When it comes to Querying Semantic Cultural Heritage Data

Béatrice Bouchou Markhoff, T .B. Nguyen, Cheikh Niang

When it comes to Querying Semantic Cultural Heritage Data

Béatrice Markhoff[1], Thanh Binh Nguyen[2⋆], and Cheikh Niang[3]

[1] Université Francois Rabelais de Tours, Laboratoire d'Informatique, France
beatrice.markhoff@univ-tours.fr
[2] Université d'Orléans, INSA CVL, LIFO, France
thanh-binh.nguyen@univ-orleans.fr
[3] Agence Universitaire de la Francophonie, Paris, France
cheikh.niang@auf.fr

Abstract. As more and more cultural institutions publish their data using the web-of-data semantic level, there is a need for novel applications for exploring, analyzing, mining and visualizing such data. A first step for these applications is to be able to query the linked open data. In this paper we survey the different existing systems for this purpose, providing examples from our experiences in the Cultural Heritage domain, and discussing some possible mutual enrichment between some of them.

Keywords: Semantic Web, Querying, SPARQL Endpoint, Federated-Query System, Ontology-Based Data Integration, Cultural Heritage

1 Introduction

When considering research papers in Digital Humanities and Cultural Heritage (CH) conferences that deal with semantic web topics, most of them report either on solutions for building ontologies and metadata for CH, or for data curation, mapping for integration, and enrichment using semantic web standards and technologies. Indeed, since the end of 2000's, many projects rely on semantic web technologies to expose CH data in the Linked Open Data (LOD [2]): a survey is given in [15], where it is shown that the commonly used process is to convert museum catalogs into open RDF triplestores, through an extract-transform-load process. Moreover, proposals described for instance in [7] rely on schema-level alignments, using existing LOD resources (DBPedia, Geonames) and reference ontologies, including the *CIDOC-CRM*. Originally designed for the semantic integration of information from museums, libraries, and archives, this later has become an extensible semantic framework that a wide range of CH resources can be mapped to [8, 19]. Based on such resources, nowadays many cultural institutions provide big Knowledge Bases (KBs) on the web, that can be used by

⋆ Supported by a PhD grant Orléans-Tours.

semantic web applications: the British Museum[4], EUROPEANA[5], the Smithsonian American Art Museum[6], the BnF[7] and so on.

There is a need now for *semantic web applications*, to help humans exploring this huge knowledge network, for performing data analysis tasks, and even for data mining, to extract new knowledge which may complete or correct the existing KBs. To this end, semantic web application designers first have to consider solutions for accessing semantic web data source(s). When analyzing research papers on querying the semantic web, lots of them were first only dealing with querying one RDF/RDFS/OWL resource, until the W3C SPARQL specifications were published (formal studies are still published recently, that deal with more expressive solutions). From the end of the 2000's, two main approaches for querying semantic web data sources can be distinguished: the first one addresses the linked-data's specificity, it is represented by the LOD Query systems, while the second one, in general motivated by users' needs, corresponds to the Ontology-Based Data Integration (OBDI) systems. LOD Query systems range from a single SPARQL Endpoint to Full-Web Query systems based on links traversal [12], and include the Federated Query systems, surveyed in [23, 20]. These latter propose a single interface to perform a query on a fixed set of SPARQL Endpoints. This is also the case for OBDI systems, which fulfill the needs of a community to access a set of sources [6, 17].

Linked Open Data sources are published on the web according to the principles introduced in [2], and now officially formalized by the W3C[8]. The LOD-initiators' vision of how applications shall use web data, called *follow-your-nose*, may be summarized as follows: data providers provide data, while data consumers discover, select and tailor data to their needs. We give a short state of the art of LOD querying approaches in Section 2. It allows us to notice that there is a need for an upper application level upon such systems, for taking into account specific user needs, and in particular the need to safely rely on a tailored view of the sources. Offering a defined view and a single access point to several sources is the purpose of data integration systems [14], and this is why we recall the principles of existing solutions based on semantic web data integration in Section 3. Building such solutions is now a well defined process when the sources' owners contribute to the integration task. But in the LOD open space, sources are not supposed to contribute to any specific integrated system, a priori. Nevertheless, we believe it is feasible to enhance semi-automatic building of mediator systems for accessing several CH data providers on the LOD, by adapting techniques used for querying the LOD, and by integrating them at the mediator level.

[4] http://collection.britishmuseum.org/
[5] http://labs.europeana.eu/api/linked-open-data-sparql-endpoint
[6] http://americanart.si.edu/collections/search/lod/about/
[7] http://data.bnf.fr/sparql/
[8] https://www.w3.org/TR/dwbp/

2 LOD Querying Systems

LOD querying approaches can be classified according to the scope of the queried data: one single source, a finite set of query-federated sources and the full web. We will present their main features following this classification.

2.1 Single SPARQL Endpoint

The cultural Knowledge Bases evoked in Introduction are central repository infrastructures, in the same way as the huge, general purpose, and multilingual KBs such as Yago or BabelNet, that are automatically constructed by harvesting public knowledge-sharing platforms [26]. It means that they collect data from different sources and integrate it into a single repository before query processing. Such central repository infrastructures are supported by Triple Store Management Systems[9], and provide in general an ontology as the schema to design queries. Notice that, in the Cultural Heritage field, in general an ontology also plays the role of global schema for integrating the collected data (see Section 3, and this is the very purpose of the CIDOC-CRM's design. The data managed by triple stores is queried in SPARQL via a SPARQL endpoint, a web service that implements the SPARQL protocol defining the communication processes as well as the accepted and output formats (e.g. RDF/XML, JSON, CSV, etc.). The major advantage of central repository infrastructures is the direct availability of locally stored data, which enables optimized query evaluation techniques. However, when harvested from independent sources, the queried data is not always up to date, which may be a serious drawback in the dynamic context of the web of data. Very few initiatives exist that tackle the challenge of allowing humans (besides applications) to explore cultural heritage Knowledge Bases, even for a single provider. To our knowledge, the best example of such solution is the ResearchSpace project[10], that uses Metaphactory, the Metaphacts end-to-end plateform[11]. The ResearchSpace project is led by the British Museum [19].

2.2 Full-Web Querying

Full-Web query systems refer to approaches where the scope of queries is the complete set of Linked Data on the Web [20]. Instead of extracting, transforming, and loading all data from a fixed set of sources before querying it, here all relevant data for a query is discovered during runtime execution. The query evaluation is initialized from a single triple pattern as starting point and, in an iterative process, relevant data is downloaded by dereferencing URIs which are used to identify Linked Data documents on the web. Parts of the query are iteratively evaluated based on downloaded data, and additional URIs are added, which are dereferenced in the next iteration step. Indeed, as an RDF

[9] https://www.w3.org/wiki/LargeTripleStores
[10] http://www.researchspace.org/
[11] http://www.metaphacts.com/application-areas/cultural-heritage

resource may be referred by multiple URIs from multiple independent sources which may use different ontologies to model their RDF knowledge bases, for instance URIs can be co-referenced via the *owl:sameAs* property, data from different sources are connected together, and as a consequence applications can potentially traverse the whole Web of Data by starting from one point. The evaluating process terminates when there are no more URI with potential results to follow. Fully relying on the Linked Data principles, the only requirement is that the needed data should correctly comply with those principles. This method potentially reaches all data on the web, and the freshest data. But, as the Web of Data is an unbounded and dynamic space, the querying evaluation may also not terminate, so practical experiments [13] actually restrict the range of queries to a finite part of the Web of Data.

2.3 Federated-Query Systems

A federated-query system refers to a unique interface for querying data from multiple independent given data sources[12], based on a federation query engine that decomposes the incoming user query into sub-queries, distributes them to data sources, and constructs the final result by combining answers from each source. The query processing in a federation framework comprises four phases performed in the following order: query parsing, data source selection, query optimization and query execution. Query parsing transforms the initial query into a set of triple patterns. Data source selection is the most studied phase, as it highly determines the overall performances of Federated Query systems. Even if no existing solution is directly based on the Full-Web query principles, *owl:sameAs* links are taken into account when determining the relevant sources containing relevant results for each triple pattern of the query, in order to avoid sending all of them to all participating sources, which is essential to avoid network overloading. Many techniques are proposed to deal with this challenge, they are categorized as index-free, index-assisted and hybrid [25, 1, 24, 22].

FedX [25] is an index-free federated engine. It sends SPARQL ASK queries to data sources at runtime to discover potential sources. Thank to the simplicity of ASK queries which return boolean values, relevant data sources who can answer parts of the query triple patterns are quickly identified, but this can become expensive when the number of triple patterns and the number of data sources grow, so a cache mechanism is used to save the relevance of each triple regarding each data sources. In contrast to the index-free fashion, index-assisted approaches as DARQ [22] rely on statistics to create indexes for all predicates and types used in the queries, concerning their presence in data sources. As using only indexed summaries and possibly out-of-date indexes does not guarantee a result set completeness, those systems must deal with indexes maintenance. Hybrid

[12] Notice that this is different from the W3C recommendation of a Federated Query extension for SPARQL 1.1, for executing queries distributed over different SPARQL endpoints by specifying the distant endpoint using the SERVICE keyword, which supposes that the query author has to manage all this low-level knowledge.

systems as ANAPSID [1] combine indexed data and ASK queries. Proposed enhancements generally consist in integrating, or dynamically querying, more knowledge about data sources. For instance, some sources are automatically discarded by making use of the URI authorities in HiBISCuS [24].

The query optimization phase aims to eliminate unnecessary data transfers between the federated-query system and the sources by (i) using caching, (ii) choosing the appropriate join method, (iii) ordering and grouping the triple patterns. The resulting execution plan is processed in the last phase, the query execution. In the Federated Query approach, queries are answered based on the up-to-date data at original sources. In the web context, this is a major advantage compared to centralized materialized approaches. This is also the case for Ontology-Based Data Integration (OBDI) systems [6, 21]. In the next section, we recall their principles and report some experiences of building and using OBDI systems in Cultural Heritage projects.

We do not know any existing web information system for Cultural Heritage based on Full-Web or Federated-Query solutions. But it is very interesting to notice that, the more Federated Query systems store information in cache or index, about their sources on the one hand, and the user queries on the other hand, the more they resemble to traditional integration systems. In particular, indexes storing the relationships between query predicates and source predicates play the same role as GAV or LAV mappings. Compared to the expressive power of OBDI systems, Federated Query systems lack the ability to compile more knowledge into the user query in order to get more complete and more correct results with respect to the user's needs. Nevertheless, the capabilities they developed for automatically harvesting knowledge, about sources and about queries, may be reused to facilitate and enhance OBDI systems building.

3 Ontology-Based Data Integration Systems

The knowledge that OBDI systems add to the user query is stored in their ontology, that is their global schema. Remember that a data integration system $\mathcal{J}$ is a triple $\mathcal{J} = \langle \mathcal{G}, \mathcal{M}, \mathcal{S} \rangle$, where [14]:

- $\mathcal{G}$ is the expected global schema.
- $\mathcal{S}$ is the source schemas, i.e. schemas of the sources where data are stored.
- $\mathcal{M}$ is the mappings between $\mathcal{G}$ and $\mathcal{S}$, i.e. a set of assertions establishing the connection between the elements of the global schema and those of the source schema.

We first briefly recall the fundamentals for defining and using a global schema, which takes here the form of a web ontology. Next, we analyze requirements for sources to become parts of a web based semantic mediation system, through the mappings, which represent a crucial part of a data integration system. Finally, we recall the principles of query architectures based on the OBDI paradigm.

3.1 Web Ontology as Global Schema

Remember that two data integration architectures exist, the *data warehouse* and the *mediation* systems. We already mentioned examples of the first type of architectures in the semantic web context: Yago and BabelNet. They rely on an ontology as the global schema. The main advantage of this architecture is the efficiency of query evaluation. Its main drawbacks are its costs in term of storage, and in terms of refreshment process, as updates performed on the original data sources must be propagated to the warehouse. On the contrary, in the mediation approach, data are kept in sources and information is retrieved dynamically from original databases, at query time. The integration is virtual in the sense that data stay in sources, but the user who interacts with the mediator, via the global schema, feels like interacting with a single database. The challenges of this solution are related to its query-answering process: when a query q_g is posed in terms of the global schema $\mathcal{G}$, the system must reformulate it in terms of a suitable set of queries q_s posed to the sources, send each computed sub-query q_{si} to the involved source $\mathcal{S}_i$, and compose the received results into a final global answer for the user. Mediation solutions fit the open, volatile and distributed web context.

Both of the data warehouse and mediation approaches require the design of a shared global schema: web ontologies play this role in the semantic web context. OBDI systems combine semantic web and data integration principles for overcoming semantic heterogeneity in order to share and efficiently reuse data among autonomous interconnected stakeholders. The following three main OBDI architectures are traditionally used [6]:

- *The single-ontology approach*, where source data is directly mapped to a global ontology. For sources having different views of the shared domain, finding a consensus in a minimal ontology commitment is known to be a difficult task.
- *The multiple-ontologies approach*, where source data is described with its own local ontology, and local ontologies are organized as a peer-to-peer system. This approach requires the construction of mappings between local ontologies and the lack of a common vocabulary between them can make this task difficult.
- *The hybrid approach*, which combines the two previous ones, using local ontologies that are mapped to a common top-level vocabulary, alleviating thereby the definition of inter-ontology mappings.

The global ontology building is a difficult task that usually requires human experts, but there already exist many resources and techniques that can facilitate it.The ResearchSpace [19] and the ARIADNE project for COINS presented in [10] are good examples of *single-ontology* OBDI systems that chose the data warehouse approach: data is mapped and transformed from the source schemata to RDF triples, compliant with the CIDOC-CRM schema. In [18], a *single-ontology* OBDI mediator is presented, where the global ontology was devised by experts, based on the CIDOC-CRM with extensions and thesaurus of the

conservation-restoration domain. A solution for automating the global ontology building in *an hybrid OBDI system* was used in the query architecture presented in [3], that was designed for a project of prosopography in the Renaissance. The European project EP-Net also well demonstrated the usefulness of Ontology-Based Data Access and OBDI for easing the access of scholars to historical and cultural data, distributed across different data sources, including public semantic resources on the web [5].

3.2 Semantic Web Source

In single-ontology and in hybrid OBDI systems, in order to provide the unified global query-interface, the mediator relies on mappings $\mathcal{M}$, between the global schema $\mathcal{G}$ and the local schemas $\mathcal{S}$. Mappings are used for rewriting the global query into a union of queries that match the local schemas. These mappings can be directed either from entities in the global schema to entities in the local sources - Global As View (GAV) mappings, or from entities in the local sources to the ones in the global schema - Local As View (LAV) mappings [14]. LAV mappings require more sophisticated inferences to resolve a query on the global schema than GAV mappings, but they make it easier to add or retrieve data sources to the mediation system. Some hybrid solutions exist, for instance in [3] the overall architecture uses both LAV mappings between the sources and the domain reference ontology and GAV mappings between the global ontology and the sources.

To participate in a web-ontology-based mediation system, a source must anchor the data that it wants to share into the global system. To this end, when the source is not yet a semantic web resource, a required step is to build an ontological representation of its data, at least for the part that should contribute to the integration process. There are several ways to expose data at the semantic level, either by using suitable tools such as Ontop [4], X3ML [16], or Karma [11], or by implementing ad hoc wrappers. For instance in [18], a solution based on Ontop is implemented for the first source (a relational DB), while an ad hoc wrapper is used to export in RDF the second source (a set of MS-Word files). Once the source can be queried in SPARQL, mappings $\mathcal{M}$ can be defined and used to implement the distributed query system.

3.3 Mediator Querying System

We focus here on a concrete implemented example, but keeping the presentation sufficiently formal to be reusable. In [3] each submitted query is reformulated based on the knowledge contained in the global ontology, and on the integration knowledge (i.e. knowledge about sources, including the mappings). In a nutshell, given the global ontology $\mathcal{O}_g$ and a set of source repositories $\mathcal{S}$, this rewriting processes a global query q_g, expressed over $\mathcal{O}_g$, by reformulating it into a union of sources queries $\mathcal{Q}_s$, after having compiled the suitable knowledge in $\mathcal{O}_g$ into q_g. More precisely, the query q_g, intended to extract a set of elements from the distributed semantic databases, is resolved following these steps:

1. Apply the *Consistent* algorithm [21] with q_g as canonical instance, for verifying its consistency w.r.t constraints expressed in $\mathcal{O}_g$.
2. Apply the *PerfectRef* algorithm [21] with q and $\mathcal{O}_g$ as input, for integrating the ontological constraints into the initial query.
3. Let $\mathcal{Q}_s$ be the union of queries resulting from Step 2, apply an adapted *MiniCon* processing in order to distribute $\mathcal{Q}_s$ to the involved sources. Each source evaluates the received queries over its semantic repository.
4. Let $ans(Q_s)$ be the set of answers received from sources, build the global answer $ans(Q_g)$.

The first step avoids evaluating queries that could only lead to an empty result. The second step is a reasoning task performed over the global schema $\mathcal{O}_g$. In [3], as positive assertions are used for expressing the GAV mappings, each obtained sub-query q_s in $\mathcal{Q}_s$ is expressed using terms of a specific source, allowing the *MiniCon* to distribute them and compute the global answer afterwards. It is interesting to notice that, compared to Federated Query systems presented in Section 2.3, the difficult challenge they face for data source selection does not exist with OBDI, because the necessary knowledge is stored in the mediator.

4 Conclusion

When it comes to querying semantic web Cultural Heritage data, this is useful to have an overview of semantic web querying systems in order to build tailored services for users. We surveyed the existing solutions for querying the semantic web, those for LOD querying on the one hand, and those based on OBDI on the other hand. It is important to notice that, in the Cultural Heritage field, existing semantic data management systems are OBDI systems, in general based on the CIDOC-CRM or some extensions, and that most of them are centralized materialized RDF data triple stores, that are queried via a SPARQL endpoint. Nevertheless, in order to go closer to The Dream of a Global Knowledge Network for Cultural Heritage [9], it is necessary to be able to also rely on the *decentralized* Linked Data that is *distributed* over the WWW.

To this end, we recalled the alternative to centralized materialized data warehouses, that traditionally exists among data integration systems: the mediators, which allow the user for querying several legacy data sources without extracting and loading data. Before that, we sketched the principles of existing Full-Web and Federated-Query systems, noticing that they have the general tendency to propose some automatic generation of knowledge, about sources and about queries, this knowledge being stored in cache, or index, or summary, or ranking table as in [13]. According to the so-called *follow-your-nose* query principle [2], these systems are not devised to offer an integrated view of several resources, on contrary they explicitly leave the data integrating effort to their users, but some of their methods for harvesting knowledge, and for optimizing query execution plans, may be re-used to improve the automation of web OBDI mediators building, and also to improve their query performances. Big KBs that exist on the semantic web, such as Yago or BabelNet, are harvested from public web

knowledge-sharing platforms, which involves information extraction techniques to produce RDF data, sometimes from natural language texts. Their creators also devised methods and techniques [26] that could be re-used for projects dedicated to Cultural Heritage.

Now that many Cultural Heritage institutions have opened their data to the semantic web level, we are convinced that building OBDI systems is the best way to develop applications for connecting Cultural Heritage data, for many different user needs. This is already done in many projects [19, 10, 18, 5]. Moreover, even though we know that centralized materialized data warehouses are the most efficient solution for operating complex computations on big data, we also believe that semantic mediators are the best solutions in order to deal with the decentralized and highly evolutive features of the web. Our survey highlights some elements for making their construction simpler and their operation more efficient, even for querying existing public semantic web resources.

References

1. M. Acosta, M.-E. Vidal, T. Lampo, J. Castillo, and E. Ruckhaus. Anapsid: An adaptive query processing engine for sparql endpoints. In *Proceedings of the 10th International Conference on The Semantic Web - Volume Part I*, ISWC'11, pages 18–34, Berlin, Heidelberg, 2011. Springer-Verlag.
2. C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.
3. B. Bouchou and C. Niang. Semantic mediator querying. In *International Database Engineering and Applications Symposium (IDEAS)*, pages 29–38. ACM, 2014.
4. D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, and G. Xiao. Ontop: Answering sparql queries over relational databases. *Semantic Web*, (Preprint):1–17, 2016.
5. D. Calvanese, P. Liuzzo, A. Mosca, J. Remesal, M. Rezk, and G. Rull. Ontology-based data integration in epnet: Production and distribution of food during the roman empire. *Engineering Applications of Artificial Intelligence*, 51:212 – 229, 2016. Mining the Humanities: Technologies and Applications.
6. I. F. Cruz and H. Xiao. Ontology driven data integration in heterogeneous networks. In *Complex Systems in Knowledge-based Environments*, pages 75–98. 2009.
7. M. Damova and D. Dannells. *Third International Conference on Software, Services and Semantic Technologies S3T 2011*, chapter Reason-Able View of Linked Data for Cultural Heritage, pages 17–24. Springer Berlin Heidelberg, 2011.
8. M. Doerr. Ontologies for cultural heritage. In *Handbook on Ontologies*, pages 463–486. Springer, 2009.
9. M. Doerr and D. Iorizzo. The Dream of a Global Knowledge Network; A New Approach. *J. Comput. Cult. Herit.*, 1(1):5:1–5:23, June 2008.
10. A. Felicetti, P. Gerth, C. Meghini, and M. Theodoridou. Integrating Heterogeneous Coin Datasets in the Context of Archaeological Research. In *Proceedings of the Workshop on Extending, Mapping and Focusing the CRM co-located with 19th International Conference on Theory and Practice of Digital Libraries (2015), Poznań, Poland, September 17, 2015.*, pages 13–27, 2015.
11. S. Gupta, P. Szekely, C. A. Knoblock, A. Goel, M. Taheriyan, and M. Muslea. Karma: A system for mapping structured sources into the semantic web. In *The Semantic Web: ESWC 2012 Satellite Events*, pages 430–434. Springer, 2012.

12. O. Hartig. Querying a web of linked data: foundations and query execution. *SIGWEB Newsletter*, 2014(Autumn):3:1–3:2, 2014.
13. O. Hartig and M. T. Özsu. Walking Without a Map: Ranking-Based Traversal for Querying Linked Data. In *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part I*, pages 305–324, 2016.
14. M. Lenzerini. Data integration: A theoretical perspective. In *PODS*, pages 233–246, 2002.
15. J. Marden, C. Li-Madeo, N. Whysel, and J. Edelstein. Linked Open Data for Cultural Heritage: Evolution of an Information Technology. In *Proceedings of the 31st ACM International Conference on Design of Communication*, SIGDOC '13, pages 107–112. ACM, 2013.
16. N. Minadakis, Y. Marketakis, H. Kondylakis, G. Flouris, M. Theodoridou, G. de Jong, and M. Doerr. X3ML Framework: An Effective Suite for Supporting Data Mappings. In *Proc. of the Workshop on Extending, Mapping and Focusing the CRM co-located with 19th International Conference on Theory and Practice of Digital Libraries (2015), Poznań, Poland, September 17, 2015.*, pages 1–12, 2015.
17. C. Niang, C. Marinica, E. Leboucher, L. Bouiller, C. Capderou, and B. Bouchou. Ontology-based data integration system for conservation-restoration data (obdis-cr). In *20th International Database Engineering & Applications Symposium, IDEAS 2016, Montreal, Canada, July 11-13, 2016*, pages 218–223, 2016.
18. C. Niang, C. Marinica, B. B. Markhoff, E. Leboucher, F. Laissus, O. Malavergne, L. Bouiller, C. Darrieumerlou, and C. Capderou. Supporting semantic interoperability in conservation-restoration domain: the parcours project. *ACM Journal on Computing and Cultural Heritage (JOCCH), Special Issue on Digital Infrastructure for Cultural Heritage*, To appear.
19. D. Oldman, M. Doerr, G. de Jong, B. Norton, and T. Wikman. Realizing lessons of the last 20 years: A manifesto for data provisioning & aggregation services for the digital humanities (A position paper). *D-Lib Magazine*, 20(7/8), 2014.
20. M. T. Özsu. A survey of rdf data management systems. *Front. Comput. Sci.*, 10(3):418–432, June 2016.
21. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. In *Journal on data semantics*, pages 133–173. Springer, 2008.
22. B. Quilitz and U. Leser. Querying distributed rdf data sources with sparql. In *Proceedings of the 5th European Semantic Web Conference on The Semantic Web: Research and Applications*, ESWC'08, pages 524–538, Berlin, Heidelberg, 2008. Springer-Verlag.
23. M. Saleem, Y. Khan, A. Hasnain, I. Ermilov, and A.-C. Ngonga Ngomo. A fine-grained evaluation of sparql endpoint federation systems. *Semantic Web*, (Preprint):1–26, 2015.
24. M. Saleem and A.-C. Ngonga Ngomo. *HiBISCuS: Hypergraph-Based Source Selection for SPARQL Endpoint Federation*, pages 176–191. Springer International Publishing, Cham, 2014.
25. A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt. Fedx: Optimization techniques for federated query processing on linked data. In *The Semantic Web - ISWC 2011 - 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I*, pages 601–616, 2011.
26. G. Weikum, J. Hoffart, and F. M. Suchanek. Ten years of knowledge harvesting: Lessons and challenges. *IEEE Data Eng. Bull.*, 39(3):41–50, 2016.