# Application of the NOK method in sentence modelling

**M. Rauker Koch*, M. Pavlić**, A. Jakupović*****
*Railway and technical school Moravice, Vrbovsko, Croatia
** Department of Informatics, University of Rijeka
*** Business Department, Polytechnic of Rijeka
marina.rauker@skole.hr, mile.pavlic@ris.hr, alen.jakupovic@veleri.hr

**Abstract - Knowledge representation is one of the areas covered by artificial intelligence. One of the methods for graphical representation of text expressed knowledge is the method NOK (Nodes Of Knowledge). NOK method enables transformation of text expressed knowledge into a graphical network of words and group of words. In this paper application of NOK method is presented. This application is based on sentences from an Aesop's Fable in Croatian ("Golden eggs in the chicken") and English ("The Goose with the golden eggs") version. In this way the applicability of this method on two natural human languages is presented, and similarities and differences that are partially conditioned by freedom of translators, and not only by differences in the syntax of the two languages, are observed.**

**Key words:** Node Of Knowledge, Text-based knowledge representation, Network-based knowledge representation.

## I. INTRODUCTION

AI studies computational modelling of human intelligence. Intelligent behaviour is conditioned by knowledge – human act based on what they know (or believe) about the world [1], [2]. One of the fields of interest of AI is knowledge representation and reasoning, which is suitable for reaching artificially intelligent behaviour. Many methods of converting knowledge into a form that can be encoded on a computer have been developed and they are applicable to different types of knowledge [3], [4].

Graphical method Nodes Of Knowledge designed for the presentation of sentences written in natural language is applied in this paper [5]. The formalization of this method is shown in [6]. The possibility of its application has been demonstrated on one text in Croatian and English. By analyzing produced graphic presentations answers to the following questions were obtained: Is there a difference in the complexity of making diagrams, which model is more complex, can answers to the same questions be read from both diagrams, regardless of differences in translation, are there any questions applicable on one model only and how can this problem be solved.

## II. REVIEW OF RELEVANT PUBLICATION

In the field of artificial intelligence different approaches to knowledge representation formalisms are defined [2], [3], [7]. Conceptual graphs [8], [9], [10], semantic networks [11], frames [12] and predicate logic [13] are traditional formalisms that eventually produced numerous other formalisms and methods for knowledge representation. Ontologies have had a significant impact in the field of knowledge representation in the past twenty years [14], [15], [16]. They emphasize the semantic components of knowledge and language [17]. Furthermore, the field of computer analysis of natural language (NLP) is also involved in the formalization of knowledge expressed in language and text [18], [19], [20], [21], [22], [23], [24], [25], [26], [27].

There are two approaches to the formalisation of text-expressed knowledge: declarative and procedural. Examples of declarative approaches are: BG (Basic Conceptual Graphs) [28], MULTINET (Multi-layered extended semantic networks) [29], HSF (Hierarchical Semantic Form) [30], RDF (Resource Description Framework) [31], and other methods based on the first order predicate calculus (FOPC). The result of the procedural approach is a rule system.

## III. RESEARCH MOTIVATION

Text as a knowledge carrier is not suitable for computer-supported knowledge dynamics (changes, paraphrasing, distribution, use, generation of new knowledge, abandonment, etc.). It should be specially formatted - formalised - in order to enable a computer to perform mathematical/logical operations. Nodes of Knowledge (NOK) method is a new method by which sentences written in natural language can be modelled. The purpose of graphical modelling sentences with NOK method is the creation of models (formal representation) of different forms of sentences (declarative, interrogatory, simple, complex, etc.) in order to verify the basic concepts of the method. Designed models and their verification represent the basis for the development of mathematical formalism for knowledge representation and reasoning based on a NOK method - first formalism is shown in [6].

The aim of this paper is to model the same text in two languages, English and Croatian, compare derived models and notice similarities and differences in the model that are conditioned by the freedom of translators.

## IV. NODES OF KNOWLEDGE

NOK method belongs to a group of semantic networks [11] where knowledge is represented as a graph [32]. Its

aim is to present textual knowledge as a network of knowledge [29].

NOK is a method for modelling by which textual knowledge is represented graphically in the form of a diagram DNOK (Diagram Nodes Of Knowledge) [33]. Sentences written in human language are converted to graphical model from which it is possible to interpret the meaning because this model kept it. NOK method consists of nodes (node, processing node and linker) and connections between them which contains the role identifier and they together form a network of knowledge.

Characteristics of the NOK method are [5]:

- simplicity – it has only three elements: nodes, links and the role identification in link
- expressiveness – ability to represent different levels of abstraction
- simplicity of reading – it is possible to start reading knowledge from any node

### A. *Nodes*

Node represents a word or group of words in a sentence that has some meaning. For example, in the sentence "The toy is very old." nodes are: The toy, is, very, old. Graphical representation of nodes and concepts in the NOK method is given in Table I.

Node name is an attribute of node and it gives it semantic meaning. The name may also consist of several words that have a meaning. To implement the method it is necessary to appoint a node name that clearly identifies it in the network. Homonyms and synonyms can be used [32], [35] but should be linked to their originals on contextual level of knowledge.

We can distinguish between [34], [35], [36]:

- Node – ordinary, static, entity, concept, term
- Context node – node of a higher degree of abstraction (group, sort, class, type)
- Data node – place where data are kept
- Process node – node that connects other nodes (using links) and with them creates a more complex presentation of knowledge. It presents knowledge that cannot be displayed by ordinary node such as connections between multiple nodes. The most common types of words in the names of process nodes are verbs.

### B. *Links*

Links connect the nodes in the network of knowledge. They do not have names, but may have role identifications that belong to the nodes [35]. The names of role identifications are questions which connect two adjacent nodes. This link is represented with line, with or without arrows. Arrows are used when connecting nodes on different levels of abstraction [33] or for direct connection of process nodes.

The cardinality of a link is always (1, 1) (1, 1) [35].

TABLE I. SYMBOLS OF NOK METHOD

| Simbol | Concept |
|---|---|
| goose | NODE - part of the knowledge in the reality or minds |
| 1. going | PROCESSING NODE - action, event, occurrence |
| to | LINKER – complex sentence of more parts |
| What? | LINK (one-way) and QUESTION (role indentifier) – connecting nodes |

## V. MODELLING OF SENTENCES

Using the NOK method sentences of natural language can be displayed by using nodes and links between them. From the resulting model, without access to the original text, answers to asked questions can be obtained. The applicability of NOK method will be shown in Croatian and English language on the example of the same Aesop's[1] fable „Zlatna jaja u kokoši“ [37]. The fable is titled "The Goose with the golden egg" in English [38].

For modelling sentences from the fables the symbols shown in Table I were used. Specifying the nodes, processing nodes and linkers, when there was a need for them, and connecting with links which were assigned the role and the meaning of nodes with questions in sentence, both graphic representation were constructed, in Croatian (Figure 1.) and in English (Figure 2.).

Fable „Zlatna jaja u kokoši“ in Croatian has the following 4 sentences:

1. „Neki gospodar imao je među mnogim svojim kokošima i jednu takovu koja je nesla zlatna jaja.
2. On je mislio, da joj u trbuhu mora biti mnogo zlata, pa je, ne razmišljajući mnogo, zakolje.
3. Ali se ljuto prevario, ne našavši u njoj ništa drugo nego li i u ostalim kokošima.
4. Tad se uzalud sam korio što je bio lud, da se polakomio za mnogim blagom, pa tako izgubio i ono što je imao.“

The same fable, "The Goose with the golden egg", in English according to [38] has sentences as follows:

1. „One day a countryman going to the nest of his Goose found there an egg all yellow and glittering.
2. When he took it up it was as heavy as lead and he was going to throw it away, because he thought a trick had been played upon him.

[1] Aesop (620 - 560. BC.), slave and storyteller who lived in Ancient Greece

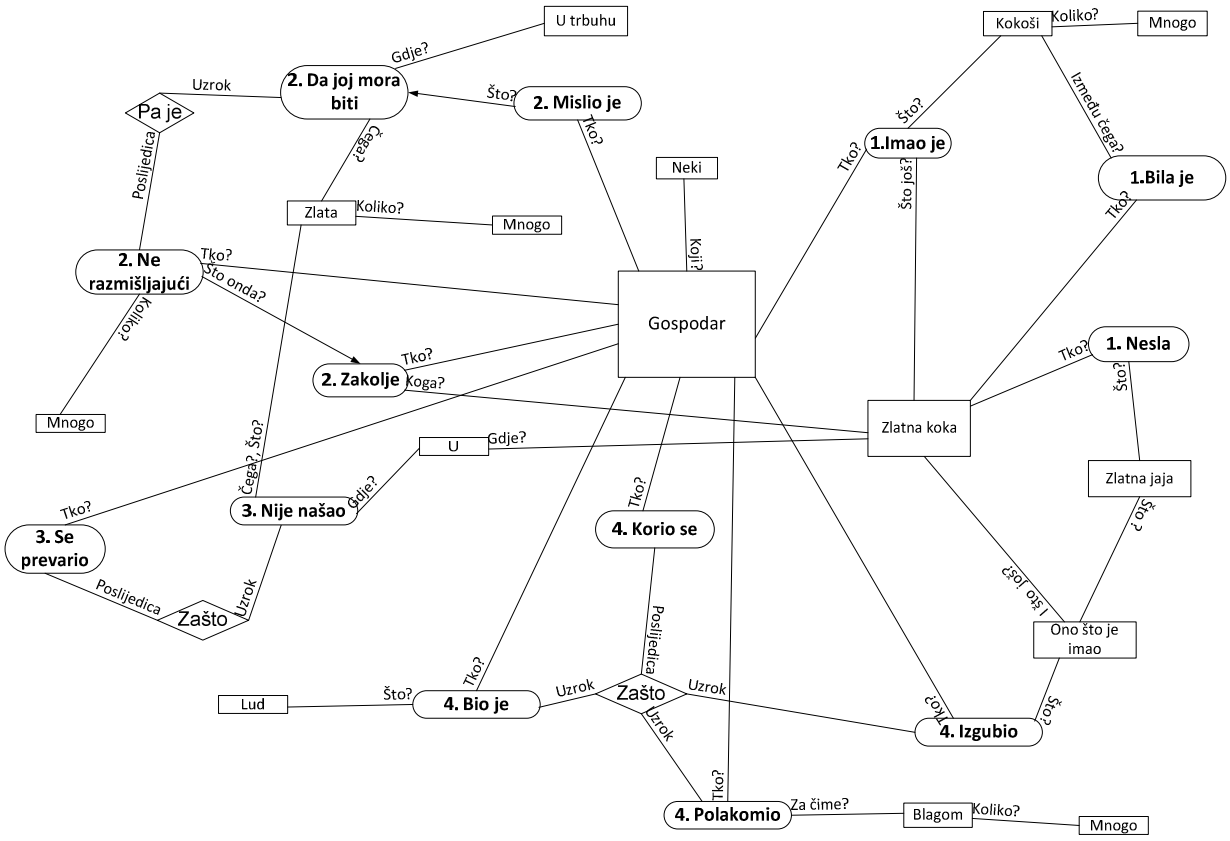


Figure 1. Representation of the fable “Zlatna jaja u kokoši” by NOK method

3. But he took it home on second thoughts, and soon found to his delight that it was an egg of pure gold.
4. Every morning the same thing occurred, and he soon became rich by selling his eggs.
5. As he grew rich he grew greedy; and thinking to get at once all the gold the Goose could give, he killed it and opened it only to find nothing.“

From the number of sentences, or words and groups of words in them, it is obvious that it will be more complex to make diagrams of the English version of the fable as it will consist of more nodes and connections between them.

In complex sentences, and between several separate clauses occurs causal - consequence relationship between certain parts that is displayed by using a linker node "why" which connects the causes and consequences. There is a possibility of multiple causes and / or consequences which determines the complexity of the model.

### C. *Interpretation rules*

In the sentence modelling the following rules are applied:

- Each noun becomes a node
- Each verb becomes a processing node
- Each node represents a single term that can consist of a group of words (eg. „zlatna jaja“, „was going to throw“)
- Each node must be linked to at least one node
- Process node is numbered by the number of sentence from which it was taken
- Pronouns are eliminated and the related network is connected with the noun to which eliminated pronouns refer.
- Related actions are connected with a rhombus, and this connection represents cause-and-effect relation.
- Node relationships are identified by questions that determine the semantic connections between nodes.

### D. *Comparison*

At first sight it is evident, by comparing the texts of the fable, that there is a difference in the translation which is determined by customs, folk tradition and the environment for which it is intended and, naturally, by the autonomy of translators. The title itself reveals a difference in choosing the animal which lays the golden eggs: it is a hen in the Croatian version and a goose in the English one.

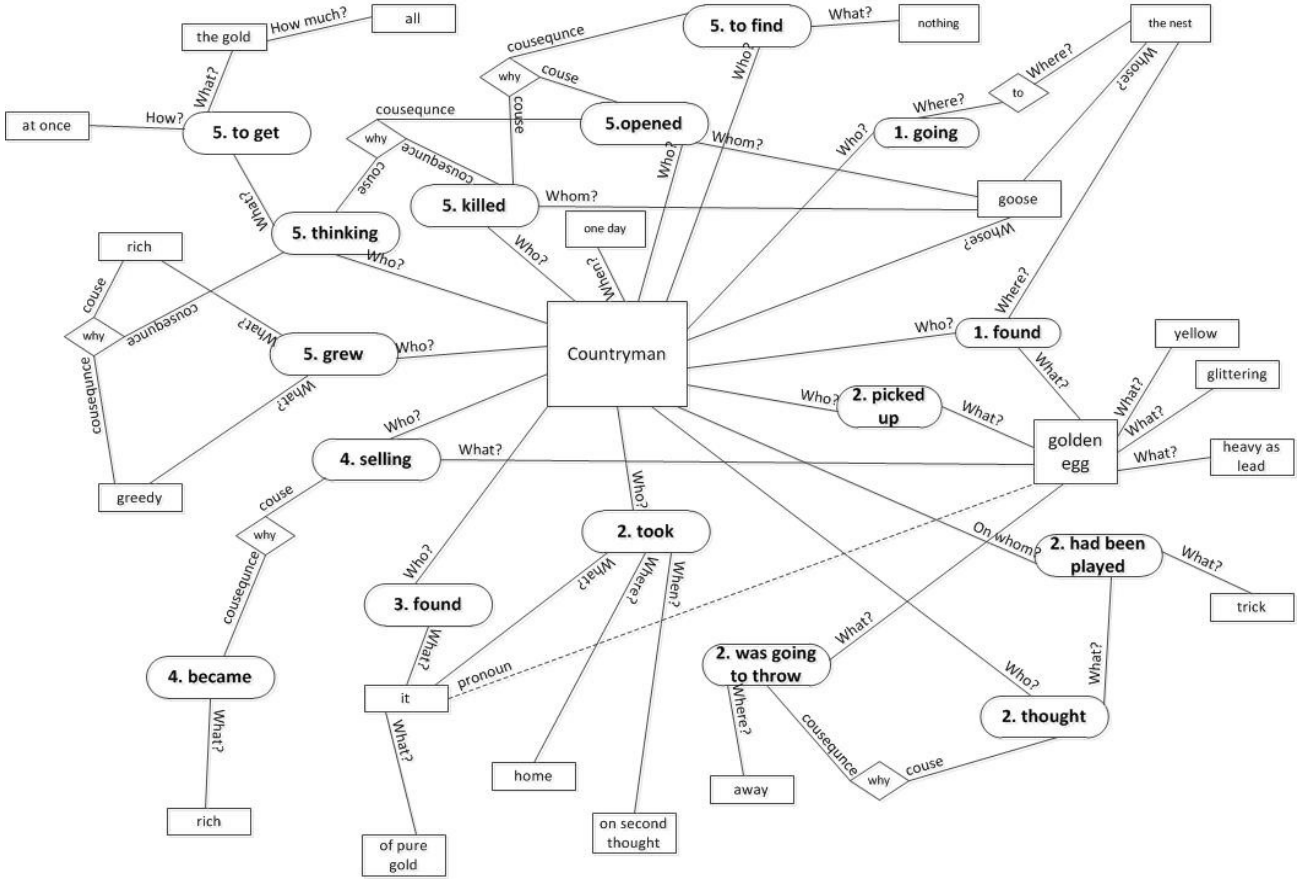


Figure 2. Representation of the fable “The Goose with the golden egg” by NOK method

From this we can conclude that translation is not a mathematically rigorous mapping of terms in one language in the same terms in other language. Also, two different translators will not translate the same text in the same way. Therefore translation does not provide a single best solution. Translators believe that their translation preserves the message and knowledge that the author wanted to convey to the reader.

Despite this, the fable has the same moral and is told in a similar way. This means that the knowledge on the high level of abstraction contained in the fable can be stored in both languages in a different way. This brings us to the conclusion that it is possible to keep text-based representation of the knowledge in a variety of models.

The diagram (Figure 1. and 2.) shows a difference in the presentation. Difference in the complexity of its construction is also evident. English translation of the fable is more complex in terms of the number of nodes and links that are displayed. The English version consists of five, and Croatian version of four sentences. Depending on the number of verbs, ie process nodes, each of these sentences can be divided into simple sentences in which difference becomes even greater. This is evident by the numbering of process nodes on the diagrams (Figure 1. and 2.) that indicate the number of the sentence where the node is.

What is significant for NOK method is getting answers to questions from the both created model. If we ignore the hen/goose difference in translation, we can get similar or the same answers to the same questions from the model, even though the arrangement of nodes in each model is different: for example, who laid the golden eggs?, who was greedy? (Figure 3.), who butchered the hen / goose? etc. On the other hand, there are questions applicable to one model only (for example, English version: who found the golden egg?, who is going to the nest?; Croatian version: who laid the golden egg?, who lost what he had?).

In the graphical representations of fables with NOK method contextual connections between the terms in the text are also recorded. In both versions, there are sentences without explicitly listed content-related terms. For example, in the following part of sentence "he killed it and opened it only to find nothing," one concludes from the context that this refers to the countryman and the goose, and the diagram supports that because the connections between nodes 5. killed, countryman and goose preserved it (Figure 2).

## VI. Conclusion

In this paper the same fable in two languages translated by two different authors was presented. For these translations a knowledge network was made, by using the NOK method.

NOK method is applicable to the modelling of a sentence written in a variety of natural human languages.

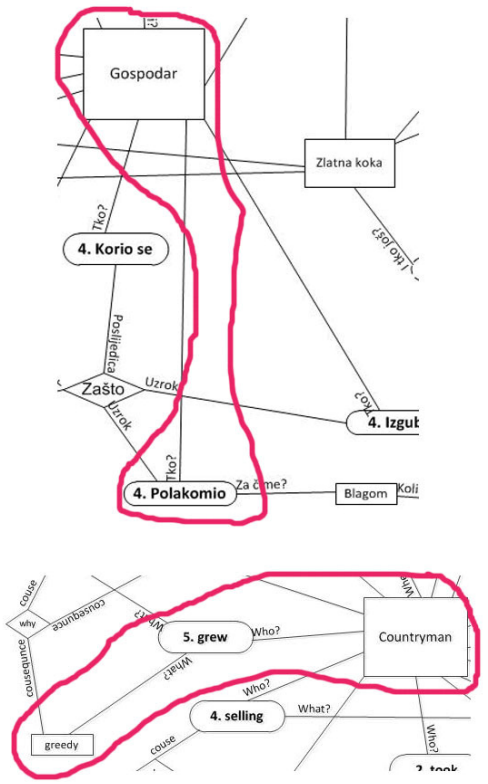


Figure 3. Answer to the question "Who grew greedy?"

Regarding the translator's autonomy, some parts of the model remain similar, but in some parts there are differences in the same story and in the same sentences.

By translating and merging of different models, not just by modelling of translation, we can get the model in different language from which it is possible to get the same meanings.

Choosing texts that are more similarly translated could provide a model with greater overlapping. Furthermore, it is possible to measure the difference in models and translations by a number of used concepts and words. It is possible to explore how NOK method can be used to translate texts and improve the process of translating one language in the DNOK and vice versa. NOK method is a new method for knowledge representation and it yet remains to investigate how it keeps knowledge and what its potential in building a computer system is.

## ACKNOWLEDGMENT

The research has been conducted under the project "Extending the information system development methodology with artificial intelligence methods" (reference number 13.13.1.2.01.) supported by University of Rijeka (Croatia).

## REFERENCES

[1] Jordan, M. I., & Russell, S., Computational Intelligence. In R. A. Wilson, & F. C. Keil (Eds.), The MIT Encyclopedia of the Cognitive Sciences (lxxiii - xc). Massachusetts: A Bradford book, 2001.

[2] Brachman, R., & Levesque, H., Knowledge Representation and Reasoning. San Francisco: Morgan Kaufmann Publishers, 2004.

[3] F. Harmelen, V. Lifschitz i B. Porter, Handbook of Knowledge Representation, Oxford: Elsevier, 2008.

[4] S. Russell i P. Norvig, Artificial Intelligence: A Modern Approach, Third edition, Prentice Hall, 2010.

[5] M. Ašenbrener Katić, „Predstavljanje znanja: pregled područja", Odjel za informatiku, Sveučilište u Rijeci, kvalifikacijski rad, 2013.

[6] Jakupović, A., Pavlić, M., & Dovedan-Han Z., Formalisation Method for the Text Expressed Knowledge. Expert Systems with Applications. In Press, Accepted Manuscript. from: http://www.sciencedirect.com/science/article/pii/S0957417414001353, 2014., 8.3.2014.

[7] Russell, S. J., & Norvig, P., Artificial Intelligence: A Modern Approach. New Jersey: Pearson Education, 2010.

[8] Sowa, J. F., Conceptual graphs for a data base interface. IBM Journal of Research and Development. 20(4). 336 - 357., 1976.

[9] Sowa, J. F., Semantics of conceptual graphs. Proceedings of the 17th annual meeting on Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics. 39 - 44. 1979.

[10] Sowa, J. F., Knowledge representation: logical, philosophical and computational foundations. Pacific Grove: Brooks/Cole Publishing Co., 2000.

[11] Quillian, R., Semantic memory. Semantic Information Processing. MIT Press. 227 - 270., 1968.

[12] Minsky, M. , Framework for Representing Knowledge. Technical Report. Cambridge. 1974.

[13] McCarthy, J. Programs with common sense. Teddington Conference on the Mechanization of Thought Processes. Teddington: National Physical Laboratory. 75 - 91. 1958.

[14] Gomez-Perez, A., Fernandez-Lopez, M., & Corcho, O., Ontological Engineering. London: Springer., 2004.

[15] Grimm, S., Hitzler, P., & Abecker, A. (2007).Knowledge Representation and Ontologies, Logic, Ontologies and Semantic Web Languages. In

[16] Gruber, T. R., A translation approach to portable ontology specifications. Knowledge Acquisition.5(2). 199 - 220., 1993.

[17] Garcia-Crespo, A., Ruiz-Mezcua, B., Lopez-Cuadrado, J. L., & Gonzalez-Carrasco, I., Semantic model for knowledge representation in e-business. Knowledge-Based Systems. 24 (2). 282 – 296, 2011.

[18] Allen, J., Natural Language, Knowledge Representation and Logical Form. In Bates, M., & Weischedel, R. (1993). Challenges in Natural Language Processing. Cambridge University Press. 1993.

[19] Allen, J., Natural Language Understanding. Boston: Addison-Wesley. 1994.

[20] Allen, J., Swift, M., & Beaumont, W., Deep Semantic Analysis of Text. Proceedings of the 2008 Conference on Semantics in Text Processing. Stroudsburg: Association for Computational Linguistics.343 - 354., 2008.

[21] Chomsky, N., Three models for the description of language. IRE Transactions on Information Theory. Volume 2. 113 – 124, 1956.

[22] Chomsky, N., On certain formal properties of grammars. Information and Control.2(2).137-167. 1959.

[23] Chomsky, N., Studies on Semantics in Generative Grammar. Janua Linguarum Series Minor. Mouton De Gruyter, 1980.

[24] Jurafsky, D., & Martin, J. H., Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. New York: Prentice Hall PTR. 2000.

[25] McQueen, T., Hopgood, A. A., Allen, T. J., & Tepper, J. A., Extracting finite structure from infinite language. Knowledge-Based Systems. 18(4-5). 135-141, 2005.

[26] Shapiro, S. C., & Rapaport, J. W., Knowledge representation for natural language processing, Natural Language Planning Workshop. New York: Northeast Artificial Intelligence Consortium. 56 - 77. 1987.

[27] Wermter, S., Riloff, E., & Scheler, G. Connectionist, statistical and symbolic approaches to learning for natural language processing. New York: Springer-Verlag., 1996.

[28] Chein, M., & Mugnier, M. L., Graph-based Knowledge Representation. London: Springer, 2009.

[29] Helbig, H., Knowledge Representation and the Semantics of Natural Language. London: Springer, 2006.

[30] Stanojević, M., & Vraneš, S., Knowledge representation with SOUL, Expert Systems with Applications: An International Journal, 33(1), 122 - 134., 2007.

[31] Tolle, K. Introduction to RDF, Johann Wolfgang Goethe – Universität Frankfurt am Main. from: http://www.dbis.informatik.uni-frankfurt.de/~tolle/RDF/DBISResources/RDFIntro.html, 2000., 8.3.2013.

[32] M. Pavlić, M. Ašenbrener, A. Jakupović, A. Meštrović, S. Čandrlić i M. Ivašić-Kos, „Inteligentni informacijski sustavi“ u Razvoj poslovnih i informatičkih sustava CASE 25, Rijeka, 2013.

[33] M. Pavlić, Development of a method for knowledge modeling 2013., Rijeka: Odjel za informatiku Sveučilišta u Rijeci, 2013.

[34] M. Pavlić, A. Jakupović, A. Meštrović: “Nodes of knowledge method for knowledge representation”, Informatol. 46, 2013., 3, 206-214

[35] M. Pavlić, A. Meštrović, A. Jakupović, “Graph-Based Formalisms for Knowledge Representation”, Proceedings of the 17th World Multi-Conference on Systemics Cybernetics and Informatics (WMSCI 2013), Vol 2. 2013. 200-204.

[36] A. Jakupović, M. Pavlić i Z. Dovedan Han, „Formalisation Method for the Text Expressed Knowledge“ neobjavljeno, 2013.

[37] Ezop, „Basne“, http://lektire.skole.hr/pdf/ezop/ezop_basne.pdf, 10.12.2013.

[38] Aesop, „The Goose with the golden egg“, http://www.bartleby.com/17/1/57.html, 10.12.2013.