

Report Semantic Web

Maxime
Philippe

This report aims to investigate the utility of querying over Linked Data, utilizing datasets like DBpedia and Wikidata. The queries examined will include both pre-defined and self-generated questions, and their answers will be explored in-depth. In particular, we will delve into a hybrid approach that combines the capabilities of Linked Data with the large language model (LLM) methods, such as ChatGPT, to provide comprehensive and accurate answers.

We use wiki data as our Linked Database. In Wikidata, wdt: provides a straightforward view of data, giving basic answers quickly. It's used for general information. On the other hand, p: offers a detailed view, showing all aspects of the data. We use wdt: for simple queries and p: for in-depth ones.

ChatGPT

We utilized ChatGPT to identify the linked database and structure our queries. However, there were instances where ChatGPT provided incorrect values for certain elements, leading to results that did not match with our expectations.

Questions

Question 1

Find that landmark article on data integration written by an Indian researcher in the 1990s.

```
SELECT DISTINCT ?item ?itemLabel ?date ?awardLabel ?work ?workTitle WHERE {  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }  
  {  
    SELECT DISTINCT ?item ?date ?award ?work ?workTitle WHERE {  
      ?item wdt:P31 wd:Q5. # nature of element is Human  
      ?item wdt:P27 wd:Q668. # country of citizenship is India  
      ?item (wdt:P106/(wdt:P279*)) wd:Q1650915. # Occupation is (subclass of) researcher  
      ?item (wdt:P101/(wdt:P279*)) wd:Q2374463. # field of activity is (subclass of) Data Science  
      ?item p:P166 ?award_statement. # awards received as statement  
      ?award_statement ps:P166 ?award. # award recieved as ?award  
      ?award_statement pq:P1686 ?work. # for work ?work  
      ?work wdt:P1476 ?workTitle. # with title ?workTitle  
      ?award_statement pq:P585 ?date. # at date ?date  
      FILTER(?date < "+2000-01-01T00:00:00Z"^^xsd:dateTime). # ?date before 2000  
      FILTER(?date > "+1989-01-01T00:00:00Z"^^xsd:dateTime). # ?date after 1989  
    }  
  }  
}
```

RESULT:

item	itemLabel	date	awardLabel	work	workTitle
Q593034	Ravindran Kannan	1 janvier 1991	prix Fulkerson	Q55671716	A random polynomial-time algorithm for approximating the volume of convex bodies

Question 2

Who is the French archer who has won an Olympic event and received the Honor Legion ?

```
SELECT DISTINCT ?item ?itemLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
  {
    SELECT DISTINCT ?item WHERE {
      ?item wdt:P31 wd:Q5. # nature of element is Human
      ?item wdt:P27 wd:Q142. # country of citizenship is France
      ?item wdt:P106 wd:Q13382355. # occupation is archery (job not hobby)
      ?item p:P1344 ?event_statement. # participate in events as statement
      ?event_statement (ps:P1344/(wdt:P31*)) wd:Q18536594. # event is (instance of) Olympic event
      ?event_statement pq:P166 wd:Q15243387. # award received is Olympic gold medal
      ?item (wdt:P166/(wdt:P279*)) wd:Q10855271. # award received is (subclass of) Knight of the Legion of Honour
    }
  }
}
```

RESULT:

item	itemLabel
Q1311057	Sébastien Flute

Question 4:

What are the 20 books that have the most pages, which is lower than 20,000, and that are written by at least one author who has a child ?

```
SELECT DISTINCT ?item ?itemLabel ?numberOfPages WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
  {
    SELECT DISTINCT ?item ?numberOfPages WHERE {
      ?item wdt:P31 wd:Q571. # nature of element is book
      ?item wdt:P50 ?author. # author is ?author
      ?author wdt:P40 _:anyValueP40. # has child exist
      ?item wdt:P1104 ?numberOfPages. # number of pages is ?numberOfPages
      FILTER(?numberOfPages < 20000) # ?numberOfPages is less than 20 000
    }
  }
}
ORDER BY DESC (?numberOfPages)
LIMIT 20
RESULT:
```

Table			20 résultats en 2895 ms	Code	Télécharger	Lien
Item	ItemLabel	numberOfPages				
Q wd:Q1943729	Systema Mycologicum	1866				
Q wd:Q111602524	Q111602524	1801				
Q wd:Q51407079	Manuel de conchyliologie et de paleontologie conchyliologique : ou, Histoire naturelle des mollusques vivants et fossiles	1486				
Q wd:Q106584307	Q106584307	1246				
Q wd:Q113367291	Q113367291	951				
Q wd:Q99227983	Q99227983	786				
Q wd:Q99912674	Principes du droit des gens	779				
Q wd:Q51393152	A monograph of the British fossil corals	736				
Q wd:Q100255508	Q100255508	735				
Q wd:Q51392595	Grundzüge der Anatomie der wirbellosen Thiere	644				
Q wd:Q112578977	Q112578977	644				
Q wd:Q51469440	Epicrisis Systematis mycologici	610				
Q wd:Q114950569	Q114950569	610				
Q wd:Q4378679	Histoire de l'Égypte ancienne	593				
Q wd:Q111594361	Q111594361	592				
Q wd:Q112766602	Q112766602	592				
Q wd:Q86907717	Q86907717	587				
Q wd:Q106583829	Q106583829	584				
Q wd:Q106407347	Q106407347	577				
Q wd:Q110249810	Dieu, la science, les preuves : l'aube d'une révolution	577				

We also find our question :

Which film has the shortest title in the history of the Cannes Film Festival's Palme d'Or winners and didn't get the award the same year as he got published?


```
SELECT DISTINCT ?item ?itemLabel ?title ?titlelength ?award_year ?earliest_publication_year WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
  {
    SELECT ?item ?title ?titlelength ?award_year (MIN(?publication_year) as ?earliest_publication_year) WHERE
  {
    ?item wdt:P31 wd:Q11424. # nature of element is Movie
    ?item wdt:P1476 ?title. # with title ?title
    ?item wdt:P577 ?publication_date. # published the ?publication_date
    ?item p:P166 ?award_statement. # award statement
    ?award_statement ps:P166 wd:Q179808. # award is Palme d'Or
    ?award_statement pq:P585 ?award_date. # given the ?award_date
    BIND(STRLEN(?title) as ?titlelength). # length of the title
    BIND(YEAR(?publication_date) as ?publication_year). # year of the publication date
    BIND(YEAR(?award_date) as ?award_year). # year of the Palme d'Or award date
  } GROUP BY ?item ?title ?titlelength ?award_year
  HAVING ( ?earliest_publication_year != ?award_year) # didn't get the award the same year as he got
published
}
}
ORDER BY ASC (?titlelength)
LIMIT 5
RESULT:
```

Table			5 résultats en 5 ms	Code	Télécharger	Lien
Item	ItemLabel	title	titlelength	award_year	earliest_publication_year	
Q wd:Q658204	Blow-Up	Blowup	6	1967	1966	
Q wd:Q105393862	TiRane	TiRane	6	2019	2021	
Q wd:Q1346745	If...	If...	6	1969	1968	
Q wd:Q1353441	Que le spectacle commence	All That Jazz	13	1980	1979	
Q wd:Q714524	Quand passent les cigognes	Летят журавли	13	1958	1957	

We compared the result with ChatGPT :

P

Which film has the shortest title in the history of the Cannes Film Festival's Palme d'Or winners and didn't get the award the same year as he got published?

 As of my last update in January 2022, the film with the shortest title among Cannes Film Festival's Palme d'Or winners is "MAS*H," directed by Robert Altman. The film won the award in 1970. However, the movie was actually released in the same year that it won the award, so it doesn't meet your criterion of not winning the award in the same year it was published.

It's worth noting that my data may be outdated, so you may want to check the most current list of Palme d'Or winners for any updates.

We can see that ChatGPT struggled to answer to this question.

Justification

For each query, we verified the results by doing the query on wikidata then checking the Wikipedia pages of the items we obtain at the end to see if they match all criterias.

Conclusion

In conclusion, our study shows the effectiveness of using Wikidata's Linked Data for complex queries. We combined this with ChatGPT for a hybrid approach, finding it useful despite some limitations. Our results were rigorously verified against Wikipedia, confirming the utility and accuracy of this method for data retrieval and analysis. This work suggests that combining Linked Data with machine learning models like ChatGPT could be promising for future research.