International Journal of Advances in Computer Science & Its Applications – IJCSIA
Volume 4: Issue 3 [ISSN 2250-3765]
Publication Date : 30 September,2014

Lexical Ontology Model Based on SBVR Representations

Gintare Krisciuniene, Lina Nemuraite, Rita Butkiene, Bronius Paradauskas

***Abstract*— The paper presents a model of lexical ontology, based on SBVR representations, which is related to domain ontology used for semantic search in Lithuanian Internet corpus. The advantage of using SBVR based lexical ontology is the support of various relations among different types of meanings and representations, considering phrases instead of single words, and possibility of transformations to (and from) ontologies. We make an assumption that such ontology model is capable to encompass features contained in current lexical ontologies, and may be used in semantic search and Natural Language Processing techniques.**

***Keywords*— ontology, semantic roles, lexical ontology, SBVR, OWL 2, FrameNet, VerbNet, WordNet**

I. Introduction

The goal of the current research is to relate lexical and domain ontologies applied in semantic search over the Internet corpus using Semantics of Business Vocabulary and Business Rules (SBVR) [1]. The semantic search is possible after accomplishment of morphological, syntactic and semantic annotating of the searchable content. The purpose of the semantic annotating is to map text fragments with ontology concepts. The mentioned task is still a great challenge even for the English language. The problem is that the same meaning in texts can have a variety of representation forms. The simplest way for defining various representations of the meaning in ontologies is to use the annotation property "label". However, representations have their own structure, so a label, even if it is multivalued and has a tag for indicating a language, is not sufficient for expressing representations [2]. In the world practice, lexical resources as WordNet [3], VerbNet [4], [5], FrameNet [6], [7], or PropBank [8], are used for relating senses of words with their representations in semantic annotating and search.

In Lithuania, the similar problems previously have not yet been solved. For the Lithuanian language, we do not have such rich lexical resources as WordNet, etc. The attempts to translate these resources have revealed that differences between English and Lithuanian languages hinder to successfully use them.

Gintare Krisciuniene[1], Lina Nemuraite[2], Rita.Butkiene[3], Bronius.Paradauskas[4]
Kaunas University of Technology, Department of Information Systems &Centre of Information Systems Design Technologies
Lithuania

The experience of WordNet and other lexical databases is invaluable, though the structures of these lexical resources often are criticized, especially due to their overlap and mutual discrepancies. Whereas it is desirable to use each of them together, there have been mappings made between these resources [9], [10]. However, since we have to start everything from the beginning, we propose to use the SBVR for creating lexical resources in Lithuanian language. Unlike ontologies, meaning and representation in SBVR are separate primary concepts: every meaning may have several representations, and each expression may have several meanings. The SBVR metamodel is similar to what is encompassed by WordNet, VerbNet, FrameNet or other lexical ontologies where various syntactic forms are related to meaning. The advantage of SBVR metamodel is the support of various relations among different types of meanings and representations, considering phrases instead of single words, and possibility of transforming to (and from) ontologies.

The proposed lexical ontology in OWL 2 [11], based on SBVR representations, is related to domain ontology and may be used in semantic search and Natural Language Processing techniques. Other important problems, which are beyond the current paper, but can be considered further in relation with the SBVR and current research, are about relating semantic representations with linguistic information and integration with existing Semantic Web ontologies presented in other languages (mostly in English).

The rest of the paper is structured as follows. Section 2 analyses related work. Section 3 presents the model of SBVR based lexical ontology. Section 4 describes patterns for lexical structures representing different types of meaning. Section 5 summarizes conclusions and envisages future research.

II. Related Work

SBVR is the most expressive knowledge model intended to represent knowledge about business concepts and business rules. The SBVR specification "provides an unambiguous, meaning-centric, multilingual, and semantically rich capability for defining meanings of the language", targeted to "exchange of the meanings of concepts and business rules between humans and tools as well as between tools without losing information about the essence of those meanings" [1]. As it is based on First Order Logics with some extensions of higher and modal logics, it yet does not have tools for inference. Part of SBVR logics can be transformed into OWL 2 [12], [13]. SBVR and OWL 2 are two related specifications that can be used for semantic search, in which SBVR Structured English serves as the human interface for OWL 2 ontologies processed by computer programs. Semantic search is not the primary purpose of the SBVR; however, some recent research supports the assumption about SBVR applicability for that purpose [14], [15].

Currently, there is a lot of research aiming at transforming SBVR business vocabulary and business rules into domain ontologies. Some prototypes e.g., [12], [13] and commercial implementations as Collibra are developed for transforming SBVR business vocabularies and business rules into OWL 2, and SBVR questions into SPARQL queries [16]. Also, there are some works for obtaining SBVR vocabularies from existing domain ontologies [17], [18]. None of the analyzed works has considered the usage of SBVR representations for representing terminological knowledge in ontologies.

Possibilities to map terminological and linguistic information to ontology concepts are of great importance [19] and have been studied using simplified [20], [14] or complex [21] linguistic models. We argue that consideration of SBVR representations has sense for representing terminological information in ontologies because SBVR metamodel is capable to encompass and systematize capabilities of existing lexical ontologies as FrameNet, VerbNet, and WordNet.

VerbNet [4] is the largest hierarchical, domain independent verb lexicon for English, organized into verb classes, described by thematic roles, argument restrictions, and frames consisting of syntactic descriptions and semantic predicates. In the Verb Net, multiple types of information are combined into single thematic roles: actor, agent, asset, attribute, beneficiary, cause, location, destination, source, experiencer, extent, instrument, material, product, patient, predicate, recipient, stimulus, theme, time, topic [5]. VerbNet 3.2 contains 8537 verbs, explicitly linked to the WordNet synsets.

The Berkeley FrameNet [6], [7] is an on-line lexical resource for English, based on frame semantics and supported by corpus evidence. Word senses are grouped into conceptual structures, called "frames", which share certain semantic properties and are related with a meaning. FrameNet does not use generalized sematic roles but defines specific roles for each conceptual event or state [22]. It does not provide links between lexical entries and the frames, and does not contain selection restrictions [7].

The WordNet [3] covers approximately 150.000 words, nouns, verbs, adjectives and adverbs. Words are grouped into sets of cognitive synonyms (synsets), which express distinct concepts with their antonyms, hyponyms, hypernyms and other relations between concepts of the WordNet. It is a widely used resource for a variety of natural language processing tasks. VerbNet presents about 11.000 verbs, related to 24.632 senses. The polysemy of verbs is thus quite high, as each verb approximately has the 2.3 senses.

PropBank is composed of a verb lexicon and a semantically annotated corpus. The lexicon contains about 3600 verbs. Each verb is represented by a frame. Each frame is composed of one or more framesets that refer to specific verb senses. The PropBank contains 5050 framesets, accompanied with sets of semantic roles, identified by generic argument labels (Arg0, Arg1, …, ArgM). The mapping between roles and labels is role set specific, i.e., a label is usually assigned to different roles across the lexicon [105]. Arg0 and Arg1 constitute the 85 % of arguments, while Arg2−5 performance drops significantly.

The study of verb syntax and semantics is the prerequisite for semantic annotating, as in fact verbs convey the core meaning of sentences. The state of the art described in a sentence is expressed through its main verb concept. Nouns and noun phrases express participants of the state of the art, while adjectives and other parts of speech are used to better specify and describe them. Participants are playing certain semantic roles in certain situations, and use some linguistic encoding of those situations. An assignment of roles is a problem, as it is impossible to unambiguously define roles that concepts are playing in different contexts. There have been several proposals in semantic role classification and grouping. Semantic roles of FrameNet and VerbNet are used more consistently, but the definition of the roles is not given in a formal manner and their semantic characteristics are unclear.

A comparison of roles in FrameNet, PropBank, VerbNet can identify pros and cons for each. For VerbNet pros is that its level of generality produces many instances for each role and has connections to predicate-logic type semantic representations. Cons – needs the better coverage of verbs and verb senses, and the clearer definitions of thematic roles. FrameNet pros are the clear definitions of roles; the cons are the need for better coverage of verbs and verb senses, and the narrow roles that can result in a sparse data problem. For PropBank, pros are the easy application of argument labels and consistency of Arg0 and Arg1 across verbs. Cons are the difficulty of making generalizations across verbs due to verb-specific numbered arguments. Our purpose is to use the experience of creating lexical resources and at least partially avoid the identified difficulties of using them, by basing on the SBVR, which can provide the formal basis for creating such resources.

III. The SBVR Based Lexical Ontology

Our research aims to create SBVR based Lexical Ontology for Lithuanian language (SLOL) that would have capabilities to systematic linking of lexical constructs and meaning. The difference from analyzed lexical ontologies is, in semantic sense, that meanings in SBVR are organized around a conceptual model of a specific problem domain. The SBVR itself provides a framework for organizing meanings by defining the essential concepts as the "state of the art", "state", "activity", "event", etc. In lexical sense, the difference is in that SBVR terms, names and verb concept wordings include not only single words but may be compound, composed of several words. SLOL is concentrated on representation of SBVR noun concepts and verb concepts, and does not take separate adjectives and adverbs (that are not included in phrases) into account. Terms and names, representing noun concepts, may be single words and noun phrases; verb concept wordings are comprised of verb phrases and noun phrases, representing semantic roles of verb concepts. SLOL uses only

SBVR synonyms and synonymous forms; other WordNet relations (hypernyms, hyponyms, etc.) are covered in SBVR based domain ontology and are based on SBVR verb concept relations (categorizations, associations, property associations, partitive verb concepts, classifications, and characterizations). SLOL does not take into account antonyms as they are non-ontological relations.

Morphological and syntactic information is outside of SBVR representations but is indispensable in Semantic Web applications. Linguistic analysis is beyond the scope of our research but SBVR based lexical ontology should be related with linguistic information as well. In SBVR vocabulary, each case of noun in singular or plural form is treated as different term, i.e. “students” and “student’s” would be synonyms of their primary (preferred) representation. This is not an adequate approach, especially for the Lithuanian language, in which nouns have seven singular and seven plural cases, verbs have various tenses, etc. Currently, we are using the lemmatizing Web service in our SBVR editor [16], which allows avoiding this problem by entering nouns, verbs and their phrases once and recognizing them in various cases in verb concept wordings and business rule expressions.

A part of SBVR metamodel, extended with some concepts for SLOL, is presented in Figure 1. Actually, SLOL and domain ontologies are separate ones, pursuing the principles of modularity [23]. In SLOL, representations are individuals of ontology classes – terms for general concepts, names for individual concepts, verb symbols for verbs, verb concept role designations for verb concept roles, and placeholders for relating verb concept role designations or terms with verb concept wordings. As it is impossible to relate SLOL ontology individuals with classes of domain ontology, we have used the punning as ontology metamodeling technique [23].

I.e., for relating domain classes with representation individuals, the single individual must be created for each class of domain ontology, with the name, exactly corresponding to the name of the class it represents. Moreover, additional classes must be created for object and data properties with exactly matching names, along with individuals for each such class. The SLOL seems a little cumbersome; however, it allows relating representations with meaning concepts through matching names of domain concepts and individuals.

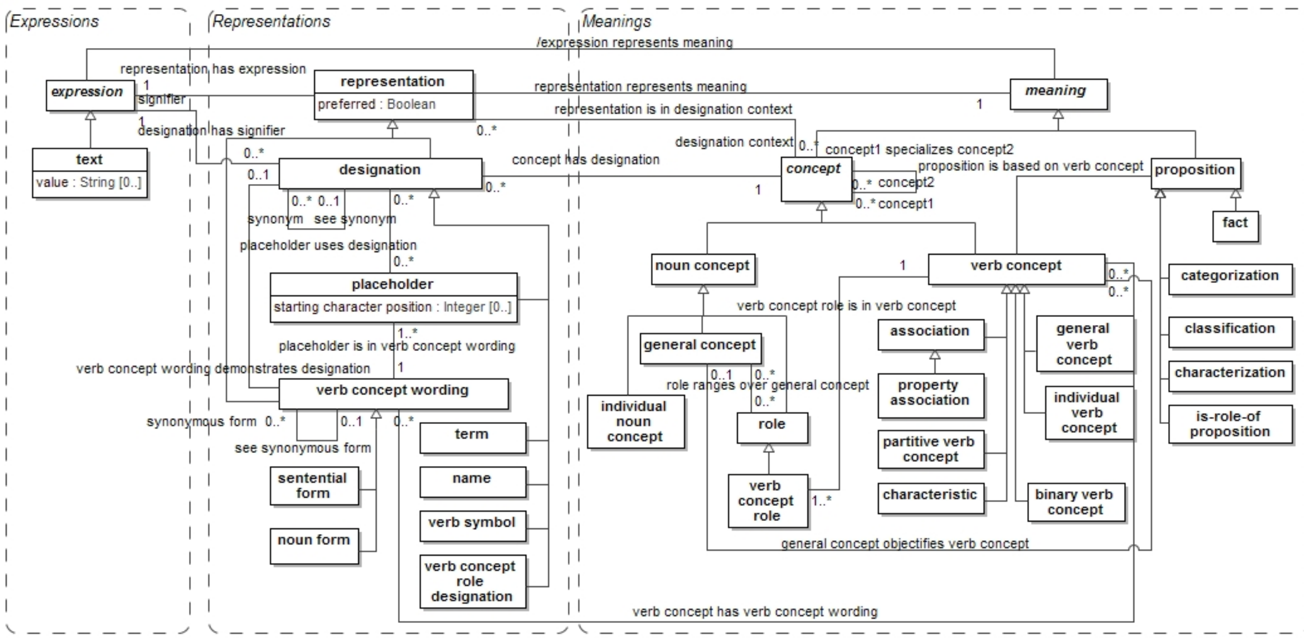

Figure 1. SBVR metamodel for SLOL, relating meanings, expressions and representations (adapted from[1])

IV. SBVR Based Lexical Patterns

For explaining SLOL constructions, we use the simple example of domain ontology (Figure 2). Let us consider the following concepts from domain ontology “dom”: general concept “person”; verb concept role “employee”, played by person in the context of verb concept “employee hold position”, role “time value”, and individual concept “Arnas Paukštė” (this is enough for representing main lexical patterns). Domain ontology constructs except individuals are presented in English language, SLOL constructs – in Lithuanian (without spaces and Lithuanian symbols). The actual words of Lithuanian texts are presented in labels of SLOL expressions.

```
Declaration(Class(dom:person))
Declaration(ObjectProperty(dom:hold__position))
Declaration(Class(cemployee))
SubClassOf(dom:employee dom:role)
Declaration(DataProperty(dom:time_value))
Declaration(NamedIndividual(dom:Arnas_Paukste))
```

We will use the typical ontology constructs for general concepts, verb symbols, roles, verb concepts, verb concept roles and individuals. Data property “preferred” with the value “true” (“false”) is specified for each preferred (synonymous) representation (it is shown only for general concepts).

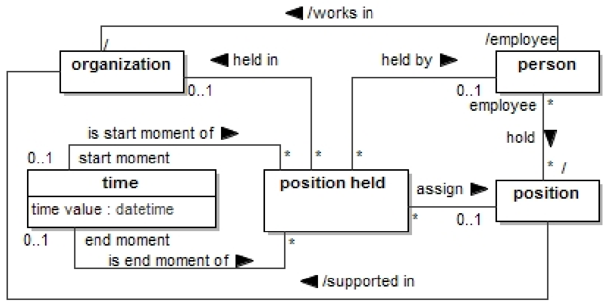

Figure 2. Example of domain ontology schema

Roles and verb concept roles will use the similar pattern as the general concept, except for the verb concept role the "vcr_des" (abbr. of "verb concept role designation") will be used instead the "term".

Pattern for terms (representing general concepts)

```
Declaration(NamedIndividual(slol:asmuo))
ClassAssertion(slol:text slol:asmuo)
Declaration(NamedIndividual(slol:asmuo__term))
ClassAssertion(slol:term slol:asmuo__term)
ObjectPropertyAssertion
(slol:representation__has__expression
  slol:asmuo__term slol:asmuo)
ObjectPropertyAssertion
 (slol:representation__represents__meaning
slol:asmuo__term dom:person)
DataPropertyAssertion(slol:preferred
 slol:asmuo__term "true"^^xsd:boolean)
```

Pattern for verb symbols (representing verb concepts)

```
Declaration(NamedIndividual(slol:uzima))
ClassAssertion(slol:text slol:uzima)
Declaration(NamedIndividual(slol:uzima_verb))
ClassAssertion(slol:verb_symbol slol:uzima_verb)
ObjectPropertyAssertion
  (slol:representation__has__expression
    slol:uzima_verb slol:uzima)
ObjectPropertyAssertion
  (slol:representation__represents__meaning
    slol: uzima_verb dom:hold__position)
```

Pattern for sentential forms (representing verb concepts)

```
Declaration(NamedIndividual
  (slol:darbuotojas__uzima__pareigas))
ClassAssertion(slol:text
  slol:darbuotojas__uzima__pareigas)
Declaration(NamedIndividual
  (slol:darbuotojas__uzima__pareigas__sentential_form))
ClassAssertion(slol:sentential_form
  slol:darbuotojas__uzima__pareigas__sentential_form)
ObjectPropertyAssertion
  (slol:representation__has__expression
slol:darbuotojas__uzima__pareigas__sentential_form
slol:darbuotojas__uzima__pareigas)
ObjectPropertyAssertion
  (slol:representation__represents__meaning
    slol:darbuotojas__uzima__pareigas__sentential_form
      dom:hold__position)
ObjectPropertyAssertion
  (slol:verb_concept_wording__incorporates__verb_symbol
    slol:darbuotojas__uzima__pareigas__sentential_form
       slol:uzima_verb)
```

Pattern for placeholders (representing verb concept roles)

```
Declaration(NamedIndividual(slol:darbuotojas))
ClassAssertion(slol:text slol:darbuotojas)
Declaration(NamedIndividual(slol:darbuotojas__vcr_des))
ObjectPropertyAssertion
(slol:representation__has__expression
  slol:darbuotojas_vcr_des slol: darbuotojas)
Declaration(NamedIndividual(slol:darbuotojas__pl_1))
ClassAssertion(slol:placeholder slol:darbuotojas__pl_1)
ObjectPropertyAssertion
(slol:placeholder__is_in__verb_concept_wording
  slol: darbuotojas__pl_1
    slol:darbuotojas__uzima__pareigas__sentential_form)
ObjectPropertyAssertion
(slol:placeholder__uses__designation
slol:darbuotojas__pl_1 slol:darbuotojas__vcr_des)
ObjectPropertyAssertion
 (slol:representation__represents__meaning
   slol: darbuotojas__pl_1 dom:employee)
DataPropertyAssertion(slol:starting_character_position
  slol:darbuotojas__pl_1 "1"^^xsd:positiveInteger)
```

SLOL allows specifying synonyms for noun concepts and verbs, and synonymous forms for verb concepts. Placeholders allow filling verb concept roles and verbs with sets of synonyms thus giving rich possibilities of combining multiple representations for expressing the same meaning. SBVR distinguishes the unique preferred representation for representing each meaning as constructs of domain ontologies (or other modelling languages) require using strict terminology. Synonyms and synonymous forms of ontology concepts are described by using the same patterns as for preferred representations, except the data property "preferred" values is "false", and the reference "see…" is given to the preferred representation.

Additional constructs for synonyms

```
ObjectPropertyAssertion(slol:see_synonym
slol:eina__verb slol:uzima_verb)
```

Additional constructs for synonymous forms

```
ObjectPropertyAssertion(slol:see_synonymous_form
 slol:darbuotojas__eina__pareigas__sentential_form
   slol: darbuotojas__uzima__pareigas__sentential_form)
```

Objectification pattern is used for representing n-ary relations. Usually, states of the art, events and other generic entities are n-ary relations, which are defined as classes in ontology. Using verbs for such class names causes inconveniences for naming relations of verb concepts with semantic roles (usually, each n-ary relation has several derived binary relations). SBVR defines relation "verb concept is objectified by general concept", which allows specifying verb concept representations and their synonyms, and relate their meaning to the corresponding general concept.

Objectification pattern

```
ObjectPropertyAssertion
(slol:verb_concept__is_objectified_by__general_concept
   dom:hold__position dom:position_held)
```

V. Conclusion and Future Works

The experimental investigation of the proposed lexical ontology SLOL, based on SBVR, has shown that it allows specifying verb concepts, various noun concepts, and their synonymous representations using SLOL. Various representations of the same concepts may be retrieved via SPARQL queries. Our research group has created the prototypical tools for editing SBVR business vocabularies, business rules and questions, and transforming them into OWL 2 and SPARQL. Also, some prototypes are developed for semantic annotating and search. For coping with grammatical complexity of Lithuanian language, we are using lemmatizing Web Services, created by the researchers of Vytautas Magnus University. Currently we cannot present the solid evidence of application of the SLOL in semantic annotating and search as we have not accumulated the sufficient lexical and semantic resources. We have created a prototype of database application, and our future work is intended for that purpose. However, we hope that our images would be useful for the Semantics Communities, which have encountered the similar problems.

Acknowledgement

The work is supported by the project VP1-3.1-ŠMM-10-V-02-008 „Integration of Business Processes and Business Rules on the Basis of Business Semantics" (2013–2015), which is funded by the European Social Fund (ESF).

References

[1] OMG, Semantics of Business Vocabulary and Business Rules (SBVR). SBVR 1.2 OMG Document Number: formal/2013-11-04 (2013)

[2] P. Cimiano, P. Buitelaar, J. McCrae, and M. Sintek, "LexInfo: A Declarative Model for the Lexicon-Ontology Interface". Journal Web Semantics: Science, Services and Agents on the World Wide Web, 9(1), March, 2011, pp. 29–51.

[3] WordNet: A lexical database for English, http://wordnet.princeton.edu/

[4] VerbNet: A Class-Based Verb Lexicon, http://verbs.colorado.edu/ ~mpalmer/projects/verbnet.html

[5] K. K. Schuler, "Verbnet: a broad-coverage, comprehensive verb lexicon" Dissertation, the University of Pennsylvania (2005)

[6] C.F. Baker, C.J. Fillmore, and J.B. Lowe, "The Berkeley FrameNet Project", in COLING '98 Proceedings of the 17th international conference on Computational linguistics, vol. 1, 1998, pp. 86–90.

[7] M.O. Dzikovska, M.D. Swifty, and J.F. Allen, "Building a computational lexicon and ontology with FrameNet", in LREC, 2004, pp. 53–60.

[8] M. Palmer, P. Kingsbury, and D. Gildea, "The Proposition Bank: An Annotated Corpus of Semantic Roles", Computational Linguistics, vol. 31 (1), 2005, pp. 71–106.

[9] T.M. Pazienza, M. Pennacchiotti, and F.M. Zanzatto, "Mixing wordnet, verbnet and propbank for studying verb relations", in Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006), 2006.

[10] L. Shi and R. Mihalcea, "Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing", in Computational Linguistics and Intelligent text processing, LNCS, vol. 3406, 2005, pp. 100–111.

[11] W3C, OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition) W3C Recommendation 11 December 2012, 2012.

[12] M. Linehan and E.H. Kendall. "Mapping SBVR to OWL2", IBM Research Report, RC25363 (WAT1303-040), 2013.

[13] J. Karpovic, L. Nemuraite, and M. Stankeviciene, "Requirements for semantic business vocabularies and rules for transforming them into consistent OWL2 ontologies", in Proceedings of Information and software technologies, 18th International Conference, ICIST 2012, Kaunas, Lithuania, September 13-14, 2012. Berlin, Heidelberg: Springer, CCIS, 2012, vol. 319, pp. 420–435.

[14] M. Kleiner, P. Albert, and J. Bézivin, "Parsing SBVR-Based Controlled Languages", in Schürr, A., Selic, B. (Eds.) Models 2009, LNCS 5795, Springer-Verlag Berlin Heidelberg, 2009, pp. 122–136.

[15] M. Selway, W. Mayer and M. Stumptne, "Configuring Domain Knowledge for Natural Language Understanding", in Proceedings of the 15th International Configuration Workshop, August 29-30, 2013, Vienna, Austria, pp. 63–70.

[16] A. Sukys, L. Nemuraite, and B. Paradauskas, "Representing and transforming SBVR question patterns into SPARQL", in Information and software technologies, Proceedings of 18th International Conference, ICIST 2012, Kaunas, Lithuania, September 13-14, 2012. Berlin, Heidelberg: Springer, CCIS, 2012, vol. 319, pp. 436–451.

[17] G. Bernotaityte, L. Nemuraite, R. Butkienė, and B. Paradauskas, "Developing SBVR vocabularies and business rules from OWL2 ontologies", in Information and Software Technologies, Proceedings of 19th international conference, ICIST 2013, Kaunas, Lithuania, October 10-11, 2013. Berlin, Heidelberg: Springer, CCIS, vol. 403, 2013, pp. 134–145.

[18] F. Gailly, "Transforming Enterprise Ontologies into SBVR formalizations", in Proceedings of the CAiSE'13 Forum at the 25th International Conference on Advanced Information Systems Engineering (CAiSE), Valencia, Spain, June 20th, 2013, pp. 122–129.

[19] T. Declerck, D. Gromann, "Combining three Ways of Conveying Knowledge: Modularization of Domain, Terminological, and Linguistic Knowledge in Ontologies", in Proceedings of the 6th International Workshop on Modular Ontologies, vol. 875, CEUR Workshop Proceedings, Graz, Austria, CEUR-WS, Aachen, 2012, pp. 28–40.

[20] A. Reymonet, J. Thomas, and N. Aussenac-Gilles. "Modelling Ontological and Terminological Resources in OWL DL", in Proceedings of ISWC'07 workshop "From Text to Knowledge: The Lexicon/Ontology Interface" (OntoLex'07), Busan, South Korea, 2007.

[21] P. Buitelaar, P. Cimiano, P. Haase, and M. Sintek, "Towards Linguistically Grounded Ontologies", in L Aroyo et al. (Eds.) ESWC 2009, LNCS 5554, Springer-Verlag Berlin Heidelberg, 2009, pp. 111–125.

[22] Y. Matsubayashi, Y. Miyao, and A.N. Aizawa, "Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet", in W. Daelemans, M. Lapata, and L. Màrquez (Eds.), EACL, The Association for Computer Linguistics, 2012, pp. 686-695.

[23] A.L. Rector, "Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL", in: K-CAP 2003, ACM, New York, 2003, pp.121–128.

[24] R. Hoekstra, Ontology Representation Design Patterns and Ontologies that Make Sense. SIKS Dissertation Series No. 2009-15, SIKS, Dutch, 2009.

About Author (s):

Phd. Stud. Gintare Krisciuniene

Prof. Dr. Lina Nemuraite

Assoc. Prof. Dr. Rita Butkiene

Prof. Dr. Bronius Paradauskas