Interlinking Big Data to Web of Data

Enayat Rajabi and Seyed-Mehdi-Reza Beheshti

Abstract The big data problem can be seen as a massive number of data islands, ranging from personal, shared, social to business data. The data in these islands is getting large scale, never ending, and ever changing, arriving in batches at irregular time intervals. Examples of these are social and business data. Linking and analyzing of this potentially connected data is of high and valuable interest. In this context, it will be important to investigate how the Linked Data approach can enable the Big Data optimization. In particular, the Linked Data approach has recently facilitated the accessibility, sharing, and enrichment of data on the Web. Scientists believe that Linked Data reduces Big Data variability by some of the scientifically less interesting dimensions. In particular, by applying the Linked Data techniques for exposing structured data and eventually interlinking them to useful knowledge on the Web, many syntactic issues vanish. Generally speaking, this approach improves data optimization by providing some solutions for intelligent and automatic linking among datasets. In this chapter, we aim to discuss the advantages of applying the Linked Data approach, towards the optimization of Big Data in the Linked Open Data (LOD) cloud by: (i) describing the impact of linking Big Data to LOD cloud; (ii) representing various interlinking tools for linking Big Data; and (iii) providing a practical case study: linking a very large dataset to DBpedia.

Keywords Big data · Linked open data · Interlinking optimization

E. Rajabi (✉)
Computer Science Department, Dalhousie University, Halifax, NS, Canada
e-mail: rajabi@dal.ca

Seyed-Mehdi-RezaBeheshti
University of New South Wales, Sydney, Australia
e-mail: sbeheshti@cse.unsw.edu.au

A. Emrouznejad (ed.), *Big Data Optimization: Recent Developments and Challenges*, Studies in Big Data 18, DOI 10.1007/978-3-319-30265-2_6

1 Introduction

The big data problem can be seen as a massive number of data islands, ranging from personal, shared, social to business data. The data in these islands are increasingly becoming large-scale, never-ending, and ever changing; they may also arrive in batches at irregular time intervals. Examples of these are social (the streams of 3,000−6,000 tweets per second in Twitter) and business data. The adoption of social media, the digitalisation of business artefacts (e.g. files, documents, reports, and receipts), using sensors (to measure and track everything), and more importantly generating huge metadata (e.g. versioning, provenance, security, and privacy), for imbuing the business data with additional semantics, generate part of this big data. Wide physical distribution, diversity of formats, non-standard data models, independently-managed and heterogeneous semantics are characteristics of this big data. Linking and analysing of this potentially connected data is of high and valuable interest. In this context, it will be important to investigate how the Linked Data approach can enable the Big Data optimization.

In recent years, the Linked Data approach [1] has facilitated the availability of different kinds of information on the Web and in some senses; it has been part of the Big Data [2]. The view that data objects are linked and shared is very much in line with the goals of Big Data and it is fair to mention that Linked Data could be an ideal pilot place in Big Data research. Linked Data reduces Big Data variability by some of the scientifically less interesting dimensions. Connecting and exploring data using RDF [3], a general way to describe structured information in Linked Data, may lead to creation of new information, which in turn may enable data publishers to formulate better solutions and identify new opportunities. Moreover, the Linked Data approach applies vocabularies which are created using a few formally well-defined languages (e.g., OWL [4]). From searching and accessibility perspective, a lot of compatible free and open source tools and systems have been developed on the Linked Data context to facilitate the loading, querying and interlinking of open data islands. These techniques can be largely applied in the context of Big Data.

In this context, optimization approaches to interlinking Big Data to the Web of Data can play a critical role in scaling and understanding the potentially connected resources scattered over the Web. For example, Open Government establishes a modern cooperation among politicians, public administration, industry and private citizens by enabling more transparency, democracy, participation and collaboration. Using and optimizing the links between Open Government Data (OGD) and useful knowledge on the Web, OGD stakeholders can contribute to provide collections of enriched data. For instance, US government data[1] including around 111,154 datasets, at the time of writing this book, that was launched on May 2009 having

[1] http://data.gov/.

only 76 datasets from 11 government agencies. This dataset, as a US government Web portal provides the public with access to federal government-created datasets and increases efficiency among government agencies. Most US government agencies already work on the codified information dissemination requirements, and 'data.gov' being conceived as a tool to aid their mission delivery. Another notable example, in the context of e-learning, provides linking of educational resources from different repositories to other datasets on the Web.

Optimizing approaches to interconnecting e-Learning resources may enable sharing, navigation and reusing of learning objects. As a motivating scenario, consider a researcher who might explore the contents of a big data repository in order to find a specific resource. In one of the resources, a video on the subject of his interests may catch the researcher's attention and thus follows the provided description, which has been provided in another language. Assuming that the resources in the repository have been previously interlinked with knowledge bases such as DBpedia,[2] the user will be enabled to find more information on the topic including different translations.

Obviously, the core of data accessibility throughout the Web can provide the links between items, as this idea is prominent in literature on Linked Data principles [1]. Indeed, establishing links between objects in a big dataset is based on the assumption that the Web is migrating from a model of isolated data repositories to a Web of interlinked data. One advantage of data connectivity in a big dataset [5] is the possibility of connecting a resource to valuable collections on the Web. In this chapter, we discuss how optimization approaches to interlinking Web of data to Big Data can enrich a Big Dataset. After a brief discussing on different interlinking tools in the Linked Data context, we explain how an interlinking process can be applied for linking a dataset to Web of Data. Later, we experiments an interlinking approach over a sample of Big Dataset in eLearning literature and conclude the chapter by reporting on the results.

2 Interlinking Tools

There exist several approaches for interlinking data in the context of LD. Simperl et al. [6] provided a comparison of interlinking tools based upon some criteria such as use cases, annotation, input and output. Likewise, we explain some of the related tools, by focusing on their need to human contribution (to what extent users have to contribute in interlinking), their automation (to what extent the tool needs human input), and the area (in which environment the tool can be applied).

From a human contribution perspective, User Contributed Interlinking (UCI) [7] creates different types of semantic links such as *owl:sameas* and *rdf:seeAlso*

[2] http://dbpedia.org.

between two datasets relying on user contributions. In this Wiki-style approach, users can add, view or delete links between data items in a dataset by making use of a UCI interface. Games With A Purpose (GWAP) [8] is another software which provides incentives for users to interlink datasets using game and pictures by distinguishing different pictures with the same name. Linkage Query Writer (LinQuer) [9] is also another tool for semantic link discovery [10] between different datasets which allows users to write their queries in an interface using some APIs.

Automatic Interlinking (AI) is another linking approach for interconnecting of data sources applied for identifying semantic links between data sources. Semi-automatic interlinking [11], as an example, is a kind of analyzing technique to assign multimedia data to users using multimedia metadata. Interlinking multimedia (iM) [11] is also a pragmatic way in this context for applying the LD to fragments of multimedia items and presents methods for enabling a widespread use of interlinking multimedia. RDF-IA [12] is another linking tool that carries out matching and fusion of RDF datasets according to the user configuration, and generates several outputs including *owl:sameAs* statements between the data items.

Another semi-automatic approach for interlinking is the Silk Link Discovery Framework [13], which finds the similarities within different LD sources by specifying the types of RDF links via SPARQL endpoints or data dumps. LIMES [14] is also a link discovery software in the LOD that presents a tool in command-line and GUI for finding similarities between two datasets and suggests the results to users based on the metrics automatically. LODRefine [15] is another tool for cleaning, transforming, and interlinking any kinds of data with a web user interface. It has the benefit of reconciling data to the LOD datasets (e.g., Freebase or DBpedia) [15]. The following table briefly summarizes the described tools and mentions the area of application for each one (Table 1).

Table 1 Existing interlinking tools description

Tool	Area
UCI	General data source
GWAP	Web pages, e-commerce offerings, Flickr images, and YouTube
LinQuer	LOD datasets
IM	Multimedia
RDF-IA	LOD datasets
Silk	LOD datasets
LIMES	LOD datasets
LODRefine	General data, LOD datasets

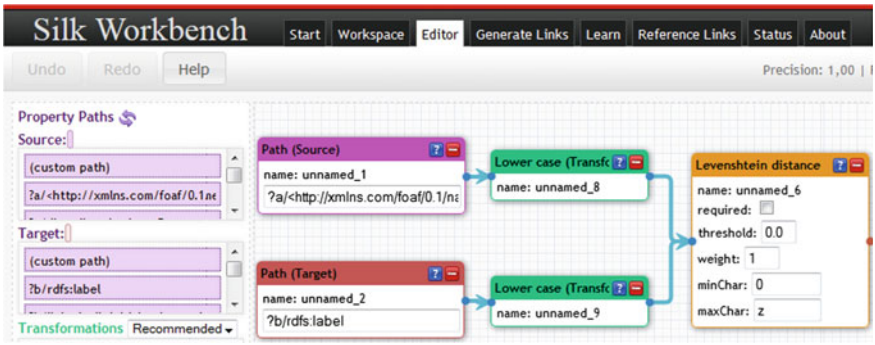

Fig. 1 Silk work-bench interface

To discuss the most used tools in Linked Data context we have selected three software and explain their characteristics and the way that they interlink datasets.

2.1 Silk

Silk [13] is an interlinking software that matches two datasets using string matching techniques. It applies some similarity metrics to discover similarities between two concepts. By specifying two datasets as input (SPARQL endpoints or RDF dumps), Silk provides as an output e.g., "sameAs" triples between the matched entities. Silk Workbench, is the web application variant of Silk which allows users to interlink datasets through the process of interlinking different data sources by offering a graphical editor to create link specifications (consider Fig. 1). After performing the interlinking process, the user can evaluate the generated links. A number of projects including DataLift [16] have employed the Silk engine to carry out their interlinking purposes.

2.2 LIMES

Link Discovery Framework for Metric Spaces (LIMES) is another interlinking tool which presents a linking approach for discovering relationships between entities contained in Linked Data sources [14]. LIMES leverages several mathematical characteristics of metric spaces to compute pessimistic approximations of the similarity of instances. It processes the strings by making use of suffix-, prefix- and position filtering in a string mapper by specifying a source dataset, a target dataset,

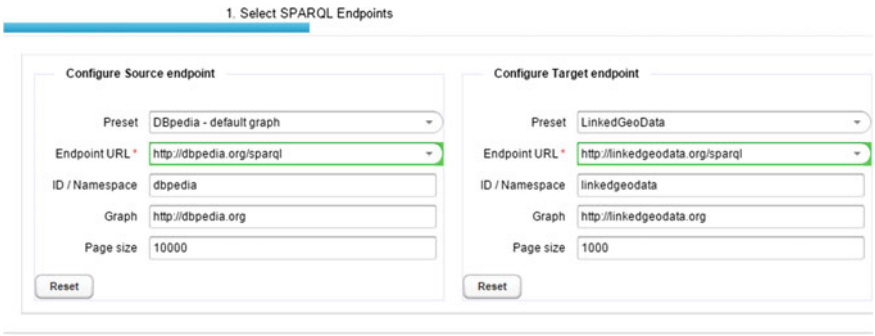

Fig. 2 LIMES web interface

and a link specification. LIMES applies either a SPARQL Endpoint or a RDF dump from both targets. A user can also set a threshold for various matching metrics by which two instances are considered as matched, when the similarity between the terms exceeds the defined value. A recent study [14] evaluated LIMES as a time-efficient approach, particularly when it is applied to link large data collections. Figure 2 depicts the web interface of LIMES (called SAIM[3]) was recently provided by AKSW group.[4]

2.3 LODRefine

LODRefine [15] is another tool in this area that allows data to be loaded, refined, and reconciled. It also provides additional functionalities for dealing with the Linked Open Data cloud. This software discovers similarities between datasets by linking the data items to the target datasets. LODRefine matches similar concepts automatically and suggests the results to users for review. Users also can expand their contents with concepts from the LOD datasets (e.g., DBpedia) once the data has been reconciled. They can also specify the condition for the interlinking. Eventually, LODRefine reports the interlinking results and provides several functionalities for filtering the results. LODRefine also allows users to refine and manage data before starting the interlinking process, which is very useful when the user dataset includes several messy content (e.g., null, unrelated contents) and facilitates the process by reducing the number of source concepts. Figure 3 depicts a snapshot of this tool.

[3] http://saim.aksw.org/.

[4] http://www.aksw.org.

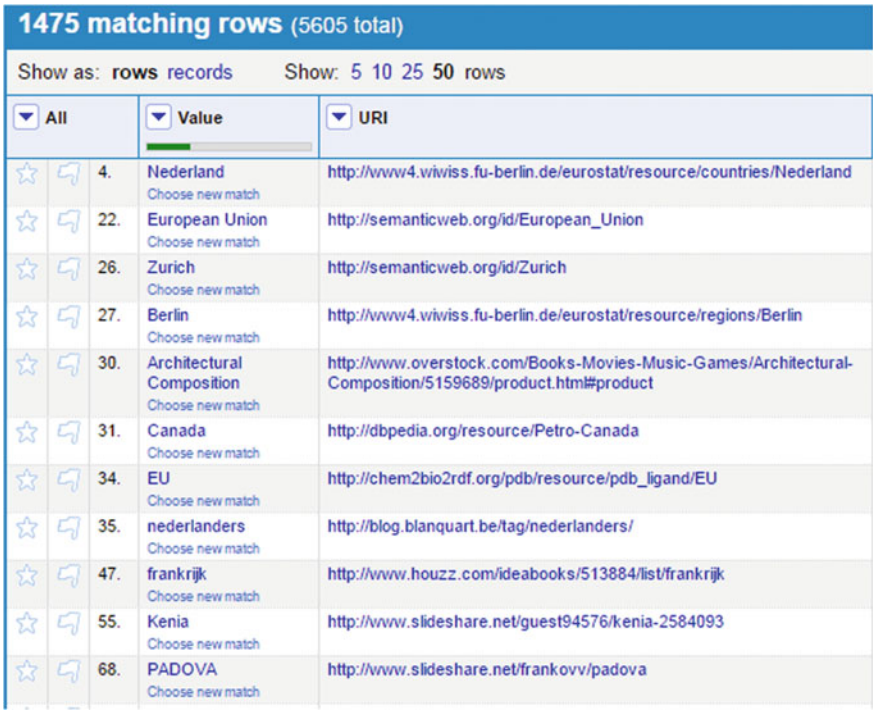

Fig. 3 LODRefine interface

3 Interlinking Process

In an ideal scenario, a data island can be linked to a diverse collection of sources on the Web of Data. However, connecting each entity, available in the data island, to an appropriate source is very time-consuming. Particularly when we face a big number of data items, the domain expert needs to explore the target dataset in order to be able to apply queries. As mentioned earlier and to minimize the human contribution, interlinking tools have facilitated the interlinking process by implementing a number of matching techniques. While using an interlinking tool, several issues such as defining the configuration for the linking process, specifying the criteria, and post-processing the output need to be addressed. In particular, the user sets a configuration file in order to specify the criteria under which items are linked in the datasets. Eventually, the tool generates links between concepts under the specified criteria and provides output in order to be reviewed and verified by users. Once the linking process has finished, the user can evaluate the accuracy of the generated links that are close to the similarity threshold. Specifically, the user can verify or reject each link recommended by the tool as the two matching concepts (see Fig. 4).

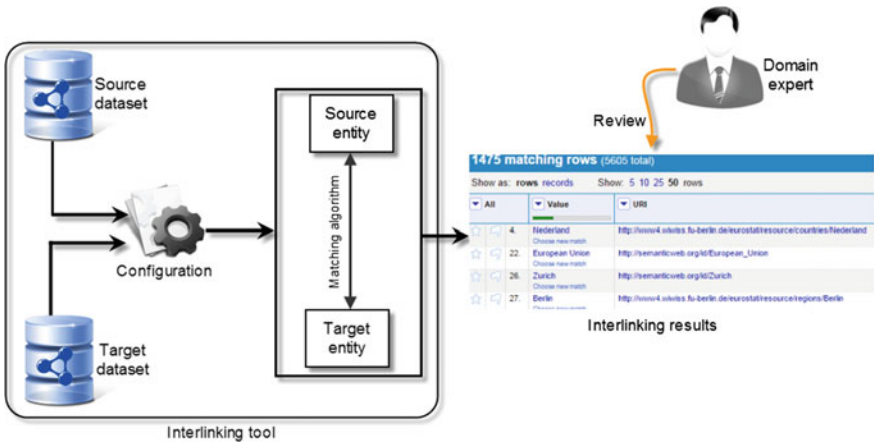

Fig. 4 The interlinking process

4 A Case Study for Interlinking

There exist a wide variety of data sources on the Web of Data that can be considered as part of the Big Data. With respect to authors' experiences on eLearning context and given that around 1,362 datasets have been registered in datahub[5] and tagged as "eLearning datasets", we selected the GLOBE repository,[6] a large dataset with almost 1.2 million learning resources and more than 10 million concepts [5]. The GLOBE is a federated repository that consists of several other repositories, such as OER Commons [17] which includes manually created metadata as well as aggregated metadata from different sources, we selected GLOBE for our case study to assess the possibility of interlinking. The metadata of learning resources in GLOBE are based upon the IEEE LOM schema [18] which is a de facto standard for describing learning objects on the Web. Title, keywords, taxonomies, language, and description of a learning resource are some of the metadata elements in an IEEE LOM schema which includes more than 50 elements. Current research on the use of GLOBE learning resource metadata [19] shows that 20 metadata elements are used consistently in the repository.

To analyze the GLOBE resource metadata, we collected more than 800,000 metadata files via OAI-PMH[7] protocol from the GLOBE repository. Some GLOBE metadata could not be harvested due to validation errors (e.g., LOM extension errors). Particularly, several repositories in GLOBE extended the IEEE LOM by adding new elements without using namespaces, which caused a number of errors

[5] http://datahub.io/.

[6] http://globe-info.org/.

[7] Open Archives Initiative Protocol for Metadata Harvesting." [Online]. Available: http://www.openarchives.org/pmh. [Accessed: 22-February-2014].

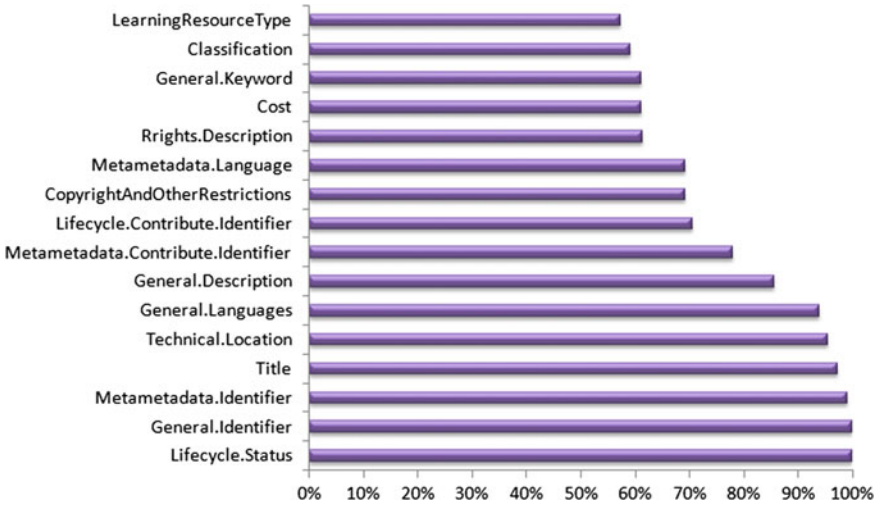

Fig. 5 The usage of metadata elements by GLOBE resources

detected by the ARIADNE validation service.[8] Later, we converted the harvested XML files into a relational database using a JAVA program in order to examine those elements that are more useful for the interlinking purpose. Figure 5 illustrates the metadata elements those used by more than 50 % of learning resources in GLOBE of which title of learning resource, as an example, has been applied by more than 97 % of the GLOBE resources. More than half (around 55 %) of the resources were in English and 99 % of the learning objects were open and free to use. English is the most prominent language in GLOBE [5] and thus the linking elements used as a source in our data scope were limited to English terms of the selected elements, which were represented in more than one language.

Several metadata elements such as *General.Identifier* or *Technical.Location* are mostly included local values provided by each repository (e.g., "ed091288" or "http://www.maa.org/") and thus could not be considered for interlinking. Additionally, constant values (e.g., dates and times) or controlled vocabularies (e.g., "*Contribute.Role"* and "*Lifecycle.Status*") were not suitable for interlinking, as the user could not obtain useful information by linking these elements. Finally, the following metadata elements were selected for the case study, as they were identified as the most appropriate elements for interlinking [20]:

- Title a learning resource (*General.Title*)
- The taxonomy given to a learning resource *(Classification.Taxon)*
- A Keyword or phrase describing the topic of learning objects *(General. Keyword)*.

[8] http://ariadne.cs.kuleuven.be/validationService/.

As the GLOBE resources were not available as RDF, we exposed the GLOBE metadata via a SPARQL endpoint.[9] We exposed the harvested metadata, which were converted into a relational database, as RDF using a mapping service (e.g., D2RQ [10]) and set up a SPARQL Endpoint in order to complete the interlinking process. As a result, the GLOBE data was accessible through a local SPARQL endpoint in order to be interlinked to a target dataset. There were 434,112 resources with title, 306,949 resources with Keyword, and 176,439 resources with taxon element all in English language.

To find an appropriate target in the Web of Data, we studies a set of datasets in datahub of which we selected DBpedia,[11] one of the most used datasets [21] and Linked Data version of Wikipedia that makes it possible to link data items to general information on the Web. In particular, the advantage of linking of contents to DBpedia is to make public information usable for other datasets and to enrich datasets by linking to valuable resources on the Web of Data.The full DBpedia dataset features labels and abstracts for 10.3 million unique topics in 111 different languages[12] about persons, places, and organizations. All DBpedia contents have been classified into 900,000 English concepts, and are provided according to SKOS [13], as a common data model for linking knowledge organization systems on the Web of Data. Hence, this dataset was selected for linking keywords and taxonomies of metadata.

When running an interlinking tool like LIMES, the user sets a configuration file in order to specify the criteria under which items are linked in the two datasets. The tool generates links between items under the specified criteria and provides output which defines whether there was a match or a similar term in order to be verified by users. Once the linking process has finished, the user can evaluate the accuracy of the generated links that are close to the similarity threshold. Specifically, the user can verify or reject each record recommended by the tool as two matching concepts. Eventually, we ran LIMES over three elements of GLOBE (title, Keyword, and Taxon) and DBpedia subjects. Table 2 illustrates the interlinking results in which more than 217,000 GLOBE resources linked to 10,676 DBpedia subjects through keywords. In respect to Taxonomy interlinking, around 132,000 resources in GLOBE were connected to 1,203 resources of the DBpedia dataset, while only 443 GLOBE resources matched to 118 DBpedia resources. The low number of matched links for the title element refers to this fact that interlinking long strings does not lead many matched resources, as most of the GLOBE metadata contained titles with more than two or three words.

The following table (Table 3) illustrates some sample results show those GLOBE resources connected to the DBpedia subjects (two results per element). Having the results and reviewing the matched links by data providers, GLOBE can be enriched with new information so that each resource is connected to DBpedia using e.g., *owl:sameAs* relationship.

[9] http://www.w3.org/wiki/SparqlEndpoints.

[10] http://d2rq.org/

[11] http://dbpedia.org.

[12] http://blog.dbpedia.org.

[13] https://www.w3.org/2004/02/skos/

Table 2 Interlinking results between GLOBE and DBpedia

Element	Globe resources#	DBpedia resources#	Total links
Title	443	118	443
Keyword	217,026	10,676	623,390
Taxon	132,693	1,203	268,302

Table 3 Sample interlinking results

Phrase	Element	DBpedia resources URI
Bibliography	Title	http://dbpedia.org/resource/Category:Bibliography
Analysis	Title	http://dbpedia.org/resource/Category:Analysis
Plutoniu	Keyword	http://dbpedia.org/resource/Category:Plutonium
Biology	Keyword	http://dbpedia.org/resource/Category:Biology
Transportation	Taxon	http://dbpedia.org/resource/Category:Transportation
Trigonometry	Taxon	http://dbpedia.org/resource/Category:Trigonometry

5 Conclusions and Future Directions

In this chapter we explained the interlinking approach as a way of optimizing and enriching different kinds of data. We have described the impact of linking Big Data to the LOD cloud. Afterward, we explained various interlinking tools used in Linked Data for interconnecting datasets, along with a discussion about the interlinking process and how a dataset can be interlinked to Web of Data. Finally, we have represented a case study where a interlinking tools (LIMES) used for linking the GLOBE repository to DBpedia. Running the tool and examining the results, many GLOBE resources could connect to DBpedia and after an optimization and enrichment step the new information can be added to the source datasets. This process makes the dataset more valuable and the dataset' users can get more knowledge about the learning resources. The enrichment process over one of large datasets in eLearning context have been presented and it was shown that this process can be extend to other types of data: the process does not depend to a specific context. The quality of a dataset is also optimized when it is connected to other related information on the Web. The previous study on our selected interlinking tool (LIMES) [14] is also showed that it is a promising software when it is applied to a large amount of data.

In conclusion, we believe that enabling the optimization of Big Data and the open data is an important research area, which will attract a lot of attention in the research community. It is important as the explosion of unstructured data has created an information challenge for many organizations. Significant research directions in this area includes: (i) Enhancing linked data approaches with semantic information gathered from a wide variety of sources. Prominent examples include the Google Knowledge Graph [22] and the IBM Watson question answering system [23]; (ii) Integration of existing machine learning and natural language processing

algorithms into Big Data platforms [24]; and (iii) High-level declarative approaches to assist users in interlinking Big data to open data. A good example of this can be something similar to OpenRefine [25] which can be specialized for the optimization and enrichment of interlinking big data to different types of open source data; e.g. social data such as Twitter. Summarization approaches such as [26] can be also used to interlinking big data to different sources.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data—The Story So Far. Int. J. Semantic Web Inf. Syst. (IJSWIS) . **5**(3) 1–22, 33 (2009)
2. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think, Reprint edn. Eamon Dolan/Houghton Mifflin Harcourt (2013)
3. Klyne, G., Carroll, J.J.: Resource description framework (RDF): concepts and abstract syntax. W3C Recommendation (2004)
4. OWL 2 Web Ontology Language Document Overview, 2nd edn. http://www.w3.org/TR/owl2-overview/. Accessed 19 May 2013
5. Rajabi, E., Sicilia, M.-A., Sanchez-Alonso, S.: Interlinking educational data: an experiment with GLOBE resources. In: Presented at the First International Conference on Technological Ecosystem for Enhancing Multiculturality, Salamanca, Spain (2013)
6. Simperl, E., Wölger, S., Thaler, S., Norton, B., Bürger, T.: Combining human and computation intelligence: the case of data interlinking tools. Int. J. Metadata Semant. Ontol. **7**(2), 77–92 (2012)
7. Hausenblas, M., Halb, W., Raimond, Y.: Scripting user contributed interlinking. In: Proceedings of the 4th workshop on Scripting for the Semantic Web (SFSW2008), co-located with ESWC2008 (2008)
8. Siorpaes, K., Hepp, M.: Games with a purpose for the semantic web. EEE Intell. Syst **23**(3), 50–60 (2008)
9. Hassanzadeh, O., Xin, R., Miller, J., Kementsietsidis, A., Lim, L., Wang, M.: Linkage query writer. In: Proceedings of the VLDB Endowment (2009)
10. Beheshti, S.M.R., Moshkenani, M.S.: Development of grid resource discovery service based on semantic information. In: Proceedings of the 2007 Spring Simulation Multiconference, San Diego, CA, USA, vol. 1, pp. 141–148 (2007)
11. Bürger, T., Hausenblas, M.: Interlinking Multimedia—Principles and Requirements
12. Scharffe, F., Liu, Y., Zhou, C.: RDF-AI: an architecture for RDF datasets matching, fusion and interlink. In: Proceedings of IJCAI 2009 IR-KR Workshop (2009)
13. Volz, J., Bizer, C., Berlin, F.U., Gaedke, M., Kobilarov, G.: Silk—A Link Discovery Framework for the Web of Data. In Proceedings of the 2nd Linked Data on the Web Workshop (LDOW2009), Madrid, Spain, 2009.
14. Ngonga, A., Sören, A.: LIMES—a time-efficient approach for large-scale link discovery on the web of data. In: Presented at the IJCAI (2011)
15. Verlic, M.: LODGrefine—LOD-enabled Google refine in action. In: Presented at the I-SEMANTICS (Posters & Demos) (2012)
16. The Datalift project: A catalyser for the Web of data. http://datalift.org. Accessed 24 Nov 2013
17. OER Commons. http://www.oercommons.org/. Accessed 17 Jun 2013
18. IEEE P1484.12.4™/D1, Draft recommended practice for expressing IEEE learning object metadata instances using the dublin core abstract model. http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce?action=AttachFile&do=get&target=LOM-DCAM-newdraft.pdf. Accessed 12 May 2013

19. Ochoa, X., Klerkx, J., Vandeputte, B., Duval, E.: On the use of learning object metadata: the GLOBE experience. In: Proceedings of the 6th European Conference on Technology Enhanced Learning: towards Ubiquitous Learning, pp. 271–284. Berlin, Heidelberg (2011)
20. Rajabi, E., Sicilia, M.-A., Sanchez-Alonso, S.: Interlinking educational resources to web of data through IEEE LOM. Comput. Sci. Inf. Syst. J. **12**(1) (2014)
21. Rajabi, E., Sanchez-Alonso, S., Sicilia, M.-A.: Analyzing broken links on the web of data: an experiment with DBpedia. J. Am. Soc. Inf. Scechnol. JASIST 2014;65(8):1721
22. Official Google Blog: Introducing the knowledge graph: things, not strings https://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html
23. Ferrucci, D.A.: Introduction to this is Watson. IBM J. Res. Dev. **56**(3.4), 1:1–1:15 (2012)
24. Beheshti, S.-M.-R., Venugopal, S., Ryu, S.H., Benatallah, B., Wang, W.: Big data and cross-document coreference resolution: current state and future opportunities (2013). arXiv:13113987
25. Open-refine—Google Refine, a powerful tool for working with messy data. http://openrefine.org/. Accessed 14 Jul 2013
26. Beheshti, S.-M.-R., Benatallah, B., Motahari-Nezhad, H.: Scalable graph-based OLAP analytics over process execution data. Distributed and Parallel Databases, 1–45 (2015). doi:10.1007/s10619-014-7171-9

Author Biographies

Dr. Enayat Rajabi is a Post-Doctoral Fellow at the NICHE Research Group at Dalhousie University, Halifax, NS, Canada. He got his PhD in Knowledge Engineering at the Computer Science department of University of Alcalá (Spain) under the supervision of Dr. Salvador Sanchez-Alonso and Prof. Miguel-Angel Sicilia. The subject of his PhD was "Interlinking educational data to Web of Data", in which he applied the Semantic Web technologies to interlink several eLearning datasets to the LOD cloud.

Dr. Beheshti is a Lecturer and Senior Research Associate in the Service Oriented Computing Group, School of Computer Science and Engineering (CSE), University of New South Wales (UNSW), Australia. He did his PhD and Postdoc under supervision of Prof. Boualem Benatallah in UNSW Australia.