UDC 004.82
https://doi.org/10.31891/csit-2022-3-2

VIKTOR SHYNKARENKO, LARYSA ZHUCHYI
Ukrainian State University of Science and Technologies

ONTOLOGICAL SUITABILITY ANALYSIS OF RAILWAY TRACKS FOR HIGH-SPEED TRAFFIC

High-speed railways are developing rapidly in the world and, in particular, in Europe, and are promising in the territory of Ukraine. The purpose of the work is to check the consistency of the existing railway characteristics with the high-speed rail transport regulations by ontological means. The ontology was developed in the Protégé and MRcube software environments in the OWL DL profile using the Extreme methodology. Ontology population is automated in OpenRefine and INCEpTION based on texts and tables semantic annotation. Data extraction is done in Tabula for tabular data resources and FineReader for drawings. The ontological support was developed according to the model of multi-level concretization employing a distinction between data and rules modules. The ontology of resources is based on a tabular knowledge representation model. The usability of the proposed approach is experimentally evaluated on the example of the sections of the «Mariupol Port-Volnovakha» railway in Ukraine, «Plovdiv-Burgas» in Bulgaria and «Coswig-ESig G» in Germany and the corresponding regulations. The ontology can be used for conventional railways (e.g., 160 km/h), as well as further integration with the railway train model can be done. The ontology was evaluated by the domain expert and using the OntOlogy Pitfall Scanner. The scientific novelty of the work lies in the integration of heterogeneous resources of railway infrastructure characteristics by ontological means. The semantic annotation and formalization of railway regulations have been employed. Railway transport data extraction and validation processes were improved. The practical significance of the work lies in the possibility of improving the safety of train traffic by checking the suitability of railway tracks to ensure the declared speeds of high-speed and conventional railways.

Keywords: conceptual modelling, knowledge bases, ontology, OWL, high-speed railway transport, railway

ВІКТОР ШИНКАРЕНКО, ЛАРИСА ЖУЧИЙ
Український державний університет науки і технологій

ОНТОЛОГІЧНИЙ АНАЛІЗ ПРИДАТНОСТІ ЗАЛІЗНИЧНИХ КОЛІЙ ДЛЯ ЗАБЕЗПЕЧЕННЯ ВИСОКОШВИДКОГО РУХУ ПОЇЗДІВ

Високошвидкісні залізниці розвиваються високими темпами у світі та, зокрема, Європі та є перспективними на території України. Мета роботи – перевірка узгодженості характеристик існуючих залізничних колій із нормативним забезпеченням високошвидкісного залізничного транспорту онтологічними засобами. Онтологія розроблена у програмних середовищах Protégé та MRcube у профілі OWL DL з використанням методології Extreme. Підготовка екземплярів онтології виконана автоматизованим чином на основі семантичного анотування текстів та таблиць у OpenRefine та INCEpTION. Витягування даних виконується в Tabula для табличних джерел даних, і FineReader – для креслень. Розробка онтологічного забезпечення виконана згідно з моделлю багаторівневої конкретизації з розмежуванням модулів даних та правил. Онтологія джерел заснована на моделі табличного представлення знань. Експериментально підтверджено придатність запропонованого підходу на прикладі ділянок залізниці «Маріуполь Порт-Волноваха» України, «Plovdiv-Burgas» Болгарії та «Coswig-ESig G» Німеччини та відповідного нормативного забезпечення. Онтологія може бути використана і для звичайних залізниць (наприклад, 160 км/год), і також в подальшому інтегрована з поїзною моделлю. Здійснено оцінку онтології експертом у предметній галузі та засобами OntOlogy Pitfall Scanner. Наукова новизна роботи полягає в інтеграції різнорідних джерел опису залізничної інфраструктури онтологічними засобами. Виконано семантичне анотування та формалізація залізничного нормативного забезпечення. Удосконалено процеси витягування та перевірки даних на залізничному транспорті. Практична значимість роботи полягає у можливості підвищення безпеки руху поїздів шляхом перевірки придатності залізничних колій для забезпечення заявлених швидкостей високошвидкісних та звичайних залізниць.

Ключові слова: концептуальне моделювання, бази знань, онтологія, OWL, високошвидкісний залізничний транспорт, залізниця

Introduction

High-speed railways are developing rapidly in the world and, in particular, in Europe. Among the changes that have undergone information support of trans-European and national transport systems, one should highlight the introduction of ontological support, which is used to integrate data and check their consistency. One example is the Rail Core Ontology [1], which allows the integration of heterogeneous sources, such as train schedules and maps within the UK railway. Using a different transport ontology, iCity, one performs data integration, such as vehicle routes and incident reports [2].

In the authors' previous works, an ontological approach was studied for formalizing and integrating data on the permitted speeds of trains, taking into account the condition [3] and the characteristics of the tracks [4] belonging to the lines from the first to the seventh category according to the State Building Norms (conventional railway) [5], presented in heterogeneous data sources: regulations, infrastructure descriptions and related calculations. They are presented in the formats of natural language texts and drawings.

Here, taking into account the trends in the development of high-speed railways, an experimental check is made on the consistency of regulatory documents for high-speed transportation and data of information systems on the example of sections of the «Mariupol Port-Volnovakha» railway of Ukraine (belongs to networks [6, 7]), «Plovdiv-

Burgas» in Bulgaria and «Coswig-ESig G» (belongs to the Berlin-Dresden section) in Germany by ontological means. We have taken into account both national and EU regulations.

Problem statement and purpose

We have developed an ontology for checking the speed of trains, including the frog type of railway switches and rail types concepts of conventional Ukrainian railways [4]. In this paper, the approach is extended taking into account the curves and gradients of high-speed railways and data of EU countries upgrading railways to increase the speed of trains.

Such a country is Bulgaria, for example, where the section Plovdiv-Burgas is being reconstructed to increase the speed of trains up to 160 km/h [8] and the section of the German railways Berlin-Dresden – up to 200 km/h. The section of the railway line adjacent to the city of Dresden currently is designed for a maximum speed of 120-160 km/h and needs further reconstruction [9].

In Ukraine, the railway infrastructure is designed for 160 km/h traffic, and the high-speed railway network is modelled in [6, 7], where its rationale is scientifically established.

The purpose of the work is to check the consistency of Ukraine and the EU countries' railway infrastructure description data with the relevant high-speed transportation regulations based on ontological means, which helps to improve the safety of train traffic.

Achieving this goal involves the following tasks:

- development of a multi-level ontology of the railway infrastructure;
- formalization of high-speed railway transportation regulations;
- railway information systems data integration;
- checking the consistency of the developed model in Protégé.

Related works

In 1958 under the Treaty of Rome, Trans-European Transport Networks were created to support economic development. Within the framework of the approach, the national infrastructure should be considered as part of the European one. This is achieved, for example, by integrating information systems and regulation data.

Data is integrated using the developed uniform data models, such as Network Statement, railML, RailTopoModel, and Register of Infrastructure (RINF). The next logical step in the natural progression of XML (railML and RINF technology) is railway transport ontologies, for example [10, 11]. A notable example is Rail Topology Ontology [12] based on RailTopoModel and enriching the UML model with description logic axioms.

To integrate the railway regulations of various countries, for example, CIM/SMGS consignment notes have been developed, which include the freight regulation, and EU TSI to describe the restrictions of railway track and other subsystems that ensure their interoperability. TSI is a regulation developed for all EU countries. Such centralization has the disadvantage that the law does not cover all aspects of rail transport, i.e. presented only in national regulations.

Another way to integrate regulations is their formalization. For example, Commission Regulation (EU) No 1299/2014 [13] and other EU documents are distributed in the Government gazette format. XML allows one to check the validity of legal documentation, including electronic signatures [14]. Its own XML publishing schema has been developed for European Union Official Journal. There are other schemes, such as AKOMA NTOSO and MetaLex. They can be enriched with LKIF ontology rules [15] for logical reasoning. For example, using the RDF dictionary LEXML it is possible to compare the legal concepts of different languages [16].

The ontology [17] includes the concepts of transport infrastructure in terms of catenary structure and trusses for their connection. A-box is the characteristics of the assembly and the coefficients determined by regulatory documents. The purpose of the ontology is to check the assembly structure consistency with the enterprise regulations. Stiffness calculation formulas are formalized using SWRL rules and associate ontology instances with a «true» literal, such as the relation mast_ isHeightValid.

Another way to formalize regulations is semantic annotation. The texts of normative documents are integrated by employing annotation so that experts search for restrictions by subject matter using SPARQL queries [18].

Thus, it is possible to supplement the integration of heterogeneous sources of railway infrastructure description with formalized transportation process regulations of various countries to check their consistency.

Methods

Consistency of information systems data and regulatory documentation is checked in a theoretical and experimental way. The experiment is carried out according to the plan:

- data preparation:
  - national railway regulations:
    - Germany Railways-Construction and Operating Regulations (Eisenbahn-Bau-und Betriebsordnung (EBO)) in PDF format [19];
    - Bulgaria regulations on the design and construction of railway lines in MS Word format [20];

- Ukraine State Building Norms (SBN) in PDF format [5];
- EU directives in PDF format:
  - Commission Regulation (EU) No 1299/2014 [13];
  - Directive 2008/57/EC [21];
- infrastructure descriptions:
  - railML files in the XML format that include the radii of the «Coswig-ESig G» section railway curves of the German railways [22];
  - Network Statement tables in PDF that include the gradients of the «Plovdiv-Burgas» section of the Bulgarian railway [23], as in [24];
  - the railway tracks longitudinal profile drawing in PDF format, the include radii of track curves and gradients of the «Mariupol Port-Volnovakha» section of Ukrainian railways [25], as in [26];
- calculation of the permitted radius in the XLS format;

- ontological support development:
  - ontology schema modules according to Fig 1;
  - data transformation into ontology instances;
  - import into the ontology of all developed modules according to the procedure in Fig 1;
- experiment execution:
  - SPARQL query results in infrastructure elements having characteristics inconsistent with the regulation restrictions for the expected speeds;
- analysis of results:
  - ontology evaluation with OntOlogy Pitfall Scanner (OOPS!) [27];
  - performed results adequacy assessment by the domain expert [28];
  - calculation of the Accuracy indicator [29] of the developed ontology.

The ontology is developed in a modular way, which facilitates its reuse [28] and subsequent maintenance, and also allows one to check the consistency of one data set with the rules of various regulations, for example, to determine the applicability of speed of trains for the Ukrainian railway network tracks separately for more and less than 200 km/h, which complies with State Building Norms regulations or EU directives. Or, on the contrary, check the consistency of the rules for the high-speed train traffic and both the Bulgarian and the German track data. A similar check is performed, for example, in the Network Statement Checker Ontology [24], which was developed to determine the compatibility of the technical characteristics of the railway tracks of different countries and international trains.

Modules correspond to three levels of abstraction (Table 1). The numbering of the ontologies in Figure 1 and Table 2 is consistent. The abstract level is represented by RDF and OWL dictionaries, lightweight ontologies containing class and relationship concepts for describing data sources and railway infrastructure, such as «table», «has part», «railway» and so on. The first one was developed using the tabular knowledge representation model [30], which is based on the application of an order relation and allows modelling a spectrum of data structures such as «set», «tuple», «class» and «table».

Table 1

Resources and railway infrastructure ontology modules according to their level of abstraction

Level	Ontology	
	Resources	**Railway infrastructure**
Abstract level	1	5
The first level of concretization	2a, 2b, 3a, 3b	6a, 6b, 7a, 7b
The second level of concretization	4a, 4b	8a, 8b

At the first level of specification, the vocabulary is enriched with data or rules, for example, restrictions on the gradient and radius of curves. In the second case, another part of the heavyweight ontology scheme is formed, which later will allow one to make inferences in Protégé to obtain new knowledge, as well as check the consistency of information systems data and regulations.

At the second level, rules and data are combined, for example, the Commission Regulation (EU) No 1299/2014 is applied to the data describing the railway infrastructure of Ukraine, Bulgaria and Germany. Thus, a knowledge base is formed that allows checking the compliance of the railway track of Ukraine or the EU countries for the movement of trains at a speed of 200 km/h in Protégé.

Concrete resources model ontologies that include data are developed separately for each type of table in the form of two modules 2a, and 2b. First, the table header and the script for its transformation are described manually. Then they are linked and the reasoner classifies the table according to its columns and the development software. If the ontology is consistent, a one uses SPARQL query to get a script that is imported along with the vocabulary and a table into OpenRefine to transform the table's data into ontology instances.

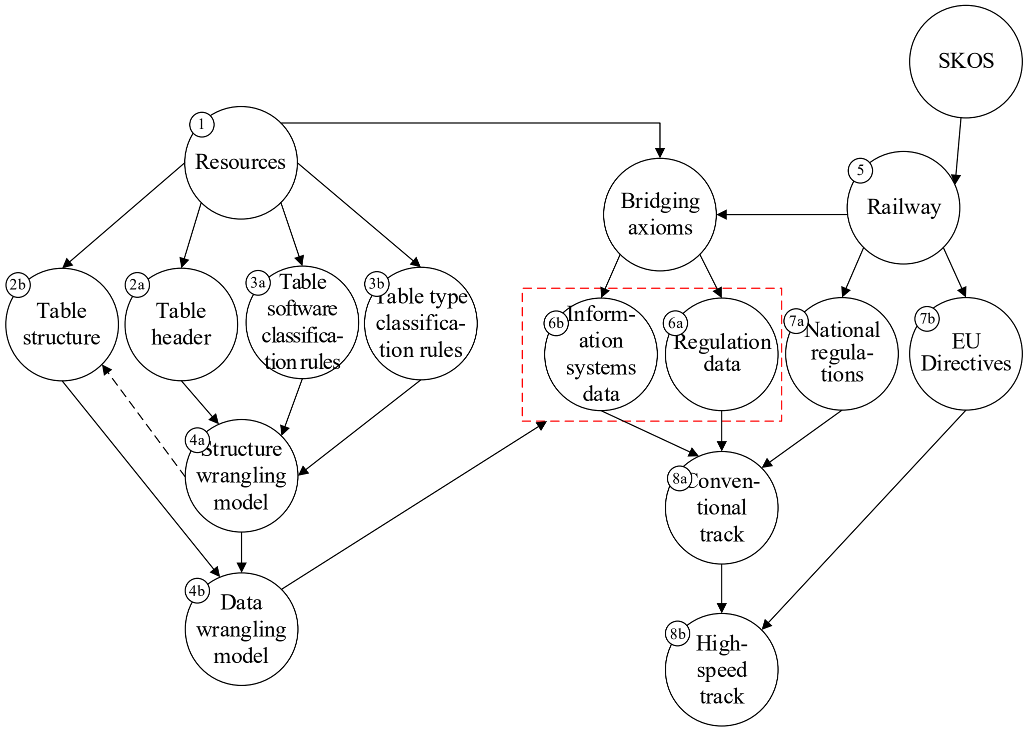


Fig 1. Modularity of the railway infrastructure ontology

In general, the experiment plan corresponds to the extreme ontology development methodology [28] with the stages of ontology development and testing using a SPARQL query. Therefore, the following article is structured as follows. The data preparation procedure is described. Next, each of the ontology modules is developed in the extreme order [28] (module, testing). The experiment ends with an evaluation of the results and conclusions.

Data preparation

The experiment is carried out using real information systems and regulations data obtained from open sources using the tools of

Table 2. Data is extracted using such methods as semantic role labelling for texts in Inception, optical character recognition for drawings in FineReader, and data extraction for PDF tables in Tabula.

Table 2

Tools used to prepare and develop resources data and ontology

Purpose of use	Software tool
The resources vocabulary development	mrCube
Heavyweight ontology and railway infrastructure vocabulary development and execution of SPARQL queries	Protégé
Semantic annotation of regulation texts	INCEpTION [31]
PDF table data extraction	Tabula
XSL, TSV, CSV format data wrangling to RDF	OpenRefine
Radius calculation	Excel
PDF files of railway section profiles of Ukraine recognition	FineReader

Semantic annotation is performed using tags based on the railway infrastructure ontology vocabulary. The input data of Inception are tags and regulations texts, and the output data is a TSV table suitable for its transforming into ontology instances in OpenRefine. Annotation allows one to link the regulation text and ontology and is performed using the concepts of the abstract railway infrastructure model ontology (Fig. 2).

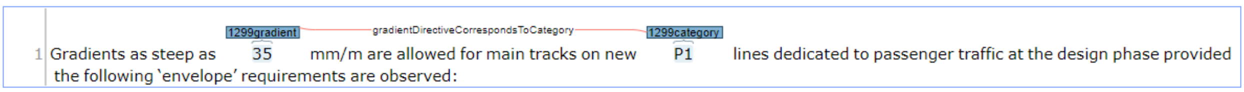


Fig. 2 Commission Regulation (EU) No 1299/2014 semantic annotation

Part of the data was prepared manually. The table for calculating the permitted radius values of German and high-speed Ukrainian railways according to the EBO formula (1) [19] is developed in Excel (

Table 3).

$$R = \frac{k \cdot V^2}{h + \Delta h} \tag{1}$$

where h – the cant – 180 mm [13], Δ h – the cant deficiency – 153 mm for 1435 mm railway track gauge [32] and 163 mm for 1520 mm [30], k – the coefficient related to the railway track gauge – 11.8 for 1435 mm track gauge [19] and 12.5 for 1520 mm [33], V – the speed of trains, R – the permitted curve radius.

Table 3

Calculation of permitted curve radii according to the railway train speed

R, m	k	V, km / h	h, mm	Δh, mm
908	11.8	160	180	153
1149	11.8	180	180	153
1418	11.8	200	180	153
1458	12.5	200	180	163

The resources model ontology implementation

Conceptual modelling of the abstract resources model ontology is developed in the form of an RDF vocabulary that unites table description concepts, infrastructure data and regulations.

The ontology includes the concepts of a table, a tuple, a value, an identifier, and scripts for transforming tables, as well as the hasPart and hasElement relationships for associating the table with the tuple and the tuple with the value. The description of the table structure allows one to determine the applicability of OpenRefine scripts for its processing.

The value is referenced by specifying the order relation between the elements of the tuples. For example, the radius calculation Table 5 has columns «r», «k», «v», «h», «deltah», and the following order relations have been developed for it: rBeforeK, kBeforeV, vBeforeH, hBeforeDeltah. This makes it possible to map the resources and railway infrastructure ontology, for example, to link values of railway tracks and gradients.

The concrete resources model ontology is generated as twelve RDF modules in OpenRefine.

The tables are represented in regulations, documents describing the infrastructure, and annotations of texts. Data on the permitted high-speed and conventional railways radius of Ukraine are represented in the tables of the SBN «The value of the smallest radius of curves» and «Indicators in the category of rail lines» [5], and on the categories of lines on the railways of the European Union – in the Commission Regulation (EU) No 1299/2014 [13] table «Performance parameters for passenger traffic».

Transformation of SBN tables [5], Excel calculations, railML, Network Statement, longitudinal track profile, Commission Regulation (EU) No 1299/2014 [13] and regulation annotations means that the tables are annotated with concrete resources model ontology concepts. OpenRefine additionally generates columns with a checked speed for all railway track description tables, as well as identifiers for the «Mariupol Port-Volnovakha» section profile elements of the Ukrainian railway.

The ontology module is tested on the example of the Bulgaria «Plovdiv-Burgas» section. After exporting data from OpenRefine and importing it to Protege, for example, the query «what values are in the «Interstation Section» column?» is done (Fig. 3).

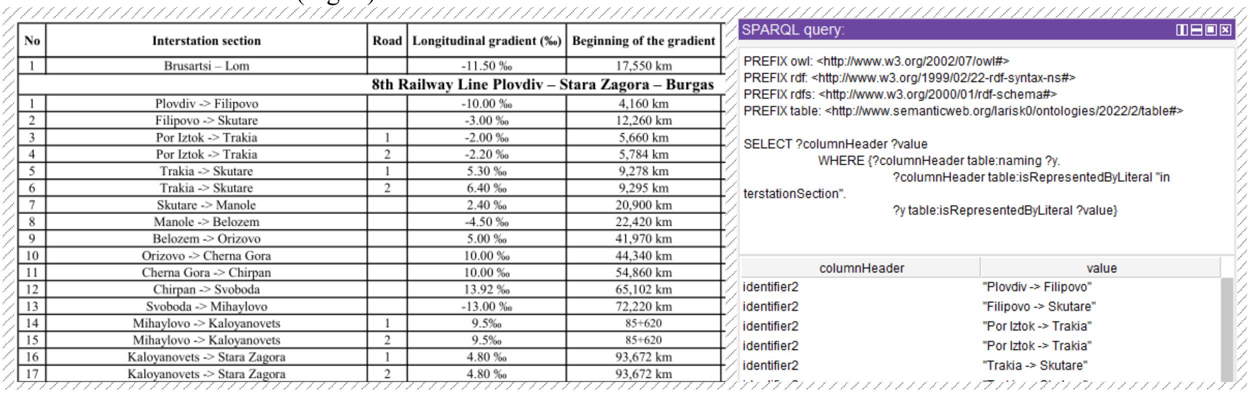


Fig. 3. Resources ontology of «Plovdiv-Burgas» section Bulgaria Network Statement «SE NRIC Maximum gradients along the railway lines» table SPARQL query

The concrete resources model ontology, which includes rules, is developed in the form of three OWL ontologies in Protege.

Tables are classified according to the features of the header tuple and development software. The hierarchy is built by logical definitions (for example, for a table for calculating the permitted curve radii):

```
calculationRadiusTable Equivalent To excelTable and hasPart some
calculationRadiusTableHeader.
```

The header tuple is a table tuple whose elements are identifiers. For example, calculationRadiusTableHeader has the following logical definition:

`calculationRadiusTableHeader Equivalent To tuple and (hasElement some RIdentifier) and (hasElement some kIdentifier) and (hasElement some vIdentifier) and (hasElement some hIdentifier) and (hasElement some deltahIdentifier) and (hasElement some gaugedentifier)`

If a table has a unique header tuple, it can be classified by the reasoner. Text annotation tables do not have a unique header tuple and should be classified manually.

The applicability of the OpenRefine script to a table is determined by the format and structure of the table and is implemented in the ontology by restrictions. For example, the curve radii calculation table data transformation script has the following restriction:

`calculationRadiusTableScript SubClass Of inverse inputOf only calculationRadiusTable.`

The concrete resources model ontology is developed in Protégé, where script and table ontology instances are linked. The reasoner checks the applicability of the script for processing the table using the restriction. If the ontology is consistent, for example, the «which tables are the input data of the scripts?» query is executed.

The ontology module is tested on the example of the SBN table «Indicators of the railway line categories» [5]. Fig. 4 shows a query to infer the type of table.

Queries on inferred facts are performed in the Protégé DL Query tab or Snap SPARQL Entailment Regime one [34] which is used in the paper.

Bridging ontology

Bridging ontology is developed in Protégé as a link between resources and railway infrastructure ontologies. The following requirements have been developed for the ontology:

reclassification of instances from resources ontology into railway infrastructure classes;

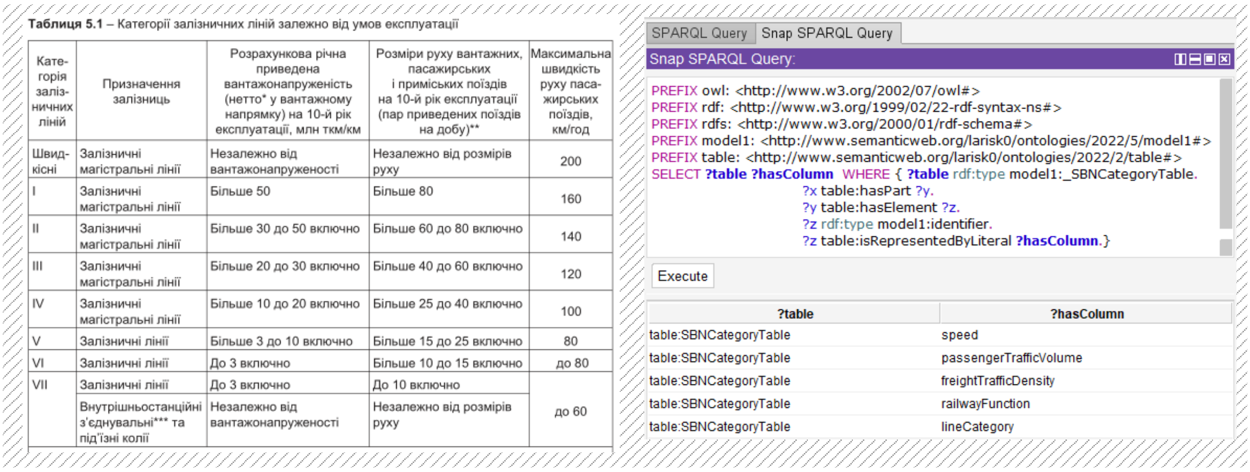
linking instances with railway infrastructure ontology relations.

The resources railway infrastructure ontology classes are connected by logical definitions like:

`trackSection Equivalent To valueClass and (inverse naming some radiusChangeIdIdentifier or inverse naming some interstationSectionIdentifier or inverse naming some profileColumn2Identifier or inverse naming some profileGradientColumn2Identifier)`

To link the characteristics of the railway track section, property chains like the following one (radii calculating table) have been developed:

`radiusRegulationCorrespondToSpeed SuperProperty Of rBeforeK o kBeforeV`



Таблиця 5.1 – Категорії залізничних ліній залежно від умов експлуатації

Категорія залізничних ліній	Призначення залізниць	Розрахункова річна приведена вантажонапруженість (нетто* у вантажному напрямку) на 10-й рік експлуатації, млн ткм/км	Розміри руху вантажних, пасажирських і приміських поїздів на 10-й рік експлуатації (пар приведених поїздів на добу)**	Максимальна швидкість руху пасажирських поїздів, км/год
Швидкісні	Залізничні магістральні лінії	Незалежно від вантажонапруженості	Незалежно від розмірів руху	200
I	Залізничні магістральні лінії	Більше 50	Більше 80	160
II	Залізничні магістральні лінії	Більше 30 до 50 включно	Більше 60 до 80 включно	140
III	Залізничні магістральні лінії	Більше 20 до 30 включно	Більше 40 до 60 включно	120
IV	Залізничні магістральні лінії	Більше 10 до 20 включно	Більше 25 до 40 включно	100
V	Залізничні лінії	Більше 3 до 10 включно	Більше 15 до 25 включно	80
VI	Залізничні лінії	До 3 включно	Більше 10 до 15 включно	до 80
VII	Залізничні лінії	До 3 включно	До 10 включно	до 60
	Внутрішньостанційні з'єднувальні*** та під'їзні колії	Незалежно від вантажонапруженості	Незалежно від розмірів руху	


```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX model1: <http://www.semanticweb.org/larisk0/ontologies/2022/5/model1#>
PREFIX table: <http://www.semanticweb.org/larisk0/ontologies/2022/2/table#>
SELECT ?table ?hasColumn  WHERE { ?table rdf:type model1:_SBNCategoryTable.
                    ?x table:hasPart ?y.
                    ?y table:hasElement ?z.
                    ?z rdf:type model1:identifier.
                    ?z table:isRepresentedByLiteral ?hasColumn.}
```

?table	?hasColumn
table:SBNCategoryTable	speed
table:SBNCategoryTable	passengerTrafficVolume
table:SBNCategoryTable	freightTrafficDensity
table:SBNCategoryTable	railwayFunction
table:SBNCategoryTable	lineCategory

Fig. 4. SPARQL query inferring the SBN table «Indicators of the railway line categories» type

The railway infrastructure model ontology implementation

The abstract railway infrastructure model ontology is developed in Protégé as a vocabulary for describing the data of information systems and regulations.

The conceptual model – ontology base – contains the concepts of a railway track section, curves, gradients, as well as relationships for checking the compliance of characteristics defined in information systems and regulatory documents, such as radiusMayCorrespondToSpeed and speedRegulationCorrespondToRadius.

The concrete railway infrastructure model ontology, which includes data, is developed in the form of twelve modules in Protégé for each type of table. A concrete sources model and bridging ontologies are imported into the module.

The railway infrastructure vocabulary and bridging ontology modules were tested on the German «Coswig-ESig G» railway section. The SPARQL query «what curve radii does the track have?» is executed (Fig. 5).

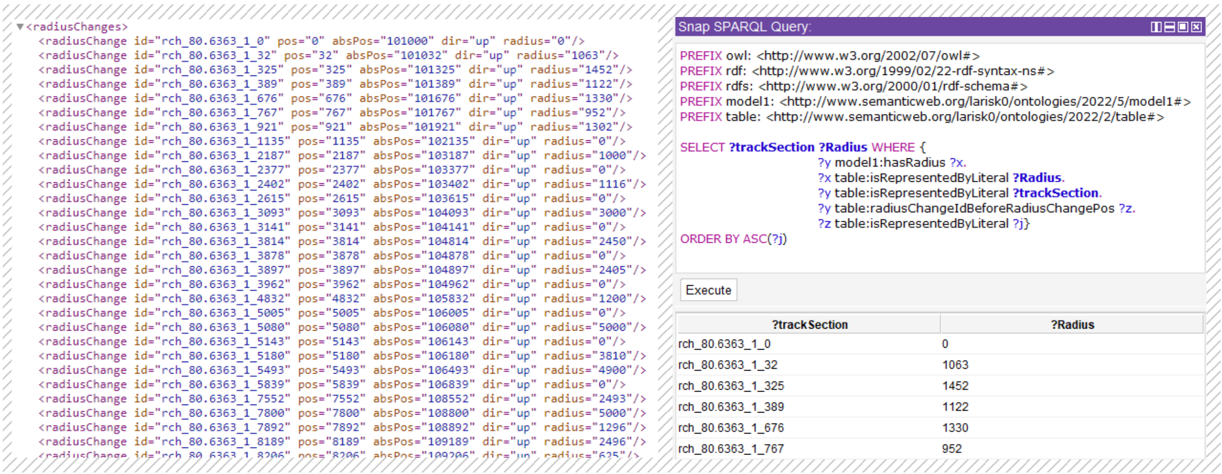


Fig. 5. SPARQL query inferring curve radius of the Germany «Coswig-ESig G» railway section

The concrete railway infrastructure model ontology, which includes rules, is developed in Protégé based on [35].

We utilize owl:hasKey construct for speedIsRepresentedByLiteral, categoryIsRepresentedByLiteral, wordIDIsRepresentedByLiteral data properties. Thus the relations become inverse functional and, for example, all speed values having the same value are connected with the relation owl:sameAs by the reasoner, which allows using Fig. 6 compositions of relations. Relation numbering indicates the execution order of the reasoning path.

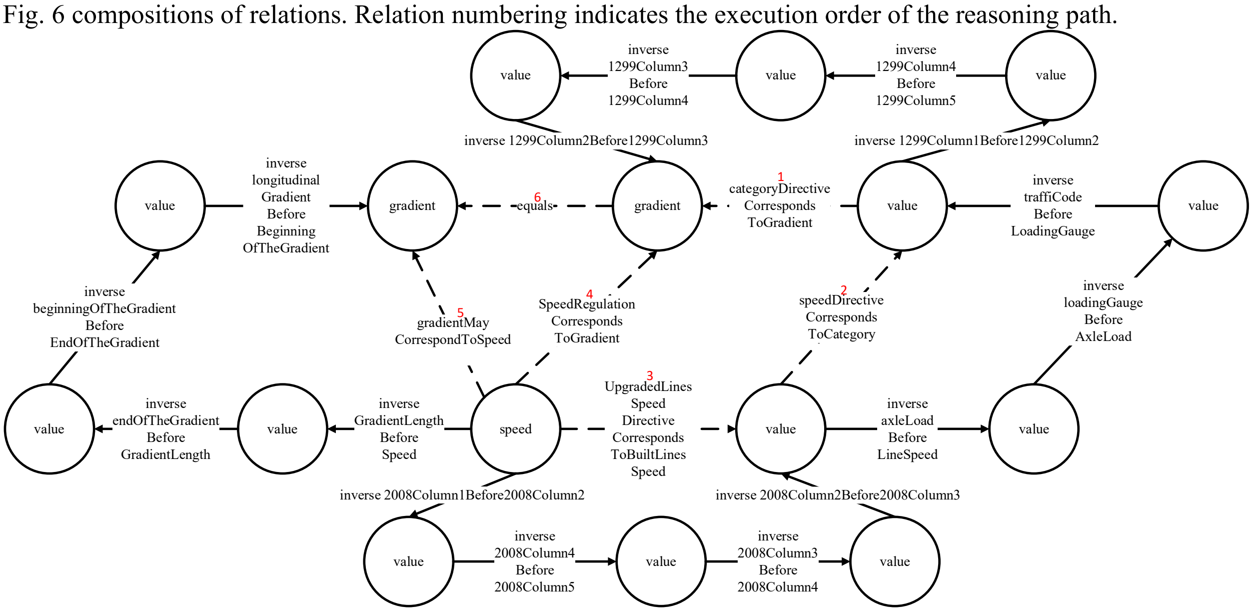


Fig. 6. Property chains that allow checking the consistency of regulations data and information systems

Relationship compositions link the annotation tables of Directive 2008/57/EC [21], Commission Regulation (EU) No 1299/2014 [13] and the Commission Regulation (EU) No 1299/2014 «Performance parameters for passenger traffic» table, since the answer to the question about permitted gradients is contained in statements about the correspondence of the track gradient and category, the track category and speed, as well as newly built and reconstructed lines speeds.

`Speed Directive Corresponds To Gradient SuperProperty Of upgraded Lines Speed Directive Corresponds To Built Lines Speed o speed Directive Corresponds To Category o category Directive Corresponds To Gradient`

The infrastructure gradient description and the regulation data are connected by the «equals» relation property chain:

`equals SuperProperty Of gradient May Correspond To Speed o inverse speed Regulation Correspond To Gradient`

After that, they are compared for belonging to the same range of values.

The concrete railway infrastructure model ontology is developed in Protégé. Reasoner checks the consistency of the speed and characteristics of the track using logical definitions.

The ontology module was tested on the «Mariupol Port-Volnovakha» section example. Using a SPARQL query, railway track curve radii are obtained that do not meet the high-speed traffic restrictions.

In Fig. 7 a) for the «Mariupol Port-Volnovakha» section, curves that do not meet the 200 km/h speed traffic restrictions were obtained. It is possible to use the ontology for speeds of 200 km/h and conventional railways (for example, 160 km/h). In Fig. 7 b), curves were obtained that do not meet the 160 km/h restrictions.

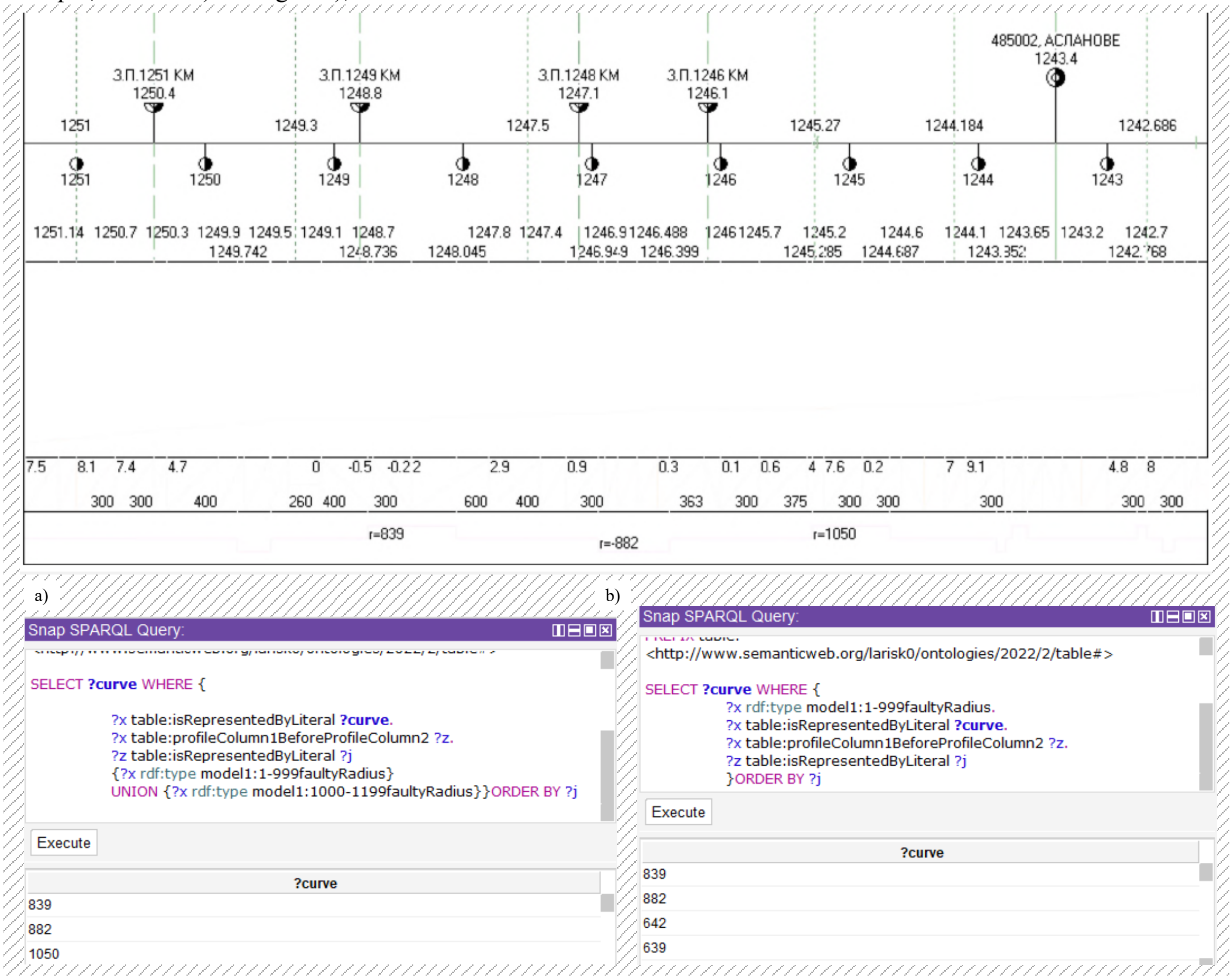


Fig. 7. SPARQL query inferring «Mariupol Port-Volnovakha» section track curves that are not consistent with 200 km/h and 160 km/h traffic restrictions

The consistency of infrastructure description data and regulations is checked using high-speed railway radius calculation tables. The following assumptions are made. Since there is no rolling stock in Ukraine of 200 km/h speed, traffic of 180 km/h is considered high-speed, in contrast to that specified in the SBN [5]. High-speed traffic is considered to be from 200 km/h for reconstructed lines [21]. Track characteristics calculated for the 1540 mm gauge [32, 33].

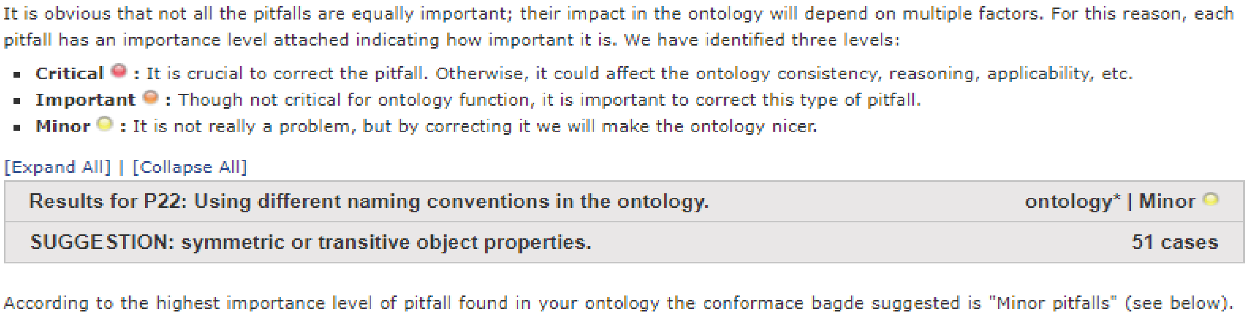
Checking the consistency of infrastructure description and regulations data is performed on the basis of SBN «Value of the smallest radius of curves» and «Indicators of the railway line categories» tables [5]. The comparison is performed according to the column of «very difficult conditions».

Ontology evaluation

Evaluation – the last stage of ontology development - is performed in two ways: manually by a domain expert and automatically by a web application.

Ontology quality is evaluated using the OOPS (OntOlogy Pitfall Scanner!) application [27], as, for example, in [36]. This method helps to check the ontology for the presence of bad practices (40 in total), like structural, functional and usability-related, categorized as critical, important and minor. For example, facts linking object properties with an "inverse" relation to themselves are unacceptable, or the absence of annotations that, in the general case, are used to improve the readability of the ontology by a person.

After importing into the ontology scanner, minor flaws (Fig. 8) were identified, presumably related to the presence of dots in the value identifiers, for example, «value2.1».


It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:

- **Critical** : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important** : Though not critical for ontology function, it is important to correct this type of pitfall.
- **Minor** : It is not really a problem, but by correcting it we will make the ontology nicer.

[Expand All] | [Collapse All]

Results for P22: Using different naming conventions in the ontology.	ontology* | Minor
SUGGESTION: symmetric or transitive object properties.	51 cases

According to the highest importance level of pitfall found in your ontology the conformace bagde suggested is "Minor pitfalls" (see below).

Fig. 8. Evaluation of the developed ontology in OntOlogy Pitfall Scanner

The fulfillment of the development requirements by the ontology is determined using SPARQL test queries, test modules, followed by an assessment of the result adequacy by a domain expert, which corresponds to the eXtreme methodology [28]. We developed queries according to the competency questions set for the ontology modules. For example, one of the aims of source ontology module development is to determine the type of a table according to its content. The ontology usability is determined by the Figure 4 SPARQL query. To assess the adequacy of the SPARQL query results, we developed a questionnaire for a domain expert and examples of questions presented in Table 4.

Table 4

Domain expert evaluation questionnaire of the ontology

Statement	True/false
In Figure 3, the values belong to the «Plovdiv-Burgas» section «SE NRIC Maximum gradients along the railway lines» table «Interstation section» column	
In Figure 4 the table description corresponds to the SBN «Indicators of the railway line categories» one [5]	
In Figure 5, the «Coswig-ESig G» section railway track curve radii are got	
In Figure 7 the section «Mariupol Port-Volnovakha» curves having 839 m and 882 m radii are not consistent with restrictions of 200 km/h speed	
In Figure 7 the section «Mariupol Port-Volnovakha» curves having 1050 m are consistent with restrictions of 160 km/h speed.	

At the same time, regression testing [28] is performed, i.e., checking the usability of modules combined in one ontology. The developed ontology contains 16 modules imported into each other according to the procedure shown in Figure 1. Thus, each subsequent module should not contradict the previous one, for example, in naming and getting the results of SPARQL queries that correspond to the ontology development aims.

Axioms are evaluated using the Accuracy indicator in [29] as the ratio of the number of axioms with the false expert's assessment to the total number of axioms. The developed knowledge base contains about 3000 axioms and 500 individuals. It is impossible to assess the entire volume, so the expert evaluates the following part of the axioms.

Inference of the types of information system and regulation tables (SBN tables [5], Excel calculation, railML [22], Network Statement [23] and Commission Regulation (EU) No 1299/2014 [13]) as in Figure 4 – 6 axioms. Inferring the consistency of the «Mariupol Port-Volnovakha» section speed and railway track curve radii – 23 axioms, the «Plovdiv-Burgas» section speed and gradients – 50 axioms. Of these, all 79 axioms were evaluated as true by the domain expert.

Discussion

Existing transport ontological studies suffer from fragmentation as they do not consider heterogeneous data [24] or foreign railways data [1]. In transport, insufficient attention is paid to the railway infrastructure regulations and annotation of texts.

We carried out approbation of the tabular representation knowledge model and integration of information systems of various railway subsystems in [3, 4].

Further integration of the developed railway infrastructure ontology can be performed with the railway train model using the bridging axiom [35]:

```
modelTrain EquivalentTo: train and hasWeight some xsd:int[> 0, 3400] and
hasLocation some modelTrack.
```

Conclusions

Methods for checking the suitability of railway tracks for high-speed traffic by annotating and formalizing regulatory documents have been developed. The usability of the approach was evaluated experimentally for the Ukraine «Mariupol Port-Volnovakha», Bulgaria «Plovdiv-Burgas» and Germany «Coswig-ESig G» railway sections.

The ontology was developed according to the Extreme methodology using the OpenRefine automated generation of instances and the Inception semantic annotation. It was evaluated by OOPS and a domain expert. Within the paper framework, we transformed data of 3000 axioms and 500 individuals, of which 79 axioms were selected and evaluated by the domain expert as true.

It is worth mentioning that the developed ontology is a heavyweight one in that it contains restrictions, rules and logical definitions, which allows one for logical inferences to obtain new knowledge and check data consistency. Because of this, it is not suitable to work with large amounts of data, especially in environments like Protégé and the Hermit reasoner we use. When importing rules for large amounts of data into an ontology, it is possible to process them in parts, such as those used for demonstration in the paper. Nevertheless, the ontology is modular, which in our case means the presence of a developed lightweight version of the ontology (vocabulary), which is of use not for consistency checking, but for data integration, which we perform in [3, 4].

In the future, we plan to enrich the ontology with the railway track defects and integrate it with the railway train model.

References

1. Tutcher J., Easton J., Roberts C. Enabling Data Integration in the Rail Industry Using RDF and OWL: the RaCoOn Ontology. Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. 2017. DOI: https://doi.org/10.1061/AJRUA6.0000859
2. Katsumi M., Fox M. An Ontology-based Standard for Transportation Planning. International Workshop on Formal Ontologies Meet Industry. 2019.
3. Shynkarenko V., Zhuchyi L. Semantic Checking of Information Support for Heterogeneous Resources of Train Speed Restrictions by Ontological Means. *Semantic Technologies for Intelligent Industry 4.0 Applications*. 2022. In print
4. Verstichel S., Ongenae F., Loeve L., Vermeulen F., Dings P., Dhoedt B., De Turck F. Efficient data integration in the railway domain through an ontology-based methodology. Transportation Research Part C: Emerging Technologies. 2011. Vol. 19(4). P. 617-643. DOI: https://doi.org/10.1016/j.trc.2010.10.003
5. SBN B.2.3-19-2008. Railway transport facilities 1520 mm track design standards (29895). URL: shorturl.at/chQX3 (06.2022).
6. Kirpa G. N., Bosov A. A., Korzhenevich I. P. On the high-speed network on the railways of Ukraine. Science and progress of transport. 2004. Vol. 4. P. 103-109.
7. Бараш Ю. С., Момот А. В. Удосконалена методика визначення економічної ефективності будівництва та експлуатації високошвидкісної магістралі в Україні. Проблеми економіки транспорту. P. 7-24. 2014. DOI: https://doi.org/10.15802/pte.v0i6.23021
8. Shynkarenko V., Zhuchyi L. Semantic checking of permitted speeds information sources of different types in railway transport. Computational Linguistics and Intelligent Systems. 2022.
9. Plovdiv-Burgas Railway Line Phase Two Rehabilitation Project, Bulgaria. URL: shorturl.at/fjp18 (19.06.2022).
10. Deutsche Bahn 2020 Integrated Report. URL: shorturl.at/LRY57 (дата звернення: 19.06.2022).
11. Skalozub V., Ilman V., Shynkarenko V. Development of ontological support of constructive-synthesizing modeling of information systems. Eastern-European journal of enterprise technologies. 2017. Vol. 6. P. 58–69. DOI: https://doi.org/10.15587/1729-4061.2017.119497
12. Skalozub V., Ilman V., Shynkarenko V. Ontological support formation for constructive-synthesizing modeling of information systems development processes. Eastern-European journal of enterprise technologies. 2018. Vol. 5. P. 55–63. DOI: https://doi.org/10.15587/1729-4061.2018.143968
13. Bischof S., Schenner G. Rail Topology Ontology: A Rail Infrastructure Base Ontology. Semantic Web. 2021. 597-612. DOI: https://doi.org/10.1007/978-3-030-88361-4_35
14. Cyrul W. Computer aided legislation. Legal Tech. Seite. 129 – 140. 2021. DOI: doi.org/10.5771/9783748922834-129
15. LKIF-Core Ontology. URL: http://www.estrellaproject.org/lkif-core/ (19.06.2022).
16. Shelar A., Minal M. A Comparative Study to Determine a Suitable Legal Knowledge Representation Format. International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques. 514-519. 2018. DOI: https://doi.org/10.1109/ICEECCOT43722.2018.9001363
17. Carrasco C. G., Carballeira F. G. High performance computing techniques applied to the design of complex railway infrastructures. dis./ Universidad Carlos III de Madrid. Madrid, 2016. 142 p.
18. Mouromtsev D. I., Shilin I. A., Pliukhin D. A., Baimuratov I. R., Rezeda R. K. Building knowledge graphs of regulatory documentation based on semantic modeling and automatic term extraction. Journal Scientific and Technical Of Information Technologies, Mechanics and Optics. 2021. P. 256. DOI: https://doi.org/10.17586/2226-1494-2021-21-2-256-266
19. Eisenbahn-Bau- und Betriebsordnung (EBO). URL: shorturl.at/cDEY7 (19.06.2022).
20. Naredba № 55 ot 29 yanuari 2004 g. Za proektirane i Stroitelstvo na zhelezopatni linii, Zhelezopatni gari, zhelezopatni prelezi i drugi Elementi ot zhelezopatnata infrastruktura. URL: shorturl.at/FGHSU (19.06.2022).
21. Commission Regulation (EU) No 1299/2014 of 18 November 2014 on the technical specifications for interoperability relating to the 'infrastructure' subsystem of the rail system in the European Union Text with EEA relevance. URL: shorturl.at/ejuY2 (19.06.2022).
22. Directive 2008/57/EC of the European Parliament and of the Council of 17 June 2008 on the interoperability of the rail system within the Community (Recast) (Text with EEA relevance). URL: shorturl.at/hqsy9 (19.06.2022).
23. East Saxony railway network by FBS. URL: https://www.railml.org/en/user/exampledata.html (19.06.2022).
24. Maximum gradients along the railway lines of SE NRIC. URL: shorturl.at/hJZ67 (19.06.2022).
25. Donetsk railway track profiles. 2020. URL: shorturl.at/etwBS (19.06.2022).
26. A device for reducing the magnetic field strength in the gap of a rail insulating joint: Pat. 150269 Russia, MPK E 01 B 11/54, LLC NTC Information Technologies, No. 2014123170/11; Appl. 06/09/2014. Published 02/10/2015
27. María P.-V., Gómez-Pérez A., Suárez-Figueroa M. C. OOPS!(Ontology Pitfall Scanner!): An on-line tool for ontology evaluation. International Journal on Semantic Web and Information Systems. 2014. Vol. 10.2. P. 7-34. DOI: https://doi.org/10.4018/ijswis.2014040102
28. Blomqvist E., Hammar K., Presutti V. Engineering Ontologies with Patterns-The eXtreme Design Methodology. Ontology Engineering with Ontology Design Patterns. 2016. P. 23-50
29. McDaniel M., Storey V. C., Sugumaran V. Assessing the quality of domain ontologies: Metrics and an automated ranking system. Data & Knowledge Engineering. 2018. Vol. 115. P. 32-47. DOI: https://doi.org/10.1016/j.datak.2018.02.001
30. Shynkarenko, V., Zhuchyi, L., Ivanov, O. Conceptualization of the tabular representation of knowledge. Computer Sciences and Information Technologies (CSIT). 2021. Vol. 2. P. 1-4. DOI: https://doi.org/10.1109/CSIT52700.2021.9648761
31. Klie J. C., Bugert M., Boullosa B., de Castilho R. E., Gurevych I. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. International Conference on Computational Linguistics: System Demonstrations. 2018. P. 5-9.
32. Patlasov O., Patlasov Ye. Unification of the cant and maximum values for cant deficiency. Technologijos ir menas technology and art. 2016. Vol. 6. P. 111–115.

33. Gailienė I., Gedaminskas M., Laurinavičius A. Approach to rational calculation of superelevation in dual gauge track. Transport. 2018. Vol. 33(3). P. 699-706. DOI: https://doi.org/10.3846/transport.2018.1577
34. Horridge M., Musen M. Snap-SPARQL: a java framework for working with SPARQL and OWL. International Experiences and Directions Workshop on OWL. 2015. Vol. 9557. P. 154-165. DOI: https://doi.org/10.1007/978-3-319-33245-1_16
35. Shynkarenko V., Zhuchyi L. Ontological harmonization of railway transport information systems. Computational Linguistics and Intelligent Systems. Vol. 2870. P. 541–554.
36. Beden S., Cao Q., Beckmann A. SCRO: A Domain Ontology for Describing Steel Cold Rolling Processes towards Industry 4.0. Information. 2021. Vol. 12(8). P. 304. DOI: https://doi.org/10.3390/info12080304

Viktor Shynkarenko **Віктор Шинкаренко**	Doctor of Engineering Sciences, Professor of Department of Computer Information Technologies, Ukrainian State University of Science and Technologies, Dnipro, Ukraine e-mail: shinkarenko_vi@ua.fm https://orcid.org/0000-0001-8738-7225	Доктор технічних наук, професор кафедри Комп'ютерних Інформаційних Технологій, Український державний університет науки і технологій, Дніпро, Україна
Larysa Zhuchyi **Лариса Жучий**	assistant of Department of Computer Information Technologies, Ukrainian State University of Science and Technologies, Dnipro, Ukraine e-mail: larisa_zhuchiy@ukr.net https://orcid.org/0000-0002-9209-7262	Асистент кафедри Комп'ютерних Інформаційних Технологій, Український державний університет науки і технологій, Дніпро, Україна