Modelling Scientific Publication Landscape with Linked Data Reactor

John Can Lokman and Ali Khalili

VU University Amsterdam

Abstract. This report documents the design and prototype of an application that is developed for the Knowledge and Data course of 2017 at Vrije Universiteit Amsterdam (VU), and also as part of the project 'Knowledge Flows in Interdisciplinary Research' of VU Network Institute. This application combines various data sources into linked data using an ontology, serializes this ontology and its instances as a triple store, and allows users to query it with a user-friendly interface. In sections 1 and 2, a description of the application and its intended users were described and its initial design is summarized. In section 3 the ontology developed for the application is detailed; in section 4, inferencing is discussed; and in the last sections, the first prototype is introduced.

Date: 27.10.2017
Keywords: semantic web, linked data, knowledge representation, SPARQL, Stardog, Protege, Python, RDF, Turtle

1 Introduction: A Description of the Application and Users

1.1

1.2 Context: Linked Data

Although the break-neck speed of innovation in science and technology make these very exciting times, we are also facing serious bottlenecks when it comes to communicating scientific and technical knowledge. Due to their inefficient nature, the current scientific communication tools— lectures, scientific papers, and data— make research and education agonizingly slow processes [20]. The heart of the matter seems to be a general stagnancy in communicating implicit knowledge to others, or a lack of participation in the effort of making the implicit explicit, so that it can be processed faster by humans and machines alike. After all, searching for an answer to a question through querying a knowledge graph with natural language is likely to be much more efficient than using the modest text-recognition and reasoning faculties of humans, and for a machine, it would be much more efficient to reason over that same graph than scanning the texts of billions of documents with an attempt to extract knowledge from them. Indeed, this bottleneck in knowledge flow is widely recognized, and solutions such as Semantic Web and linked data has been proposed and acknowledged as the next step in World Wide Web's evolution [4]. These are relatively old propositions, however, and their adoption still seems limited after more than 20 years. Awe-inspiring initiatives such as DBpedia and adoption by giant data creators like Twitter aside, the semantic web technology remains inaccessible (or simply undesirable) to average Web citizen / developer / designer today. The Web is full of dead SPARQL endpoints, and technologies such as REST APIs is being preferred over triple stores.

One of the major reasons for slow adoption rate of semantic web technologies could be their inaccessibility. Like with any new technology, semantic web is in a technical-state and generally lacks user-friendly interfaces. For instance, the industry standard ontology editor, Protege is disliked by many but used anyway due to lack of a better alternative; and tools for semantic web are spread out throughout the web. Besides the shortcomings of existing standard tools even in the most basic aspects (e.g., both Protege and Stardog missing RDF validators, and Protege 5 not even giving an error message upon encountering a .turtle file with a faulty line), as of today, there is no integrated development environment for semantic web IDE—something that could streamline the messy linked data workflow from data to ontology, to triple store, and to query interface. Therefore, development of tools that are better designed from both a software development and user-experience perspective (e.g., Neo4j [19]) are urgently needed to overcome linked data and semantic web initiatives' own bottleneck: lack of good tools. Linked Data Reactor (LD-R), a project at Vrije Universiteit Amsterdam, is such a candidate, and it could significantly lower the threshold for using and producing linked data. The current project aims to utilize the LD-R framework and demonstrate this point on usability with a prototype, while also using the framework to create an analysis platform for a scientific problem: knowledge flows and interdisciplinarity in research.

1.3 Domain: Interdisciplinarity

Interdisciplinarity in research is generally seen as desirable, and it is likely to be an important factor that can bring about new perspectives and solutions to our increasingly sophisticated and multi-faceted research pursuits today. However, the impact of interdisciplinarity —or to put simply, the effect of diversity of research in an article, journal, or institute— on the scientific quality and merit is a matter of debate, and there does not seem to be conclusive findings. Some authors suggest that 'distance' between disciplines may play a critical role in the effectiveness of interdisciplinarity [12,31], and some claim a 'U-shaped' relationship [29], while the discussion also includes various other theories and findings [30,2]. The ongoing debate and possible impact of results on policy making invites more studies in this direction, and due to the reliance of common research questions in this field on bibliometric data, it provides a suitable domain to investigate for the current project.

1.4 Goal

The current project aims to create a linked-data interface that allows users to query a database of scientific publications in order to get detailed metadata concerning publication patterns . Through the interface, it is particularly aimed to allow a data-driven investigation of interdisciplinary collaboration patterns. Some research questions, alongside other possible exploratory ones that may arise in future, are as following:

- Are there any collaboration patterns that are biased towards a certain disciplines? For instance, when medical researchers and computer scientists collaborate on research projects, do they tend to publish their research on journals that belong to one of their respective disciplines (i.e., a medical journal versus a computer science journal)?
- Does interdisciplinarity of a research project affect its impact (e.g., as measured with number of outgoing citations)?
- Are there fields of research that seem to yield higher-impact results when they collaborate?
- Does being an interdisciplinary researcher lead to more publications?
- Are there any publication patterns that can explain researcher career trajectories. For instance, do variables such as interdisciplinarity of an author's lifetime research, number of overall collaborations with other researchers (i.e., network size), and other similar variables affect career-related variables of researchers, such as influence (e.g., as measured by number of citation) or tenure attainment?

This project is part of the 10-month research project 'Knowledge Flows in Interdisciplinary Research' [27], and as the investigation progresses over the next few months, the current linked-data interface is expected to shed light to these research questions, and also motivate new ones.

Methodology

In order to explore the effects of interdisciplinarity and discover scientific collaboration patterns, we aim to model the domain of of scientific publications using an ontology, and populate it using a bibliometric databases such as Elsevier's Pure [9], RISIS [23], and Web Of Science [6]. After being parsed and integrated with an ontology, this data will be made available in a triple store, which will be connected to a linked data query engine and & interface, Linked Data Reactor (LD-R) [13]. For the purposes of the current project, and bearing in mind time limitations, the Pure database of VU was implemented in the prototype detailed in this report.

1.5 Users

The intended initial user base for the linked data interface are the researchers involved in the Knowledge Flows in Interdisciplinary Research project: Dr. Ali

Khalili, Dr. Sascha Friesike, Prof. Peter van den Besselaar; and Academy Assistants Frederik König and John Can Lokman. As this research trajectory continues in the following years, more researchers and students who are involved in proceeding —or similar— projects in both Vrije Universiteit Amsterdam and other universities can be expected to join the user base.

2 Design

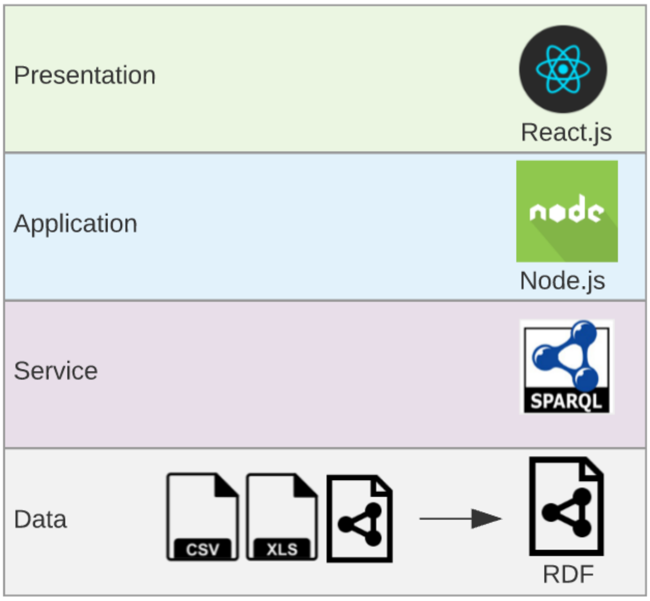

Fig. 1. The technologies used for the project, distributed to four layers of application design.

2.1 Data Layer

In the background, the application will be based on an ontology of scientific collaboration, which will model the landscape of scientific publications (i.e., scientific journals) and collaborations between authors and scientific fields. The ontology will be populated with instances by incorporating data from multiple sources such as RISIS [23] and VU Research Portal of Vrije Universiteit Amsterdam—a service that is powered by Pure [9], Scopus [10], and Elsevier

Fingerprint Engine [8]. More data sources that contain meta-scientific information could later be added as the project progresses. If such additional data sources do not come in RDF format, necessary transformations will be done using appropriate scripts or methods. (For a summary of application design, see Figure 1.)

2.2 Service Layer

The ontology and its instances will be serialized as a triple store using Stardog and will be hosted locally (and at later stages of the overarching project, online) as linked open data.

2.3 Application Layer

The application layer and SPARQL wrapper is implemented using Javascript (in particular Node.js and React.js), as part of Linked Data Reactor Framework [13].

2.4 Presentation Layer

Presentation layer is implemented using mainly HTML and React.js as part of Linked Data Reactor Framework [13]. As the current application is primarily aimed for research purposes, in future, it will evolve towards being a highly accessible linked data browser designed with 'What you see is what you query' principle [14] (see Figure 3). A short walkthrough of the presentation layer as the first concepts of the prototype's design, and also as a demonstration of future directions is provided below.

Introduction and Step 1: Datasets Users are greeted with an introduction page explaining the purpose of the application, and then will proceed to a page that shows them the datasets that will be used for their queries in the next page. This page will be visually similar to the one in Figure 2, and the users will likely be given the option to include or exclude databases with checkboxes.

Step 2: Linked Data Browser After selecting datasets to query, users proceed to a linked data browser that is similar to Figure 3. In this interface, users are allowed to explore the data in a visual and accessible way.

Although the time constraints on the current course will likely will not allow incorporation of additional interfaces, a few possible ideas that may be realized after the course are summarized below.

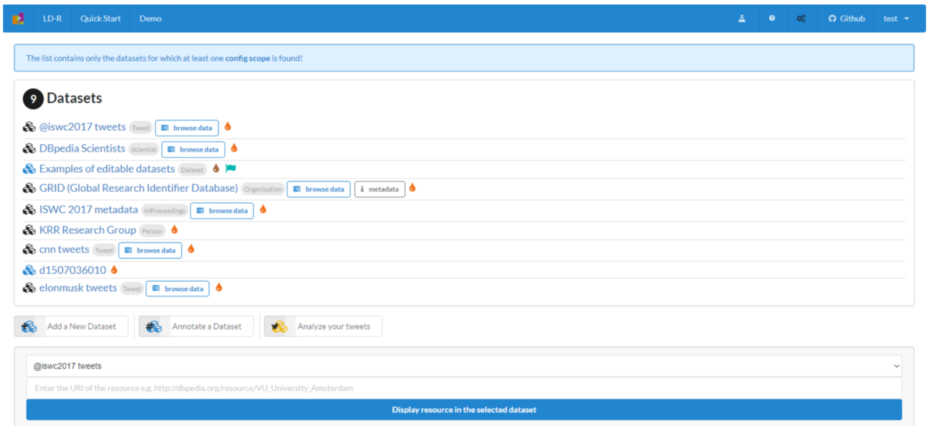

Fig. 2. An example page that lists available datasets. Screenshot taken from Linked Data Reactor [13]

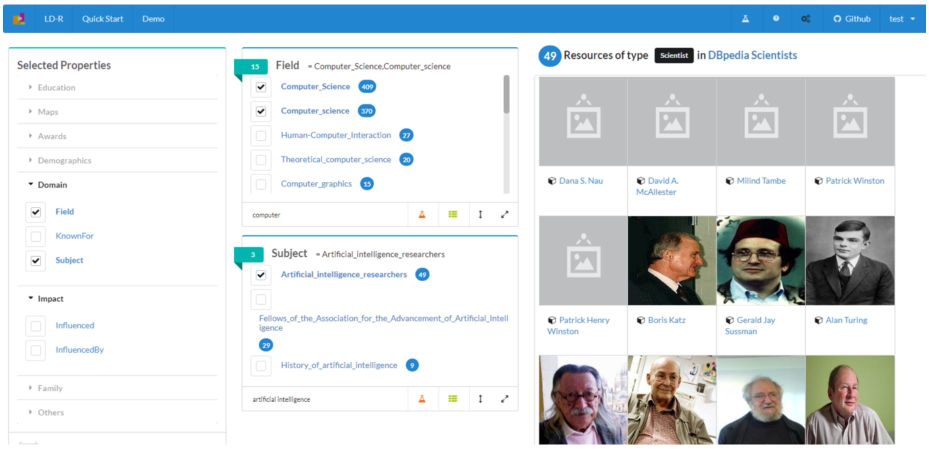

Fig. 3. A screen that resembles the implemented linked data browser for the current project. Screenshot taken from Linked Data Reactor [13].

Possible Future Interface/Component - Search Engine In an alternative interface to that of Linked Data Reactor, a Google-like search engine built to search Meta-science could be used to greet users (see Figure 4). This search engine could allow searching for multiple research fields, researchers, and research projects. Alternatively, such a search box could also be added to the regular Linked Data Reactor browser in Figure 3.

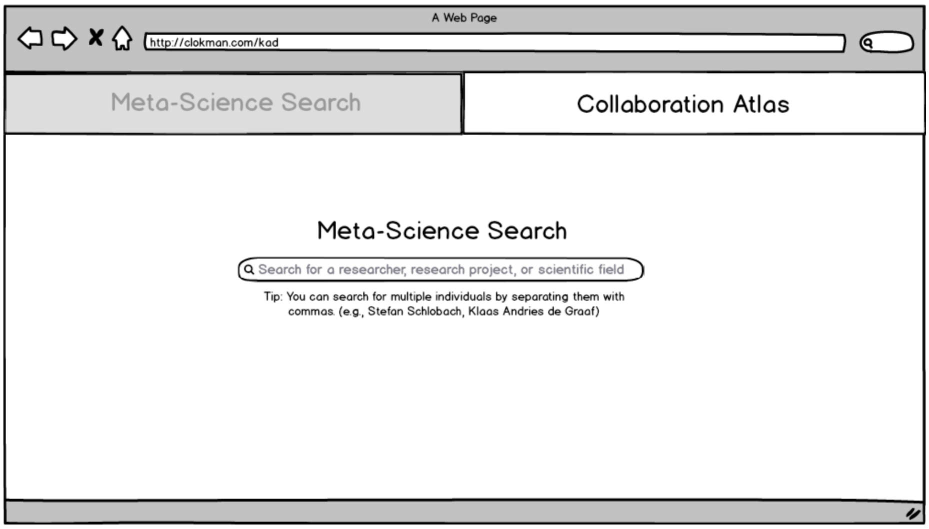

Fig. 4. A concept search engine to query a meta-scientific ontology and the instances it contains.

Possible Future Interface/Component - Linked-data-driven Visualizations The results of the search query could be returned as various interactive visualizations depending on the type of search queries entered (see Figures 5, 6, and 7). Such visualizations could be built using libraries such as D3 (for graphics) and React.js (for other interface elements and operations, such as removing keywords from search query on-the-go—i.e., without requiring users to go back to the initial search page).

3 Domain Modeling

3.1 Domain

In order to explore the effects of interdisciplinarity and discover scientific collaboration patterns, a domain model was created, which consists of properties and resources that are related to scientific publications, such as type of publication (e.g., journal, book), authors, topics, scientific fields, and so on. Since scientific collaborations of today's academia generally result in some sort of publication, and there are ample records regarding to these publications, they are highly likely to be good probes to measure of research and collaboration patterns. Therefore, modelling the landscape of scientific publications via an ontology, and populating it with instances from various datasets is viewed as a valid approach.

The instances for the ontology is currently drawn from a bibliometric database, Elsevier's Pure [9], and more instances will be later added from RISIS [23], and

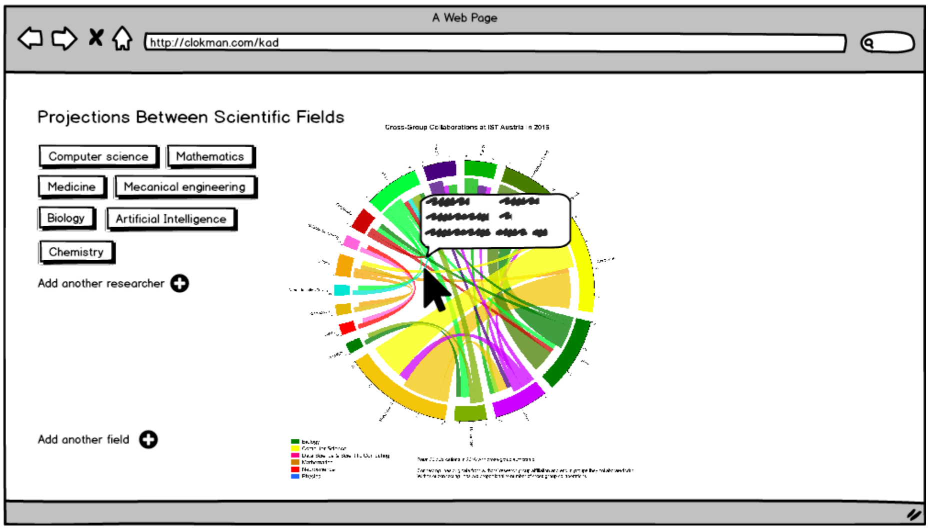

Fig. 5. A concept interactive visualization that could be generated as a response to a search query that relates to scientific fields or topics.

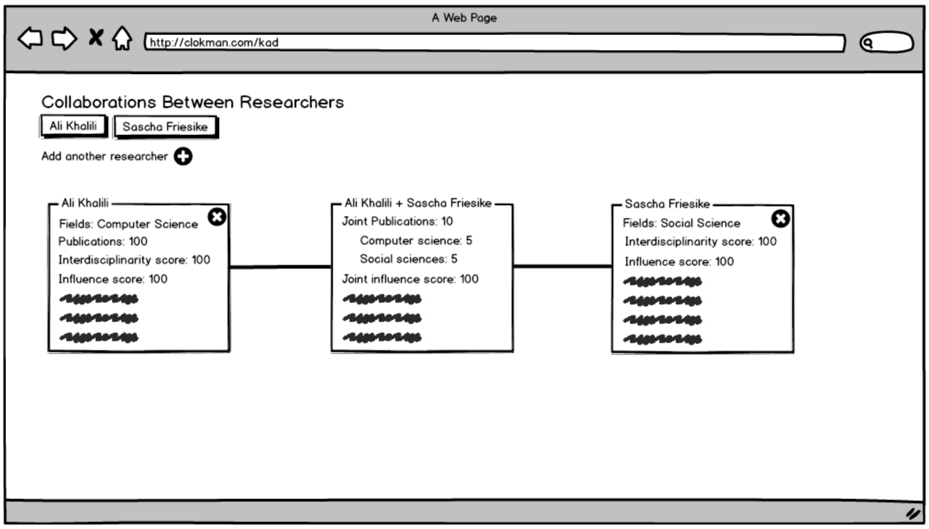

Fig. 6. An interactive visualization that shows collaboration between two researchers as well as other related information for each researcher individually.

Web Of Science [6]. In interdiciplinarity research, like other meta-scientific research topics, this bibliometric approach is an often used and considered an

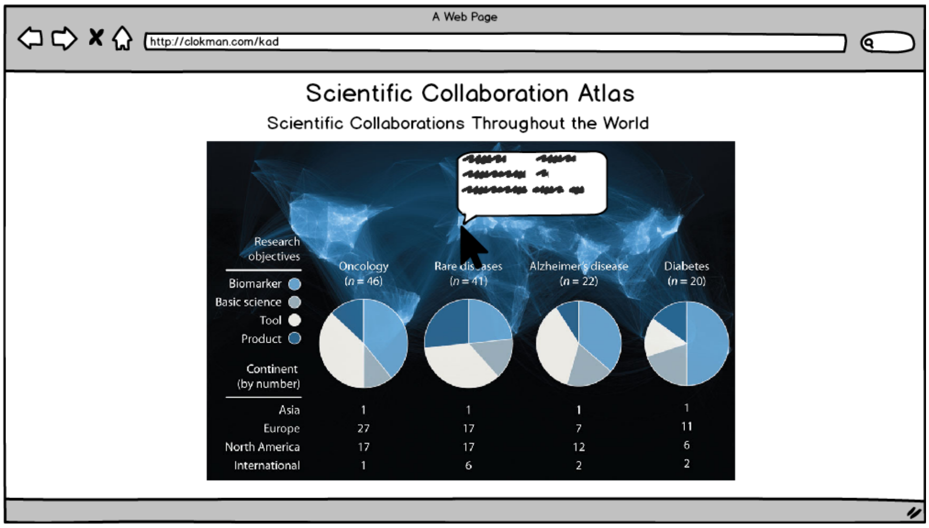

Fig. 7. A concept interactive visualization that maps collaborations based on author locations and research topic. Image taken from [15], and background visualization from [3].

effective method [24,18,32,21,5]. The methodology used in the current project is detailed further in the proceeding section.

3.2 Methodology and Ontology Building Process

Ontology Creation and Revisions In order to create a model of the domain, a first prototype of an ontology was developed during the past month as part of the course, and this model was revised and improved through project meetings. As the ontology progressed —and I gained more experience— methodological changes occurred and the ontology was significantly changed between revisions. Most notably, the range and class restrictions that were often applied with 'rdfs:range' and 'rdfs:domain' properties were entirely removed due to the reasoning errors and inflexibility they lead to, and also as per expert recommendations (now, more delicate range and domain restrictions are applied through equivalency and subclass relationships where needed). The latest version of the ontology features more sophisticated class definitions (see Fig. 8) through equivalency statements and this results in a more stable ontology and more reliable inferences. Besides the technical advancements, the structure of the ontology was updated based on project meetings. Therefore, the current version consists of a more comprehensive, accurate, and stable model of the domain.

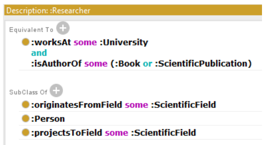

Fig. 8. A class definition that resembles natural language, made by using equivalency and subclass assertions.

Populating the Ontology with Instances In order to populate the ontology with individuals, a bibliography database was initially obtained from VU's *Pure* service [9]. Due to database being in .bib format, and due to lack of sufficiently high quality .bib to RDF conversion packages in Python, a parser and .ttl converter was programmed using Python (and 'pybtex' package). Although the data preparation part of the ontology creation process had cost most of the available time in this way, the time investment was seen as necessary due to *Pure* dataset being an important element for the project. These scripts can be reached at '`https://github.com/clokman/KAD/tree/master/Prototype_Final`'. The reader is encouraged to view the Python scripts as they would likely be able to demonstrate a good degree of understanding of the concepts of the course and may be relevant for evaluation (related files will also be attached together with the submission of this report).

It should also be noted that due to practical concerns (i.e., lack of time and computing power), a truncated version of the actual Pure database was used for importing instances, and the number of instances imported to the ontology was limited to roughly 375 instances (124 KB in size). The original *Pure* bibliography contains more than 1.5 million lines (about 100 MB), and it will be used later in the project when more time and computational resources are available.

Each instance in the .bib file consisted of bibliographic information regarding to a document, and had varying elements (e.g., while some publication instances had a lot of detail, some only had the basic information such as author, title and year). Therefore, the parser script was programmed to be able to deal with missing variables on a regular basis.

Adding External Classes Although a few example scientific domains (e.g., computer science) were used as placeholders during previous assignments, a comprehensive and accurate domain map of scientific fields were necessary for the purposes of the overarching project (although less necessary for the current stage of the project and the course). Therefore, as a convenient and trustworthy way of categorizing fields of science, ' Web Of Science Category Terms' [26] was added to the ontology as classes through parsing and conversion to .ttl. (This Python script is named 'd_web-of-science-categories.py' in the GitHub repository and is also among submission files.)

3.3 Conceptualization and Realization

The current version of the ontology includes 28 classes for describing the domain of scientific publications and academic research output in general:

- *Publications*: 9 classes including journal articles, conference proceedings, and books (fig. 10).
- *Persons*: 6 classes for authors, researchers, editors, and collaborators (fig. 11).
- *Institutions*: 8 classes for organizations such as universities and publishers 12
- *Scientific fields*: A superclass with over a hundred subclasses for scientific fields imported from Web of Science, this will likely be the main way of defining scientific areas of researchers, journals, and institutions in the future iterations of the app (see fig. 13).

Through using properties (fig. 14), a conceptual network between these classes has been established. 39 object properties describe:

- Citations and publications
- Work and collaboration associations

A visual summary and explanations of the structure of the ontology can be seen in figures 9 through 14.

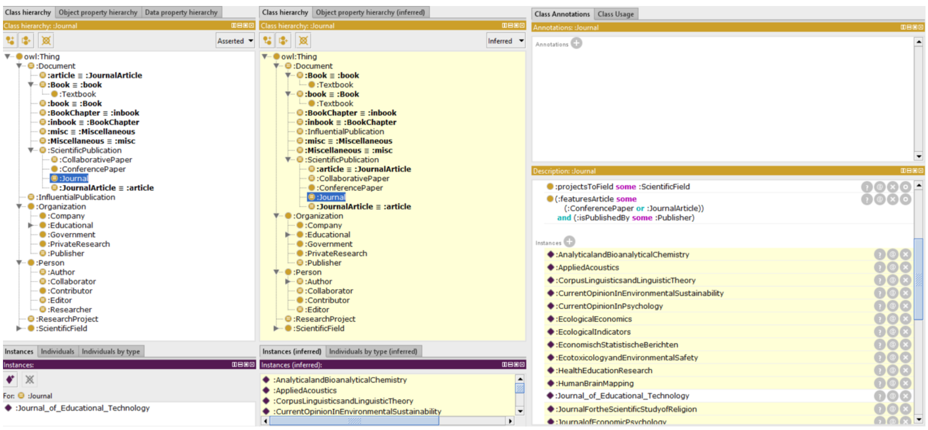

Fig. 9. A screenshot from Protege for a more 'formal' view of the ontology compared to the visualizations in other figures.

Future Enhancements for the Ontology Although instances and classes were successfully imported to the ontology, the external files worked with were

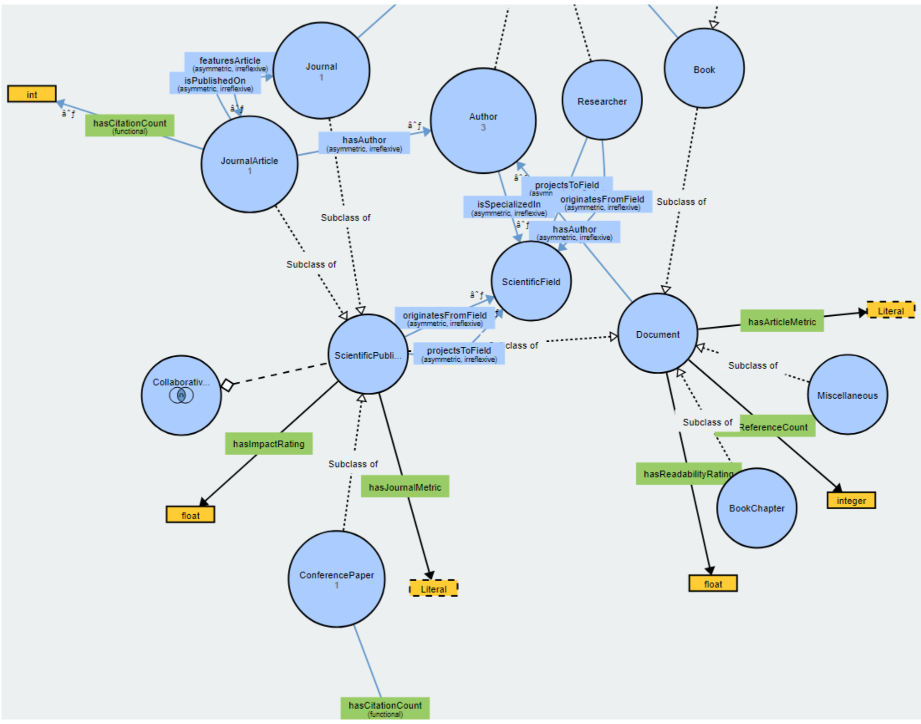

Fig. 10. A model of documents and scientific publications in the ontology. Visualizations are produced with VOWL [16].

not in RDF or similar format, and therefore, entities from them were given their own namespaces. Future versions of the ontology could be more connected to existing ontologies (in the current version of ontology, there are only one or a couple of such external links from existing ontologies [besides basic elements like rdf:, rdfs: and owl:] , such as 'foaf:knows').

Furthermore, although the ontology contains a detailed branch for scientific fields, the instances are not yet mapped to these fields (though instances are linked to *topics*) . Mapping topics (i.e., publications) to fields of science is likely to be a month or two-long project on its own, and in order to facilitate insights about interdisciplinary research, a very necessary step in the immediate future of the project.

3.4 Inferencing

The ontology used plenty of class restrictions and has been able to make meaningful inferences on the imported VU-Pure data. For instance, fig. 15 shows imported articles serving as a clue for inferring which journal they belong to. In the earlier versions of the ontology, there were, in fact, more inferences being

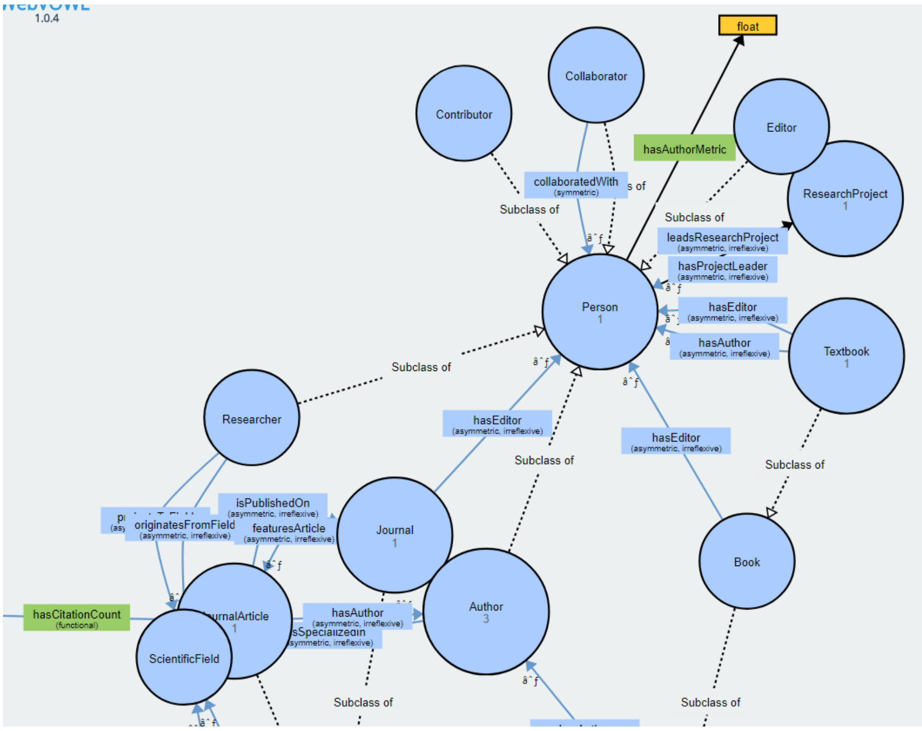

Fig. 11. Part of the ontology that is showing person-related classes and properties.

made due to somewhat more liberal class definitions being in use (fig. 16, also see fig. 14). As the external data tuned better and better to the ontology with the development of the Python scripts used to prepare them, the bibliographical information imported also became more detailed over time, and general inferences (e.g., "all things that has an author is a document") were replaced by more precise assertions that came with the imported file (e.g., "this_instance has type article"). In future, more experimentation could be made to increase the number of inferences, although well-prepared and pre-aligned files (e.g., by adding class equivalency information in Python during data parsing, rather than doing this in Protege) may, once again reduce the need for inferencing for crucial information in future work as well. And unfortunately, as mentioned before the prototype does not yet use the *scientific field* class (i.e., it is not yet specified what article belongs to which scientific field) or *organization* class (i.e., institution) in inferences . These (crucial) features will be implemented after completion of the course as part of the research project, and inference will likely play an important role in inferring scientific fields from the topics (i.e., keywords) of articles.

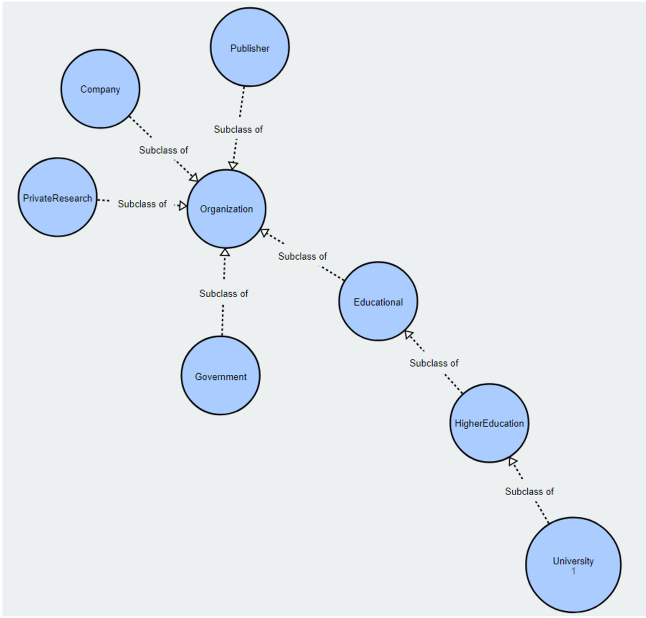

Fig. 12. Organization (i.e., institution) part of the ontology.

4 Data (Re)use and Queries

4.1 Data Sources

In its current state, the application uses the following four data sources, for the associated reasons:

1. **The ontology of the domain of scientific publications** (namespace 'sr') that is developed during this course (which forms the 'internal' SPARQL endpoint for Milestone 3): Created in order to model scientific collaboration and publication, this ontology is at the heart of the project. Currently, it is flexible and responsive to conceptual changes that may occur in early phases of research, and in future, it will serve as a platform on which data sources from various places can be integrated on, and enable their communication with each other.
2. **VU-Pure database** (namespace 'pvu') for populating the ontology with instances (previously integrated with the ontology): Pure's holds detailed records of researchers of Vrije Universiteit Amsterdam, and because has been more accessible due to practical reasons (i.e., already being in our possession), it was seen as a good starting point before other datasets are added.
3. **Web Of Science categories** (namespace 'wsc') for additional classes (also previously integrated with the ontology): Because a good model of scientific

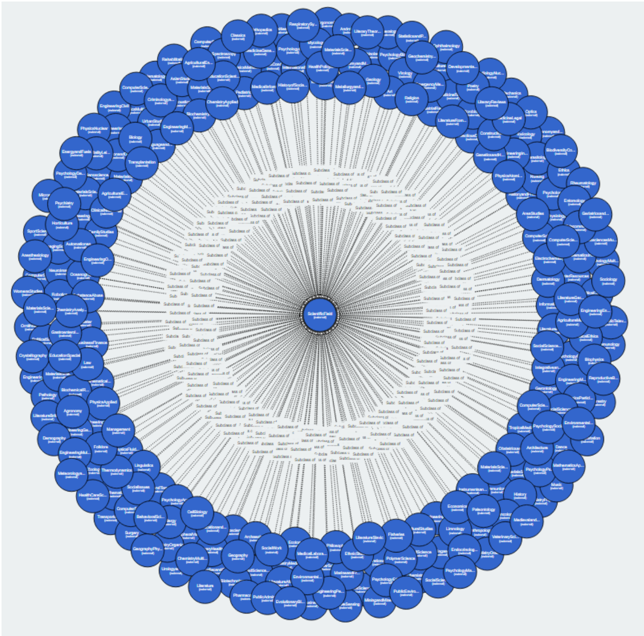

Fig. 13. The superclass 'ScientificField' in the middle, and scientific fields as subclasses on the sides. Subclasses parsed, transformed to Turtle format, and imported to the ontology from Web of Science Category Terms [26].

fields is essential for a study that aims to understand collaboration patterns between disciplines, we decided to integrate into our ontology Web of Science Category Terms [26], a popular and established way of categorizing scientific fields. As mentioned, these categories are made part of the combined ontology, but not yet mapped to scientific publications.

4. **A database of universities of the world** (separate ontology/triple store) , which is being queried from dbpedia's SPARQL endpoint. Initially a demo database that comes from Linked Data Reactor (LD-R) Framework, due to its relevance to the current project, this database is kept on the server (and modified in order to switch from the then non-functional 'live.dbpedia.org' to 'dbpedia.org' domain). This database will likely to be integrated with the current ontology in future, as it would be efficient and beneficial to be able to name universities in the world without having to build a new ontology.

4.2 Producing the Data through Parsing, Querying, and Inferencing

The Pure and Web of Science databases were integrated with the current ontology through using a self-made Python parser, and then joined together afterwards with the support of inferencing and Protege. The last dataset (i.e., uni-

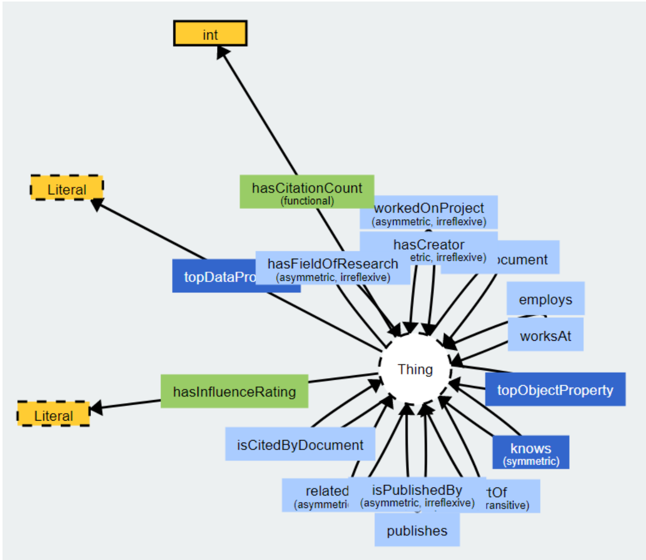

Fig. 14. Some properties in the ontology. At times, broad range and domain of these properties allowed these to be more involved in inferencing.

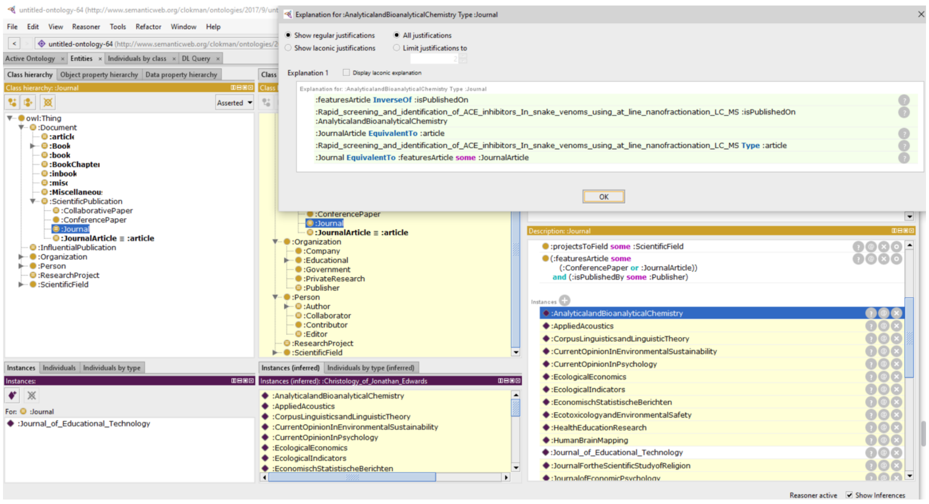

Fig. 15. The inferences on the external data. The journal instances are assigned through the interpretation rules (the reasoning can be seen on the 'Explanation' window).

versities data), however, is not joined this way, and is integrated to the current

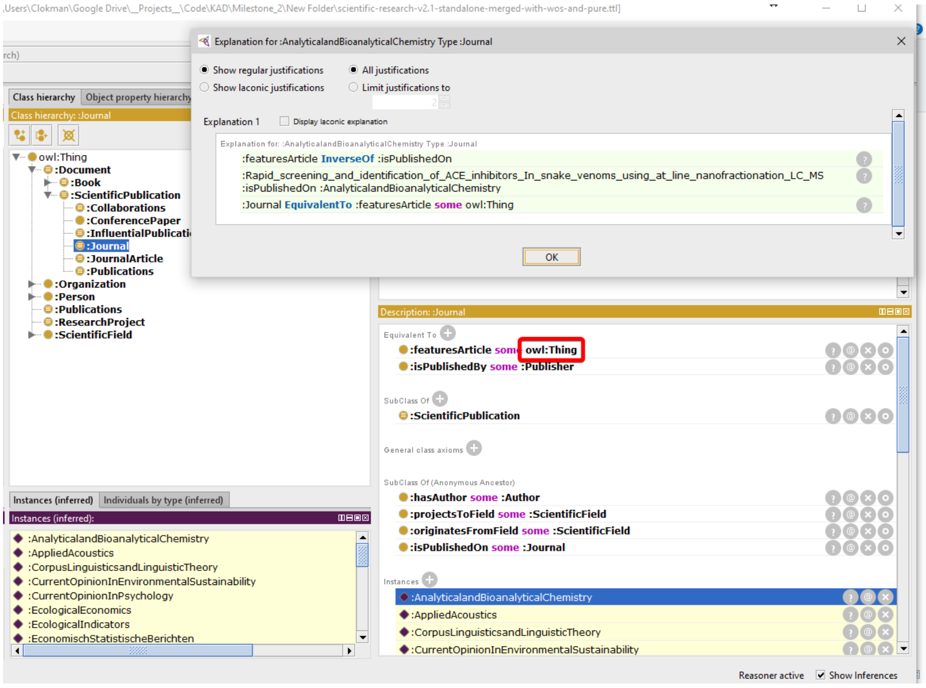

Fig. 16. A previously set, more liberal class definition that allowed a higher number and variety of inferences on the external data (the reasoning can be seen on the 'Explanation' window) , but is highly possible that it also led to less accurate categorizations.

application with SPARQL queries (fig. 17). In all stages, however, inferencing has been helpful in reducing the effort needed to explicitly specify every possible relationship between entities, which would be highly unfeasible without inferencing. Indeed, the significant difference the inferencing made for the current database is evident from queries like the ones in fig 18, 19, and 22 which, returns respectively three authors, only one publication, and nothing when inferencing is not on. And when it's on, although the number of returned results are much higher, they are still modest. This is because the development is still being carried out with a truncated version of the Pure dataset, which has an order of magnitude larger number of instances. Thus, when (or if, at least during the course, given the limited computing power) the full Pure dataset is added as instances of the ontology, the number of inferred relationships and instances can be expected to increase dramatically. And finally, another place where inference is being utilized, with the help of LD-R framework, is the application's visual query interface (fig. 20). This visual interface has the potential to make otherwise complex queries (see fig. 21) intuitive, and help linked data workflows to become more efficient and user friendly.

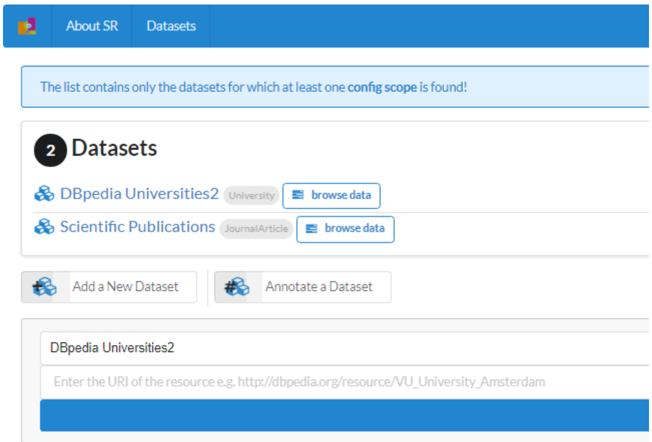

Fig. 17. The two SPARQL endpoints that are connected to the current prototype.

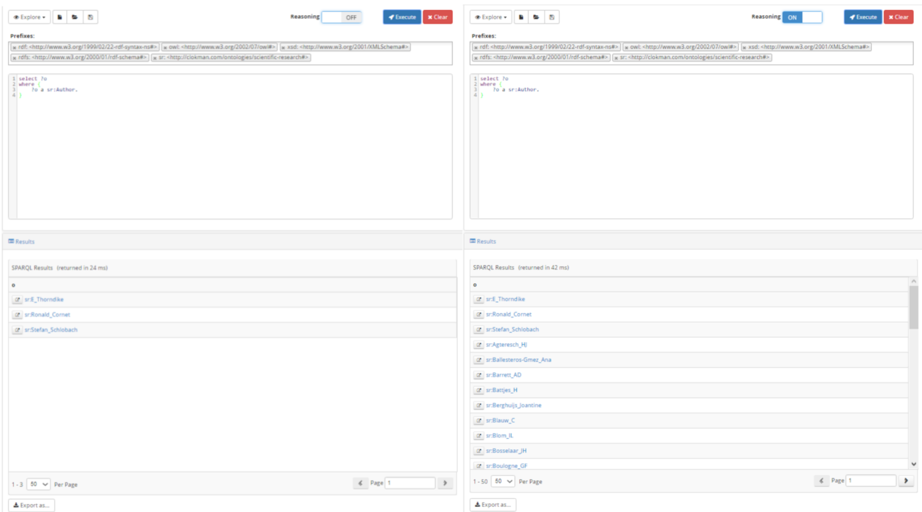

Fig. 18. Two sreenshots showing the difference inferencing makes for this dataset. The non-reasoner version on the left returns only three hard-coded authors from the beginning of the course, while the version with the reasoner returns many more instances.

5 The Prototype

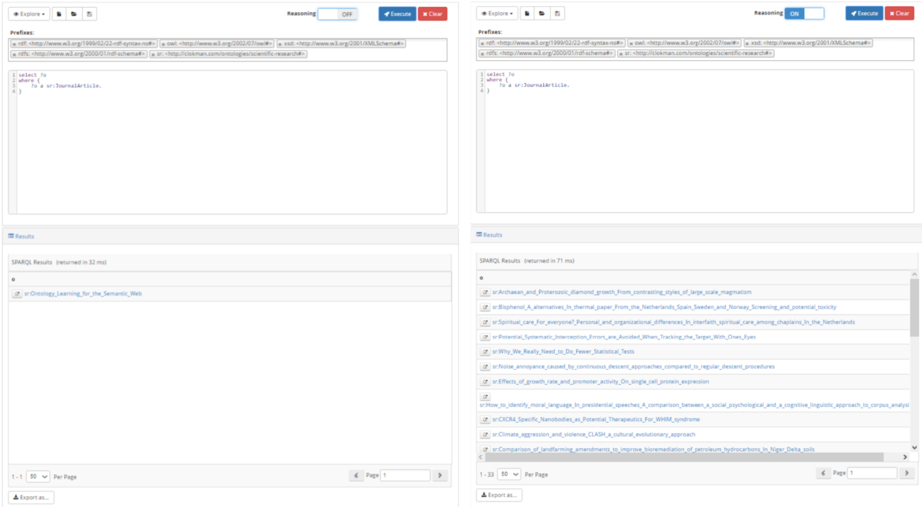

Fig. 19. Another example of using inferencing versus not using it. Although the query asks all the journal papers (and there are certainly more than one instance in the dataset) the non-reasoning query returns one instance, which is a hard coded instance from the first prototypes of the ontology. Switching the reasoner on leads to many more results.

Videos As a continuation of this report, a short, video has been produced in order to demonstrate the development behind the prototype, and its user interface in action. To document a complex and feature-rich application with more than four layers of development (Python, Protege, Stardog, LD-R), unfortunately, about five minutes proved to be the minimum required time. However, in case the two-minute limit is not a negotiable one, a shorter (although still three-minutes long) version that omits the development and simply focuses on front-end of the application is also included. However, the full version is highly recommended, as it is highly likely to clarify, support, and therefore complement this report to a much greater extent.

- Full video: Development workflow and user interface :`https://vimeo.com/240286405`
- Short version: User interface: `https://vimeo.com/240290153`

GitHub Repository

- The code, data, ontologies, and other files of the prototype application can be reached at: `https://github.com/clokman/KAD/tree/master/Prototype_Final`
- Additionally, the all output produced during the Knowledge and Data course can be reached at: `https://github.com/clokman/KAD`

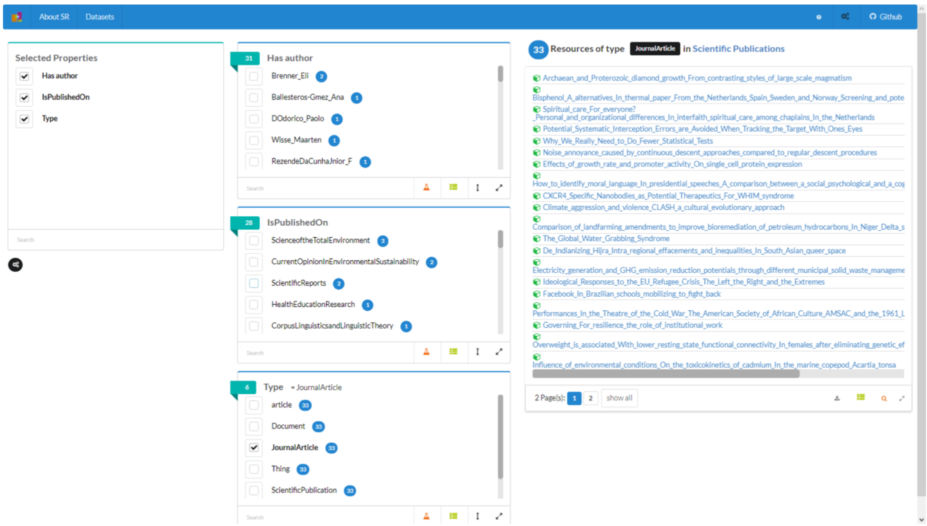

Fig. 20. An example of the inferencing used in the prototype for (relatively) complex queries, and a preview of the intended final deliverable of the course: an accessible application that disambiguates data and makes complex queries look intuitive.

6 Additional Output

– There have also been numerous illustrations and/or infographics created during the past two months of the project in order to aid with understanding and communication of the concepts covered in the course. These illustrations can be entertaining and perhaps could also be useful for teaching or communication. If they are seen so, they can be re-used freely. The illustrations can be found at `https://github.com/clokman/KAD/tree/master/__Illustrations__` .

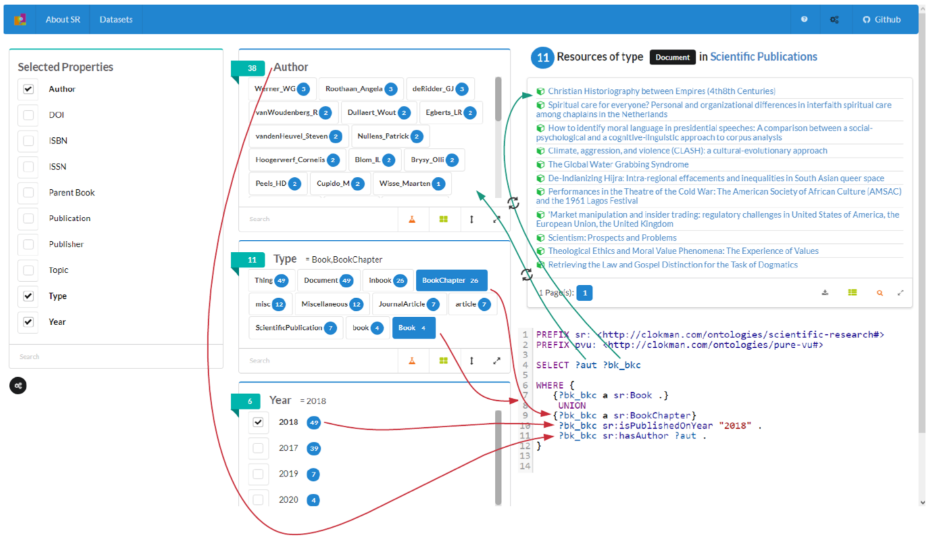

Fig. 21. The visual query logic in LD-R, and the translation of its interface state to a SPARQL query. Red arrows indicate query variables (˜input) while green arrows indicate what the query returns. LD-R supports such complex queries, and also performs inferencing. The above query, for instance, does not return any results when inferencing is turned off (see fig. 22).

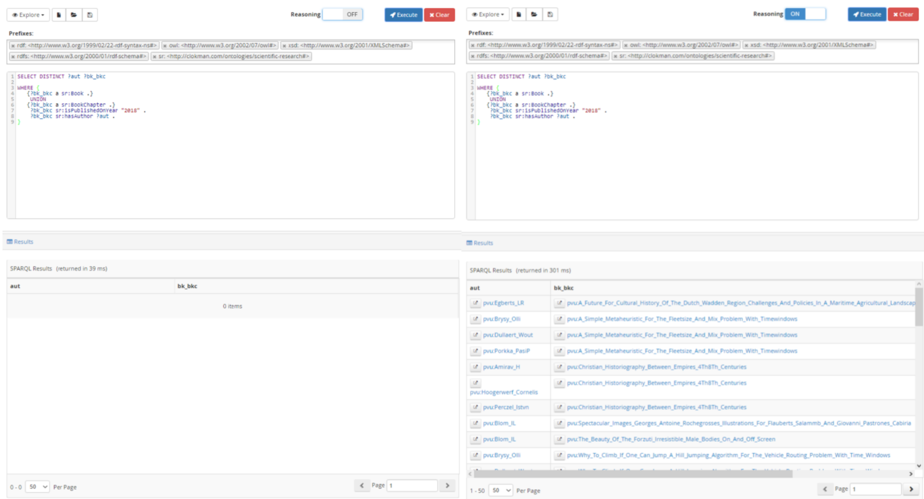

Fig. 22. The query in fig. 21 with inferencing on and off. Investigation revealed that the empty results when the reasoner is off is likely due to class equivalency axioms (e.g., "pvu:inbook sr:equivalentClasses sr:BookChapter" not being interpreted as to allow mapping instances of 'pvu' (Pure-VU) classes to 'sr' (Scientific Research, the main student ontology) classes. As a result, sr:classes ended up with no instances (or a couple of hard-coded example instances) as if Pure-VU dataset was not imported. (Note: These two versions of the query is not demonstrated on Linked Data Reactor interface, as turning off the reasoner of LD-R lead to problems on a more fundamental level (i.e., LD-R not being able to read any data at all). The absence of this situation in Protege [which was set to use Fact++ reasoner] and Stardog's query editor [above] indicates that the reasoning types may be different between these applications, and that some definitions necessary for LD-R's reasoning engine may not be present without reasoning.)

References

1. Ang, H.M., Kwan, Y.H.: Bibliometric analysis of journals in the field of geriatrics and gerontology. Geriatrics & Gerontology International 17(2), 357–360 (feb 2017), https://doi.org/10.1111%2Fggi.12880
2. Barry, A., Born, G., Weszkalnys, G.: Logics of interdisciplinarity. Economy and Society 37(1), 20–49 (feb 2008), https://doi.org/10.1080%2F03085140701760841
3. Beauchesne, O.H.: Map of scientific collaboration between researchers. http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/ (2011), http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/, accessed on Fri, October 13, 2017
4. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284(5), 34–43 (may 2001), https://doi.org/10.1038%2Fscientificamerican0501-34
5. Cardona, M., Marx, W.: Vitaly L. Ginzburg: A Bibliometric Study. In: On Superconductivity and Superfluidity, pp. 217–232. Springer Berlin Heidelberg, https://doi.org/10.1007%2F978-3-540-68008-6_7
6. Clarivate Analytics: Web of Science. http://wokinfo.com (2017), http://wokinfo.com, accessed on Wed, October 18, 2017
7. van Eck, N.J., Waltman, L.: Software survey: VOSviewer a computer program for bibliometric mapping. Scientometrics 84(2), 523–538 (dec 2009), https://doi.org/10.1007%2Fs11192-009-0146-3
8. Elsevier: Elsevier Fingerprint Engine. https://www.elsevier.com/solutions/elsevier-fingerprint-engine (2017), https://www.elsevier.com/solutions/elsevier-fingerprint-engine, accessed on Fri, October 13, 2017
9. Elsevier: Pure. https://www.elsevier.com/solutions/pure (2017), https://www.elsevier.com/solutions/pure, accessed on Fri, October 13, 2017
10. Elsevier: Scopus. https://www.scopus.com/home.uri (2017), https://www.scopus.com/home.uri, accessed on Fri, October 13, 2017
11. Fitzgerald, D., Callard, F.: Social Science and Neuroscience beyond Interdisciplinarity: Experimental Entanglements. Theory Culture & Society 32(1), 3–32 (jun 2014), https://doi.org/10.1177%2F0263276414537319
12. Jensen, P., Lutkouskaya, K.: The many dimensions of laboratories' interdisciplinarity. Scientometrics 98(1), 619–631 (sep 2013), https://doi.org/10.1007%2Fs11192-013-1129-y
13. Khalili, A., Loizou, A., van Harmelen, F., Andries de Graaf, K., Albert Merono-Penuela, Pek van Andel, P.v.: Linked Data Reactor. http://ld-r.org (2017), http://ld-r.org, accessed on Sat, October 14, 2017
14. Khalili, A., Merono-Penuela, A.: WYSIWYQ – What You See Is What You Query (2017), http://research.ld-r.org/papers/wysiwyq.pdf, accessed on Sat, October 14, 2017
15. Lim, M.D.: Consortium Sandbox: Building and Sharing Resources. Science Translational Medicine 6(242), 242cm6–242cm6 (jun 2014), https://doi.org/10.1126%2Fscitranslmed.3009024
16. Lohmann, S., Negru, S., Haag, F., Ertl, T.: VOWL 2: User-Oriented Visualization of Ontologies. In: Lecture Notes in Computer Science, pp. 266–281. Springer International Publishing (2014), https://doi.org/10.1007%2F978-3-319-13704-9_21
17. Morooka, K., Ramos, M.M., Nathaniel, F.N.: A bibliometric approach to interdisciplinarity in Japanese rice research and technology development. Scientometrics 98(1), 73–98 (sep 2013), https://doi.org/10.1007%2Fs11192-013-1119-0

18. Mugabushaka, A.M., Kyriakou, A., Papazoglou, T.: Bibliometric indicators of interdisciplinarity: the potential of the Leinster–Cobbold diversity indices to study disciplinary diversity. Scientometrics 107(2), 593–607 (feb 2016), https://doi.org/10.1007%2Fs11192-016-1865-x
19. Neo4j, Inc.: Neo4j, the world's leading graph database - Neo4j Graph Database. https://neo4j.com (2017), https://neo4j.com/, accessed on Sat, October 28, 2017
20. Olah, C., Carter, S.: Research Debt. Distill 2(3) (mar 2017), https://doi.org/10.23915%2Fdistill.00005
21. Perianes-Rodriguez, A., Waltman, L., van Eck, N.J.: Constructing bibliometric networks: A comparison between full and fractional counting. Journal of Informetrics 10(4), 1178–1195 (nov 2016), https://doi.org/10.1016%2Fj.joi.2016.10.006
22. Prathap, G.: Quantity quality, and consistency as bibliometric indicators. Journal of the Association for Information Science and Technology 65(1), 214–214 (sep 2013), https://doi.org/10.1002%2Fasi.23008
23. RISIS Consortium: RISIS: Research Infrastructure for Science and Innovation Studies. http://risis.eu (2017), http://risis.eu, accessed on Fri, October 13, 2017
24. Roessner, D., Porter, A.L., Nersessian, N.J., Carley, S.: Validating indicators of interdisciplinarity: linking bibliometric measures to studies of engineering research labs. Scientometrics 94(2), 439–468 (oct 2012), https://doi.org/10.1007%2Fs11192-012-0872-9
25. Stardog Union: Stardog: the Enterprise Knowledge Graph. http://www.stardog.com (2017), http://www.stardog.com, accessed on Fri, October 13, 2017
26. Thomson Reuters: Web of Science Category Terms. http://images.webofknowledge.com/ (2017), http://images.webofknowledge.com, accessed on Wed, October 18, 2017
27. VU Network Institute: Academy Assistants and Projects. http://www.networkinstitute.org/academy-assistants/academy-projects-17/ (2017), http://www.networkinstitute.org/academy-assistants/academy-projects-17/, accessed on Fri, October 13, 2017
28. Wang, J., Thijs, B., Gllnzel, W.: Interdisciplinarity and Impact: Distinct Effects of Variety Balance and Disparity. SSRN Electronic Journal (2014), https://doi.org/10.2139%2Fssrn.2548957
29. Wang, J., Thijs, B., Glänzel, W.: Interdisciplinarity and Impact: Distinct Effects of Variety Balance, and Disparity. PLOS ONE 10(5), e0127298 (may 2015), https://doi.org/10.1371%2Fjournal.pone.0127298
30. Yegros-Yegros, A., Rafols, I., D'Este, P.: Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. PLOS ONE 10(8), e0135095 (aug 2015), https://doi.org/10.1371%2Fjournal.pone.0135095
31. Zhang, L., Rousseau, R., Glänzel, W.: Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. Journal of the Association for Information Science and Technology 67(5), 1257–1265 (feb 2015), https://doi.org/10.1002%2Fasi.23487
32. Zulueta, M.A., Bordons, M.: A global approach to the study of teams in multidisciplinary research areas through bibliometric indicators. Research Evaluation 8(2), 111–118 (aug 1999), https://doi.org/10.3152%2F147154499781777612