В ходе разработки алгоритма преобразования запроса на естественном языке в запрос на языке SPARQL пришлось столкнуться с рядом проблем, потребовавших дополнительного исследования.

В рамках одной онтологии возможно одновременное наличие различных именований одного и того же предиката, что не позволяет делать унифицированные по структуре запросы на языке SPARQL для запросов на естественном языке даже c одинаковой структурой. Например, в онтологии DBpedia в информации о городе предиката связывающий данный город с количеством жителей может иметь одно из следующих именований: «population», «populationTotal», «p», «pop2010census» (пример представлен в таблице 1). Причем предикат «p» при описании города может использоваться как в значении «количество населения», так и – в значении «название района города». Выше описанная проблема делает затруднительным программное построение запроса на языке SPARQL.

Таблица 1 – Предикаты «Население» у разных городов

|  |  |
| --- | --- |
| **Город** | **Предикат** |
| Оттава | population |
| Москва | populationTotal |
| Ульяновск | p |
| Северодвинск | pop2010census |

Проанализировав работы других исследователей можно выделить два применяемых способа решения описанной выше проблемы. Первый заключается в создании собственной онтологии, зачастую на основе данных других онтологий, но со своими классами объектов и предикатами. Поскольку такая онтология создается непосредственно для целевой системы, в нее закладывается необходимая семантическая структура и система имен, что позволяет унифицировать генерацию запроса на языке SPARQL. Но данный подход требует больших затрат времени на разработку собственной онтологии, поэтому в рамках данной работы применятся не будет.

Также в проектах, целью которых является обращение к системе LOD на естественном языке, применяются онтологии, использующие строго описанный набор классов и предикатов. Примером такой онтологии является YAGO, система классов и типов которой основана на наборе классов и предикатов Schema (специальная онтология, описывающая логическую структуру семантической связанности объектов реального мира). Данная онтология предоставляет возможность генерации унифицированных по структуре запросов на языке SPARQL. Но онтология YAGO содержит большое количество слабо связанных данных, т.е. некоторая часть данных об объектах реального мира просто не указана или указана не корректно. Например, в описании городов связь со страной, в которой находится данных город, указывается с помощью предиката комментария, значение которого является строка приблизительно следующего содержания: «город в России» или «Казахстанский город». Хотя, придерживаясь принципов LOD, необходимо было сделать предикат, например, с название «страна», значением которого являлась бы ссылка (URI) на описание необходимой страны.

Подводя итог выше написанному, в рамках данной работы для устранения проблемы с неоднозначностью именования предикатов будет применяться следующий подход разрешения имен: имена предикатов целевой онтологии должны быть связываться с именами предикатов, использующихся в К-представлении запроса на естественном языке.

1. Towards the Representation of Hashtags in Linguistic Linked Open Data Format Thierry Declerck Piroska Lendvai
2. Small in Size, Big in Precision: A Case for Using Language-Specific Lexical Resources for Word Sense DisambiguationSteven Neale, Joao Silva and Antonio Branco
3. Accessing Linked Open Data via A Common Ontology Kiril Simov Atanas Kiryakov
4. A Semantic Based Question Answering System for Thailand Tourism Information Alisa Kongthon, Sarawoot Kongyoung, Choochart Haruechaiyasak and Pornpimon Palingoon
5. Tirad M. Almalahmeh, Sameem Abdul Kareem, Mansoor A. Abdulgabber Semantic recommender system with natural language interface: malaysian tourism industry // International Journal of Scientific & Engineering Research. - 2013. - Volume 4, Issue 9. - С. 1059-1065.
6. Neli Zlatareva, Devansh Amin. Natural Language to SPARQL Query Builder for Semantic Web Applications // Journal of Machine Intelligence and Data Science. - 2021. - Volume 2. - С. 44-53.