

Building English Premier League (EPL) Knowledge Graph and Game Prediction

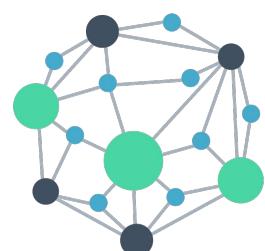
Jae Young Kim, Dong-Ho Lee



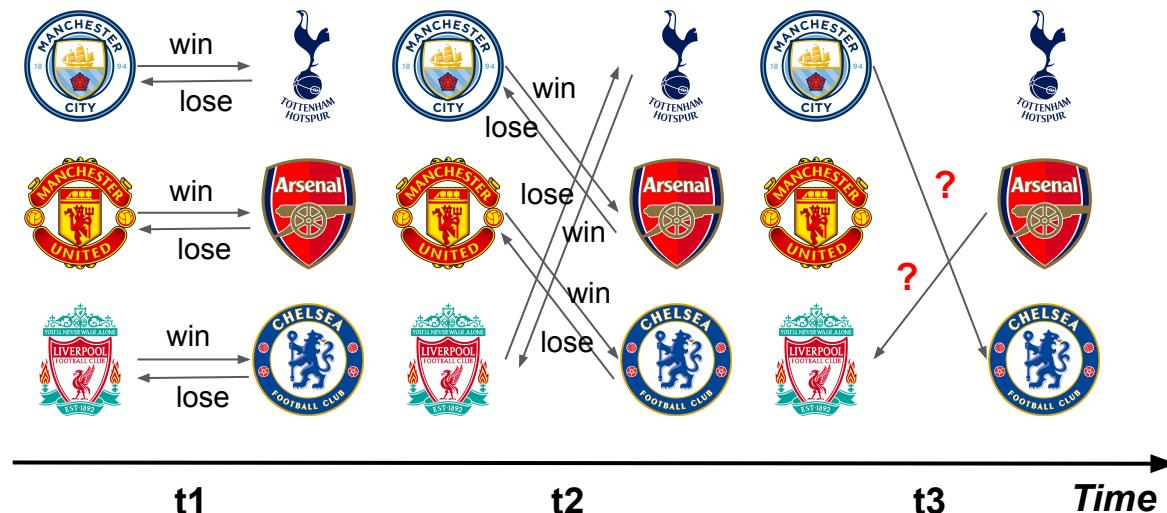
Goal / Objective

Building EPL Knowledge Graph and Game prediction through temporal information

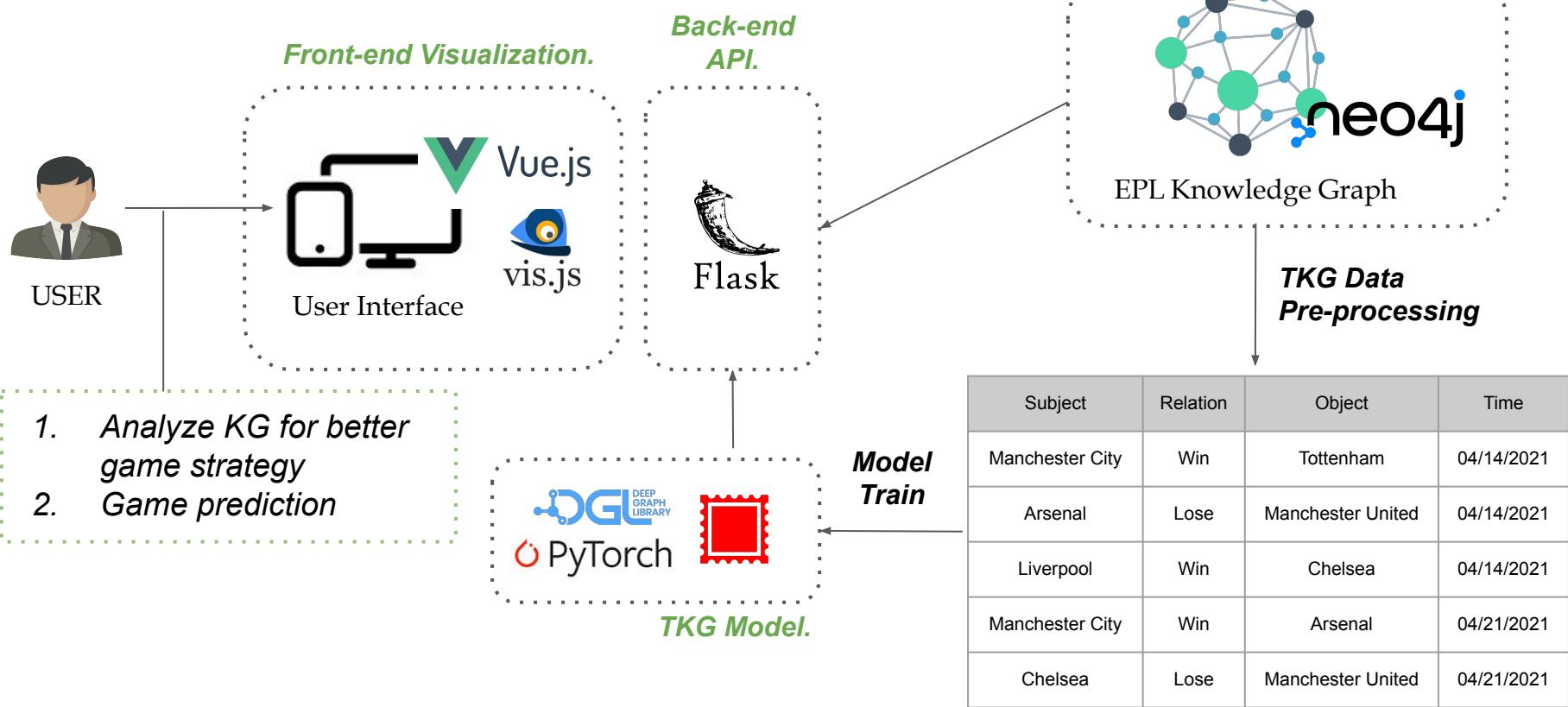
KG Construction
: EPL Premier League



Temporal KG Forecasting (Extrapolation)
: What is the next event would be ? (s, r, ?) / (s, ?, o) / (? , r, o)



Technical Challenge



1. Analyze KG for better game strategy
2. Game prediction



Overview

1. Knowledge graph construction

- a. Data source & Crawling
- b. Entity Resolution
- c. Knowledge graph structure & database

2. Knowledge graph analysis

- a. Query engineering
- b. Front-end visualization

3. Temporal knowledge graph

- a. Data preprocessing
- b. Data statistics & Task
- c. Evaluation
- d. Front-end visualization

Data Source

worldfootball.net

Score: 3:1

Goals:

- 1:0 Emile Smith Rowe (L) (right footed shot) (Bakayoko Salako)
- 2:0 Pierre-Emerick Aubameyang (left footed shot) (Emile Smith Rowe)
- 2:1 Bakayoko Salako (right footed shot)
- 2:2 Heung-min Son (right footed shot) (Heung-min Son)

Height: 183 cm
Weight: 78 kg
Position: Right Winger
Foot: both

Player	Position	Nationality
Aaron Ramsdale	GK	England
Tom Lawrence	RB	England
Ben White	CB	England
Gabriel	LB	Colombia
Tahith Chong	DM	England
Thierry Henry	FW	France
Bukayo Saka	RW	England
Martin Ødegaard	MF	Norway
Emile Smith Rowe	CM	England
Granit Xhaka	CM	Switzerland
Pierre-Emerick Aubameyang	ST	Germany
Heung-min Son	RW	South Korea
Eric Dier	CB	England
Matt Doherty	LB	Ireland
Emerson Royal	DM	Brazil
Marcus Rashford	ST	England

Whoscored.com

Tottenham Fixtures

- W CL 26-08-21 Tottenham 3:0 Pacos de Ferreira
- W EPL 29-08-21 Tottenham 1:0 Watford
- L CL 11-09-21 Crystal Palace 1:0 Tottenham
- D CL 16-09-21 Rennes 2:2 Tottenham
- L EPL 19-09-21 Tottenham 0:3 Chelsea
- W EFLC 22-09-21 Wolverhampton W... 2:2 Tottenham
- EPL 26-09-21 Arsenal vs Tottenham
- CL 30-09-21 Tottenham vs Mura

Squad

Son Heung-Min

Name: Son Heung-Min
Current Team: Tottenham
Age: 29 years old (08-07-1992)
Nationality: South Korea

Full Name: Heung-Min Son
Shirt Number: 7
Height: 183cm
Positions: Midfielder (Centre, Left, Right), Forward

Specific information of teams & players

Team & player's game records

Wikipedia

Tottenham Hotspur F.C.

Founded: 1882
Stadium: White Hart Lane
Homepage: [www.spurs.co.uk](http://spurs.co.uk)

Ref: New score View history Search Wikipedia

Son Heung-min

Name: Son Heung-min
Born: 15 July 1992
Place of birth: Gyeongju, South Korea
Nationality: South Korea
Height: 183 cm
Weight: 78 kg
Position: Right Winger
Foot: both

Full Name: Heung-Min Son
Shirt Number: 7
Height: 183cm
Positions: Midfielder (Centre, Left, Right), Forward

Wikipedia infobox for organization

Datahub EPL Data

Date	HomeTeam	AwayTeam	FTHG	FTDG	PTG	HTG	HTDG	HTR	Referees	HD	HST	HT	HF	AF	HY	AV	HR	BSGH	BSGD	
01/08/2019	Leicester	Watford	2	1	1	0	0	H	A.Martinez	8	13	8	4	11	1	2	1	0	1	3.8
01/08/2019	Everton	Sheffield United	1	0	1	1	0	H	J.Barker	10	10	6	6	9	11	9	1	2	0	2.5
01/08/2019	Crystal Palace	2	0	2	0	1	0	M.Dean	10	10	6	6	9	11	9	1	2	0	3.4	
01/08/2019	Huddersfield	Chelsea	0	3	0	0	2	A	C.Kanevsky	6	13	1	4	9	8	2	2	1	0	6.5
01/08/2019	Newcastle	Tottenham	1	2	1	1	2	A	M.Abdullah	15	2	5	11	12	3	2	2	0	5.9	
01/08/2019	West Ham	Leeds United	0	1	0	0	0	H	A.Martinez	10	10	6	6	9	11	9	1	2	0	3.2
01/08/2019	Everton	2	0	2	1	1	0	C.Pavard	11	6	4	5	8	7	3	0	1	0	3.3	
01/08/2019	Man City	0	2	0	0	1	0	M.Oliver	9	17	3	8	11	9	2	2	0	0	4	
01/08/2019	Everton	4	0	4	0	2	0	A.Taylor	18	18	8	8	2	14	9	2	0	0	1.9	
01/08/2019	Leeds United	0	0	0	0	0	0	A.Martinez	10	10	6	6	9	11	9	1	2	0	3.3	
01/08/2019	Cardiff	Newcastle	0	0	0	0	0	D.Chowdhury	12	12	1	6	14	5	1	2	0	1	3.1	
01/08/2019	Chelsea	Ajax	3	2	0	2	0	D.Martinez	24	15	11	6	12	9	5	0	2	0	1.8	
01/08/2019	Leicester	Leeds United	0	0	0	0	0	A.Martinez	10	10	6	6	9	11	9	1	2	0	3.2	
01/08/2019	Leicester	Wolves	2	0	1	2	0	H.M.Dean	4	11	3	3	10	8	1	1	2	0	2.94	
01/08/2019	Tottenham	Rotherham	3	1	0	0	0	A.Taylor	25	10	11	3	5	3	0	0	0	1.38		
01/08/2019	West Ham	Brunswick	1	2	1	0	0	H.G.Skelton	11	12	5	5	14	10	4	2	0	0	3.8	
01/08/2019	West Ham	Sheffield United	0	1	0	0	0	A.Taylor	10	10	6	6	9	11	9	1	2	0	3.3	
01/08/2019	Brunswick	Leeds United	1	3	1	1	0	P.Tarney	8	9	3	6	9	18	9	1	2	0	2.45	
01/08/2019	Huddersfield	Man City	6	1	3	1	0	A.Martinez	32	9	14	1	9	4	10	5	2	0	1.1	
01/08/2019	Cardiff	Tottenham	0	0	0	0	0	M.Oliver	8	16	2	6	6	12	9	1	1	1	1.9	
01/08/2019	Everton	Leeds United	2	0	2	0	0	L.Hoban	17	11	5	3	12	10	6	1	0	0	3.5	
01/08/2019	Wolverhampton	Everton	2	2	0	0	0	D.Chowdhury	5	14	1	4	8	10	7	0	1	0	3.7	
01/08/2019	Wolverhampton	Cardiff	0	0	0	0	0	M.Oliver	5	14	1	4	8	10	7	0	1	0	3.7	
01/08/2019	Liverpool	Brighton	1	0	0	0	0	C.Kanevsky	23	4	3	2	8	14	8	1	1	0	1.16	

Structured knowledge
(Tabular data of games)

Crawling



Scrapy



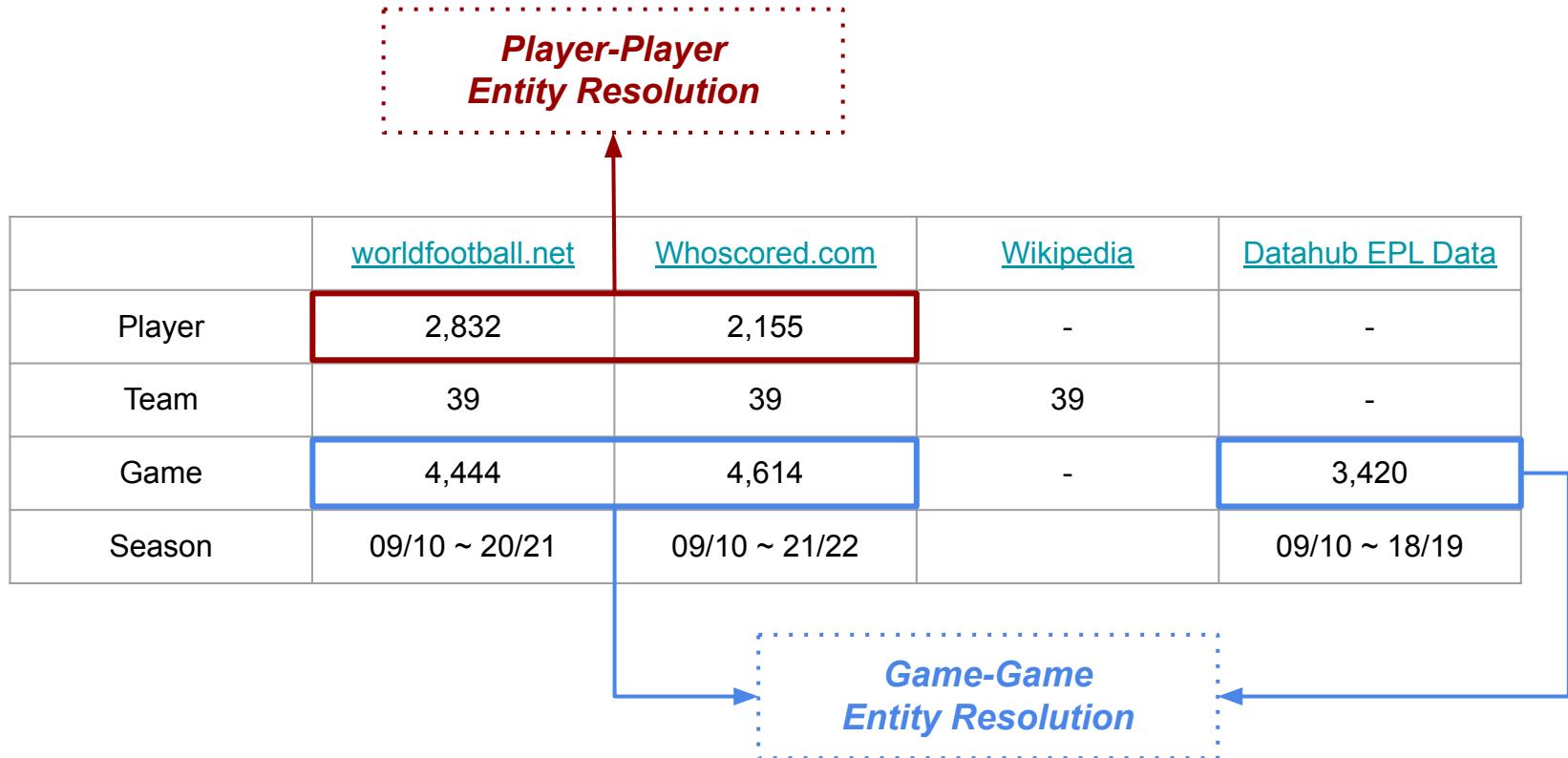
Selenium⁴

Dynamic page Crawling

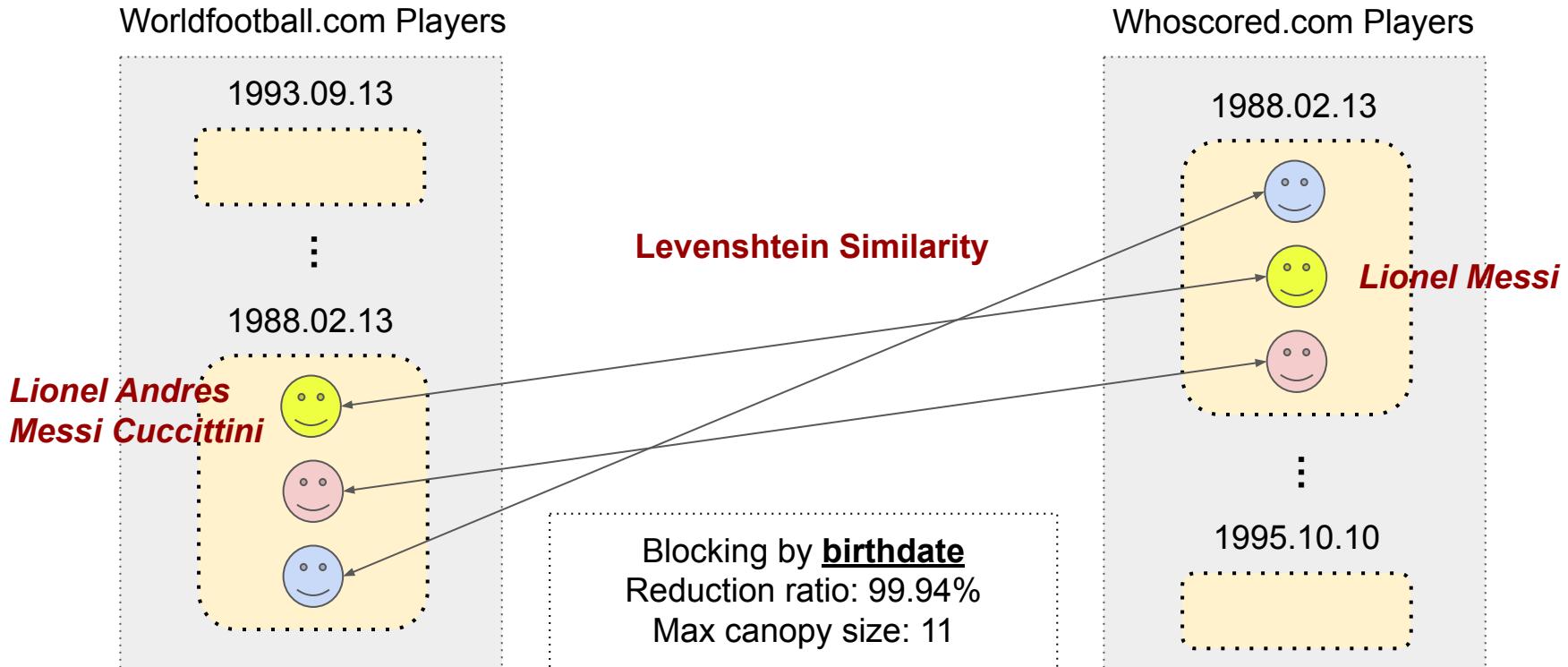
CSV file
(Structured)

	worldfootball.net	Whoscored.com	Wikipedia	Datahub EPL Data
Player	2,832	2,155	-	-
Team	39	39	39	-
Game	4,444	4,614	-	3,420
Season	09/10 ~ 20/21	09/10 ~ 21/22		09/10 ~ 18/19

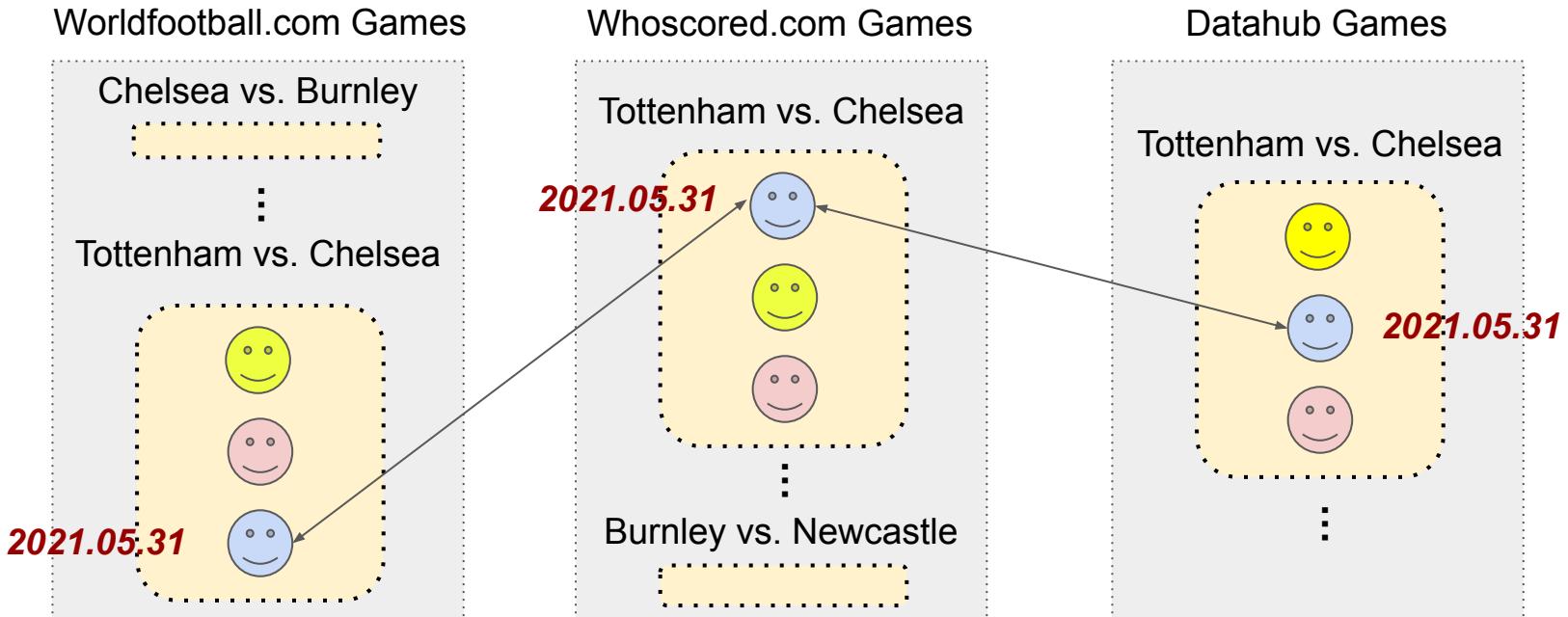
Entity Resolution



Player Entity Record Linkage



Game Entity Record Linkage

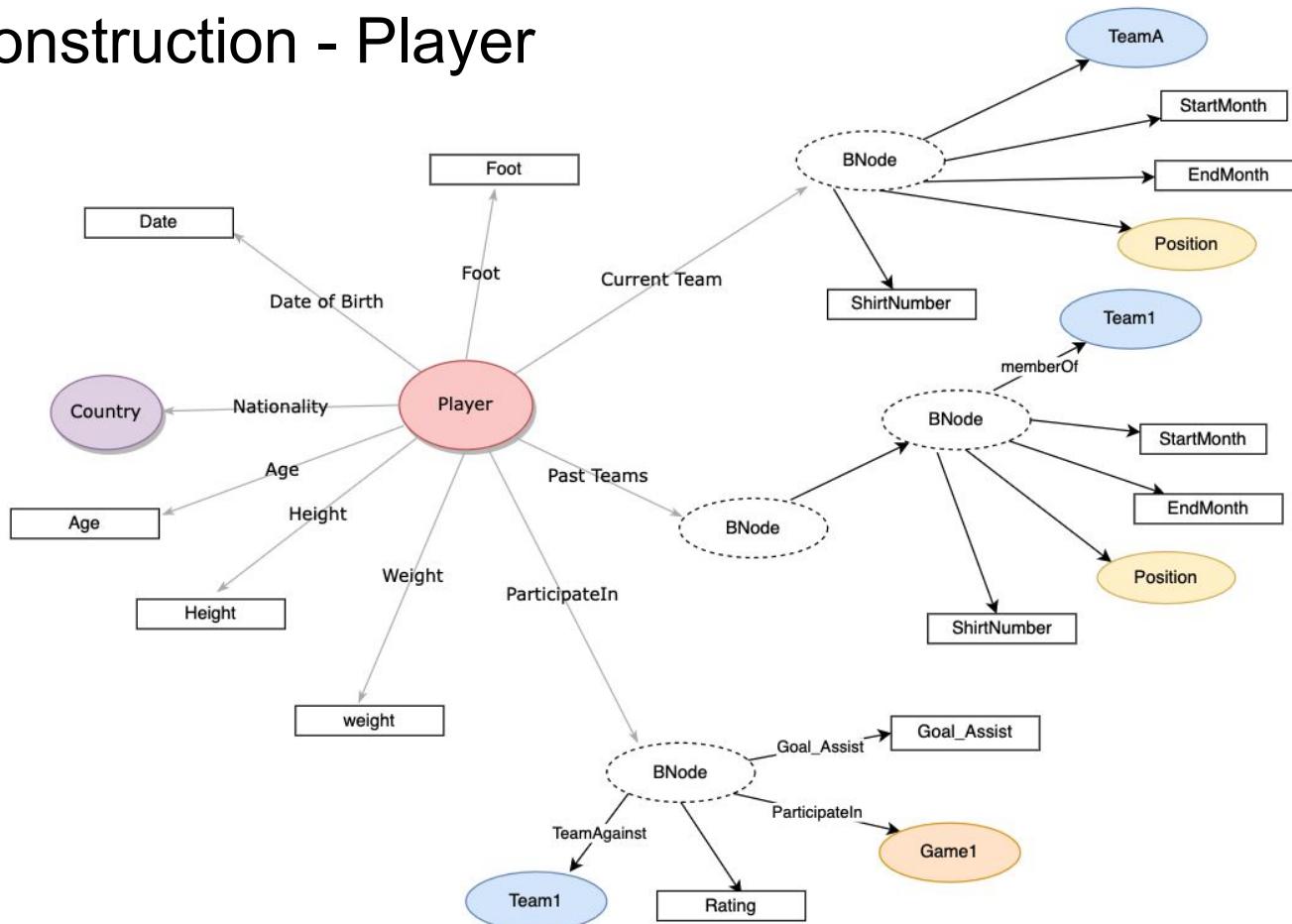


Blocking by Home team vs. Away team / Match by Date

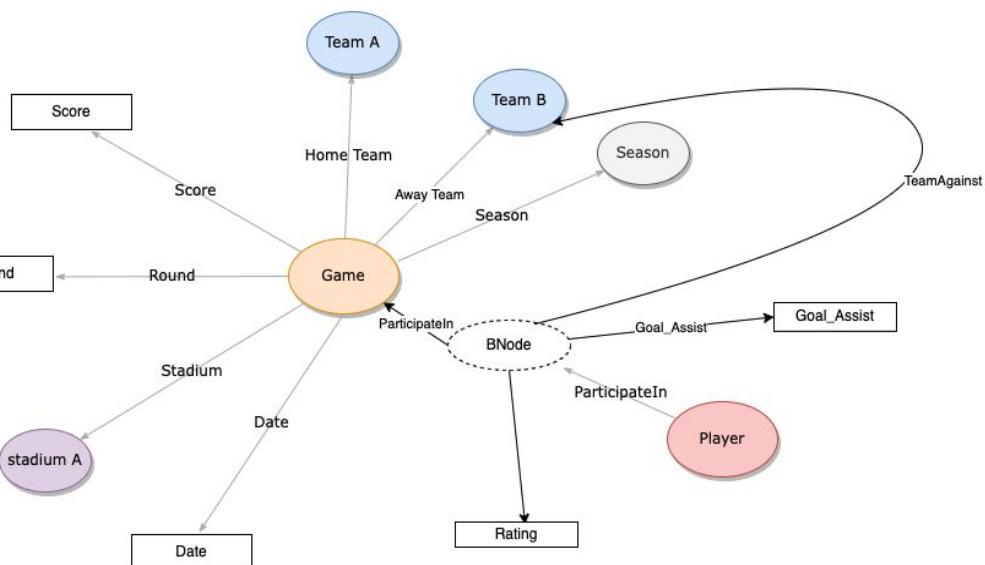
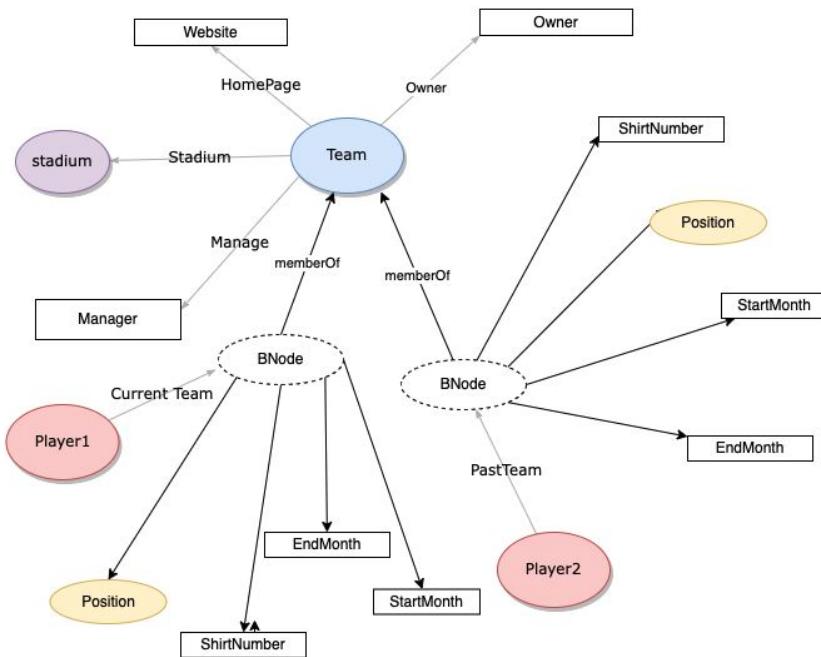
Reduction ratio: 99.86% , 99.87%

Max canopy size: 27, 23

KG Construction - Player



KG Construction - Team & Game



KG Database Management

RDFlib -> .ttl

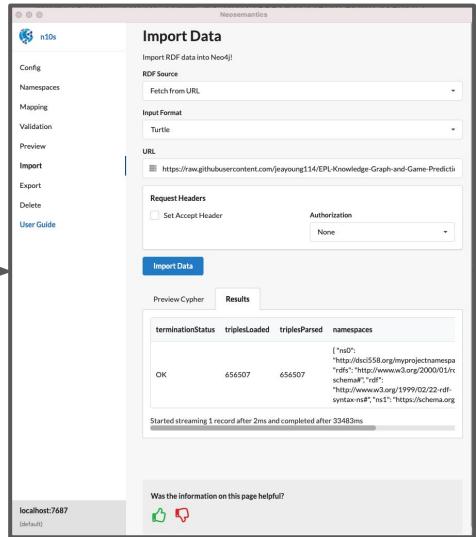
```
@prefix my_ns: <http://dsci558.org/myprojectnamespace#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <https://schema.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

my_ns:0_14_2012_08_19 a my_ns:event ;
    my_ns:dateTime "2012-08-19T13:30:00"^^xsd:dateTime ;
    my_ns:awayTeam my_ns:Chelsea_FC ;
    my_ns:homeTeam my_ns:Wigan_Athletic ;
    my_ns:roundof 1 ;
    my_ns:scoreOf "0:2" ;
    my_ns:seasonOf my_ns:2012_2013 .

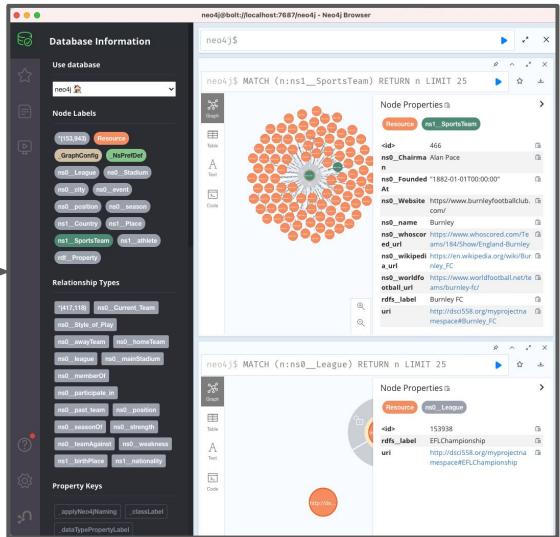
my_ns:0_21_2010_08_14 a my_ns:event ;
    my_ns:dateTime "2010-08-14T15:00:00"^^xsd:dateTime ;
    my_ns:awayTeam my_ns:Blackpool_FC ;
    my_ns:homeTeam my_ns:Wigan_Athletic ;
    my_ns:roundof 1 ;
    my_ns:scoreOf "0:4" ;
    my_ns:seasonOf my_ns:2010_2011 .

my_ns:0_27_2011_08_13 a my_ns:event ;
    my_ns:dateTime "2011-08-13T15:00:00"^^xsd:dateTime ;
    my_ns:awayTeam my_ns:Norwich_City ;
    my_ns:homeTeam my_ns:Wigan_Athletic ;
    my_ns:roundof 1 ;
    my_ns:scoreOf "1:1" ;
    my_ns:seasonOf my_ns:2011_2012 .
```

Neosemantics



Neo4J



Query Engineering

Question	Query Template
What is the information of <u>player</u> ?	Information of [Player]
What is the information of <u>team</u> ?	Information of [Team]
Which teams are participated in EPL <u>~~~~ season</u> ?	List of teams in [Season]
Where is the birthplace of <u>player</u> ?	[Player]'s birthplace
What is the play style of <u>player</u> ?	Play style of [Player]
How was the performance of <u>player</u> against other teams ?	[Player]'s match records against each team
Who have specific <u>skill</u> in <u>team</u> ?	Players who has [Skill] in [Team]
Who are the best players in <u>team 1</u> against <u>team 2</u> ?	Best players in [Team1] against [Team2]



Front-end Visualization - Template Selection

DSCI558 - EPL KNOWLEDGE GRAPH

SEARCH TKG PREDICT

Search

- Information of [Player]
- Information of [Team]
- List of teams in [Season]
- [Player]'s birthplace
- Play style of [Player]
- [Player]'s match records against each team
- Players who has [Skill] strength in [Team]
- Best players in [Team1] against [Team2]

Template Selection

Search

Best players in [Team1] against [Team2]

Best players in [Team1] against [Team2] against SEARCH

Template text fields

Front-end Visualization - KG Analysis

Search

Best players in [Team1] against [Team2]

Best players in Tottenham hotspur  against Chelsea  SEARCH



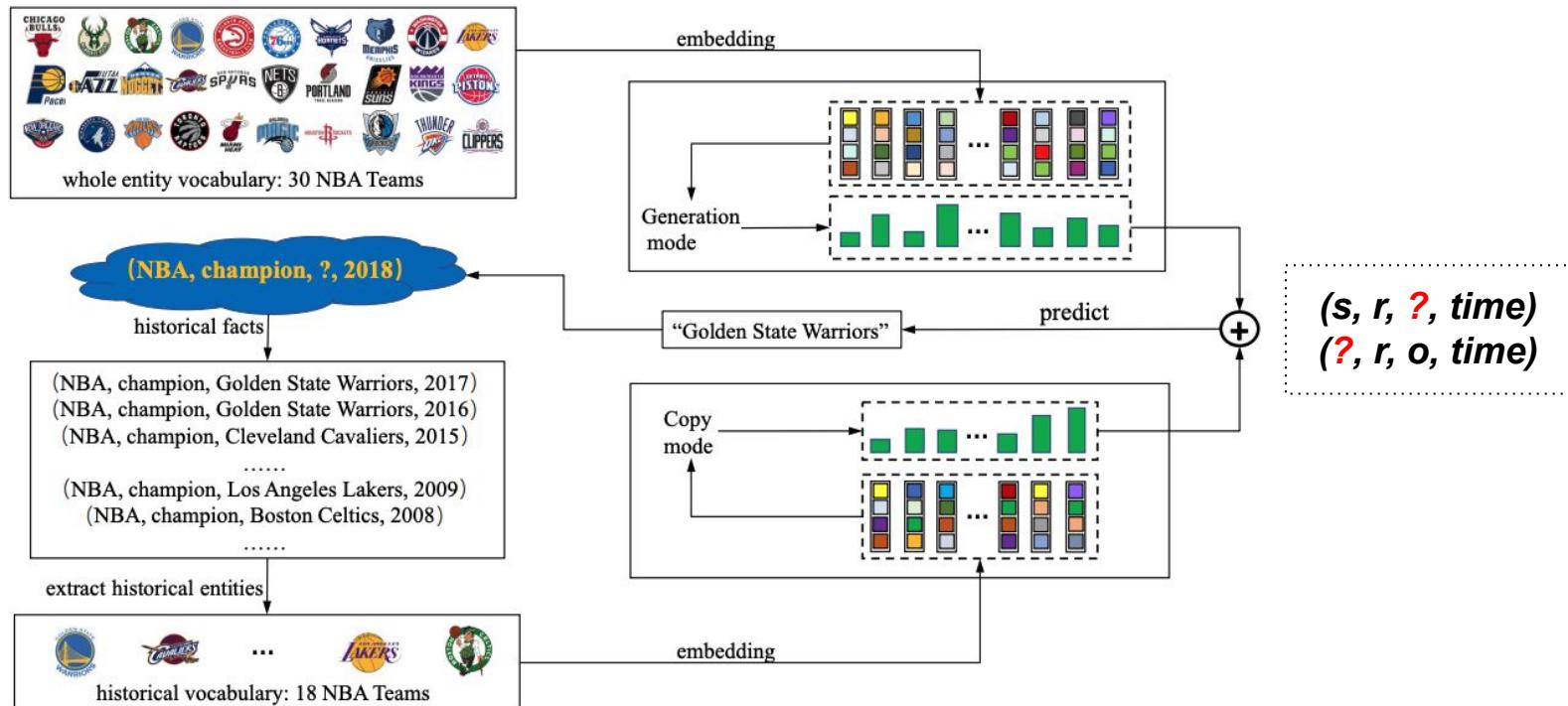
Avg_rating	Name
7.6	Giovani Lo Celso
7.416666666666666	Dele Alli
7.1454545454545455	Harry Kane
7.125	Davinson Sánchez
6.925	Matt Doherty

Template text fields

Query results

Temporal Knowledge Graph (TKG) Overview

CyGNet (AAAI 2021)



Data preprocessing for TKG

Game Records

Subject	Relation	Object	Season	Round
Chelsea FC	win	Hull City	2009/2010	1
Everton FC	lose	Arsenal FC	2009/2010	1
Liverpool FC	win	Stoke city	2009/2010	2
⋮				
Chelsea FC	draw	Manchester United	2021/2022	13
Leicester City	win	Watford FC	2021/2022	13
Brentford FC	win	Everton FC	2021/2022	13

Diagram illustrating the data preprocessing flow:

```

graph LR
    A[Game Records] --> B[Relation_to_idx / Entity_to_idx]
    B --> C[Training Data]
    B --> D[Validation Data]
    B --> E[Testing Data]
    C --> D
    C --> E
    D --> E
  
```

The original "Game Records" table is processed by a mapping function (`Relation_to_idx / Entity_to_idx`) to produce three datasets:

- Training Data:** Contains 18 rows.
- Validation Data:** Contains 8 rows.
- Testing Data:** Contains 13 rows.

⋮ indicates continuation between the validation and testing datasets.

Subject	Relation	Object	Time
18	1	19	1
8	0	9	1
13	1	6	2
⋮			
18	2	10	468
30	1	32	468
38	1	8	468

Train 80% / Valid 10% / Test 10%
 Train time < Valid time < Test time

TKG Performance

# Entity	# Relation	# Train	# Valid	# Test	# Total
38	3	3,848 (2009/2010 round 1 ~ 2018/2019 round 34)	482 (2018/2019 round 35 ~ 2020/2021 round 4)	482 (2020/2021 round 5 ~ 2021/2022 round 13)	4,812
		Hits@1	Hits@3	Hits@5	Hits@10
Subject Prediction		0.2714	0.5219	0.6785	0.8288
Object Prediction		0.2881	0.5741	0.7035	0.8622
Total		0.2797	0.5480	0.6910	0.8455
		Hits@15	Hits@15	MRR	
		0.9102	0.9311	0.4447	0.4711
		0.9207	0.4579		

(Tottenham hotspur, Win, ?, 2021/2022 round 3)

Chelsea 0.02
 Manchester City 0.01
 ...
 Burnley 0.3
 Newcastle 0.2

Rank among 38 entities.
 "Is correct prediction
 appeared in top k ?"

TKG Inference

# Entity	# Relation	# Train	# Valid	# Test	# Total
38	3	4,331 (2009/2010 round 1 ~ 2020/2021 round 4)	481 (2020/2021 round 4 ~ 2021/2022 round 13)	260 (2021/2022 round 14 ~ 2021/2022 round 38)	5.072

(Tottenham hotspur, ?, Brentford FC, 2021/2022 round 14)

Brentford FC's rank in prediction

(Tottenham hotspur, **win**, ?, 2021/2022 round 14)

2

(Tottenham hotspur, **draw**, ?, 2021/2022 round 14)

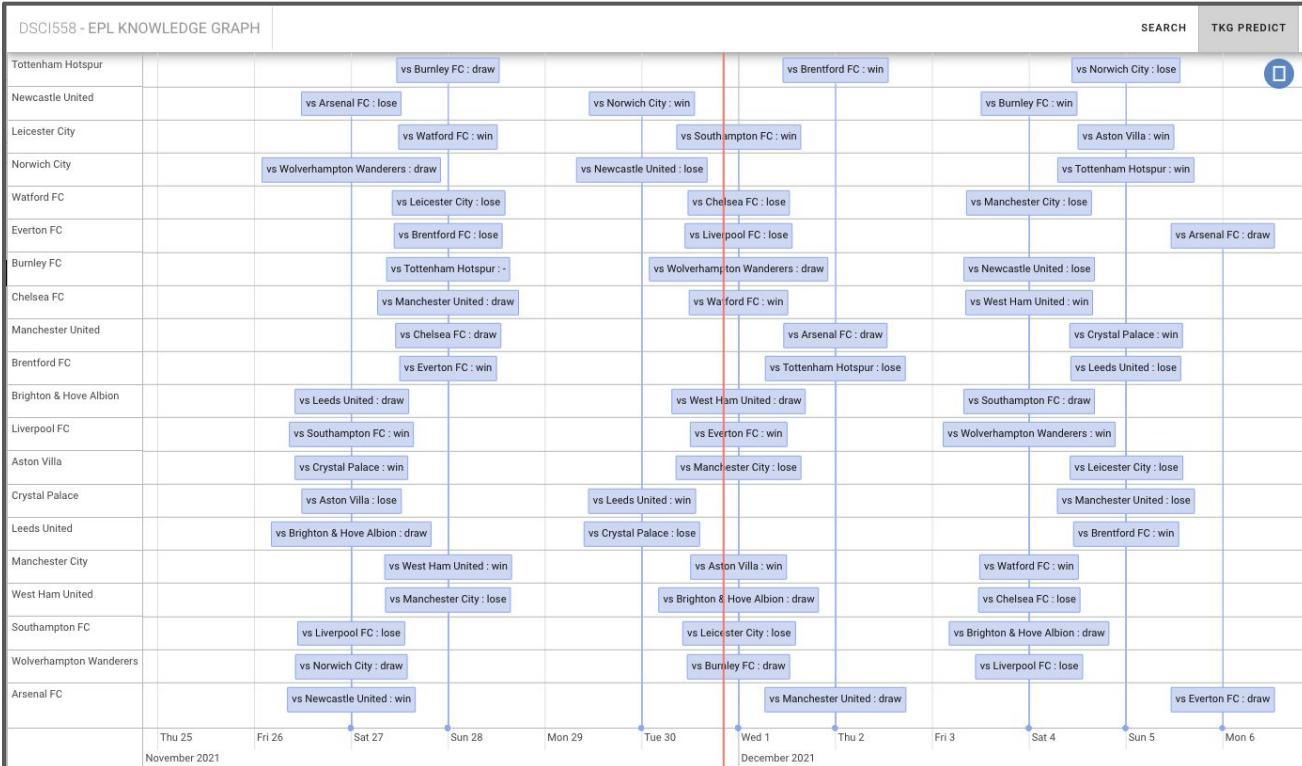
9

(Tottenham hotspur, **lose**, ?, 2021/2022 round 14)

27

**Tottenham hotspur
will **win** over
Brentford FC**

Front-end Visualization - TKG Prediction



Previous records



Forecast