Cairo University
Faculty of Computers & Artificial Intelligence
Computer Science Department
Year 2021 – 2022
First Term

Semantic Web Project.

Deadline of assignment submission is 4/1/2022.

The students per group should not exceed 3 students.

Projects Goal.

The final goal is to be able to

1- Create ontology using different resources

2- Use SPARQL to process that knowledge.

Project 1: Football League Ontology

Do the following-:

1- Transform data from football database to RDF using the Jena framework+ Portege. In order to do that, it is necessary to:
   a. Choose a namespace for the IRIs
   b. Decide what data needs to be transformed into classes, resources, properties or Literals
   c. Process the data to feed into the model,

   There are 3 CSV files, one file contains statistics about players, another one contains statistics about teams and the other file contains statistics about matches as in the snapshots.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	full_name	age	league	season	position	Current Cl	minutes_p	minutes_p	minutes_p	nationality	appearanc	appearanc	appearanc	goals_ove	goals_hom
	Aaron Cre	31	Premier League	2018/2019	Defender	West Ham	1589	888	701	England	20	11	9	0	0
	Aaron Leni	33	Premier League	2018/2019	Midfielder	Burnley	1217	487	730	England	16	7	9	1	1
	Aaron Moo	30	Premier League	2018/2019	Midfielder	Huddersfie	2327	1190	1137	Australia	29	15	14	3	1
	Aaron Ran	30	Premier League	2018/2019	Midfielder	Arsenal	1327	689	638	Wales	28	14	14	4	2
	Aaron Rov	20	Premier League	2018/2019	Forward	Huddersfie	69	14	55	England	2	1	1	0	0
	Aaron Wai	23	Premier League	2018/2019	Midfielder	Crystal Pal	3135	1605	1530	England	35	18	17	0	0
	Abdelhami	24	Premier League	2018/2019	Midfielder	Huddersfie	49	0	49	Morocco	2	0	2	0	0
	Abdoulaye	28	Premier League	2018/2019	Midfielder	Watford	3062	1566	1496	France	35	18	17	5	3
	Aboubakai	26	Premier League	2018/2019	Forward	Fulham	687	468	219	France	13	8	5	3	1
	Adalberto	23	Premier League	2018/2019	Forward	Watford	0	0	0	Venezuela	0	0	0	0	0
	Adam Dav	32	Premier League	2018/2019	Midfielder	Liverpool	465	189	276	England	13	6	7	0	0
	Adam Mas	27	Premier League	2018/2019	Defender	Watford	1003	463	540	Italy	14	7	7	0	0
	Adam Smit	29	Premier League	2018/2019	Defender	AFC Bourn	2073	1051	1022	England	25	12	13	1	1

Cairo University
Faculty of Computers & Artificial Intelligence
Computer Science Department
Year 2021 – 2022
First Term

timestamp	date_GMT	status	attendance	home_team_name	away_team_name	referee	Game Week	Pre-Match	Pre-Match	home_ppg	away_ppg	home_tea	away_tea	total_goal	tota
1.53E+09	Aug 10 201	complete	74439	Manchester Unite	Leicester City	Andre Marriner	1	0	0	1.89	1.32	2	1	3	
1.53E+09	Aug 11 201	complete	51749	Newcastle United	Tottenham Hotspur	Martin Atkinson	1	0	0	1.32	1.74	1	2	3	
1.53E+09	Aug 11 201	complete	10353	AFC Bournemouth	Cardiff City	Kevin Friend	1	0	0	1.53	0.74	2	0	2	
1.53E+09	Aug 11 201	complete	24821	Fulham	Crystal Palace	Mike Dean	1	0	0	1.11	1.53	0	2	2	
1.53E+09	Aug 11 201	complete	24121	Huddersfield Tow	Chelsea	Chris Kavanagh	1	0	0	0.47	1.58	0	3	3	
1.53E+09	Aug 11 201	complete	20051	Watford	Brighton & Hove Albi	Jonathan Moss	1	0	0	1.42	0.68	2	0	2	

-

2- Deploy RDF Files to apache Jena FUSEKI

3- Define the necessary SPARQL queries to extract the necessary data, such as:

a- Find top 10 goal scorers.

b- List all players given particular nationality.

c- List all players who play in a specific team.

d- List all matches played with a specific referee.

e- List all team matches with the result given particular team.

The TA May ask you to write other queries.

Project 2:

The final goal is to integrate information coming from different sources, deploy a SPARQL

Endpoint to query the data, and obtain answers and information related to their

Needs. Specifically, you have to combine data from TER lines and schedules

(https://ressources.data.sncf.com/explore/dataset/sncf-ter-gtfs/) with data coming from Wikidata.

A user should be able to do the following:

1- Search for a combination of trains that allow him/her to travel from one city to another.

2- Have the ability to get information about relevant places and events near the stops along the travel extracted from Wikidata or DBpedia.

Do the following-:

1. Transform data from TER lines and schedules to RDF using the Jena framework. In In order to do that, it is necessary to:

a. Choose a namespace for the IRIs

b. Decide what data needs to be transformed into either classes, resources,

Properties or literals

c. Process the data to create the triples

2. Deploy RDF Files to apache Jena FUSEKI.
3. Define the necessary SPARQL queries to extract the necessary data, such as:
   a. Find the city that that is closer to a stop, according to their latitude and longitude Values
   b. Find the possible trains to go from one city to another, and the stops
   c. Find all stop areas with their stop points.

   Note that city include stop areas and stop areas include stop points

   + Cities data included

* **The required files are attached**.