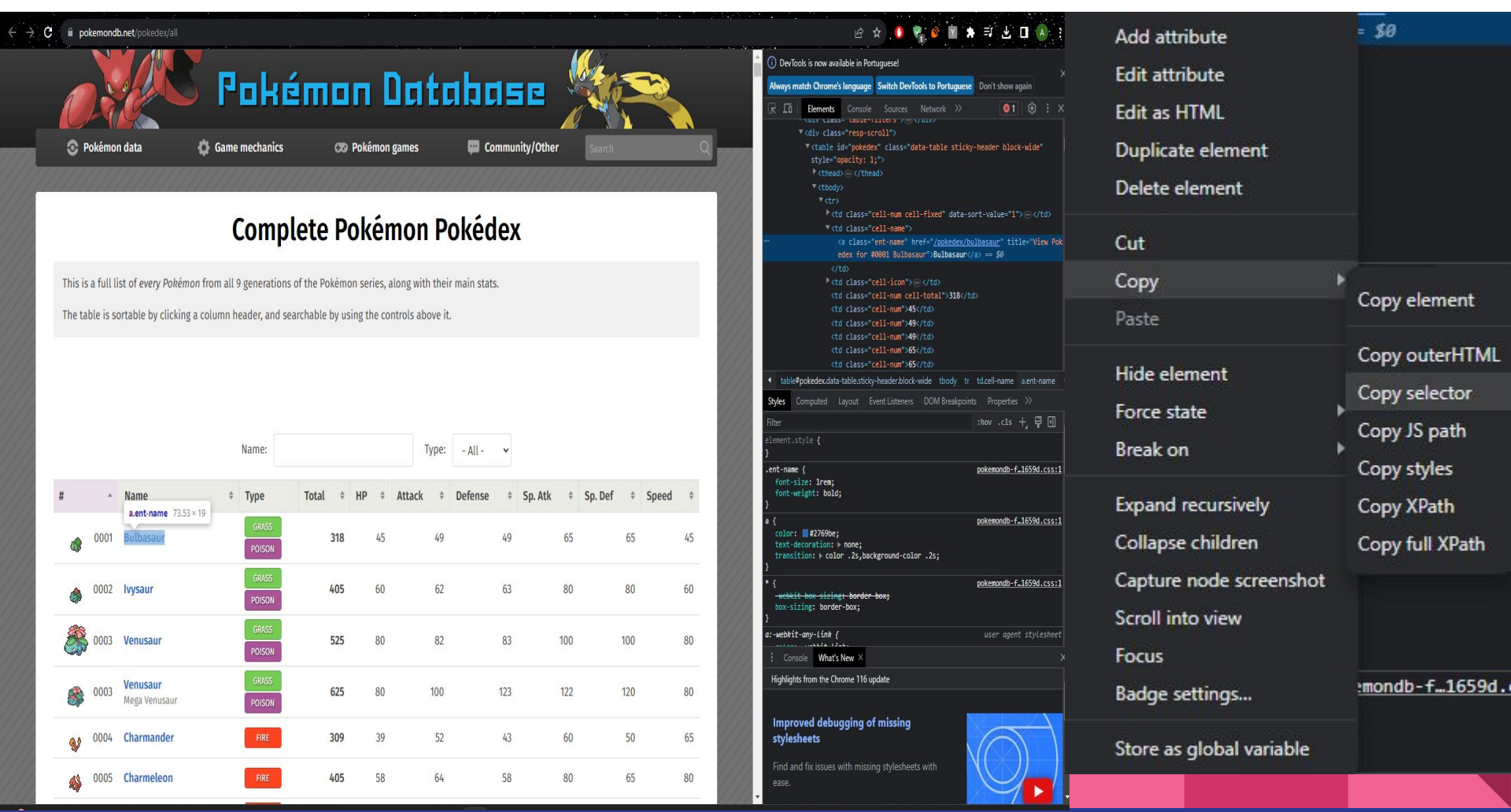
Neste trabalho implementamos um crawler que rasteja através de uma url que no caso é a de uma pokédex para extrairmos dados como id, nome, habilidade, evoluções e tipos



Como

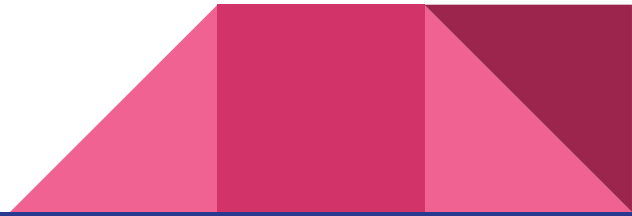
Através de seletores css que pegamos inspecionando elemento nas partes do site que desejamos como o id de um pokémon, e utilizando o método response do scrapy nós conseguimos pegar o conteúdo desses seletores




pokemondb.net/pokedex/all
Pokémon Database
Pokémon data
Game mechanics
Pokémon games
Community/Other
Search
Complete Pokémon Pokédex
This is a full list of every Pokémon from all 9 generations of the Pokémon series, along with their main stats.
The table is sortable by clicking a column header, and searchable by using the controls above it.
Name:
Type:
- All -
Name Type Total HP Attack Defense Sp. Atk Sp. Def Speed
a.ent-name 73.53 × 19
0001 Bulbasaur GRASS POISON 318 45 49 49 65 65 45
0002 Ivysaur GRASS POISON 405 60 62 63 80 80 60
0003 Venusaur GRASS POISON 525 80 82 83 100 100 80
0003 Venusaur Mega Venusaur GRASS POISON 625 80 100 123 122 120 80
0004 Charmander FIRE 309 39 52 43 60 50 65
0005 Charmeleon FIRE 405 58 64 58 80 65 80
DevTools is now available in Portuguese!
Always match Chrome's language
Switch DevTools to Portuguese
Don't show again
Elements Console Sources Network
table#pokedex.data-table.sticky-header.block-wide tbody tr td.cell-name a.ent-name
Styles Computed Layout Event Listeners DOM Breakpoints Properties
Filter
element.style {
}
.ent-name {
font-size: 1rem;
font-weight: bold;
}
a {
color: #2769be;
text-decoration: none;
transition: color .2s,background-color .2s;
}
* {
box-sizing: border-box;
}
a:-webkit-any-link {
Console What's New
Highlights from the Chrome 116 update
Improved debugging of missing stylesheets
Find and fix issues with missing stylesheets with ease.
Add attribute
Edit attribute
Edit as HTML
Duplicate element
Delete element
Cut
Copy
Paste
Hide element
Force state
Break on
Expand recursively
Collapse children
Capture node screenshot
Scroll into view
Focus
Badge settings...
Store as global variable
Copy element
Copy outerHTML
Copy selector
Copy JS path
Copy styles
Copy XPath
Copy full XPath

Como

No último slide há um exemplo de como o seletor css do nome de um pokémon é retirado.



Exemplo

Este é o nosso seletor do nome do pokémon

```
name_s  = "#main > h1::text"
```

E aqui nós pegamos o valor atrelado a esse seletor

```
name = response.css(name_s)
```

através de um yield nós “jogamos” esses dados em csv ou json

```
yield {'Id': id.get(), 'Nome': name.get(), 'Type': type.getall(), 'Altura': height.get(), 'Peso': weight.get(), "Urls das Habilidades": ability_url,"Nomes da
```

Exemplo

Esse é um exemplo do resultado do crawler para um único pokémon no caso o Bulbasaur

{"Id": "0001", "Nome": "Bulbasaur", "Type": ["Grass", "Poison", "Cerulean City"], "Altura": "0.7\u00a0m (2\u203204\u2033)", "Peso": "6.9\u00a0kg (15.2\u00a0lbs)", "Urls das Habilidades": ["/ability/overgrow", "/ability/chlorophyll"], "Nomes das Habilidades": ["overgrow", "chlorophyll"], "Evolu\u00e7\u00f5es": ["Bulbasaur", "Ivysaur", "Venusaur"]},

Tratamento de dados

O nome das habilidades pegamos através das urls das habilidades de um dado pokémon, para isso splitamos a / para pegar as palavras e adicionamos em um array só as palavras relacionados ao nome delas

```
Hab = []
for i in range(len(ability_url)):
  Hab.append(ability_url[i].split("/"))

nameHab = []
for i in range(len(Hab)):
  nameHab.append(Hab[i][2])
```