

Web Source 1: https://www.futhead.com/18/players/?bin_platform=ps

- Futhead is a website which collects data regarding all the players in the FIFA 18 Ultimate Team league. FIFA 18 is a soccer video-game in which players from all around the world build teams and compete. The data collected consists of various player attributes like name, age, in-game quality statistics etc.

Web Source 2: https://www.easports.com/fifa/ultimate-team/fut/database/results?position_secondary=LF,CF,RF,ST,LW,LM,CAM,CDM,CM,RM,RW,LWB,LB,CB,RB,RWB

- EA Sports is the official maker of the game FIFA 18. They have an official database which consists of all the in-game player attributes and statistics.

Data Extraction:

The data from both the websites was extracted using web crawling and scraping.

On the futhead website, crawling involved following the pagination and following the link to each player card details. This was done in parallel asynchronously. Following the pagination just involved parsing the HTML <a> tags and no JavaScript functionality was required.

On the EA Sports website, the data was presented in a single list and more elements of the list could be populated on click on a list-item which called a JavaScript method. Hence, a library which provided JavaScript capabilities was required. It was not possible to extract data in parallel for the same reason as each on-click event rendered more page area. Each list-item was a link to the corresponding player details.

Entity:

FIFA 18 Ultimate Team player details – name, country, overall rating, league, position, club, nation, date of birth, etc.

Table A: 3788 tuples

Table B: 9724 tuples

Tools Used:

python – open source scripting language.

scrapy – python library for web crawling and HTML DOM parsing.

unicodedata – python library for converting Unicode strings to ASCII strings.

selenium – python library for web crawling and HTML DOM parsing with JavaScript capabilities.