SPARQL Endpoints and Web API (SWApi)

Pasquale Lisena[1][0000−0003−3094−5585] Albert
Meroño-Peñuela[2][2222−−3333−4444−5555]

[1] EURECOM, Sophia Antipolis, France
`pasquale.lisena@eurecom.fr`
[2] King's College London, UK
`albert.merono@kcl.ac.uk`

Abstract. The success of Semantic Web technology has boosted the publication of Knowledge Graphs in the Web of Data, and several technologies to access them have become available covering different spots in the spectrum of expressivity: from the highly expressive SPARQL to the controlled access of Linked Data APIs, with GraphQL in between. Many of these technologies have reached industry-grade maturity. Finding the trade-offs between them is often difficult in the daily work of developers, interested in quick API deployment and easy data ingestion. This tutorial covers this in-between technology space, with the main goal of providing strategies and tools for publishing Web APIs that ensure the easy consumption of data coming from SPARQL endpoints. Together with an overview of state-of-the-art technologies, the tutorial focuses on two novel technologies: SPARQL Transformer, which allows to get a more compact JSON structure for SPARQL results, decreasing the effort required by developers in interfacing JavaScript and Python applications; and grlc, an automatic way of building APIs on top of SPARQL endpoints by sharing queries on collaborative platforms. Moreover, recent developments are presented to combine the two, offering a complete resource for developers and researchers. Hands-on sessions are proposed to internalise those concepts with practical exercises.

Keywords: API · Semantic Web · Web Development

1 Introduction

A crucial factor in the adoption of the Web of Data consists in the possibility of obtaining access to the published resources. However, this access is not always simple, constrained by languages and templates that are sub-optimal for application development. As a consequence, recent ongoing initiatives such as EasierRDF and LDFlex are strongly pushing the proposal of new solutions for making the Web of Data developer-friendly [2, 5].

In order to bridge the gap between research and real-world applications, this tutorial intends to face two crucial problems in using data from the Semantic Web and Knowledge Graphs. First, going beyond the standard output of SPARQL triple-stores, we show solutions for aggregating data, parsing, and give

to the output a ready-to-use JSON structure. Second, we discuss how to detach SPARQL queries from the application logic, wrapping them in convenient Web APIs automatically.

The goal of this tutorial is to give participants sufficient knowledge for:

1. understanding the landscape of mature, production-ready and industry-grade technological solutions for publishing Web APIs on top of RDF datasets;
2. republishing RDF data through Web APIs, to increase the use and adoption of RDF datasets, even outside the Semantic Web community;
3. retrieving data from SPARQL endpoints in a more practical format to be used in small proofs-of-concept, living software or interactive notebooks (i.e. Jupyter), minimising the effort for accessing data.

2   Detailed Description

This half-day tutorial is decomposed into two parts, consisting each of slides and hands-on exercises:

– The first part focuses on data reshaping and merging. After summarising the main issues related to RDF data consumption by developers, we introduce different solutions coming from the literature, such as RDFJS [1], LDflex [11], GraphQL-based strategies [10] and others. We then explain the fundamentals of SPARQL Transformer [6, 7] and present in-depth details about its querying and templating features, its parsing capabilities, and its merging strategy, giving to the audience a complete picture of the library. Examples of usage in JavaScript and Python are shown. Some hands-on exercises are proposed in order to make the public play with the application, making use of the SPARQL Transformer playground.
– The second part covers the publication of Web APIs on top of SPARQL endpoints. Different specifications – e.g. smartAPI [4], Linked Data API specification, OpenAPI – and services – e.g. BASIL [3] – for describing RESTful APIs are introduced to the audience. Then, we show the grlc application [8] and explain the different parts of the framework: the GitHub repository, the self-generated UI and the automatically exposed API. The attendees are then guided to publish their own API in the second hands-on session. Finally, we introduce the integration of SPARQL Transformer in grlc and we conclude the tutorial with a summary of the covered topics.

3   Intended audience and level

This tutorial is directed to anyone who works with RDF data and SPARQL, and in particular is interested in providing quick access to the data and/or Web APIs on top of them. We assume that most participants are familiar with basic Semantic Web technologies (RDF, SPARQL), have some developer experience (e.g. JSON), and have some notions of HTTP and REST APIs. The tutorial aims

to provide them a general know-how and practical information for developing Web APIs on top of SPARQL endpoints.

For the hands-on session, the participants are expected to use a personal computer with a web browser installed; and to have a personal GitHub account (free).

4   Tutorial Material

The tutorial material consists of slides and pointers to online application, which is permanently available at https://api4kg.github.io/swapi-tutorial/, with the main pointers summarised in Table 1.

Table 1. List of resources used for the Tutorial

Resource	Web Address
Tutorial Website	https://api4kg.github.io/swapi-tutorial
SPARQL Transformer	https://github.com/D2KLab/sparql-transformer
Playground	https://d2klab.github.io/sparql-transformer/
grlc	https://github.com/CLARIAH/grlc
Demo	http://grlc.io/

An extended presentation of the topics introduced in the tutorial can be found in [9].

5   Biography of Presenters

Pasquale Lisena (http://pasqlisena.github.io/) is a researcher in the Data Science department at EURECOM, Sophia Antipolis (France), working on Knowledge Graphs and information extraction in the domain of digital humanities. He obtained his PhD in Computer Science from Sorbonne University of Paris in 2019, with a thesis on music representation and recommendation under the supervision of Raphaël Troncy. He was part of several national and international project, such as DOREMUS, SILKNOW and Odeuropa, actively contributing in developing domain-specific knowledge graphs and realising knowledge-driven AI technologies. His experiences include tutorials in conferences of the field (K-CAP 2017, ESWC 2018 and 2021, ISWC 2020, TheWebConf 2021). He is teaching in the Engineering program at EURECOM as assistant teaching in the Web Interaction course and as main lecturer in the Introduction to Data Bases course.

Albert Meroño Peñuela (https://www.albertmeronyo.org/) is an Assistant Professor in Computer Science at King's College London, and works on Knowledge Graph construction, access, and use in digital humanities workflows, where he has built large (>10 billion triples) Knowledge Graphs for history and musicology. Previously he has been a postdoctoral researcher at the Knowledge Representation & Reasoning group of the Vrije Universiteit Amsterdam;

the lead engineer for structured data in CLARIAH (https://clariah.nl/), a 32M
EUR Dutch project for integrating Digital Humanities data; and has served in
CLARIAH's Technical Board as coordinator of the LOD Interest Group. His
automated Linked Data API construction methods (http://grlc.io) are today
used in Elsevier, TNO, the eScience Center, and 3200 other users. He has been
teaching BSc and MSc courses at the VU since 2013 on Knowledge Representa-
tion, AI, and Digital Humanities; and has supervised the BSc/MSc theses of 33
students.

References

1. Bergwinkl, T., Luggen, M., elf Pavlik, Regalia, B., Savastano, P., Verborgh, R.:
   Interface Specification: RDF Representation, Draft Report. Tech. rep., W3C (2017)
2. Booth, D., Chute, C.G., Glaser, H., Solbrig, H.: Toward Easier RDF. In: W3C
   Workshop on Web Standardization for Graph Data. Berlin, Germany (2019)
3. Daga, E., Panziera, L., Pedrinaci, C.: A basilar approach for building web apis on
   top of sparql endpoints. In: Maleshkova, M., Verborgh, R., Stadtmüller, S. (eds.)
   Proceedings of the Third Workshop on Services and Applications over Linked APIs
   and Data. vol. 1359, pp. 22–32 (2015), http://oro.open.ac.uk/44026/, co-located
   with the 12th Extended Semantic Web Conference (ESWC 2015)
4. Dumontier, M., Dastgheib, S., Whetzel, T., Assis, P., Avillach, P., Jagodnik, K.,
   Korodi, G., Pilarczyk, M., Schürer, S., Terryn, R., Verborgh, R., Wu, C.: smar-
   tAPI: towards a more intelligent network of Web APIs. In: Proceedings of the 25th
   conference on Intelligent Systems for Molecular Biology and the 16th European
   Conference on Computational Biology (Jul 2017)
5. Gandon, F., Michel, F., Corby, O., Buffa, M., Tettamanzi, A., Faron Zucker, C.,
   Giboin, A., Cabrio, E., Villata, S.: Graph Data on the Web: extend the pivot don't
   reinvent the wheel. In: W3C Workshop on Web Standardization for Graph Data.
   Berlin, Germany (2019)
6. Lisena, P., Meroño-Peñuela, A., Kuhn, T., Troncy, R.: Easy Web API Devel-
   opment with SPARQL Transformer. In: 18th International Semantic Web Con-
   ference (ISWC), In-Use Track. pp. 454–470. Auckland, New Zealand (2019),
   https://doi.org/10.1007/978-3-030-30796-7_28
7. Lisena, P., Troncy, R.: Transforming the JSON Output of SPARQL Queries for
   Linked Data Clients. In: WWW'18 Companion: The 2018 Web Conference Com-
   panion. ACM, Lyon, France (2018), https://doi.org/10.1145/3184558.3188739
8. Meroño-Peñuela, A., Hoekstra, R.: grlc makes GitHub taste like linked data APIs.
   In: European Semantic Web Conference. pp. 342–353. Springer (2016)
9. Meroño-Peñuela, A., Lisena, P., Martínez-Ortiz, C.: Web data apis for
   knowledge graphs: Easing access to semantic data for application de-
   velopers. Synthesis Lectures on Data, Semantics, and Knowledge **12**(1),
   1–118   (2021).   https://doi.org/10.2200/S01114ED1V01Y202107DSK021,
   https://doi.org/10.2200/S01114ED1V01Y202107DSK021
10. Taelman, R., Vander Sande, M., Verborgh, R.: Bridges between GraphQL and
    RDF. In: W3C Workshop on Web Standardization for Graph Data. Berlin, Ger-
    many (2019)
11. Verborgh, R., Taelman, R.: LDflex: A Read/Write Linked Data Abstraction for
    Front-End Web Developers. In: 19th International Semantic Web Conference
    (ISWC). pp. 193–211. Springer International Publishing, Athens, Greece (2020)