

Postępy prac projektu indywidualnego - cz. 2

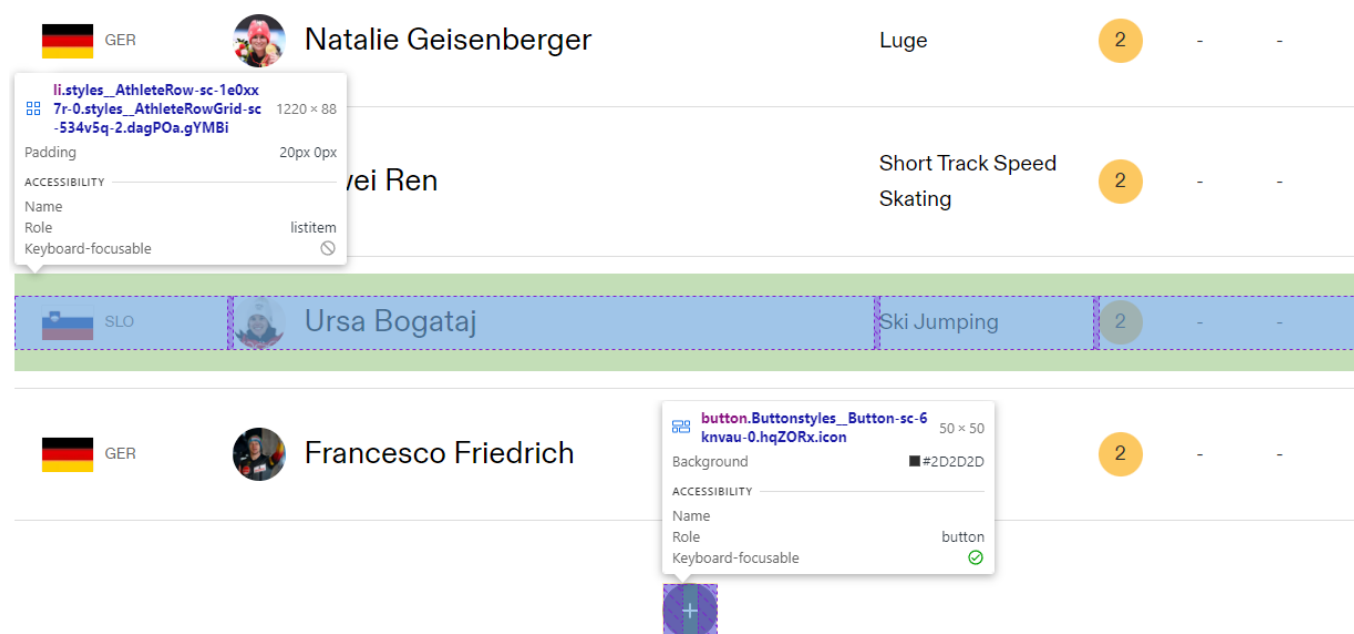
Michał Banaszczyk

6 czerwca 2022

1 Napotkane problemy

1.1 Dynamiczna zawartość strony z danymi zawodników

Na kolejnej stronie IO tym razem z danymi medalistów poszczególnych Igrzysk również wykorzystano dynamiczne doładowywanie elementów w trakcie przeglądania, zamiast załadowania wszystkiego na raz. Mianowicie jest to lista zawodników, która przy każdym doładowaniu powiększa się o kolejnych dziesięciu atletów. Guzik do ładowania dalszej części listy (oczywiście) jest niszczone i później tworzony nowy. Poniżej przedstawiono działanie aplikacji.



Zrzut ekranu 1: Fragment listy zawodników

```
> document.querySelector('[class^="Buttonstyles__Button"]')
< <button data-cy="more-button" type="button" class="Buttonstyles__
> document.querySelectorAll('[class^="styles__AthleteRow"]').length
< 20
> document.querySelector('[class^="Buttonstyles__Button"]').click()
< undefined
> document.querySelector('[class^="Buttonstyles__Button"]')
< null
> document.querySelectorAll('[class^="styles__AthleteRow"]').length
< 30
```

Zrzut ekranu 2: Zachowanie strony przy doładowywaniu kolejnych zawodników

Ciągłe pojawianie się i znikanie guzika pozwoliło jednak na ponowne zastosowanie observera do obserwacji mutacji dokumentu, w oczekiwaniu na ponowne pojawienie się tego guzika. Później, gdy guzik przestaje się już pokazywać po załadowaniu całej zawartości callback wywoływany przez timeout scrapuje całą tabelę i rozwiązuje promise, który zwraca te dane ze skryptu ewaluacyjnego do skryptu w Node. Moduł ten pobiera dane medalistów z tegorocznych Igrzysk zimowych w Pekinie, natomiast łatwo można go dostosować do pobierania danych z dowolnych Igrzysk.

1.2 Brak spójności nazw państw uczestniczących i przypisów

Na poszczególnych podstronach z niewiadomych na pierwszy rzut oka zamiast nazw państw uczestniczących figurują trzyliterowe kody. Te z kolei nie są oficjalnymi kodami państw, a wymyślonymi specjalnie na potrzeby IO (np. *SUI* dla Szwajcarii zamiast oficjalnego *CHE*). Dodatkowo w IO brały udział państwa już nie istniejące lub różne sportowe organizacje pozapaństwowe. Z tego względu jedynym wyjściem ujednolicenia zbioru było zmapowanie wszystkich występujących kodów i zamiana ich na pełne nazwy państw lub organizacji. Analogicznie zastąpiono puste wyniki przypisami wyjaśniającymi ich brak.

1.3 Czas scrapowania wyników wszystkich dyscyplin

Ilość danych do pobrania i zapisania wraz z wyżej opisaną niespójnością danych sprawiały, że jakiegokolwiek formatowanie zbioru w trakcie pobierania okazało się praktycznie niemożliwe, rzucając cały czas błędami. Z tego względu pobieranie i formatowanie podzielono na dwa moduły: `scoreRetriever.js` służył do pobrania surowych danych, a `scoreFormatter.js` - do ich formatowania. Pozwoliło to na sprawniejsze naprawianie błędów wynikających z niespójności danych na bieżąco. Moduł pobierający dodatkowo może pobierać dane fragmentami, dzięki czemu przy jakimkolwiek błędzie nie traci się dotąd zapisanych danych.

Podobny podział zastosowano przy skrapowaniu medalistów. Czas scrapowania tego zbioru danych nie jest aż tak długi jak w przypadku zbioru wszystkich dyscyplin, natomiast zwyczajnie pomaga to w utrzymaniu przejrzystości i zrozumiałości kodu oraz zapewnia wyeliminowanie tzw. "złych zapachów kodu" takich jak rozdmuchiwanie funkcjonalności poszczególnych modułów.

2 Postępy

2.1 Zbiór danych z poszczególnych dyscyplin sportowych

Po wykonaniu wyżej opisanego scrapowania i formatowania otrzymano pełen zbiór wyników wszystkich nowożytnych Igrzysk Olimpijskich. Poniżej, fragment tego zbioru. Wyszczególniono tylko cztery z ośmiu kolumn - pierwsze pięć od prawej były wynikiem scrapowania opisanego przy poprzednim opisie postępu prac, a wszystkie osiem nie zmieściłoby się też na stronie.

Event	Place	Country	Athlete	Score
Men's Alpine Combined	1	Austria	Johannes Strolz	151.43
Men's Alpine Combined	2	Norway	Aleksander Aamodt Kilde	152.02
Men's Alpine Combined	3	Canada	James Crawford	152.11
Men's Alpine Combined	4	Switzerland	Justin Murisier	152.29
Men's Alpine Combined	5	Austria	Marco Schwarz	152.71
Men's Alpine Combined	6	Israel	Barnabas Szollos	153.18
Men's Alpine Combined	7	Austria	Raphael Haaser	153.27
Men's Alpine Combined	8	Canada	Broderick Thompson	154.20
Men's Alpine Combined	9	Canada	Brodie Seger	155.03
Men's Alpine Combined	10	Italy	Christof Innerhofer	155.80
Men's Alpine Combined	11	Liechtenstein	Marco Pfiffner	156.40
Men's Alpine Combined	12	Ireland	Jack Gower	157.74
Men's Alpine Combined	13	Monaco	Arnaud Alessandria	164.20
Men's Alpine Combined	14	Ukraine	Ivan Kovbasnyuk	164.43
Men's Alpine Combined	15	Kosovo	Albin Tahiri	168.29
Men's Alpine Combined	16	People's Republic of China	Yangming Zhang	169.72
Men's Alpine Combined	17	People's Republic of China	Mingfu Xu	180.40
Men's Alpine Combined	-	Czech Republic	Jan Zabystran	Didn't finish
Men's Alpine Combined	-	France	Alexis Pinturault	Didn't finish
Men's Alpine Combined	-	Switzerland	Loic Meillard	Didn't finish

2.2 Zbiór danych wszystkich medalistów poszczególnych Igrzysk

Po wykonaniu wyżej opisanego scrapowania i formatowania otrzymano pełen zbiór medalistów Zimowych Igrzysk Olimpijskich w Pekinie, 2022. Poniżej, fragment tego zbioru.

Team	Athlete	Sport	G	S	B	Games	Birth
Norway	Johannes Thingnes Boe	Biathlon	4	0	1	3	1993
ROC	Alexander Bolshunov	Cross Country Skiing	3	1	1	2	1996
Norway	Marte Olsbu Røiseland	Biathlon	3	0	2	1	1990
Netherlands	Irene Schouten	Speed skating	3	0	1	2	1992
Norway	Therese Johaug	Cross Country Skiing	3	0	0	3	1988
France	Quentin Fillon Maillet	Biathlon	2	3	0	2	1992
Norway	Johannes Hoesflot Klaebo	Cross Country Skiing	2	1	1	2	1996
Norway	Tarjei Boe	Biathlon	2	1	1	4	1988
Netherlands	Suzanne Schulting	Short Track Speed Skating	2	1	1	1	1997
Norway	Joergen Graabak	Nordic Combined	2	1	0	3	1991
China	Ailing Eileen Gu	Freestyle Skiing	2	1	0	1	2003
Austria	Johannes Strolz	Alpine Skiing	2	1	0	1	1992
Germany	Johannes Ludwig	Luge	2	0	0	2	1986
Sweden	Nils Van Der Poel	Speed skating	2	0	0	2	1996
USA	Lindsey Jacobellis	Snowboard	2	0	0	5	1985
Germany	Thorsten Margis	Bobsleigh	2	0	0	3	1989
Germany	Natalie Geisenberger	Luge	2	0	0	4	1988
China	Ziwei Ren	Short Track Speed Skating	2	0	0	2	1997
Slovenia	Ursa Bogataj	Ski Jumping	2	0	0	2	1995
Germany	Francesco Friedrich	Bobsleigh	2	0	0	3	1990