**ONTOLOGY FOR STORING INFOMRMATION FOR INTEGRATION OF DISTRIBUTED HETEROGENEOUS SOURCES**

**Юрченко Богдана Олегівна**

студентка

Національний технічний університет України

"Київський політехнічний інститут імені Ігоря Сікорського"

м. Київ, Україна

danarossa14@gmail.com

**Introductions.** The task of storing metadata for integration with external systems was initially set in the context of the development of a system for collecting and managing the information on international cooperation. The system being developed shall simplify the task of analysis of reports on international cooperation by providing means of automatization of the data collection from external systems. The system shall allow users to set up jobs to fetch data from external sources such as relational database management systems, NoSQL databases, and public owl files. Earlier on different standards of storing the metadata were analyzed. Namely Dublin Core, schema.org, etc. None of the already existing standards was admitted as suitable for the task. So it was decided to develop an ontology to store the required information needed for integration and data mapping.

**Aim.** The aim of the work is to analyze existing approaches to ontology based integration and to develop an ontology to store integration configurations as well as the mapping and aggregation rules from external data schemes to the target ontology.

**Materials and methods.** Many works are written on the development of systems to integrate different sources. In the paper “Ontology-based integration of data sources” [1] a approach to the integration of heterogeneous data sources with a local to global ontology mapping is explored. In the work “Designing a System for Integration of Macroeconomic and Statistical Data Based on Ontology” [2] an ontology-based system for integration of macroeconomic and statistical data was developed. The described system was designed to in a semi-automatic way. Meaning that the system promths the user to upload a table with data and then manually map columns and rows to ontology entities.

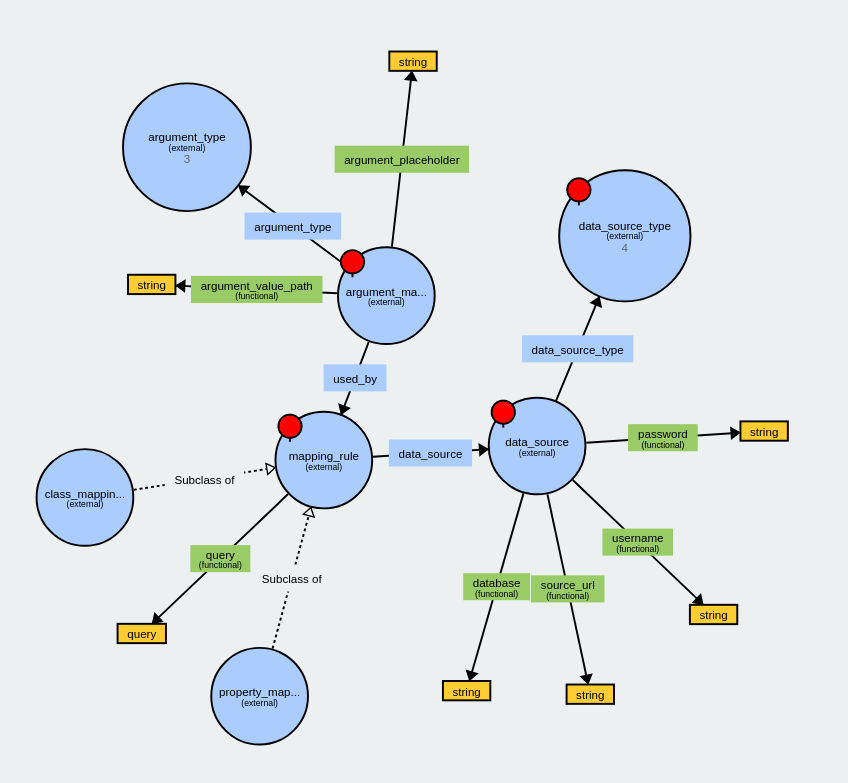
The goal of the developing system is to integrate a target ontology with external systems via a set of predefined mapping rules to omit the manual data upload and repetitive mapping. The additional ontology shall store the mapping and aggregation rules to perform the data collection automatically by initializing the connection with external sources, hense reducing the amount of manual work.

**Results and discussion.** In this work, an ontology for the connection and mapping configuration was developed.

As seen in figure 1, the ontology contains such main classes as mapping\_rule with subclasses class\_mapping\_rule and property\_mapping\_rule, which will allow exporting from external systems new individuals of specific classes as well as values of properties of already existing individuals. Class data\_source holds all properties of a specific data source. Namely, URL, login, a password for authorization, and the name of a database, that will be used in case of a database connection (and not a public owl file, where only a URL is required). Class data\_source\_type defines the type of a data source and specifies the driver to be used to connect. Individuals of argument\_mapping\_rule class specify the mapping rules of specific arguments for placeholders in the query, specified under the referenced mapping\_rule. Each argument can be of different types, and the type of the argument in the ontology is defined via a relation with argument\_type class.

The query is a query written in supported by the data source query language, be it SQL, SPARQL, or MongoDB Query Language. The query should return a single value in case if used by a property\_mapping\_rule or a table in case of class\_mapping\_rule.

The query defines the data preprocessing before it gets inserted into the target ontology.



**Figure 1. The diagram of main classes and properties of the created ontology**

The system will have a predefined set of supported argument types and data source types (MongoDB, PostgreSQL, Oracle Database, etc.). Such design creates 2 important points of extension, that will allow improving the system more easily in the future.

**Conclusions.** In the work, different approaches to collecting data from distributed heterogeneous sources were analyzed. An ontology was developed that allows the storage of information for a successful connection with an external system of any of the following types: public owl files, relational databases, and NoSQL databases. The ontology stores the mapping rules and connection configuration that in their turn allow data preprocessing and aggregation prior to insertion. The ontology can be filled using an ontology editor. The ontology can be used by a system to automate information collecting.

**References:**

1. M. Gagnon, "Ontology-based integration of data sources," 2007 10th International Conference on Information Fusion, Quebec, QC, Canada, 2007, pp. 1-8, doi: 10.1109/ICIF.2007.4408086.
2. Korableva O.N., Kalimullina O.V., Mityakova V.N. (2019) Designing a System for Integration of Macroeconomic and Statistical Data Based on Ontology. In: Arai K., Bhatia R., Kapoor S. (eds) Intelligent Computing. CompCom 2019. Advances in Intelligent Systems and Computing, vol 998. Springer, Cham. https://doi.org/10.1007/978-3-030-22868-2\_12