Ontology for data science research results reuse

Aritha Kumarasinghe[1,*,†] and Marite Kirikova[1,†]

[1] *Institute of Applied Computer Systems, Riga Technical University, 6A Kipsalas Street, Riga, LV-1048, Latvia*

Abstract

Data Science is the science that relates to the extraction of knowledge and information from data. As the amount of data we produce increases, data science projects have become a very popular endeavor in recent years, accompanied by an increased interest in research relating to data science resources such as the data sources, algorithms, technologies, and visualizations as well as the application domains of these data science resources. The amalgamation of the results gained by data science projects can be a complex process that can be time and labor-intensive. This research seeks to reduce the project complexity by proposing an ontology that can represent data science (research) project based on domain-specific (data science) project attributes that can represent all conceivable aspects of a data science project.

Keywords

Ontology, Data Science, Research Results Reuse, Project Attributes

1. Introduction

Data science aims to clean, prepare, and analyze different data sets to extract meaning from data [1]. With the increased number of applications of data science in different sectors/domains such as social housing, shipping, and automotive retail [1], to name a few, there is an increased amount of knowledge being produced by these projects relating to how data science resources can be applied in data science projects and/or domains. How this knowledge from projects can be accumulated for reuse in future projects is the issue that will be the focus of this paper.

One solution for this problem is the use of a knowledge graph which is a knowledge representation that can effectively organize and represent knowledge [2]. This solution was showcased in our previous work related to a knowledge graph for reusing research knowledge on related works in data analytics [3]. The presented knowledge graph utilized a star-like ontology based on analytics project attributes, as a schema. The 18 defined analytics project attributes were based on an initial literature review and represented different aspects of data analytics projects such as the data analysis algorithm(s) used, the data set(s) used, and the analysis software(s) or tool(s) used, etc.

BIR-WS 2024: BIR 2024 Workshops and Doctoral Consortium, 23rd International Conference on Perspectives in Business Informatics Research (BIR 2024), September 11-13, 2024, Prague, Czech Rep.

* Corresponding author.

† These authors contributed equally.

Balasuriyage-Aritha-Dewnith.Kumarasinghe@rtu.lv (A. Kumarasinghe); marite.kirikova@rtu.lv (M. Kirikova)

0009-0000-5915-8919 (A. Kumarasinghe); 0000-0002-1678-9523 (M. Kirikova)

The star-like ontology was defined based on a triple structure that considers the subject to be the data analytics project, the object to be the data analytics project attribute relating to a specific aspect of a data science project, and the predicate to be the relationship type defined based on the data analytics project attribute. This structure meant that each attribute type needed to be represented as a class in the ontology with a corresponding property defined for the class representing the data analytics project, relating it to the data analytics project attribute value. This meant that when additional analytics project attributes were defined, it increased the complexity (number of classes) of the ontology on a class level, and when the ontology changed, the process related to changing the ontology was complex (editing many "one-level" ontology classes and their properties). Additionally, this ontology, like any other ontology, relied on a static representation of the data analytics domain and assumed that the user is only interested in the aspects of the data analytics project that are represented by the analytics project attribute types. This work seeks to resolve the issue by making an ontology that can be relatively easily modified. The previous ontology was constructed inductively by considering the data analytics projects. The one proposed in this paper is built to represent what can be considered as the already established(standard) aspects of a data science project defined based on the data science body of knowledge; it also seeks to allow additional aspects to be introduced (on an instance level) based on the users' needs or the results produced by data science projects. We accomplish this by proposing an ontology that does not seek to represent the domain of data science but instead to represent projects within said domain. As there is no specific body of knowledge in data analytics available, we chose the body of knowledge that represents data science thus, we decided to increase the scope of the ontology to data science with data analysis being considered a knowledge area under this domain as defined in the Data Science Body of Knowledge (DS-BoK) by the EDISON project [4], but this extension of the scope is rather formal because the coverage of attributes in the previous ontology are specific mostly to the data analysis aspects of data science. So, instead of data analytics, we have a data science project attribute defined as a class with the individual data science project attributes defined as instances of that class. We also propose a novel method with which data science attributes can be defined based on the Data Science Competence Framework (CF-DS), also created by the EDISON project [5] to represent all currently recognized aspects of projects in the data science domain, with the possibility to define more data science project attributes based on research projects as instances of the ontology.

To ensure that the created ontology conforms to existing knowledge engineering practices we followed the Ontology Definition 101 methodology [6]. Section 2 of this paper outlines the definition process of the proposed ontology based on the steps of the Ontology Definition 101 methodology. Section 3 demonstrates how this created ontology can be applied for data science knowledge reuse. The final Section 4 concludes the paper and outlines what future research work can be done.

2. Ontology Definition Process

In this section, the ontology definition process is briefly presented.

The Ontology Definition 101 methodology [6] was used as a basis for ontology definition. It states the following steps within their knowledge-engineering methodology:

1. Determine the domain and scope of the ontology.
2. Consider reusing existing ontologies.
3. Enumerate important terms in the ontology.
4. Define classes and the class hierarchy.
5. Define the properties of classes.
6. Define the facets of the slots.
7. Create Instances.

In the remainder of this section, sub-sections are organized to represent the results related to each of the steps mentioned above, with Step 1 discussed in Section 2.1, Step 2 reflected in Section 2.2, and Steps 3-6 introduced in Section 2.3. Step 7 is presented in Section 2.4, which also presents the final version of the ontology.

2.1. Domain and Scope of the Ontology

The domain of the proposed ontology is projects within data science that aim to clean, prepare, and analyze different data sets for extracting meaning from data [1]. The goal of the ontology is to accumulate and preserve knowledge that is produced in data science projects. To encapsulate this knowledge, the ontology will represent the different kinds of resources used within the domain of data science based on data science project attribute types (such as algorithms used, data sources used, and visualizations used) and based on these attribute types the knowledge presented by the ontology will change. It should be noted that this ontology does not seek to directly represent the data science domain but instead to represent projects within this domain (Fig. 1) via the data science attribute type that represents the resources within the data science domain. A partial representation of the data science domain is possible through the resources represented.

To represent data science projects, the ontology will try to answer the following four competency questions (that are defined in the Ontology Definition 101 methodology as 'One of the ways to determine the scope of the ontology is to sketch a list of questions that a knowledge base based on the ontology should be able to answer'):

1. What data science projects have been/are being completed?
2. What data science application domains have the data science projects been completed in?
3. What type of data science project attributes represent the resources that are used in data science projects?
4. What attributes represent a completed/ongoing data science project?

Based on the competency question relating to the type of data science project attributes as well as the one relating to the application domains of data science resources, further competency questions such as 'Given a data science application domain (representing the application domains of data science resources) what machine learning algorithms can be used(representing the resources that are produced within the data science domain)?' can be inferred. Basing the ontology on data science project attribute types (e.g., algorithms used) and data science project attributes (e.g., Decision Trees) separately allows for more domain-specific knowledge to be inferred from any knowledge graph that utilizes this ontology as a schema.

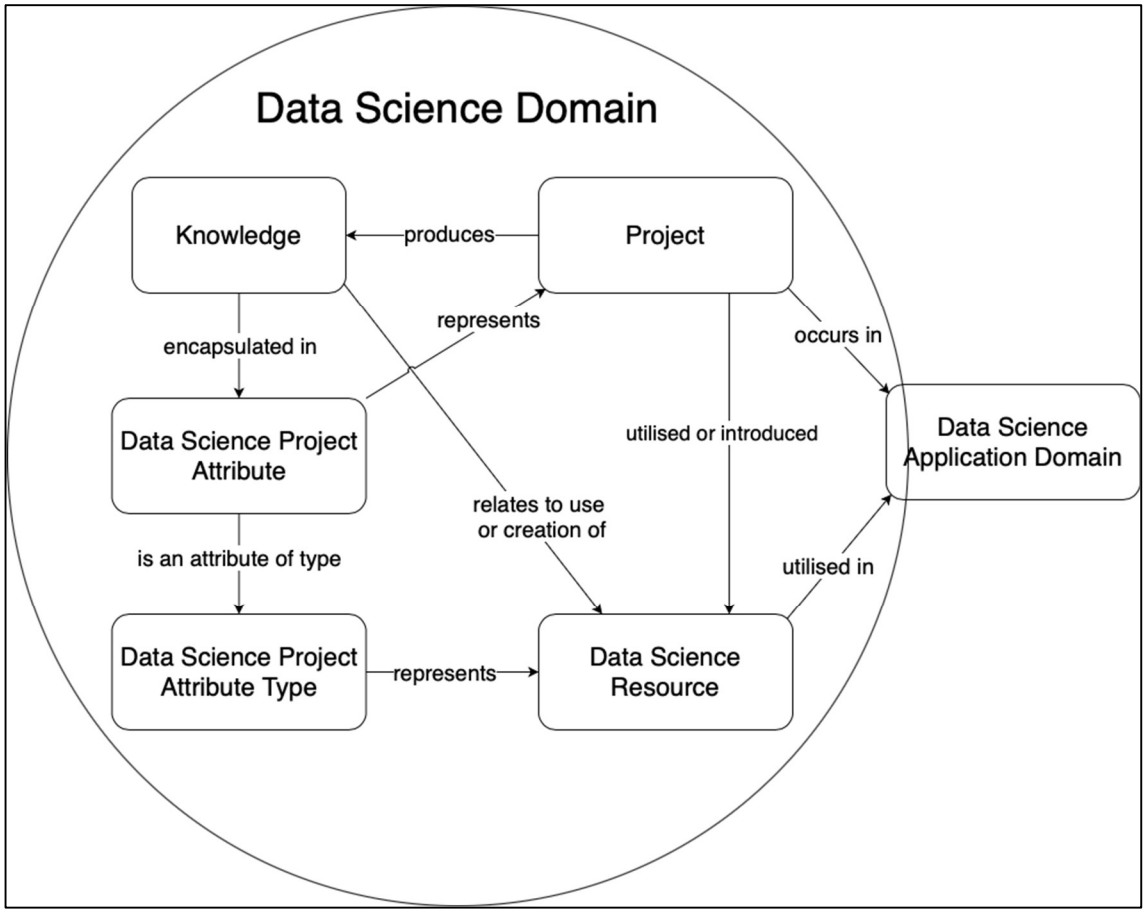

Figure 1: Proposed solution for the accumulation of knowledge produced within data science projects.

2.2. Reusing Existing Ontologies

In the authors' previous work [3] an ontology for reusing research knowledge on related work in data analytics based on 18 analytics projects was proposed and used. This work concerns an ontology with a similar purpose but expands to the domain of data science, which is larger than data analytics.

In the domain science data science, several ontologies have been proposed and some of them are considered in this section.

There is an ontology [7] for Big Data Analytics as a Service (MBDSaaS) based on the declarative (sub)model proposed in the TrustwOrthy model-awaRE Analytics Data platform (TOREADOR) project, intending to aid 'incompatibility management and the creation of OWL-S descriptions enabling different approaches for the selection tasks'. This ontology concerns such aspects as (1) Data preparation, all activities aimed to prepare data for analytics; (2) Data representation, how data are represented and representation choices for each analysis process; (3) Data analytics, the analytics to be computed; (4) Data processing, how data are routed and parallelized; and (5) Data visualization and reporting, an abstract representation of how the results of analytics are organized for display and reporting. A similar ontology is proposed within Intelligent Big Data Analytics as an intermediary between the abstract tasks in workflows of data mining to automate the data mining process [8] but based on the Cross Industry Standard for Process Mining (CRISP-DM). Another noteworthy ontology within data analytics is an ontology-based framework relating to the recommendation of an analysis method [9]. This work proposes an ontology based on data sources and analysis methods and demonstrates the value of ontology-based applications. The proposed ontologies do have the potential to further expand on data analysis method-related aspects of the ontology we propose.

To compare with our ontology used in the previous work [3], the above-mentioned approaches define data analytics in a much narrower sense. In our approach, the data analytics project practically included all the above-mentioned aspects; however, for instance, the ontology in [7] goes to a higher level of detail regarding each of the aspects while, in our case, there is no further classification of individuals. The ontology in [8] is more complex and might be harder to apply to information that is available in the scientific/project literature regarding specific data science projects. It would also be harder to maintain such ontology given that this ontology could change on a class level, whereas the ontology proposed in this paper would practically act as an OWL-based data schema that would only change on an instance level.

Thus, the question is about the granularity of the ontology. We might assume that higher granularity of ontology might give additional opportunities in knowledge amalgamation; however, as was already mentioned, the level of detail available in scientific works or project documentation does not always allow us to go to that level of detail. Also, the higher the level of detail, the more often reconsidering an ontology itself might be needed. Thus, the open question is what level of detail might be useful in amalgamating knowledge in the data science domain and what frameworks or initiatives might be used to maintain the ontology used to refer to the work in the respective domain.

As shown above, there are many applications of ontologies within the domain of data science and the proposed ontologies have different purposes. To our knowledge, there have not yet been ontologies proposed for the reuse of knowledge within the data science domain. In this paper, we, based on our experience with scientific work in data analytics, propose an ontology that is based on the skills and knowledge units defined within the EDSF (EDISON Data Science Framework) Competency Framework for Data Science (CF-DS) [5].

The final ontology can be used to reduce the complexity related to knowledge reuse in data science/data science projects.

2.3. Defining Classes and Class Properties

Within the previous sections, the following terms have been recurrent:

1. Data Science Project – Projects that that aim to clean, prepare, and analyze different data sets for extracting meaning from data
2. Data Science Project Attribute Type – Representation of the type of attributes that would represent a data science project resource such as the data mining algorithm used within the project [10]. When defining these attribute types, it is important to consider those that have already been recognized as well as newer technologies within the domain of data science. These two types of data science attributes can be represented as standard and custom data science attribute types.
3. Data Science Project Attribute – An attribute that represents a single or multiple data science project such as decision trees [11] which would be of the attribute type data mining algorithms used.
4. Data Science Application Domain – This is the domain in which data science is being applied such as social housing, shipping, and automotive retail[1]. It should be noted that, to our understanding, there is no existing taxonomy of the application domain of data science; therefore, the classification of a data science project to an application domain is at the user's discretion.

These four classes (and two subclasses) will be sufficient for accumulating data science knowledge based on data science project attributes and the class properties shown in Fig. 2.

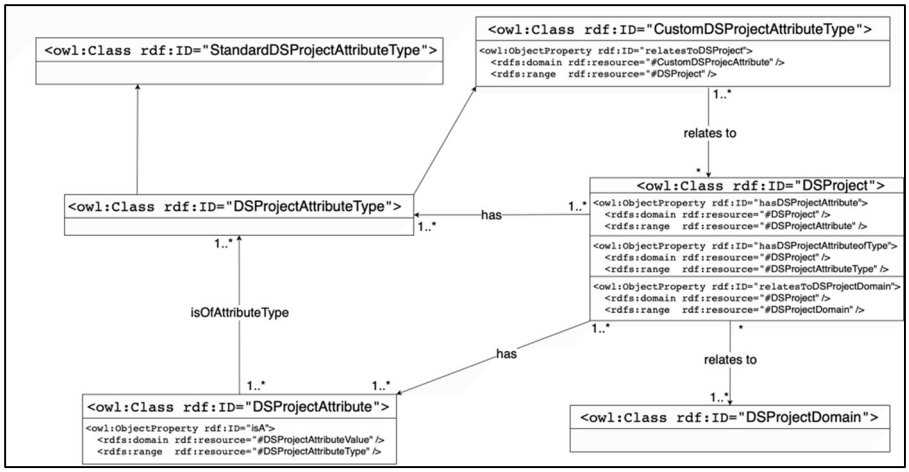

Figure 2: UML class diagram representing the Ontology for Data Science Projects.

OWL and RDFS are used for the definition of classes and class properties given that this ontology is meant to be used for a knowledge graph, created using RDF, as was the case in

the authors' previous work [3]. However, unlike the previously defined ontology, the ontology proposed in this paper allows the data science project attribute types to be defined on an instance level (thereby reducing the complexity of the ontology in relation to the number of classes) and does not require maintenance as would be the case for most of the other ontologies. This is because the conceptualization of the data science domain is done through data science project attributes, which are represented at an instance level and not the class level. Specifics of how this ontology can be utilized to store knowledge from completed or ongoing projects are demonstrated through the definition of instances for this ontology in Section 3.

Before the utility of the ontology can be demonstrated, the standard data science project attribute types need to be defined in such a way that the already recognized aspects of the data science domain must be represented; this is done in the next subsection.

2.4. Data Science Project Attribute Types Definition Process

The data science project attribute types definition process is done based on the Data Science Competency Framework (CF-DS) Release Two [5] defined as part of the EDISON Data Science Framework (EDSF) (the result of the EU-funded EDISON project), which is a collection of documents that defines the Data Science Profession which includes the aforementioned CF-DS and the Data Science Body of Knowledge (DS-BoK) [4].

Fig. 3 is a model that is meant to give the reader an understanding of elements within the CF-DS and to represent the knowledge and skills defined in the CF-DS. The CF-DS utilizes keywords to represent the skills and knowledge with keywords followed by a number; examples of the values are also shown in the diagram.

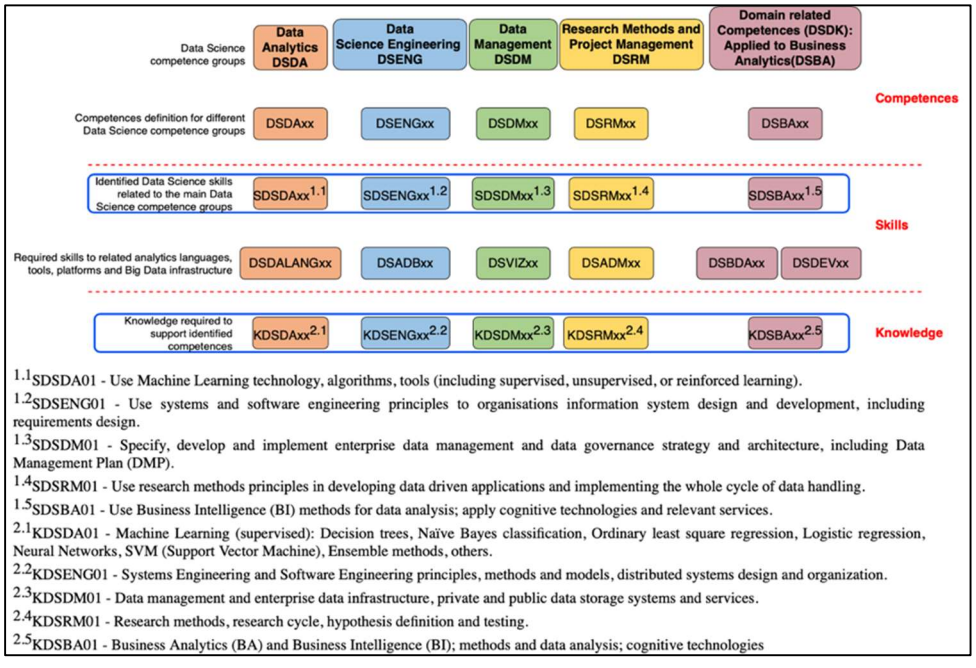

Figure 3: Competency groups, competencies, skills, and knowledge units identified/defined in EDSF CF-DS [5].

The base knowledge body does not directly address the research results in data science; rather, it reflects the knowledge that is needed to achieve them. Therefore, the attribute types of data science projects can be defined only indirectly through the concepts available in the chosen body of knowledge. In this work, we consider only the relationships between knowledge and skill (excluding those relating to the analytics languages, tools, platforms, and Big Data infrastructure) based on the preestablished notation that knowledge is a prerequisite of skillful action [12].

Unlike the previous work that limited the scope of the ontology to data analytics, in this work, the scope represented in the ontology is expanded upon by defining data science project attribute types in such a way that they map the skills defined in the CF-DF to the knowledge units defined in the CF-DF in a way that each knowledge area has at least one associated skill ensuring that all knowledge required for data science related skills are represented in the ontology. This would allow for the representation of the data science domain as a whole and the encapsulation of knowledge relating to all established aspects of a data science project.

To ensure that the defined data science project attribute types are accurate and are traceable to CF-DS knowledge units/topics and the skills, the data science project attributes are defined in an X_Y_Z format (Fig. 4), where X represents (Data Science)Domain Specific Key Words present found in both the Knowledge topics/units required and the Skills(such as Data Mining, Supervised Machine Learning, and Predictive Analytics), Y represents the (Data Science) Domain Specific Resources (such as Techniques, Tools, and Algorithms), and Z represents Actions Verbs(such as used, implemented, and developed) that are defined based on the action words(such a use, implement, and develop) that are mentioned in the CF-DS skills. This provides a formal meta-structure for data science project attribute types that were missing in the previously defined data analytics project attributes, and it enables the systemization of the data science attribute definition process.

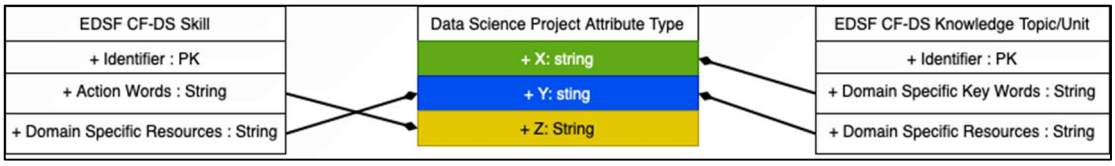

Figure 4: UML Diagram representing the structure of Data Science Project Attribute Type and it's relationship between the EDSF CF-DS Skill, and EDSF CF-DS Knowledge unit/topic.

Based on the defined Data Science Project Attribute Type Structure the EDSF CF-DS Skills, and EDSF CF-DS Knowledge unit/topics were manually parsed to recognize the relevant Keywords and Action Verbs. An example of how this was accomplished can be seen in Fig. 5, which shows how three data science project attributes were defined to map the knowledge units KDSDA01, KDSDA02, and KDSDDA03 to the skill SDSDA01.

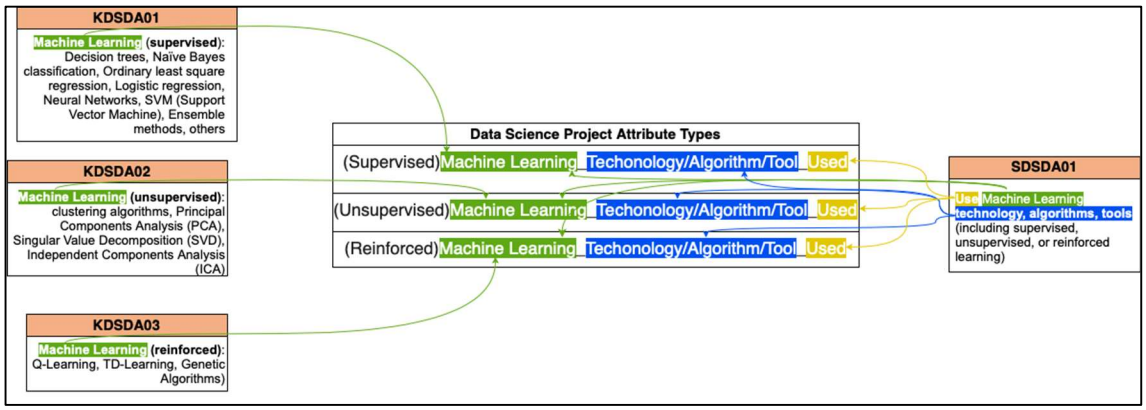

Figure 5: Data Science Project Attribute Types definition process demonstrated in relation to the EDSF CF-DS Knowledge unit/topics KDSDA01, KDSDA02, KDSDDA03, and the skill SDSDA01.

Fig. 5 shows that the defined DS Project Attribute Types have additional text within brackets; this text is introduced to provide additional specificity for the domain-specific keywords and was defined based on the Skills or Knowledge Areas/Topics.

Utilizing this Data Science (DS) Project Attribute Types Definition Process, 77 Data Science Attribute Types were defined, mapping all knowledge areas to at least one skill, thus representing all currently established aspects of data science projects and providing the possibility to define more data science attributes for capturing data science research results than were discovered by bottom-up approach in our previous work. All the skills themselves have at least one corresponding DS Project Attribute Type, with the only exception being SDSENG12 – Use of Recommender or Ranking system, but as this skill relates to the Recommender and Ranking Systems, the authors took the liberty to consider these systems as information systems which allowed to map this skill to KDSENG10 – Information Systems, collaborative systems by defining two DS Project Attribute Types: (i) (Information)Recomender_System_Used and (ii) (Information)Ranking_System_Used

The table containing a list of all identified unique keywords, resources, and action words demonstrating the relationships between the EDSF CF-DS Skills, Knowledge, and the defined data science project attributes is available in a GitHub repository [16]. It should be noted that, in some cases, the data science project attribute types have missing action verbs and/or resources due to limited text in either the knowledge unit/topic (e.g., KDSDA13 – Optimisation) or the skills (e.g., KDSDA14 – Optimisation). In some cases, the authors applied placeholders Y and Z, which were used to maximize the number of aspects of the data science domain represented by the data science project attributes: 16 (20%) of the DS project attributes are missing a Domain Specific Resource, and 3 of them are also missing a Domain Specific Action Verb. These missing values simply reduce the specificity of the defined DS Project Attribute Types while still providing a (limited) representation of this aspect of the data science project.

When comparing the newly defined data science project attribute types with the previously defined data analytics project attribute types [3] (of which there are only 18), it

is possible to conclude that these attribute types are general to the data science domain, whereas those previously defined and not present in the new ontology were specific to data analytics (which now is considered as a sub-domain). This is evidenced, for instance, by specific attribute types relating to data visualization and interactive results (dashboards) created, which here are not represented due to not being a knowledge unit related to data visualization (although it is represented as a skill related to tools and software) in the CF-DS.

3. Utility of the Created Ontology

Given the original intention of the ontology of being used as a schema for a knowledge graph, the UML schema (Fig. 6) was defined for a knowledge graph using OWL classes and subclasses as well as object properties.

The UML class diagram shows the classes and subclasses defined for this ontology, here the class(`rdf: ID="DSProjectAttributeType"`) was defined for the DS (Data Science) Project representing the resources available in the data science domain. This will organize the knowledge gained within a previously completed data science project and will relate to instances of the class(`rdf: ID="DSProjectAttribute"`) through an 'isAnAttributeOfType' relationship, for instance:

- 'DSProject ABC'(instance of `<owl:Class rdf:ID="DSProject">`)
  hasDSProjectAttrbiuteValue(property of `<owl:Class rdf:ID=" DSProject ">`)
  'Decision Trees' (instance of `<owl:Class rdf:ID="DSProjectAttributeValue">`)
- 'Decision Trees' (instance of `<owl:Class rdf:ID="DSProjectAttribute">`)
  isAnAttributeOfType(property of `<owl:Class rdf:ID="DSProjectAttribute">`)
  '(Supervised)MachineLearning_Technology/Algorithm/Tool_Used'(instance of `<owl:Class rdf:ID="DSProjectAttribute">`)
- 'DSProject ABC'(instance of `<owl:Class rdf:ID="DSProject">`)
  hasDSProjectAttributeOfType(property of `<owl:Class rdf:ID=" DSProject">`)
  '(Supervised)MachineLearning_Technology/Algorithm/Tool_Used'(instance of `<owl:Class rdf:ID="DSProjectAttributeType">`)

The three triples mentioned above represent the relationship between the data science project, data science project attribute type, and data science attribute in the form of DSProject ABC, (Supervised) MachineLearning_Technology/Algorithm/Tool_Used', Decision Tree. Similarly, other data science project attributes can represent project features such as Natural Language Processing_Method_Used, Data Preparation/Data Preprocessing_Method_Used, Performance/Accuracy_Metric_Used, etc. The additional RDF triples mentioned below relate the data science project to a data science application domain.

- 'DSProject ABC'(instance of `<owl:Class rdf:ID="DSProject">`)
  relatesToDSProjectDomain(property of `<owl:Class rdf:ID=" DSProject">`)
  'Health Care' (instance of `<owl:Class rdf:ID="DSProjectDomain">`)

This new RDF triple, combined with the reasoning capabilities of knowledge graphs realized through rule-based reasoning, allows for inferring what resources can be used within a specific data science application domain. The rule-based inference is realized in this ontology through the use of a SWRL [13] rule:

- DSProject(?project) ^ hasDSProjectAttribute(?project, ?tool) ^ relatesToDSProjectDomain(?project, ?domain) -> canBeUsedInDSApplicationDomain(?tool, ?domain)

To demonstrate the use of this ontology from a practical standpoint, we implemented a knowledge graph that will store, and present knowledge acquired from a single project within the data science domain [14] that introduces SatelliteBench, a framework for satellite image extraction and vector embeddings generation and it's utility in creating predictive models for poverty, education, and dengue prediction. The information provided in this project is presented using Protégé [15] (Fig. 6).

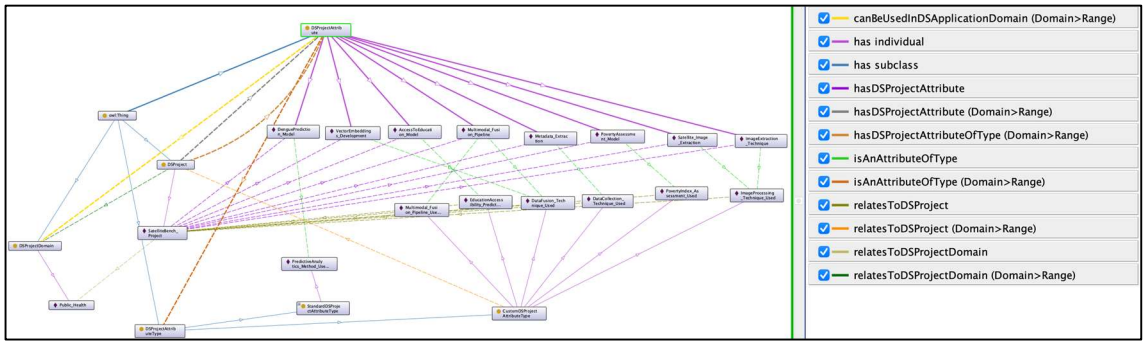

Figure 6: OntoGraf view of Data Science Project Ontology with instances representing a data science project [14].

The following Data Science Attribute instances were defined with the corresponding Data Science Project Attribute Types(with one standard attribute type, with the others being custom):

1. Dengue Prediction Model – PredictiveAnalytics_Method_Used(Standard Type)
2. Vector Embeddings Development – DataFusion_Technique_Used
3. Access to Education Model – EducationAccessibility_PredictiveModel_Used
4. Multi model Fusion Pipeline – MultiModel_Fusionpipeline_Used
5. Meta data Extraction – DataCollection_Technique_Used
6. Poverty Assessment Model – PovertyIndex_Assesment_Used
7. Satellite Image Extraction – ImageProccesing_Technique_Used
8. Image Extraction Technique – ImageProccesing_Technique_Used

The fact that all these resources can be utilized in the domain of Public Health can also be inferred using the SWRL rule that was defined relating the DSProjectDomain and the

DSProjectAttribute instances. Most new instances required the definition of Custom data science project attribute types.
It should be mentioned that these instances were defined using ChatGPT 4o (accessed 15th of August, 2024) with a query that outlined the structure of the ontology (including definitions of object properties as given in Section 2.3), provided instances of Standard Project Attributes(mentioned in Section 2.4) , and an attachment of the pdf version of the research article[14] combined with a request to present the result as instances in RDF/XML format.

This demonstrates the possibility of utilizing this ontology with LLMs or other advanced text parsing technologies to automate the accumulation and presentation of knowledge gained from completed or ongoing data science projects. The reliability of LLMs for the production of the instances for the defined data science project ontology requires further research, but this work demonstrates how domain-specific attribute types defined in a format that mimics natural language allow formulated queries to be used by LLMs easily.

The flexibility of the ontology enabled through the class representing data science project attribute types allows the user to define data resources they are interested in and to disregard the instances they are not interested in (demonstrated by the fact that of the 77 standard attribute types defined, only one 'PredictiveAnalytics_Method_Used' was needed to represent the project, whereas 7 additional custom attribute types needed to be defined). This flexibility also allows automation of the knowledge accumulation process based on the information that is available with the project reports (in this case being a research article published as a result of this data science project), and also to represent resources that are not widely used (represented by the custom data science project attributes such as DataFusion_Method_Used, PovertyIndex_Assessment_Used, etc. (Fig. 5)) or have recently been introduced by a research project.

4. Conclusions and Future Works

This work outlines the definition of an ontology that can be used to facilitate the reuse of knowledge acquired through the completion of data science projects. This ontology is based on data science project attributes (a concept of project attribute was introduced for data analytics in an author's previous work [3]) that are meant to represent various aspects of data science projects concerning the various kinds of resources (e.g., machine learning algorithms) that are available within the domain of data science represented by the data science attribute types (machine learning algorithms used). This shifts the goal of the ontology from representing a domain to representing projects within that domain in relation to the resources available within that domain. An ontology was created using the Ontology Definition 101 Methodology with the data science project attribute and attribute type as classes within this ontology. This paper also introduces a method that can be used to systematically define Data science attribute types based on the EDISON Data Science Competence Framework [5] to represent all currently recognized aspects of data science.

The utility of the data science project ontology was demonstrated by representing the knowledge acquired from a single data science project [14], presenting knowledge such as the Dengue Prediction Model as a Predictive Method Used and the Multimodal Fusion

Pipeline as a Data Fusion Technique Used. Through the use of a SWRL rule that relates the defined data science project attributes to a data science application domain, it is possible to infer the application domain of these data science resources which in this case was public health. Also demonstrated is how using this ontology (with instances of the types of project attributes within the domain of data science) in tandem with advanced text parsing technologies such as LLMs, it is possible to automate the knowledge accumulation process through the use of project reports (which in the case of the discussed project [14] was a single research article).

This paper demonstrates how shifting the goal of the ontology for domain representation to the representation of projects within a domain can simplify the ontology itself as well as reduce the complexity related to the maintenance of the ontology by making it more instance-centric than class-centric.

Future works can demonstrate the further utilization of the ontology-defined data science project attributes for knowledge representation and automation of the knowledge accumulation process.

References

[1] G., Vicario, S., Coleman, "A review of data science in business and industry and a future view." Applied Stochastic Models in Business and Industry 36, no. 1, 6-18, (2020).

[2] X., Chen, S., Jia, and Y., Xiang, "A review: Knowledge reasoning over knowledge graph." Expert systems with applications 141, 112948, (2020).

[3] A., Kumarasinghe, M., Kirikova, "Knowledge Graph for Reusing Research Knowledge on Related Work in Data Analytics." In International Conference on Advanced Information Systems Engineering, pp. 186-199. Cham: Springer Nature Switzerland, (2024).

[4] Y., Demchenko, A., Manieri, A., Belloum, " EDISON Data Science Framework: Part 2. Data Science Body of Knowledge (DS-BoK) Release 2", (2017). URL: https://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_ds-bok-release2-v04.pdf

[5] Y., Demchenko, A., Manieri, A., Belloum, "EDISON Data Science Framework: Part 1. Data Science Competence Framework (CF-DS) Release 2", (2017): URL: https://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release2-v08_0.pdf

[6] N.F., Noy, D.L., McGuinness, 2001. Ontology development 101: A guide to creating your first ontology.

[7] D., Redavid, R., Corizzo, D., Malerba, "An OWL Ontology for Supporting Semantic Services in Big Data Platforms," 2018 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, USA, 2018, pp. 228-231, doi: 10.1109/BigDataCongress.2018.00039.

[8] T. H., Akila, S., Siriweera, I. Paik, B. T. G. S., Kumara, "Onotology-based service discovery for intelligent Big Data analytics," 2015 IEEE 7th International Conference on Awareness Science and Technology (iCAST), Qinhuangdao, China, 2015, pp. 66-71, doi: 10.1109/ICAwST.2015.7314022.

[9] G., Henriques, D., Stacey, "An Ontology-Based Framework for Analysis Recommendation," 2014 IEEE International Conference on Bioinformatics and Bioengineering, Boca Raton, FL, USA, 2014, pp. 277-282, doi: 10.1109/BIBE.2014.70.
[10] F., Provost, T., Fawcett, 2013. Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp.51-59.
[11] N., Ye, 2013. Data mining: theories, algorithms, and examples. CRC press.
[12] Stanley, Jason. “Know how.” OUP Oxford, 2011.
[13] I., Horrocks, P.F., Patel-Schneider, H., Boley, S., Tabet, B., Grosof, M., Dean, 2004. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member submission, 21(79), pp.1-31.
[14] D., Moukheiber, D., Restrepo, S.A., Cajas, et al., A multimodal framework for extraction and fusion of satellite images and public health data. Sci Data 11, 634 (2024). doi: 10.1038/s41597-024-03366-1.
[15] M.A., Musen, The Protégé project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. doi: 10.1145/2557001.25757003.
[16] A., Kumarasinghe, Ontology for Data Science Research Results Reuse (Version 1) [Computer software]., (2024). https://github.com/ArithaRTU/Ontology-for-Data-Science-Research-Results-Reuse.git