

Homework 1: Crawling

INF 558 BUILDING KNOWLEDGE GRAPH

Runqi Shao

9418534943

Task 1

Website URL: <https://www.transfermarkt.com/>

Description: the website provides news, rumors, history of soccer player transfers and their performance statistics all over the world.

Task 2

Sample webpage:

<https://www.transfermarkt.com/neymar/profil/spieler/68290>

See screenshot next page

Interesting information: Age, Nationality, Agents, Current club, Position, Contract information, Current market value, Highest market value, Transfer history, Performance Data.

Task 3

1. I use a crawler to get the results.

2. Seed:

<https://www.transfermarkt.com/spieler-statistik/wertvollstespieler/marktwertetop>

3. I use a page classifier, which is a url regex. It contains a regex which matches the page I want and the seed url. Any page does not match the regex will be discarded.

type: url_regex

parameters:

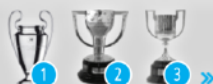
regular_expressions: [

"https:\\\\www\\.transfermarkt\\.com\\.*\\profil\\spieler\\.[^\\W]*",

"https://www.transfermarkt.com/spieler-statistik/wertvollstespieler/marktwertetop"

]

#10 Neymar



Paris SG
Ligue 1
League level: **First tier**
In team since: **Aug 3, 2017**
Contract until: **30.06.2022**

NEW ARRIVAL

Born/age: **Feb 5, 1992 (25)**
Place of birth: **Mogi das Cruzes**
Nationality: **Brazil**
Height: **1,75 m**
Position: **Left Wing**
Agents: **Mjf Publicidade e ...**

100,00 Mill. €
Last change: **Jun 5, 2017**

PROFILE PERFORMANCE DATA MARKET VALUE TRANSFERS RUMOURS NATIONAL TEAM NEWS HONOURS

CAREER

Current club
Paris Saint-Germain

Position
Left Wing

Nationality
Brazil

Contract until
30.06.2022



CURRENT JERSEYS



[Go to jersey shop](#)

CURRENT SHOE

Nike Mercurial Vapor XI Neymar FG



1/5 [To the shop](#)

PLAYER DATA

Complete name: **Neymar da Silva Santos Júnior**
Date of birth: **Feb 5, 1992**
Place of birth: **Mogi das Cruzes**
Age: **25**
Height: **1,75 m**
Nationality: **Brazil**
Position: **Striker - Left Wing**
Foot: **both**
Player's agent: **Mjf Publicidade e Promoções**
Current club: **Paris Saint-Germain**
In the team since: **Aug 3, 2017**
Contract until: **30.06.2022**
Outfitter: **Nike**
Shoe model: **Nike Mercurial Vapor XI Neymar FG**
since: **Jul 10, 2017**
Social media: [Twitter](#) [Facebook](#) [Instagram](#)

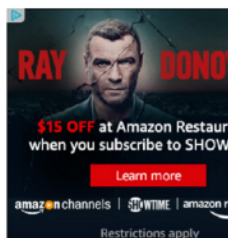
Detailed position

Main position:
Left Wing
Other position(s):
Centre-Forward
Secondary Striker



Market value development

Current market value:
100,00 Mill. €
Last change:
Jun 5, 2017
Highest market value:
100,00 Mill. €
New 77 2015



INTERNATIONAL MATCH CAREER

Order	National team	Debut	Age	Goals
10	Brazil	Aug 10, 2010	77	52
11	Brazil U20	-	7	9
11	Brazil U17	Oct 24, 2009	3	1
10	Brazil Olympic Team	Jul 26, 2012	14	8

[Go to the national player profile](#)

COMPARABLE PLAYERS

Player	Age	Club	Market value
Eden Hazard Left Wing	26	Chelsea	75,00 Mill. €

COMPARE NEYMAR WITH ...

search for players

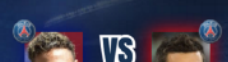
[Compare players](#)

DISCUSSIONS...

Forum	Link to thread
FC Barcelona	Link to thread

WHOM DO YOU PREFER?

Which player do you prefer...



TRANSFER HISTORY

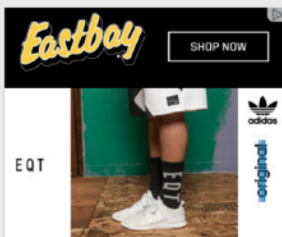
All the player's transfers.

Season	Date	Moving from	Moving to	MV	Transfer fee
17/18	Aug 3, 2017	FC Barcelona	Paris SG	100,00 Mill. €	222,00 Mill. €
13/14	Jul 1, 2013	Santos FC	FC Barcelona	50,00 Mill. €	88,20 Mill. €
08/09	Jan 1, 2009	Santos FC U20	Santos FC	-	-

Total transfer proceeds: **310,20 Mill. €**

JUNIOR CLUBS

Portuguesa Santista (1999–2003)



PERFORMANCE DATA 17/18

Competition	Goals	Assists	Minutes	Goals per minute	Assists per minute
Ligue 1	3	3	90	0,03	0,03
Total 17/18:	3	3	90	0,03	0,03

[Complete performance data](#)

Distribution of size of crawled pages:

