**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 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 develop an ontology that will store integration configurations as well as 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”[2] a approach to the integration of heterogeneous data sources with a local to global ontology mapping is explored. In the work [1] 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 picture 1, the ontology contains such main classes as .....

This will allow performing data preprocessing before it gets inserted into the target ontology.

In the work a

**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. 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
2. 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.

The article describes a process of designing software for the aggregation of data (macroeconomic and statistical indicators) from distributed heterogeneous sources and their analysis based on the previously developed ontology of innovation activity and economic potential. The software includes a data aggregation system (supporting the user’s markup process of PDF, HTML, and XLS documents and texts for further automated collection), an ontology automatic replenishment system and a system of semantic search for data subsets according to certain criteria[1].

The semantic web is built on ontologies. And the ontologies are usually used to form knowledge systems. Unlike the semantic search that uses machine learning, the ontologies use manually declared links and relations between terms. Such an approach allows engineers to extract hidden knowledge from the data set but it also requires a lot of input work (ontology formation) before the system can spot anything useful [1].

The user uploads a table of information in a file of one of the following formats: xlsx, html, xml, pdf. The system parses the file in search of a table. When a table is found it is displayed and the user is prompted to map all the rows and columns in the table to classes and their properties in the target ontology respectively. Once the mapping gets done, the system inserts the data into the target ontology. The system automates the information insert and thus simplifies the process of information preparation before the analysis. But the system still requires repetitive manual work (preparation of the file, mapping so on).

The topic of integration of heterogeneous data is very crucial in the realm of data analysis. Many works are written on the approaches applicable to the integration of different ontologies. The overall approach is to design a global ontology.