

We have a database with basketball statistics, and we are looking for someone who can write software that grabs the so called box-scores from websites with basketball results.

We already have developed a system with several programs that do this for various basketball leagues. These scraper programs roughly all have the following approach:

- 1) from a given start page, it finds all the links that point to pages with game schedules
- 2) from those game schedules it follows all the links that lead to the box-scores of those games
- 3) it parses that page, grabs the available box-score data and returns it in a json record.
- 4) it also grabs information about the players playing the game (playerid and characteristics)

You can test programs with:

<http://scraper.basketball-database.com/scraper7/public/gcscape/pctest>

You can access the example source by typing:

GC Scraper Page – Mozilla Firefox

htpasswd How to Use Apache H 26.09.2021 - F GC Scraper P GC Scraper X scraper.baske

scraper.basketball-database.com/scraper7/public/gcs

Meest bezocht VGV7519 - modem NPO 1 - Live tv Gids voor vandaag >> Andere bladwijzers

Test Scrapers

source

lpar

spar

season

proxy url

user

password

external game id:

external player id:

Or you can access one of the existing scrapers like this one:

GC Scraper Page – Mozilla Firefox

htpasswd How to Use Apache H 26.09.2021 - F GC Scraper P GC Scraper X scraper.baske + v

scraper.basketball-database.com/scraper7/public/gcs ☆

Meest bezocht VGV7519 - modem NPO 1 - Live tv Gids voor vandaag >> Andere bladwijzers

Test Scrapers

source ABA

lpar http://www.aba-liga.com/

spar

season 2022

proxy url http://scraper.basketball-database.com/dev1

user

password

show all games show all teams show conferences show specials show help

external game id: get game details

external player id: get player details

Source is a unique key which identifies the scraper. A scraper gets two parameters, lpar and spar. These enable the scraper software to deal with sources that provide data for different leagues and seasons. Lpar contains league dependent info and spar contains the season dependent stuff. This is an example of such a scraper. It is calling the FIBA_LS scraper for luxemburg:

GC Scraper Page – Mozilla Firefox

htpasswd - Ma X How to Use SS X Apache HTTP X GC Scraper Page X scraper.basketball X

scraper.basketball-database.com/scraper7/public/gcs

Meest bezocht VGV7519 - modem NPO 1 - Live tv Gids voor vandaag >> Andere bladwijzers

Test Scrapers

source	FIBA_LS
lpar	LUX
spar	;/webcast/LUX/;https://livestats.dcd.shared.geniussports.com/u/LUX/2144713
season	2022
proxy url	http://scraper.basketball-database.com/dev1
user	
password	

[show all games](#) [show all teams](#) [show conferences](#) [show specials](#) [show help](#)

external game id: [get game details](#)

external player id: [get player details](#)

Make sure that you deliver the results in the right json format as you find in the example.