SPARQL Query Construction with Monitoring Service for Endpoints

Atsuko Yamaguchi[1], Yasunori Yamamoto[1], Kouji Kozaki[2], Kai Lenz[3], Hiroshi Masuya[4,3], and Norio Kobayashi[3,4]

[1] Database Center for Life Science (DBCLS), ROIS,
178-4-4 Wakashiba, Kashiwa, Chiba, 277-0871 Japan
{atsuko, yy}@dbcls.rois.ac.jp
[2] The Institute of Scientific and Industrial Research (ISIR), Osaka University,
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
kozaki@ei.sanken.osaka-u.ac.jp
[3] Advanced Center for Computing and Communication (ACCC), RIKEN,
2-1 Hirosawa, Wako, Saitama, 351-0198 Japan
{kai.lenz, norio.kobayashi}@riken.jp
[4] BioResource Center (BRC), RIKEN,
3-1-1, Koyadai, Tsukuba, Ibaraki, 305-0074 Japan
hmasuya@brc.riken.jp

Abstract. Many databases in life sciences have been provided in Resource Description Framework (RDF) with their SPARQL endpoints. To support a user in constructing SPARQL queries, we developed a service called SPARQL Builder. Although SPARQL Builder makes a user write a query for obtaining data from LOD, empty results or incorrect data might be retrieved by the query if the SPARQL endpoint for the query is not alive or provides out-of-date data. To avoid such queries from SPARQL Builder, YummyData which monitors SPARQL endpoints and gathers information about alive rate, last update, response time, etc. was used. By connecting YummyData to SPARQL Builder, our system can always obtain a SPARQL query for SPARQL endpoints with high usability.

Keywords: SPARQL, RDF, life-science databases

1 Introduction

SPARQL Builder (http://www.sparqlbuilder.org/) is a web application that enables users to access life science data provided as Linked Open Data by assisting them in writing SPARQL queries for the data [1]. Although SPARQL is a standardized query language for searching Linked Open Data and there are many Web APIs, called SPARQL endpoints, that accept SPARQL queries, a user may have difficulties in writing a SPARQL query because the user requires knowledge of the linked data schema including data classes and terms in a wide variety of published life science data in advance. Our tool assists users to build a SPARQL

query for SPARQL endpoints of life science databases without knowledge of data schemata or SPARQL by providing an intuitive graphical user interface (GUI).

Our system use metadata extracted from SPARQL endpoint in advance, to display data schema quickly. Because point in time when a user accesses SPARQL Builder is different from the time when metadata was extracted, some SPARQL endpoints might be down or out-of-date. Therefore, our system requires a monitoring service for SPARQL endpoints. In this paper, we used a monitoring service called YummyData [2] and tried it to connect to SPARQL Builder. By obtaining the latest states of SPARQL endpoints from YummyData, SPARQL Builder always outputs a SPARQL query by which a user can obtain data from reliable SPARQL endpoints.

2 Method and Result

To support a user in constructing a SPARQL query, SPARQL Builder first displays classes that appear in life science RDF datasets. All of the classes are associated with SPARQL endpoints. YummyData has a Web API called Umaka API [3] providing a list of SPARQL endpoints with their information about accessibility and usability. Using the API, we added a filter of SPARQL endpoints into SPARQL Builder, to select SPARQL endpoints with alive rate ≥ 80 and execution time ≤ 10 seconds. Because the alive rate corresponds to rate that returns 200 as http status code, it does not show that it is accessible as a SPARQL endpoint. On the other hand, because execution time shows the response time for some SPARQL query at a particular time, it does not show that the site is stable. Therefore, we used these two criteria for selecting SPARQL endpoints. Because YummyData obtains its data from SPARQL endpoint everyday, by using the filter, SPARQL Builder can dynamically obtain reliable SPARQL endpoints at that time.

Currently, although we uses YummyData as just a filter for SPARQL endpoints, we are considering the use of the data in YummyData to weight classes, properties, and SPARQL queries, to construct queries for more valuable data. Because the numbers of classes and properties are very large, an appropriate weighting method should be very useful to display them for users.

Acknowledgments. This work was supported by the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST).

References

1. Yamaguchi A., Kozaki K., Lenz K., Wu H, Kobayashi N.: An Intelligent SPARQL Query Builder for Exploration of Various Life-science Databases, CEUR Workshop Proceedings 1279, The 3rd International Workshop on Intelligent Exploration of Semantic Data (IESD 2014), Riva del Garda, Italy.
2. YummyData: http://yummydata.org/
3. Umaka REST API: http://d.umaka.dbcls.jp/api/specifications