

\$ sciendo

DOI: 10.2478/jolace-2021-0003

Corpus-based analysis of semi-automatically extracted artificial intelligence-related terminology

Jurgita Mikelionienė & Jurgita Motiejūnienė Kaunas University of Technology, Lithuania jurgita.mikelioniene@ktu.lt, jurgita.motiejuniene@ktu.lt

Abstract

Artificial Intelligence (AI), as a multidisciplinary field, combines computer science, robotics and cognitive science, with increasingly growing applications in many diverse areas, such as engineering, business, medicine, weather forecasting, industry, translation, natural language, linguistics, etc. In Europe, interest in AI has been rising in the last decade. One of the greatest hurdles for researchers in automated processing of technical documentation is large amounts of specific terminology. The aim of this research is to analyse the semi-automatically extracted artificial intelligence-related terminology and the most common phrases related to artificial intelligence in English and Lithuanian in terms of their structure, multidisciplinarity and connotation. For selection and analysis of terms, two programmes were chosen in this study, namely *SynchroTerm* and *SketchEngine*. The paper presents the outcomes of an AI terminological project carried out with *SynchroTerm* and provides an analysis of a special corpus compiled in the field of artificial intelligence using the *SketchEngine* platform. The analysis of semi-automatic term extraction use and corpus-based techniques for artificial intelligence-related terminology revealed that AI as a specialized domain contains multidisciplinary terminology, and is complex and dynamic. The empiric data shows that the context is essential for the evaluation of the concept under analysis and reveals the different connotation of the term.

Keywords: artificial intelligence, domain-specific corpus, semi-automatic term extraction, terminology, collocates

1 Introduction

Artificial Intelligence (AI) is a highly multidisciplinary and active field, combining computer science, robotics and cognitive science, with potentially transformative applications in many other areas, such as science, engineering, industry, linguistics, translation and society. The history of the development of AI has been wavy since its beginning in the middle of 20th century. Together with the development of AI technologies, machine learning in the field of linguistics, including terminology, lexicography and semantics, has evolved as well. Investments and interest in AI boomed in the first decades of the 21st century when machine learning was successfully applied to solving many problems in academia and industry. Interest in AI in Europe has increased in the last decade. In 2019, the Lithuanian Government officially confirmed the *Lithuanian Artificial Intelligence Strategy* (Ministry of the Economy and Innovation, online). The *Strategy* presents the strengths and weaknesses of AI in Lithuania, including ethical and legal core principles for the development and use of artificial intelligence, the position of Lithuania in the AI ecosystem, integration of artificial intelligence systems in economic sectors, the future vision of artificial intelligence, etc.

One of the greatest difficulties in automated processing of technical documentation is large amounts of specific terminology that is characteristic to such domains. As AI is constantly improving its skills, it can be presumed that it should have abundant and perhaps well-established terminology. Nevertheless, the analysis of AI terminology similarly to any other domain-specific terminology poses two major challenges: how to identify domain-specific terms in the documents and how efficiently process them. Nowadays, terminology and terminography have a very strong relation with computer technologies. While some tools are very popular, others are not convenient for certain languages, especially with reference to annotated corpora and automatic terms extraction tools. For the purposes of this research, two programmes, namely *SynchroTerm* and *SketchEngine*, were chosen for selection and analysis of terms. The choice was made based on the research carried out by Havrylenko (2019) who compared seven programs for automatic extraction of terms and found out that *SketchEngine* demonstrated better results than other tools and that this online platform was promising for compiling monolingual and bilingual lexicographic resources.

Identification and extraction of terms from texts or corpora is now a well-known and widely explored step in the processes of building terminology, constructing dictionaries, creating translation memories, etc. (Foo, 2012). The aim of this research is to analyse the semi-automatically extracted artificial intelligence-related terminology and the most common phrases related to artificial intelligence in English and Lithuanians in terms of structure, multidisciplinarity and connotation.

2. Analysis of semi-automatically extracted AI-related terminology using *SynchroTerm*

The first part of the research paper presents the outcomes of an AI terminological project carried out with *SynchroTerm*, a bilingual term extractor developed by *Terminotix* (Terminotix, 2020). *SynchroTerm* allows users to extract terms and create terminology records from source and target document pairs and translation memories. The tool is compatible with 30 languages, including English and Lithuanian, and it automatically extracts source terms, their equivalents and their contexts from file pairs into any format¹. For the pilot analysis of AI terminology extraction, the document *Building Trust in Human-Centric Artificial Intelligence* from *The Communication from the Commission*² (published in 2019) was chosen; and in this research, it plays the role of a domain-specific corpus. The main task at this stage of the research was to select potential AI terms in order to determine the semantic core of the multidisciplinary area and find out the main fields of terminology in this area. *Synchroterm* provided a list of 150 English AI-related multi-word term candidates, the number of their occurrences and the equivalents in the Lithuanian language. After clicking on a term, the context was given in the bottom section (see Fig. 1).

In the next step, the record was created and transferred to the term base. The list was thoroughly inspected and checked manually. Despite the conclusion made in a similar research that the quality of terms retrieved by using *SynchroTerm* tool was notably higher than in those extracted by *memoQ* or *SDL MultiTerm Extract* (Havrylenko, 2019, p. 31), only 30 grammatically correct and semantically meaningful phrases were designated as AI terminological units or AI-related phrases and chosen for the analysis in this research. All phrases are mentioned in this paper from a structural point of view as well as in an attempt to find out which areas of terminology and vocabulary form the framework of concepts related to artificial intelligence.

The study also investigated the structural characteristics of the selected terms and determined the morphological and syntactic differences of the structure of multi-word

¹ The translators in European Commission use this program too, and our colleagues from the Lithuanian Language Department in the General Directorate of Translation kindly helped us to get the list of term candidates for this research because this software is commercial and the university has not obtained the licenses yet.



terms in English and Lithuanian. The analysis of the list of term candidates in both English (source extraction) and Lithuanian (target extraction) documents showed that the English abbreviation *AI* was always translated into Lithuanian as a two-word combination *dirbtinis intelektas* like all other multi-word terms with *AI*, e.g., *AI systems* (*dirbtinio intelekto sistemos*), *AI ethics* (*dirbtinio intelekto etika*).

iource Extraction - 1/150					Target Extraction			
Expression	Occurren	N-Gram	PNG	<u>^</u>	Expression	96	Occurren	
ai systems	24	2	0		dirbtinio intelekto sistemu	59	10	
high-level expert	13	3	0		turéty būti	55	34	
level expert group	13	3	0		dirbtinio intelekto	42	100	
human-centric	12	2	0		dirbtinio intelekto sistemos	41	10	
ai high-level expert group	11	5	1		be to	32	14	
human-centric ai	8	3	1					
ethics guidelines	8	2	1					
european ai	7	2	0					
piloting phase	7	2	1					
trustworthy ai	7	2	0	~				
ource Selection, with 2 forms	Target Selection - 10 Occurrences				-			
AI systems				•	dirbtinio intelekto sistemų			
Translation Units								
In addition to safeguarding priv Turi būti užtikrinta ne tik privatu The quality of the data sets usee Didžiausią įtaką <u>dirbtinio intelek</u>	acy and personal dat imo ir asmens duom I is paramount to the to sistemy veiksmin bould be ensured; it i abelling, and a descri	ta, requirem eny apsaug e performan gumui turi r s important iption of the	ents must be a, bet ir aukš ce of <u>AI syste</u> baudojamų d to log and de algorithm u	fulfilled t ta dirbtini uomenų r bocument t sed) that y	inkinių kokybė. [1] oth the decisions made by the systems.	, as well as the e		

Fig. 1: Synchroterm snapshot

The most frequent term in the source entry list found 24 times was *AI systems*. The Lithuanian equivalent in the target entry list was *dirbtinio intelekto sistemų*, found only 10 times. This is obvious evidence of the complexity when comparing morphologically different languages. These 10 Lithuanian occurrences were mainly expressed in the genitive case. Other plural cases that were not recognised by the tool included 4 datives, 4 accusatives and 6 nominatives. These findings show the usage of words (in a multi-word term) in different grammatical cases in the Lithuanian language that is a synthetic language with an old and rich inflectional system. All the cases were the equivalents for one English plural term *AI systems* (see Table 1).

EN	LT
	Pl. genitive: dirbtinio intelekto sistem-ų (10);
Pl. AI systems (24)	Pl. nominative: dirbtinio intelekto sistem-os (6)
	Pl. accusative: dirbtinio intelekto sistem-as (4)
	Pl. dative: dirbtinio intelekto sistem-oms (4)

Tab. 1: Distribution of the lemma *AI systems* (*dirbtinio intelekto sistemos*) in different grammatical cases

Obviously, phrases with main concept AI (*Artificial Intelligence*) prevail in this term candidate list, which formed the semantic core of the analysed domain-specific area, the boundaries of which were determined by the text register and genre. Thus, the phrases under analysis are binary, i.e., they consist of the head noun phrase (two-word term) *AI* (*dirbtinis intelektas*) and its modifiers. Depending on the grammatical and functional features of the words modifying the basic concept, three most common structural types of English term candidates were distinguished. The first type included terms with one attributive adjective modifier. The Lithuanian equivalents did not always have identical



or very similar structures, e.g., *ethical AI* (*etiškas dirbtinis intelektas*), *trustworthy AI* (*patikimas dirbtinis intelektas*), etc. Due to grammatical features of the Lithuanian language, an adjective is sometimes translated as a noun, e.g., *European AI* (*Europos dirbtinis intelektas*). The second type included the candidate terms where *AI* (*dirbtinis intelektas*) was a modifier of the noun or the noun phrase, e.g., *AI development* (*dirbtinio intelekto plėtojimo*), *AI ethics* (*dirbtinio intelekto etikos*), *AI ethics guidelines* (*dirbtinio intelekto etikos gairės*). The third group of terms included English terms with the noun as a modifier of the term *AI* (*dirbtinis intelekto vartojimas*), *key requirements for AI* (*reikalavimai dirbtiniam intelektui*). The Lithuanian equivalents were often of the second type structure mentioned above. The fact that the Lithuanian language uses fewer functional words is also determined by the above-mentioned flexural nature of the Lithuanian language.

The three phrases in this list of *AI* term candidates were different in their form, e.g., *human-in-the-loop (HITL), human-on-the-loop (HOTL), human-in-command (HIC)*. These terms had equivalents in descriptive forms in the Lithuanian texts. For example, *human-in-the-loop* had an equivalent *žmogus dalyvauja procese* (that means 'human participates in the process'); *human-on-the loop* had an equivalent *žmogus prižiūri procesą* (that means 'human looks after the process'); and *human-in-command* had an equivalent *žmogus valdo procesą* (that means 'human rules over the process').

The abbreviation *AI4EU* in the analysed document did not have terminological structural equivalents in the Lithuanian version. Instead, the full description was used (*dirbtinis intelektas Europai*). Abbreviations of such type are not common in Lithuanian scientific discourse and are considered to be unofficial.

Although two-, three or four-word terms were common in the primary list of term candidates, many phrases were general collocations (wide range, brings together, next *steps, setting up, high-performance*) or phrases that could be used in domains other than artificial intelligence, especially in the politic discourse, e.g., European Parliament Parlamentas), fundamental rights (pagrindinių teisių), (Europos international organisations (tarptautinės organizacijos), non-EU countries (ne ES šalys); General Data Protection Regulation (Bendruoju duomenų apsaugos reglamentu), etc. This was a group of important terms related to human rights and personal data protection. Some candidates in this list were not semantically complete terms as the main component of the multi-word term was missing, e.g., human-centric (*i žmogų orientuotas*). The candidate term was another phrase, i.e., human-centric AI (*i žmogų orientuotas dirbtinis intelektas*). As mentioned above, the comparison of the two lists showed that the program did not always accurately determine the equivalent. It suggested another phrase in the same sentence; therefore, manual checking of the list was necessary. For example, the piloting phase is not suinteresuotieji subjektai (Eng. stakeholders). The list of potential terms showed that the document did not solely refer to AI as a technology, but emphasized its relation to the society, ethics and impact on humans. All this shows that the humanities, social sciences, especially human rights and personal data protection related terminology, form a significant part of the multidisciplinary AI terminology system.

3. Domain-specific corpus based collocational analysis of the concept *dirbtinis intelektas* (*Artificial Intelligence*)

In the descriptive terminology, context plays a vital role in the analysis of a term, i.e., a term is assumed as a lexical unit depending on its context. According to Maynard and Ananiadou (2000, n.d.), "Terminological contexts, however, are vast information sources waiting to be tapped, and are even more useful for domain-specific applications than for



general language ones". A wider and more diverse context is necessary for deeper analysis of AI terminology. Cabré et al. (2012) mentioned the importance of creating a domainspecific corpus. The researchers claim that terminologists need to become familiar with the domain language, perform terminology and keywords extractions, conduct statistical analysis and obtain complementary information about the terms, including semantic, syntactic and collocational clues (ibid.) Terminology research based on corpus analysis has been done in many different fields, such as social sciences, humanities (Yuliawati et al., 2018; Rimkutė, 2012; Markievicz and Rimkutė, 2013; Biel et al., 2018), information technologies (Izwaini, 2005), etc. Therefore, the method of corpus based term recognition and contextual analysis of their structure used in this study is reliable and suitable for exploring the terms of such a rapidly developing field as artificial intelligence.

In order to have a more comprehensive view of the use of the basic concept, a special Lithuanian corpus *AI_DI* in the field of artificial intelligence using the *SketchEngine* platform was compiled. For this purpose, concordances, word lists and other elements that show real usage, frequency and structure and reveal an evaluation of a concept that is still completely unknown were used in this project (see Fig. 2).

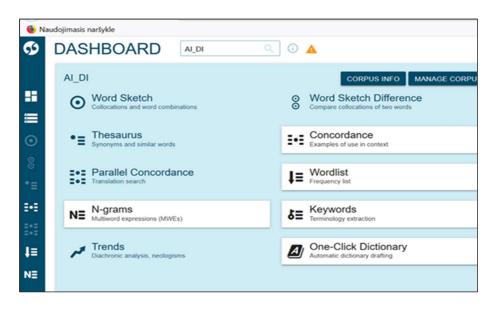


Fig. 2: Lithuanian internet corpus AI_DI (in SketchEngine)

Although the corpus *AI_DI* consists of about 1 million word forms, it is quiet representative because it was collected for this research from different artificial-intelligence-related articles, websites of universities, popular sciences journals, educational platforms, etc. This focus corpus was not annotated, so only the key words can be checked, which on their high quantity had been used to study the representation of AI in Lithuanian press and media.

In total, 493 occurrences of the term *dirbtinis intelektas* (*artifficial intelligence*) were found in the Lithuanian corpus: 101 nominative, 306 genitive, 11 dative, 47 accusative and 28 instrumental cases. No terms were found in the locative case. As this corpus was not annotated, the meaning and structure modifiers of the term *artificial intelligence* (dirbtinis intelektas) could be seen by the KWIC or by the token to the left or right of KWIC. Thus, the research also used the collocational analysis method. Smirnova and Rackevičienė (2018, p. 90) with reference to the research carried out by Atkins et al. (2003) claimed that the collocational analysis enabled the observation of the whole variety of co-occurring words and established the predominant co-occurrence patterns

and described their meanings based on their contextual environment. The collocational method may be used not only for collocations and collocates but for the extraction of multi-word terms as well. According to Nakagawa (2001), this methodology is based on the assumption that complex terms are made of existing simple terms. The part of the noun or adjective phrases from the collocation lists are terms or potential terms. The extraction of the phrases including the complex term *dirbtinis intelektas (AI)* as the head term with left and right collocates in this corpus was partly manual. Their wider context was analysed using the concordance tool (see Fig. 3) that was very suitable for lexical aspect of terminological analysis (Petrauskaitė, 2010).

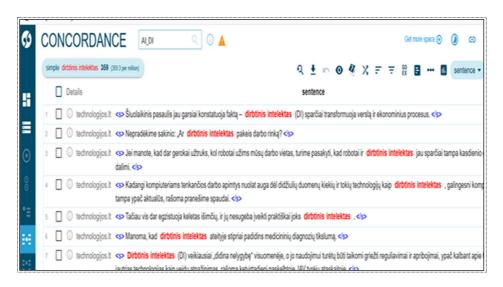


Fig. 3: Concordance from SketchEngine

After analysing all 493 concordances with the Lithuanian term *dirbtinis intelektas* (*AI*) with collocates, 4 main types of multi-word collocations were distinguished, including multi-word terms or candidate terms, namely noun collocates next to the nude concept in the first position from the right (1R (noun)), adjective collocates in the first position next to the analysed concept *artificial intelligence* (*AI*) from the left (1 L (adjective)) or in the second or in the third position from the right (2R, 3R (adjective)) and verb collocates from the right and in the first position (1 R (verb)).

Dirbtinio intelekto / AI + 1R (noun) collocate. The majority of phrases were combined with the noun on the right. Most of them were combined with abstract, verbal nouns with the suffixes *-imas* or *-ymas*. This confirmed the fact that the names of actions with suffices are the most frequent category for compiling terms in the Lithuanian language (Keinys, 2005 [1975]: 23), e.g.: dirbtinio intelekto panaudojimas (AI usage), dirbtinio intelekto kūrimas (AI creation), dirbtinio intelekto pritaikymas (AI application), dirbtinio intelekto plėtojimas (AI development). Artificial intelligence is a frequently discussed phenomenon, which was evident by the phrase dirbtinio intelekto klausimas (AI question). Among these phrases, there are collocations that were included in SynchroTerm list of potential AI terms, such as dirbtinio intelekto sistema (AI ethics), dirbtinio intelekto strategija (AI strategy), dirbtinio intelekto sistema (AI system), dirbtinio intelekto technologija (AI technology). Possessive case relationships in this situation are very strong, the main component dirbtinis intelektas (AI) is used in the genitive case (dirbtinio intelekto) and depends on the collocate that is most often used in the nominative case.

1 L (adjective) collocate + *dirbtinis intelektas / Al.* The most typical adjective or adjective phrases for the Lithuanian language are in the position of the left collocate. In other words, they are prenominal modifiers or attributes of noun phrase; and the collocate is combined with the main component. In this research, many terms were compiled from an adjective or a participle and a noun. This principle was also applied to the term dirbtinis intelektas (AI). Thus, one more adjective on the left side was not part of the term but part of the modifier with semantic connotation that mainly had a positive meaning and reflected progress, e.g., pažangus, sąmoningas, pažengęs, etiškas dirbtinis intelektas (advanced, conscious, progressive, ethical AI). One of the most interesting findings was that almost all adjectives could be used to describe a person as well, while AI is described as human-level AI. Thus, the word intelektas (intelligence) as an object had 2 modifiers, dirbtinis (artificial) as a terminological modifier and the second one, which modifies the meaning of the two-word term dirbtinis intelektas (see Fig. 4) and has a positive connotation.

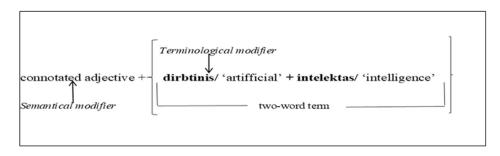


Fig. 4: Structure of the term dirbtinis intelektas (artificial intelligence) with modifiers

Dirbtinis intelektas / AI + 2R or 3R (adjective) collocate. The adjective in the position of the second or third right collocate performed the predicative function and was part of a compound predicate. In this group, *dirbtinis intelektas* was a subject, while different forms of the verb *to be* used with an adjective had the function of a predicate. The most frequent terms were *yra patikimas* (*trustworthy*) and *gali būti naudingas* (*useful*). Another example was *dirbtinis intelektas nėra <u>savatikslis</u> (AI is not self-directed*). The analysis revealed that the term AI with adjectives, regardless of their position next to the analysed concept as collocates, was perceived as ambiguous, e.g., *itin pažangus dirbtinis intelektas galėtų būti nenuspėjamas* (*highly advanced artificial intelligence could be unpredictable*). In this example, two different connotations were noticed, i.e., in the first case, the modifier was positive (*highly advanced*), while in the second case, it was negatively connoted (*unpredictable*).

Dirbtinis intelektas / AI + 2 R (verb) collocate. When considering verbal collocations of the term AI, the connotations were predominantly positive. The context demonstrated huge potential and important possibilities of artificial intelligence, e.g., *Jis [DI] gali padėti nustatyti grėsmes kibernetiniam saugumui ir sukčiavimo atvejus (It [AI] can help detect cybersecurity threats and fraud*); *Dirbtinis intelektas treniruojasi rašyti lietuviškus romanus (Artificial Intelligence is learning to write Lithuanian novels*); *Dirbtinis intelektas turi didžiulį potencialą pakeisti ligos diagnozavimą ir valdymą (Artificial Intelligence has great potential to change the diagnosis and management of the disease*). The context in concordances may as well reflect the fact that verbal phrases indicate a perceived threat, doubt and other negative emotions, e.g., *"Dirbtinis intelektas gali reikšti žmonių giminės pabaigą," – pasakė Hokingas BBC laidoje (Professor Steven Hocking in 2014 told the BBC: "The development of full artificial intelligence could spell the end of the human race"*);

Dirbtinis intelektas žmonijai kelia didžiausią egzistencinį pavojų (Artificial Intelligence poses the greatest existential danger to humanity).

Thus, the study of different structural types of collocations with the term *artificial intelligence* revealed the structure of multi-word terms. Collocations and free phrases were most common to the Lithuanian language. The informative content showed that artificial intelligence was most often mentioned in social sciences and humanities in the context of societal, scientific and technological progress. On the contrary, with regards to technological aspects of robotics, information technology, etc. artificial intelligence was rarely discussed.

Conclusions

The analysis of semi-automatic term extraction use and corpus-based techniques for artificial intelligence-related terminology reflected the fact that AI as a specialized domain was complex and dynamic and contained multidisciplinary terminology. In texts aimed at a wide audience of readers, AI was mainly considered in the context of ethics and protection of personal data as well as the progress of society, science and technology.

The Lithuanian two-word term *dirbtinis intelektas* (*AI*) was commonly used in the genitive case with the noun collocate to the right and served a descriptive function. The second most frequent phrase was used with an adjective on the left and indicated the desired characteristics of artificial intelligence. The empiric data showed that the context was essential for the evaluation of the concept under analysis, and revealed the connotation of the term. It was also determined that the term *artificial intelligence* (*AI*) and the Lithuanian equivalent *dirbtinis intelektas* possessed both positive and negative semantic connotations, whereas the modifiers and short phrases were predominantly positive.

Further and more extensive research is needed in order to provide a more detailed terminological analysis. The corpus of artificial intelligence could be expanded and annotated, especially for technical texts. A more detailed and comprehensive analysis of the terms and term candidates should be carried out and an analysis of the terms of artificial intelligence as translation equivalents could be composed in further research.

References

- Biel, Ł., Biernacka, A., & Jopek-Bosiacka, A. (2018). Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations. In: Marino S., Biel Ł., Bajčić M., Sosoni V. (Eds.), *Language and Law*. Springer. DOI: https://doi.org/10.1007/978-3-319-90905-9_14.
- Cabré, M. T., Amor Montané, M., & Nazar, R. (2012). *Corpus-based Terminology Processing*. Workshop session of the Terminology and Knowledge Engineering Conference (TKE 2012). Available on: http://terminus.iula.upf.edu/tke2012.
- European Commission (2021). *Building Trust in Human-Centric Artificial Intelligence.* Available on:

https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/bui lding-trust-human-centric-artificial-intelligence

- Foo, J. (2012). Computational Terminology: Exploring Bilingual and Monolingual Term Extraction. In *Linköping Studies in Science and Technology*. Linköping: Linköping University. Available on: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-75243.
- Havrylenko, V. (2019). Comparison of Automatic Systems of Term' Extraction. *Advanced Linguistics* 4; 29–33. DOI: https://doi.org/10.20535/2617-5339.2019.4.189753.

Izwaini, S. (2005). A corpus-based study of the translation of IT terms. *ESP Across Cultures* 2; 76–94. Available on: https://www.researchgate.net/publication/320546551_Corpus-based_Study_of_IT_Terms.

Keinys, S. (2005). Priesaginė lietuviškų terminų daryba (Suffixal Word Formation in Lithuanian Terminology). Dabartinė lietuvių terminologija (Modern Lithuanina Terminology), pp. 21–76. Vilnius: Lietuvių kalbos institutas.

- Markievicz, I., & Rimkutė, E. (2013). Švietimo ir mokslo terminų žodynas ir ontologija (Ontology and dictionary of science and education terms). *Terminologija*, 20, pp. 56– 74. Vilnius: Lietuvių kalbos institutas.
- Maynard, D., & Ananiadou, S. (2000). Terminological acquaintance: The importance of contextual information in terminology. *Proc. Int. Workshop on Computational Terminology for Medical and Biological Applications;* 19–28. Available on: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.6894.
- Ministry of the Economy and Innovation of Lithuania (online). Lithuanian ArtificialIntelligenceStrategy.Availableon:https://eimin.lrv.lt/uploads/eimin/documents/files/DI_strategija_ENG(1).pdf
- Nakagawa, H. (2001). Experimental Evaluation of Ranking and Selection Methods in Term Extraction. In Bourigault, D., Jacquemin, Ch., & L'Homme, M.-C. (Eds.), *Recent Advances in Computational Terminology* (pp. 303–325), John Benjamins Publishing Company.
- Petrauskaitė, R. (2010). Lietuvių kalbos kolokacijos (Collocations of the Lithuanian language). Kaunas: VDU. Available on: https://talpykla.elaba.lt/elaba-fedora/objects/elaba:4603559/datastreams/MAIN/content.
- Rimkutė, E. (2012). Automatinis švietimo ir mokslo terminų nustatymas lingvistiniais metodais (Automatic identification of science and education terms using linguistic methods). *Terminologija*. Vilnius, 19, 54–69.
- Smirnova, O., & Rackevičienė, S. (2018). Corpus-Driven Analysis of Multi-Word Terms Including the Word 'Risk' in English, French and Lithuanian. *Terminologija*, 25, pp. 86– 106. Vilnius: Lietuvių kalbos institutas. Available on: http://lki.lt/wpcontent/uploads/2018/12/Terrminologija_25_maketas.pdf.
- Terminotix (2020). *Synchroterm: Bilingual term extraction software*. Available on: https://terminotix.com/index.asp?content=item&item=7&lang=en
- Yuliawati, S., Suhardijanto, T., & Hidayat, R. S. (2018). A Corpus-based Analysis of the Terminology of the Social Sciences and Humanities. *IOP Conference Series: Earth and Environmental Science*, 175(1), DOI: https://doi.org/10.1088/1755-1315/175/1/012109.

Contact

Jurgita Mikelionienė & Jurgita Motiejūnienė Kaunas University of Technology K. Donelaičio St. 73 44249 Kaunas Lithuania jurgita.mikelioniene@ktu.lt, jurgita.motiejuniene@ktu.lt