Creative Commons Attribution – NonCommercial 4.0 International License

Original scientific paper
https://doi.org/10.31784/zvr.10.1.4
Received: 2. 11. 2021.
Accepted: 20. 2. 2022.

AUTOMATION OF THE CONVERSION OF NATURAL LANGUAGE TO FORMALIZED NODE OF KNOWLEDGE RECORD

Marina Rauker Koch
Professor of Education and Informatics, Senior Lecturer, Polytechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia; e-mail: mrauker@veleri.hr

Sanja Čandrlić
PhD, Associate Professor, Faculty of Informatics and Digital Technologies, University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia; e-mail: sanjac@inf.uniri.hr

Martina Ašenbrener Katić
PhD, Assistant Professor, Faculty of Informatics and Digital Technologies, University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia; e-mail: masenbrener@inf.uniri.hr

ABSTRACT

The paper describes the improvement of the system for automatic question answering based on the knowledge expressed in natural language sentences. The system is implemented using a relational database. This system will be the basis for the development of a web application for obtaining answers to the posed questions. In order for natural language sentences to be entered into a relational database, they must be prepared and formally recorded. The development of a question-answering system is based on the application of the conceptual framework Node of Knowledge (NOK), whose formalized record is suitable for input to a relational database from which answers to questions can be obtained. This paper presents an application for automated conversion of English sentences into a formalized record. The application was tested on one hundred simple English sentences, and the results of the automated conversion were compared with those obtained by processing the same sentences manually.

Key words: *Node of Knowledge (NOK), Formalized Node of Knowledge (FNOK), Knowledge Representation, Question-Answering (QA) system*

1. INTRODUCTION AND RELATED WORK

Knowledge representation is an essential area of artificial intelligence (Sowa, 2000; Morgenstern & Tomson, 2000) and deals with methods of formal recording of knowledge in a computer. Many

methods applicable to different types of knowledge have been developed to convert knowledge into a form appropriate for coding using a computer (Van Harmelen, et al., 2008; Brachman & Levesque, 2004; Russell & Norvig, 2010). Natural language processing (NLP) is the automation of human natural language processing, which can also be defined as semi-automated processing (Elallaoui, et al., 2018). NLP is related to linguistics, because natural language processing tools enable automated linguistic analysis (Lash, et al., 2012; Güneş & Aydemir, 2020). Linguistic analysis refers to the use of language and the way specific words and phrases mirror human thinking (Amirhosseini, et al., 2018; Hale, et al., 2021). To make NLP more accessible to people in fields other than linguistics and computer science, tools that spread NLP literacy are being developed (Libbrecht, et al., 2020; Baglini & Hjorth, 2021).

There are many different methods for representing knowledge, and one of them is the conceptual framework Node of Knowledge (NOK) which consists of the method Node of Knowledge (NOK) and formalism for graphical representation (Diagram Node of Knowledge, DNOK), formalism for textual knowledge representation (Formalized Node of Knowledge, FNOK) and formalism for the question formalization (Formalized Node of Knowledge Question, QFNOK). The NOK was developed to prepare natural language sentences for entry into a relational database. With its DNOK formalism, the NOK conceptual framework presents knowledge as a graph. Its aim is to present textual knowledge as a network of knowledge (Pavlic, et al., 2013). The formalization of the NOK method is shown in (Jakupovic, et al., 2014; Pavlic, et al., 2015). This method analyses natural human language, sentences, words and their meanings, as well as related word strings that create more complex phrases (Asenbrener Katic, et al., 2015; Asenbrener Katic, et al., 2017). The NOK method first converts knowledge into a model and then searches for knowledge in the model. After that, it is able to answer questions related to this knowledge (Tomljanovic, et al., 2014).

Compared to other methods, the NOK method stands out for its ease of reading and implementation and the possibility of wide application (Jakupovic, et al., 2013). NOK finds its foundation in the model of entities and relationships (ER), which is also observed in the notional model. In this model the relationships between notions are notions themselves (Vykhovanets, 2021), in contrast to NOK where the relationships between nodes are questions. The ability to customize the NLP method of Natural Language Query Interface (NLQI) for entries into a relational database is described in (Frost & Peelar, 2019; Stoica, et al., 2020; Walelign, et al., 2021).

The application of the NOK method and DNOK and FNOK formalisms to complex sentences in fables is shown in (Rauker Koch, et al., 2014; Rauker Koch, et al., 2017). Previous research of the conceptual framework of NOK has shown that it is possible to enter sentences and questions of natural language in a relational database using the FNOK formalism (Candrlic, et al., 2019) and QFNOK, as well as through the Question answering system (QA system obtains results or answers to questions). The research has shown that a QA system is language-independent if conversion rules are followed (Candrlic, et al., 2020).

In order to transform all words from natural language sentences into formalized records, it is necessary to devise a language metamodel. To do this, it is necessary to analyse all types of words for each individual natural language and define rules for converting sentences in natural language

into formalized records. So far, detailed analyses and rules for converting nouns (Asenbrener Katic, et al., 2021.), verbs (Asenbrener Katic, et al., 2018) and adjectives (Pavlic, et al., 2017) have been created, and are used in this paper.

The aim of this paper is to show the functioning of an application created in the Python programming language for automated conversion of natural language sentences into a formalized NOK record (FNOK) and to test the performance of the obtained FNOK records through the entry into a relational database and querying. A valid FNOK record is the one for which at least one answer to a question is obtained that corresponds in content and semantics to the correct answer that a person would give, based on the same initial knowledge.

Previous research, in which sentences in FNOK were converted manually, has shown that this record can be entered into a relational database (Candrlic, et al., 2019) and answers to questions can be retrieved from it, which is the basis of a QA system.

The use of NOK and the application that will be able to automatically read textual knowledge from different types of files, including web pages, and convert it into a formalized FNOK record, will enable further development of QA system based on a relational database, and it will be possible to implement it on the web.

2. RESEARCH MOTIVATION

In the research conducted so far, the conversion of natural language sentences into a formalized FNOK record was performed manually, but there was a need for automated conversion of natural language sentences into a formalized record suitable for entry into a relational database and entry into the QA system.

From the database created in this way, by using SQL queries, the answers to the questions asked for the selected sentences can be obtained.

In this paper, our aim is to test the efficiency of the PSA - FNOK application developed in Python (described in (Pavlic, et al., 2015; Jakupovic, et al., 2014)) by converting simple sentences of natural language into FNOK records and their applicability in the existing QA system based on a relational database. Simple sentences used in this paper consist of the following word types: nouns, verbs, pronouns, adjectives, adverbs, prepositions, and numbers, but do not include these: conjunctions, interjections, and two types of pronouns (relative and interrogative).

3. METHODOLOGY

The research is based on 100 sentences published in (Candrlic, et al., 2019) for which the FNOK record was manually written following the rules for converting natural language sentences into formalized records. New research improves this procedure and introduces automated conversion of FNOK records with the help of PSA - FNOK application.

The objective of this research is to show that it is possible to automatically convert English language sentences into a FNOK record and enter them into the QA system.

The rest of this section describes the PSA - FNOK application and its use, as well as the procedure for comparing the results obtained by manual and automated sentence processing.

3. 1 PSA - FNOK application

The PSA - FNOK application was created in Python and converts English language sentences into a formalized NOK record (FNOK). Figure 1. shows the user interface of the application.

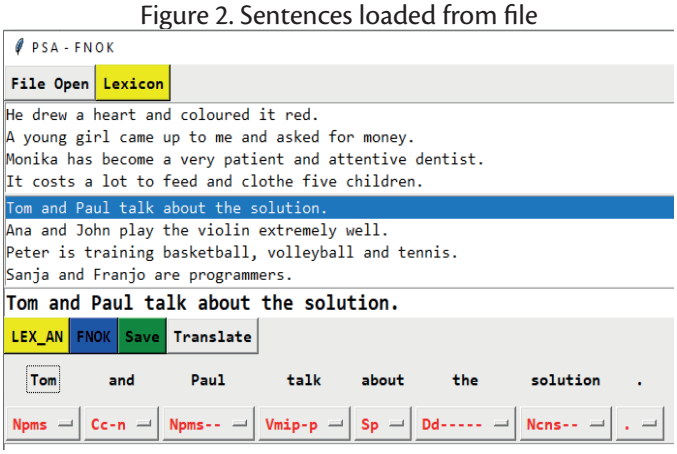
Figure 1. User interface of the PSA - FNOK application

Source: Authors

Sentences can be loaded from a text file or typed directly into the application. This version is prepared for testing, and in the future the goal is to implement it in a web application that will be able to download sentences from texts on the web.

The first step is to enter a sentence into the application, which can be done in two ways: by loading a series of sentences from a text file and selecting the desired sentence (Figure 2) or by typing in the sentence directly.

Figure 2. Sentences loaded from file

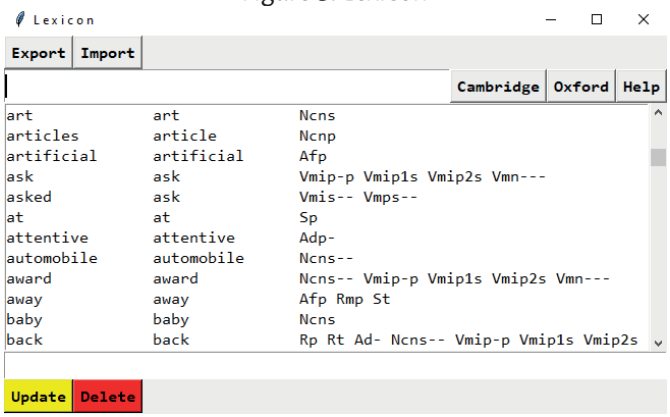


Source: Authors

The next step is the lexical analysis of the sentence which starts by using the command LEX_AN. For lexical analysis, a pre-created Lexicon (Figure 3) is used. The lexicon is a text file in which words and their forms are written with their corresponding attributes. It is based on the results of the project "MULTEXT-East" for the English language (Ide, et al., 2018; Chiarcos & Erjavec, 2011; Erjavec,

2017). The MULTEXT-East project (Multilingual Text Tools and Corpora for Eastern and Central European Languages) started in 1995, lasted two years and developed a standardized language resource for six languages of Central and Eastern Europe and English (Dimitrova, et al., 1998).

Figure 3. Lexicon



Source: Authors

In the case when a particular word form does not exist in the lexicon, it is added to it using the Update command. In the same way, a word not found in the Lexicon is added, which is most often the case with personal names.

The first column of the Lexicon contains the word form, the second contains the basic word form, and the third contains groups of attributes that define possible ways of using it.

The first letter in each attribute group indicates the word class in the English language:

- N Noun
- V Verb
- A Adjective
- P Pronoun
- D Determiner
- R Adverb
- S Apposition
- C Conjunction
- M Numeral

The first letter is followed by the other word attributes. For example, nouns have 5 different groups of attributes: type, gender, number, animacy and prefix, as well as verbs: type, form, tense, person and number. Tables 1 and 2 show the attributes of nouns and verbs.

Table 1. Noun attributes

Ord. No.	Group	Attributes
1	type	**c** common **p** proper **x** collective
2	gender	**m** male **f** female **n** neutral
3	number	**s** singular **p** plural **0** uncountable
4	animacy	**0** inanimate **l** animate **m** material **a** abstract
5	prefix	**c** consonant **v** vocal

Source: Authors

Table 2. Verb attributes

Ord. No.	Group	Attributes
1	type	**m** main **a** auxiliary **o** modal **b** base
2	form	**i** indicative **c** conditional **n** infinitive **p** participle
3	tense	**p** present **s** past **i** imperfect **f** future I
4	person	**1** first **2** second **3** third
5	number	**s** singular **p** plural

Source: Authors

For example, the word "kiss" (Figure 4) has 5 groups of attribute properties defined in the Lexicon, depending on its meaning. "Kiss" in a sentence can, depending on the context, appear both as a noun and a verb. In the first place in Lexicon, the group of attributes that most often appears in sentences is entered. The order of the attribute groups can be changed in the Lexicon at any time or the appropriate group can be selected during the lexical analysis.

Figure 4. Attributes of the word "kiss"

```
kiss            kiss            Ncns-- Vmip-p Vmip1s Vmip2s Vmn---
```

Source: Authors

The first group of attributes describes the word "kiss" as a noun whose attributes are common, of the neuter gender, singular. Dashes are entered for the remaining two attributes, which are not defined for this word. Attributes are marked using dashes whenever they are not defined. The remaining groups of attributes describe this word as a main verb that appears in the form of an indicative in the present, but in different persons and numbers. Since only the present verb tense is included in the attributes, it remains to refine the Lexicon with other verb tenses as well, which will easily be done when the need arises.

The described attributes in the PSA - FNOK application are associated with each word during lexical analysis (command LEX_AN), and are important when converting to FNOK record because they determine the hierarchy and questions contained in FNOK record in accordance with the rules of conversion.

Figure 5 shows a part of the code from the application that defines the questions for the FNOK record according to the attributes.

Figure 5. Code used to determine questions according to word attributes

```
PA   = [ (r"D[di]"          ,  '"art?"'      ),
         (r"St"             ,  '"when?"'     ),
         (r"Sp"             ,  '"where?"'    ),
         (r"Sm"             ,  '"how?"'      ),
         (r"N[cp][fm]"      ,  '"who?"'      ),
         (r"N[cpx]n"        ,  '"what?"'     ),
         (r"(Dg|Ds|N[cp]\')",  '"whose?"'    ),
         (r"Np[fm]"         ,  '"who?"'      ),
         (r"Pg"             ,  '"what?"'     ),
         (r"Pp"             ,  '"who?"'      ),
         (r"Px"             ,  '"whom?"'     ),
         (r"Pw"             ,  '""'          ),
         (r"Ab"             ,  '"from what?"'),
         (r"A[df]p"         ,  '"how?"'      ),
         (r"A[dfx]"          ,  '"what?"'     ),
         (r"As"             ,  '"whose?"'    ),
         (r"Mc"             ,  '"how_many?"' ),
         (r"Md"             ,  '"when?"'     ),
         (r"Mr"             ,  '"what?"'     ),
         (r"Rm"             ,  '"how?"'      ),
         (r"Rf"             ,  '"how_often?"'),
         (r"Rd"             ,  '"how_much?"' ),
         (r"Rt"             ,  '"when?"'     ),
         (r"R(p|pp)"        ,  '"where?"'    ), ]
```

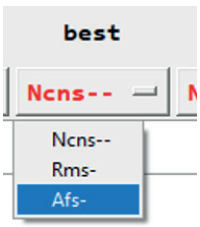
Source: Authors

3. 2 Usage of PSA - FNOK application

The example of the sentence "Honesty is the best policy." will show the functioning of the PSA - FNOK application.

The first step after entering a sentence into the application, by directly entering or loading it from a file, is to start the lexical analysis with the LEX_AN command, which associates each word with attributes from the first group of attributes associated with that word in the Lexicon. In most cases, the associated attributes correspond to the type and role of the word in the sentence, but at this point it is necessary to check whether this really is the case. The lexical analysis of the test sentence is shown in Figure 6, where it can be seen that the analysis marked the word "best" as a noun, which is not the case in this sentence, as this word appears as an adjective.

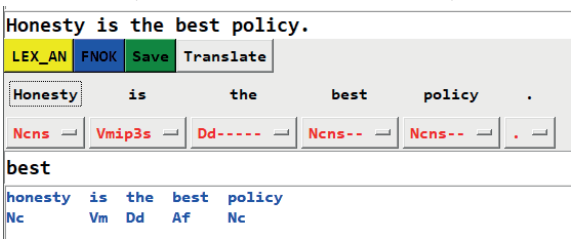
Figure 6. Changing the attribute of the word "best"



Source: Authors

Clicking on the button with the attributes of the word "best" opens all the attributes entered in the Lexicon from which the appropriate one is selected. In this case, a group of attributes is selected that describes this word as an adjective (Figure 7).

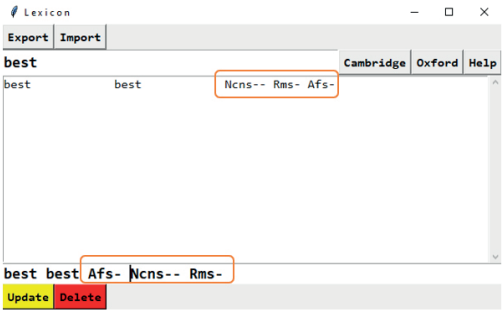
Figure 7. Lexical analysis of the sentence "Honesty is the best policy."



Source: Authors

Since the word "best" most often appears in sentences as an adjective (Afs–), and less often as a noun (Ncns––) and an adverb (Rms–), it is desirable to change the order of attributes in the Lexicon so that the automated lexical analysis of the sentence is as accurate as it can be and with as little intervention as possible; attributes of adjectives (Afs–) should be placed in the first place for application in future lexical analyses of sentences in which "best" appears. The changed order of attributes for the word "best" is shown in Figure 8.

Figure 8. Changing the order of attributes in Lexicon



Source: Authors

The next step is to start the conversion to FNOK record by clicking on the FNOK button (Figure 9).

Figure 9. Generated FNOK record for sentence "Honesty is the best policy."

Source: Authors

In this way, for all 100 sentences, automatic FNOK records were generated and compared with FNOK records created manually. The sentences that have identical FNOK records to those created manually are considered to meet the test requirements. Those sentences that have different FNOK records will be entered into a relational database and queries will be initiated, and if at least one correct answer to the question is obtained for the sentence, it will be considered that the automatic translation was successful.

The results of the comparison and testing are described in the next section.

4. RESEARCH RESULTS

The observed 100 sentences of the English language were converted into a FNOK record automatically using the PSA - FNOK application and compared with previously manually created FNOK records.

The comparison has shown that 37 sentences have completely identical FNOK records created manually to the ones created with the help of the application, and we can claim that these sentences meet the set conditions.

FNOK records were not completely identical for 33 sentences. However, entering them in the relational database and asking questions resulted in at least one correct answer, so it can be considered that the automatic conversion was successful. A more detailed analysis and comparison of these FNOKs have shown that there is a difference in the order of words in the record, which shows the flexibility of the NOK method in its application and the entire QA system.

A larger difference was observed for 30 sentences when comparing FNOK records. It was assumed that they would not be appropriate for the current structure of the QA system and entry into the relational database, so as expected they did not give correct answers to the questions, while 7 sentences could not be loaded at all into the QA system.

In this testing process, it is significant to mention that for 93 sentences the FNOK records were suitable for entry in the relational database.

The analysis of FNOK records has revealed problems that remain to be solved in future research:

- complex verb tenses
- negation
- two-word names that are considered a whole in a semantic context (e.g. New York)
- some questions from manually converted sentences have not yet been implemented in the PSA – FNOK application and it will be necessary to expand word analysis and, if necessary, define new attributes in the Lexicon that would be associated with these questions.

Figure 10 shows a comparison of a part of the sentences (from 61 to 75) among which the sentences from all three previously described cases are visible. The first column displays sentences in English language, the second column displays their FNOK records obtained manually, and the third displays FNOK records obtained with the help of the application PSA – FNOK. In the third column, identical FNOK records obtained automatically are marked in green, and records that are not identical, but give answers to the questions asked, are marked in yellow.

Figure 10. A part of compared sentences

61.	The boy is very young.	is ("who?" boy ("art?" the), "what?" young ("how?" very))	is ("who?" boy ("art?" the), "what?" young ("how?" very))
62.	Peter has an office in London.	has ("who?" Peter, "what?" office ("art?" an), "where?" London("where?" in))	has ("who?" Peter, "where?" London ("where?" in ("what?" office ("art?" an))))
63.	Rob has an office 200 meters from the sea in Dubrovnik.	has ("who?" Rob, "what?" office ("art?" an), "where?" sea ("how many?" meters ("how many?" 200), "where?" from, "art?" the), "where?" Dubrovnik ("where?" in))	has ("who?" Rob, "what?" Dubrovnik ("where?" in ("what?" sea ("art?" the ("where?" from ("what?" meters ("how_much?" 200 ("what?" office (an))))))))
64.	Tom is quite a good boy.	is ("who?" Tom, "what?" boy ("what?" good ("how?" quite, "art?" a)))	is ("who?" Tom, "what?" boy ("what?" good ("art?" a ("how?" quite))))
65.	I work hard.	work ("who?" I, "how?" hard)	work ("who?" I, "how?" hard)
66.	Noa gave this book to a friend.	gave ("who?" Noa, "what?" book ("which?" this), "who?" friend ("who?" to, "art?" a))	gave ("who?" Noa, "who?" friend ("art?" a ("who?" to ("what?" book ("what?" this)))))
67.	Maya has some interesting books.	has ("who?" Maya, "what?" books ("how many?" some, "what?" interesting))	has ("who?" Maya, "what?" books ("what?" interesting ("how many?" some)))
68.	I often go to the cinema.	go ("who?" I, "when?" often, "where?" cinema ("where?" to, "art?" the))	go ("when?" often ("who?" I), "where?" cinema ("art?" the ("where?" to)))
69.	London is the capital of Great Britain.	is ("what?" London, "what?" capital ("art?" the) ("what?" Great Britain ("what?" of)))	is ("what?" London, "what?" Britain ("what?" Great ("where?" of ("what?" capital ("art?" the)))))
70.	Mary is Lucy's friend.	is ("who?" Mary, "what?" friend ("whose?" Lucy's))	is ("who?" Mary, "what?" friend ("whose?" Lucy's))
71.	Father is a good man.	is ("who?" father, "what?" man ("art?" a, "what?" good))	is ("who?" Father, "what?" man ("what?" good ("art?"a)))
72.	Beauty is not forever.	is not ("what?" beauty, "what?" forever)	is ("what?" beauty, "how?" forever ("what?" not))
73.	Honesty is the best policy.	is ("what?" honesty, "what?" policy ("what?" best ("art?" the)))	is ("what?" honesty, "what?" policy ("what?" best ("art?" the)))
74.	Mary's father is old.	is ("who?" father ("whose?" Mary's), "what?" old)	is ("who?" father ("whose?" Mary's), "what?" old)
75.	It is six o'clock.	is ("what?" it, "what time?" o'clock ("what time?" six))	is ("what?" it, "what time?" o'clock ("what time?" six))

Source: Authors

The final results of the previous analysis which monitored the success of the PSA - FNOK application for automated conversion of English sentences to FNOK are (Table 3):

- for 37% of the sentences, FNOK is identical to the manually created sentence,
- 93% of the FNOK records were successfully entered into the QA system (37 identical FNOK records + 33 similar FNOK records + 23 FNOK records with a larger difference)
- for 70% of the sentences, the QA system gave the correct answer to at least one question asked.

Table 3. Research results

		Quality of FNOK record		
		Identical	**Similar**	**Large difference**
Answers received from the QA system	**Correct answer**	37/100	33/100	
	Wrong answer			23/100
	Unable to retrieve			7/100

Source: Authors

Thus, the objective posed in this study has been achieved.

The distribution of automated conversion of English sentences in FNOK shows that only 7% of sentences were not successfully entered into the QA system. Of the 93 sentences that were entered into the QA system, 70 sentences (75%) gave the correct answer to at least one question asked.

As there is room for improvement, in addition to upgrading the PSA – FNOK application, it is necessary to analyse the possibilities of upgrading the existing relational database and QA system and adapting them to automatically obtained FNOK records. The final goal of this research is the implementation of QA in a web application with the aim of achieving automated analysis of texts on various web pages and obtaining answers to questions asked in natural language.

5. CONCLUSIONS

This paper presents the application PSA – FNOK created in the Python programming language and intended for lexical analysis of sentences and their conversion into a formalized NOK record (FNOK). Its testing was conducted over a hundred sentences of English language. All sentences were automatically converted into a FNOK record using the application, and the resulting records were compared with existing FNOK records created manually and tested by the QA system. The comparison found that 37 sentences had identical manual and automatically generated FNOK records. The remaining 63 FNOK records required additional verification by entering them in the QA system. The entry was successful for another 57 sentences, where at least one correct answer to the questions asked was obtained for 37 sentences. We can conclude that the test success rate is 70% for obtaining answers to questions and 93% for entering automatically obtained FNOK records into the QA system. A more detailed analysis and comparison of FNOKs has shown that the difference in word order did not affect the performance, which shows the flexibility of the NOK method in its application, as well as the whole QA system.

The objective set in this research has been achieved. The research demonstrated that it is possible to automatically convert English-language sentences into a FNOK record and enter them into the QA system.

The limitation of this research lies in a rather small set of sentences containing only some word types. Our future research will include a larger set of complex sentences.

In the continuation of the research, the automatically obtained FNOK records will be analysed in detail in order to determine the reasons for unsuccessful entries into the QA system or inability to obtain answers to questions. In the FNOK records, possible problems have been identified that remain to be solved in future research: complex verb tenses, two-word names and the absence of some questions and word attributes to be connected in the Lexicon and PSA – FNOK application.

In addition to improving the PSA – FNOK application, it is necessary to consider the possibility of upgrading the existing QA system and relational database with the goal of implementing this system in a web application in order to achieve automated analysis of texts on various websites and obtain answers to questions.

This work has been fully supported by the University of Rijeka under projects uniri-drustv-18-73 and uniri-drustv-18-140.

REFERENCES

Amirhosseini, M. H., Kazemian, H. B., Ouazzane, K. & Chandler, C. (2018) "Natural Language Processing approach to NLP Meta model automation", *2018 International Joint Conference on Neural Networks* (IJCNN) pp. 1-8.

Asenbrener Katic, M., Candrlic, S. & Pavlic, M. (2017) "*Comparison of Two Versions of Formalization Method for Text Expressed Knowledge*", Communications in Computer and Information Science 716, Ustron, Poljska: Springer International Publishing, pp. 55-66.

Asenbrener Katic, M., Candrlic, S. & Pavlic, M. (2018) "Modeling of verbs using the node of knowledge conceptual framework" *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* pp. 1022-1027.

Asenbrener Katic, M., Pavlic, M. & Candrlic, S. (2015) "*The representation of database content and structure using the NOK method*". Procedia Engineering, 100, pp. 1075-1081.

Asenbrener Katic, M., Candrlic, S. & Pavlic, M., (2021) Nouns in the Conceptual Framework "Node of Knowledge". *Technical Gazette* 28, 6(2021), 2088-2093.

Baglini, R. & Hjorth, H. (2021) "Natural Language Processing 4 All (NLP4All): A New Online Platform for Teaching and Learning NLP Concepts", *Proceedings of the Fifth Workshop on Teaching* NLP, June, pp. 28-33.

Brachman, R. & Levesque, H. (2004) "*Knowledge representation and reasoning*", The Morgan Kaufmann Series in Artificial Inteligence, Morgan Kaufmann.

Candrlic, S., Asenbrener Katic, M. & Jakupovic, A. (2020) "*Preliminary Multi-lingual Evaluation of a Question Answering System Based on the Node of Knowledge Method*", Lecture Notes in Networks and Systems, San Francisco, SAD: Springer, doi:10.1007/978-3-030-12388-8_69, p. 998–1009.

Candrlic, S., Asenbrener Katic, M. & Pavlic, M. (2019) "A system for transformation of sentences from the enriched formalized Node of Knowledge record into relational database", *Expert Systems with Applications*, 115, pp. 442-464.

Chiarcos, C. & Erjavec, T. (2011) "OWL/DL formalization of the MULTEXT-East morphosyntactic specifications", *Proceedings of the 5th Linguistic Annotation Workshop*, pp. 11-20.

Dimitrova, L., Erjavec, T., Ide, N., Kaalep, H. J., Petkevic, V., & Tufiş, D. (1998) "Multext-east: Parallel and comparable corpora and lexicons for six central and eastern european languages", COLING 1998 Volume 1: *The 17th International Conference on Computational Linguistics*.

Elallaoui, M., Nafil, K. & Touahni, R. (2018) "Automatic transformation of user stories into UML use case diagrams using NLP techniques", *Procedia computer science*, 130, pp. 42-49.

Erjavec, T. (2017) "*MULTEXT-East. Handbook of Linguistic Annotation*", Springer, Dordrecht, pp. 441-462.

Frost, R. A., & Peelar, S. M. (2019) "*A New Data Structure for Processing Natural Language Database Queries*", WEBIST (pp. 80-87).

Güneş, T. & Aydemir, F. B. (2020) "Automated goal model extraction from user stories using NLP", 28th International Requirements Engineering Conference (RE) (pp. 382-387). IEEE., August, pp. 382-387.

Hale, J., Campanelli, L., Li, J., Bhattasali, S., Pallier, C., & Brennan, J. (2021). Neuro-computational models of language processing. Annual Review of Linguistics..

Ide, N., Priest-Dorman, G., Erjavec, T. & Varadi, T. (2018) "MULTEXT-East", Available at: https://nl.ijs.si/ME/V6/msd/html/msd-en.html [Acsessed 15. 8. 2021].

Jakupovic, A., Pavlic, M. & Dovedan Han, Z. (2014) *Formalisation Method for the Text Expressed Knowledge*. Expert system with applications, pp. 5308-5322.

Jakupovic, A., Pavlic, M., Mestrovic, A. & Jovanovic, V. (2013) "Comparison of the Nodes of Knowledge method with other graphical methods for knowledge representation", *36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1004-1008.

Lash, A., Murray, K. & Mocko, G. (2012) "*Natural language processing applications in requirements engineering*", American Society of Mechanical Engineers, pp. 541-549.

Libbrecht, P., Declerck, T., Schlippe, T., Mandl, T., & Schiffner, D. (2020) "NLP for Student and Teacher: Concept for an AI based Information Literacy Tutoring System", CIKM (Workshops)

Morgenstern, L. & Tomson, R. H. (2000) "*Teaching knowledge representation: Challenges and proposals*", Ann Arbor 1001, pp. 48109-2210.

Pavlic, M., Dovedan Han, Z. & Jakupovic, A. (2015) *Question Answering with Conceptual Framework for Knowledge-Based System Development "Node of Knowledge".* Expert Systems with applications, pp. 5264-5286.

Pavlic, M., Dovedan Han, Z., Jakupovic, A., Asenbrener Katic, M., & Candrlic, S. (2017) "Adjective representation with the method Nodes of Knowledge", *40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1221-1226). IEEE.

Pavlic, M., Jakupovic, A. & Mestrovic, A. (2013) "Nodes of knowledge method for knowledge representation", *Informatologija*, svez. 46, pp. 206-214.

Rauker Koch, M., Asenbrener Katic, M., & Pavlic, M. (2017) "Fable representation in FNOK and DNOK formalisms using the NOK conceptual framework", *Annals of DAAAM and Proceedings of the International DAAAM Symposium* (pp. 439-445).

Rauker Koch, M., Pavlic, M., & Jakupovic, A. (2014) "Application of the NOK method in sentence modelling", *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1176-1181)

Russell, S. & Norvig, P. (2010) "*Artificial Intelligence: A Modern Approach*", Third Edit. Prentice Hall.

Sowa, J., 2000. *Knowledge representation: logical, philosophical and computational foundations*. s.l.:Brooks/Cole Publishing Co..

Stoica, A., Pu, K. Q. & Davoudi, H. (2020) "NLP Relational Queries and Its Application", *020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, August, pp. 395-398.

Tomljanovic, J., Pavlic, M., & Asenbrener Katic, M. (2014) "Intelligent question—Answering systems: Review of research", *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1228-1233

Van Harmelen, F., Lifschitz, V. & Porter, B. (2008) "*Handbook of Knowledge Representation*", Oxford: Elsevier.

Vykhovanets, V. S. (2021) "The Notional Model of Knowledge Representation", *Journal of Physics: Conference Series* (Vol. 1864, No. 1). IOP Publishing., May.

Walelign, S., Tesfaye, D. & Kebebew, T. (2021) "Modelling and Designing of NLP Interface to Database for Afaan Oromoo", *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)* ISSN, pp. 2347-5552.

Creative Commons Attribution – NonCommercial 4.0 International License

Izvorni znanstveni rad
https://doi.org/10.31784/zvr.10.1.4
Datum primitka rada: 2. 11. 2021.
Datum prihvaćanja rada: 20. 2. 2022.

AUTOMATIZACIJA PRETVORBE PRIRODNOG JEZIKA U FORMALIZIRANI ZAPIS FNOK

Marina Rauker Koch
Profesor pedagogije i informatike, viša predavačica, Veleučilište u Rijeci, Vukovarska 58, 51 000 Rijeka, Hrvatska; *e-mail*: mrauker@veleri.hr

Sanja Čandrlić
Dr. sc., izvanredna profesorica, Fakultet informatike i digitalnih tehnologija, Sveučilište u Rijeci, Radmile Matejčić 2, 51 000 Rijeka, Hrvatska; *e-mail*: sanjac@inf.uniri.hr

Martina Ašenbrener Katić
Dr. sc., docentica, Fakultet informatike i digitalnih tehnologija, Sveučilište u Rijeci, Radmile Matejčić 2, 51 000 Rijeka, Hrvatska; *e-mail*: masenbrener@inf.uniri.hr

SAŽETAK

U radu je opisano unaprjeđenje sustava za automatsko odgovaranje na pitanja na temelju znanja iskazanog rečenicama prirodnog jezika. Sustav se implementira koristeći relacijsku bazu podataka. Ovaj sustav bit će temelj za razvoj web aplikacije za dobivanje odgovora na postavljena pitanja. Da bi se rečenice prirodnog jezika mogle upisati u relacijsku bazu podataka, treba ih pripremiti i formalno zapisati. Razvoj sustava za odgovaranje na pitanja temelji se na primjeni konceptualnog okvira Node of Knowledge (NOK) čiji je formalizirani zapis prikladan za upisivanje u relacijsku bazu podataka iz koje je moguće dobiti odgovore na postavljenja pitanja. U ovom radu pokazana je primjena aplikacije za automatsku pretvorbu rečenica engleskog jezika u formalizirani zapis. Aplikacija je testirana na primjeru sto jednostavnih rečenica engleskog jezika, a rezultati automatske pretvorbe uspoređeni su s rezultatima dobivenim ručnom obradom istih rečenica.

Ključne riječi: *Node of Knowledge (NOK), Formalizirani Node of Knowledge (FNOK), Predstavljanje znanja, Sustav pitanja i odgovora (QA)*