STORING INFORMATION VIA NATURAL LANGUAGE ADDRESSING – A STEP TOWARD MODELING HUMAN BRAIN MEMORY

Krassimira Ivanova, Stefan Karastanev

Abstract. *Our main goal in this paper is to propose a new approach for modeling human brain memory by storing information using only the names but not pointers. This approach gives us new point of view to the human brain processes. In other hand, it simplifies and speeds up the corresponded computer programs. It is called "Natural Language Addressing" (NLA). This approach is a possibility to access information using natural language words or phrases as direct addresses of the information in the computer memory. For this purpose the internal encoding of the letters is used to generate corresponded address co-ordinates.*

Keywords: *modeling human brain memory, ontologies, addressing, natural language addressing*

ACM Classification Keywords: *H.2 Database Management; H.2.8 Database Applications*

Introduction

The definition of the concept "intelligence" was given in [Mitov et al, 2010]. It follows from the General Information Theory (GIT) [Markov et al, 2007] and especially from the "Theory of Infos" [Markov et al, 2009].

The intelligence is a synergetic combination of:

– ***(primary) activity for external interaction***. This characteristic is basic for all open systems. Activity for external interaction means possibility to reflect the influences from environment and to realize impact on the environment;
– ***information reflection and information memory***, i.e. possibility for collecting the information. It is clear; memory is basic characteristic of intelligence for "the ability to learn";
– ***information self-reflection***, i.e. possibility for generating "secondary information". The generalization (creating abstractions) is well known characteristic of intelligence. Sometimes, we concentrate our investigations only to this very important possibility, which is a base for learning and recognition. The same is pointed for the intelligent system: "To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories";
– ***information expectation*** i.e. the (secondary) information activity for internal or external contact. This characteristic means that the prognostic knowledge needs to be generated in

advance and during the interaction with the environment the received information is collected and compared with one generated in advance. This not exists in usual definitions but it is the foundation-stone for definition of the concept "intelligence";

– ***resolving the information expectation***. This correspond to that the "intelligence is the ability to reach ones objectives". The target is a model of a future state (of the system) which needs to be achieved and corresponding to it prognostic knowledge needs to be "resolved" by incoming information.

In summary, the intelligence is creating and resolving the information expectation [Mitov et al, 2010].

The concept "intelligence" is a common approach for investigating the natural and artificial intelligent agents. It is clear; the reality is more complex than one definition.

Presented understanding of intelligence is important for realizations of the intelligent computer systems. The core element of such systems needs to be possibility for creating the information expectation as well as the one for resolving it. The variety of real implementations causes corresponded diversity in the software but the common principles will exist in all systems. Summarizing, the artificial system is intelligent if it has:

– Activity for external interaction;
– Information reflection and information memory;
– Possibility for generalization (creating abstractions);
– Information expectation;
– Resolving the information expectation.

Following the definition of the concept "intelligence" given in the General Information Theory (GIT) [Markov et al, 2007], the five main problems to be solved by the science “Artificial Intelligence” were defined in [Ivanova, 2013]. They are to develop more and more “smart”:

– sensors and actuators - to realize external interaction;
– memory structures - to learn;
– generalization algorithms - to make abstractions;
– prognostic knowledge generation - to create information expectation;
– resolving the information expectation - to reach objectives.

In this paper we discuss the second problem – “smart” memory structures. Our hypothesis is that we may build model of storing information which is similar but not the same as human brain memory.

The structure of this paper is as follows. In the next chapter we will remember the features of Natural Language Addressing (NLA), than we will discuss storing ontological information by NLA, and finally some practical experiments and implementations will be outlined. Paper ends with main conclusions and propositions for future work.

Natural Language Addressing

The main idea of Natural Language Addressing (NLA) was presented in [Ivanova et al, 2012; Ivanova et al, 2013a].

We usually make difference between human aspect of the concept "address" and its computer "understanding", where the basic element is an index couple: (name, address). In different sources the "name" is called "key", "concept", etc. The address usually is given by any "number", "pointer", etc.

The NLA approach is based on using human representation of the address (by natural language words) as computer memory address. NLA uses computer encoding of name (concept) letters as address of connected to it information. This way no indexes are needed and high speed direct access to information elements is available. It is similar to natural order addressing in a dictionary book where no explicit index is used but the concept by itself locates definition. For instance, let have the next concept and corresponded definition:

"London: The capital city of England and the United Kingdom, and the largest city, urban zone and metropolitan area in the United Kingdom, and the European Union by most measures."

In computer memory, for example, the definition may be stored at address "FF084920". The index couple is:

("London", "FF084920"),

i.e. at memory address "FF084920" the main text, "The capital … measures." will be stored. To read/write the main text, firstly we need to find name "London" in the index and after that to access memory address "FF084920" to read/write the definition. If we assume that name "London" in the computer memory is encoded by six numbers (letter codes), for instance by using ASCII encoding system London is encoded as (76, 111, 110, 100, 111, 110), than we may use these codes as direct address to memory, i.e. ("London", "76, 111, 110, 100, 111, 110").

One may remark that above we have written two times the same name and this is truth. Because of this we may omit this couple and index, and read/write directly to the address. For human this address will be shown as "London", but for the computer it will be "76, 111, 110, 100, 111, 110".

From other point of view, the array (76, 111, 110, 100, 111, 110) may be assumed as co-ordinates of point in multidimensional (in this case – six dimensional) information space and the definition can be stored in this point.

It is clear, the words have different lengths and, in addition, some phrases may be assumed as single concepts. This means that we need a tool for managing multidimensional information spaces with possibility to support all needed dimensions in one integrated structure.

The independence of dimensionality limitations is very important for developing new intelligent systems aimed to process high-dimensional data [Gladun, 2003]. To achieve this, one need information models and corresponding access methods to cross the boundary of the dimensional limitations and to obtain the possibility to work with information spaces with variable and practically unlimited number of dimensions. Such possibility is given by the Multi-Dimensional Information Model (MDIM) [Markov, 2004] and corresponded to it Multi-Dimensional Access Method (MDAM) [Markov, 1984]. Its advantages have been demonstrated in many practical realizations during more than thirty years. In recent years, this kind of memory organization has been implemented in the area of intelligent systems memory structuring for several data mining tasks and especially in the area of association rules mining [Markov et al, 2012].

The program realization of MDIM is called Multi-Domain Access Method (MDAM). For a long period, it has been used as a basis for organization of various information bases. There exist several realizations of MDAM for different hardware and/or software platforms [Markov, 2004].

The brain and the computer memories from point of view of NL-Addressing

The idea of NL-addressing was born under the influence of the models of the natural and artificial neuron networks. In general, they may be modeled by graphs. Special kinds of such graphs are ontologies. They represent knowledge for given domain and maybe assumed as partial models of human brain memory. Because of this, storing ontologies is important problem to be solved effectively. A possible solution is using the NLA for storing graphs [Ivanova et al, 2012; Ivanova et al, 2013b] and, respectively, ontologies [Ivanova et al, 2013a; Ivanova et al, 2013d]. It is clear; the human brain does not create indexes. The information processing in the brain looks like our model for NL-addressing. It is very interesting to provide research in this area.

Series of experiments aimed to show the possibilities of NL-addressing to be used for NL-storing of small, middle-size and large datasets were provided.

The experiments began with analyzing the easiest case: NL-storing dictionaries, and after that, NL-storing of thesauruses was realized. An experiment with WordNet thesaurus was provided using the program WordArM based on NL-addressing [Ivanova et al, 2013c].

Special attention was given to NL-storing ontologies [Ivanova et al, 2013d]. Experiments were provided with (both real and artificial) middle-size and large RDF-datasets as well as so called BigData [Markov et al, 2014]. Estimation of the results from series of experiments had shown that storing time:

(1) *Depends* on number of elements in a dataset’s instances;

(2) *Not depends* on number of instances in the dataset.

The second is very important for multi-processing because it means linear reverse dependence on number of processors. The same feature we may discover in human brain memory. It works in practically equal speed independently of the amount of stored information. What is important – the NL-addressing models only the possibilities of the left human brain hemisphere, i.e. – the linguistic storing and accessing the information.

The NLA storing time varies between 2.2 and 2.5 milliseconds.

The next graphic illustrates independence from size of the datasets. On Figure 1 the storing time of 100 millions data instances is illustrated. The regularity is the constant time for storing independently of the number of already stored instances. [Markov et al, 2015]

This time does not depend on the size of data sets, i.e. of the number of instances. For comparison with traditional indexed databases with logarithmic complexity, the graphic of function ***log n*** in comparison with graphic of NLA approach (RDFArM realization) is shown in Figure 2. [Markov et al, 2015]

The information processing in the human brain looks like our model for NL-addressing. Neurons react for less than 4 ms independently of the amount of stored information (Figure 3) [Forehand, 2009].

NL-ArM has similar to the neurons' individual time for reaction – 2.5 – 3 ms. In addition; it seems that the human brain does not create indexes. A very important feature of NLA is that the database can be updated dynamically without recompilation of the database as it looks like in the human brain.

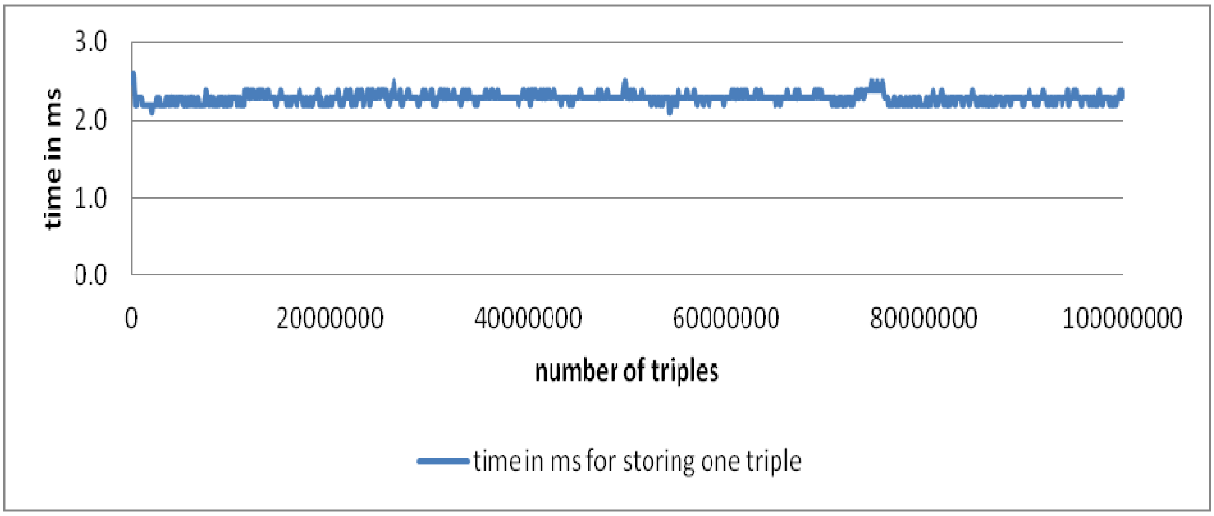

Figure 1. Storing time for 100 millions instances

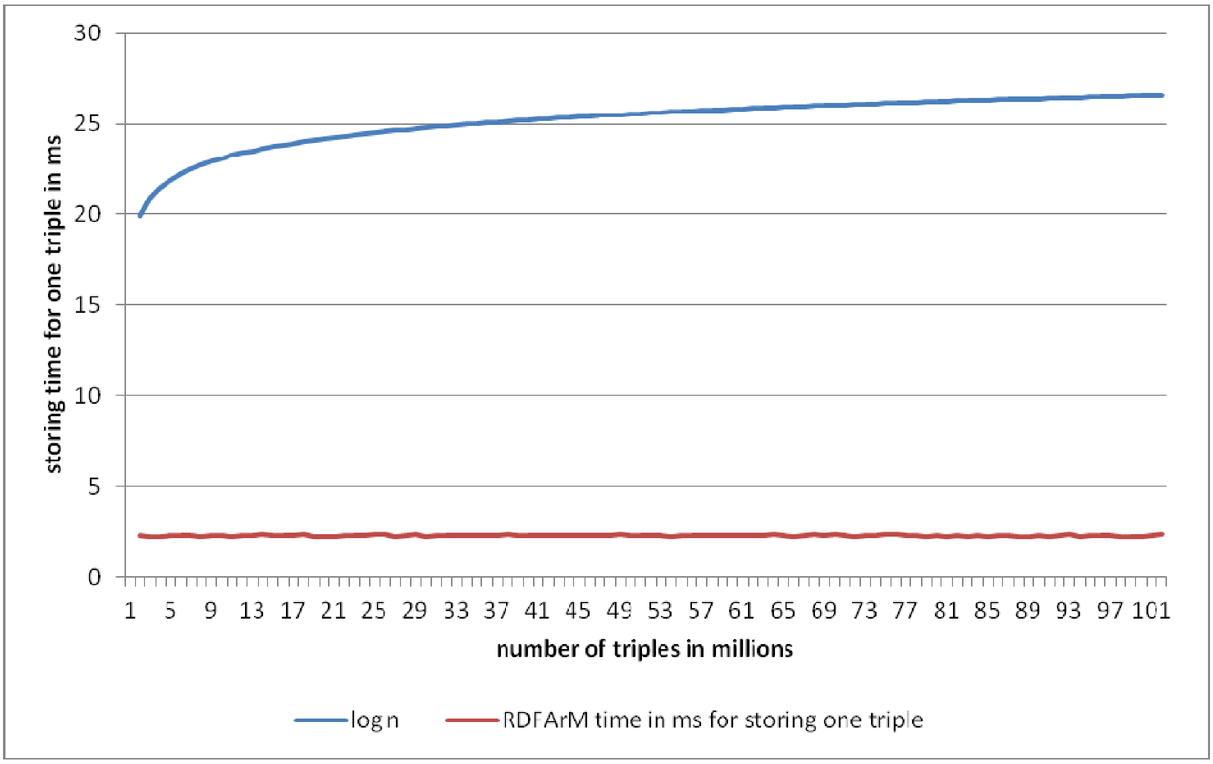

Figure 2. Comparison of log n and average time in ms for NLA storing one instance

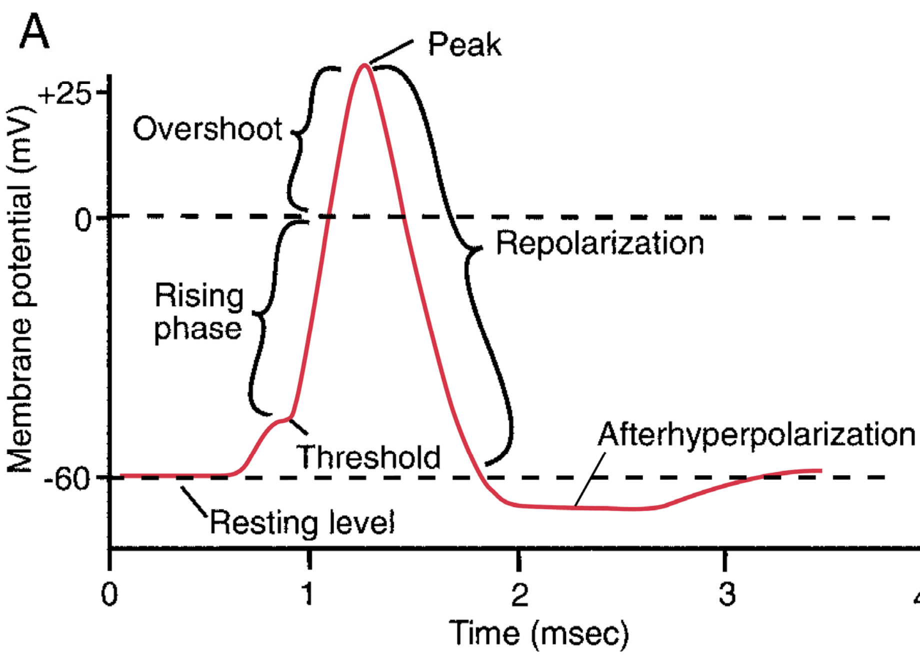

Figure 3. The phases of a neuron action potential:

Depolarization to threshold, the rising phase, overshoot, peak, repolarization, afterhyperpolarization, and return to the resting membrane potential. [Forehand, 2009]

Practical Aspects of Implementation and Using of NL-Addressing

NL-addressing is approach for building a kind of so called "post-relational databases".

The NL-addressing software realization is a part of an instrumental system for automated construction of ontologies "ICON" ("Instrumental Complex for Ontology designatioN") which is under development in the Institute of Cybernetics "V.M. Glushkov" of NAS of Ukraine [Palagin et al, 2011].

The work with ontologies is supported by a tool called OntoArM. Using it, the ontological elements can be organized in ontological graph spaces with variable ranges. There is no limit for the ranges of the spaces. Every ontological element may be accessed by a corresponding multidimensional space address (coordinates) given via word or phrase. Therefore, we have two main constructs of the physical organizations of OntoArM – ontological spaces and ontological elements. A separated ontology may be represented by OntoArM in one single archive. In addition, the NL-addressing permits accessing the equal names in different ontologies without any additional indexing or using of pointers, identification and etc. Only the NL-words or phrases are enough to access all information in all existing ontologies (respectively, graphs) [Ivanova et al, 2013b].

Conclusion

Storing information via natural language addressing as a step toward modeling human brain memory was discussed in this paper.

We remembered the features of Natural Language Addressing (NLA) and discussed storing ontological information by NLA, Some practical experiments and implementations were outlined.

The starting point of our work is the understanding of intelligence as a synergetic combination of: (1) *(primary) activity for external interaction; (2) information reflection and information memory*; (3) *information self-reflection*; (4) *information expectation;* and (5) ***resolving the information expectation***.

In this paper we discuss the second characteristic connecting the problem of developing the "smart" memory structures. Our hypothesis is that we have built a model of storing information which is similar but not the same as human brain memory.

We have implemented our new approach for storing ontologies using only the names but not pointers and this way to simplify and to speed up the corresponded computer programs. The approach is called "Natural Language Addressing" (NLA). This approach is a possibility to access information using natural language words or phrases as direct addresses of the information in the computer memory. For this purpose the internal encoding of the letters is used to generate corresponded address co-ordinates. The tool OntoArM, based on the idea of NLA approach, and experiments based on it were outlined.

What we gain and loss using NL-Addressing?

The loss is additional memory for storing internal NL-address structures. But the same if no great losses we will have if we will build balanced search threes or other kind in external indexing. It is difficult to compare with other systems because such information practically is not published.

The benefit is in two main achievements: (1) High speed for storing and accessing the information; (2) The possibility to access information immediately after storing without recompilation the database and rebuilding indexes.

NL-addressing looks like the information processing in the human brain. NLA has similar to the neurons' individual time for reaction and similar possibility to be in work condition permanently without timeouts for recompilation of the database after storing new information. The amount of stored information by NLA is practically not limited. All these characteristics of NLA give a good starting point for modeling human brain memory.

Acknowledgements

The paper is published with partial financial support from the "Scientific Research Fund" of University of Telecommunications and Posts, Sofia, Bulgaria, by the research project “Methods for modeling and evaluating of information services”.

Bibliography

[Forehand, 2009] Forehand, Cynthia J. "The action potential, synaptic transmission, and maintenance of nerve function." Medical Physiology: Principles for Clinical Medicine (Rhoades RA, BD, eds., Lippincott Williams & Wilkins, a Wolters Kluwer business, ed) (2009): 38-64. http://www.web.uwa.edu.au/__data/assets/pdf_file/0007/786652/booklet3.pdf

[Gladun, 2003] Gladun, V. P “Intelligent systems memory structuring”, International Journal Information Theories and Applications, 10(1), 2003, pp. 10–14.

[Ivanova et al, 2012] Kr. Ivanova, V. Velychko, Kr. Markov. Storing RDF Graphs using NL-addressing. In: G. Setlak, M. Alexandrov, K. Markov (ed.), Artificial Intelligence Methods and Techniques for Business and Engineering Applications. ITHEA® 2012, Rzeszow, Poland; Sofia, Bulgaria, ISBN: 978-954-16-0057-3, pp. 84 – 98.

[Ivanova et al, 2013a] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Introduction to the Natural Language Addressing”, International Journal "Information Technologies & Knowledge" Vol.7, Number 2, 2013, ISSN 1313-0455 (printed), 1313-048X (online), pp. 139–146.

[Ivanova et al, 2013b] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, “Introduction to Storing Graphs by NL-Addressing”, International Journal “Information Theories and

Applications", Vol. 20, Number 3, 2013, ISSN 1310-0513 (printed), 1313-0463 (online), pp. 263 – 284.

[Ivanova et al, 2013c] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Dictionaries and Thesauruses Using NL-Addressing", International Journal "Information Models and Analyses" Vol.2, Number 3, 2013, ISSN 1314-6416 (printed), 1314-6432(online), pp. 239- 251.

[Ivanova et al, 2013d] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Ontologies by NL-Addressing", IVth All–Russian Conference "Knowledge-Ontology-Theory" (KONT-13), Novosibirsk, Russia, 2013, ISSN 0568 661X, pp. 175 – 184.

[Ivanova, 2013] Krassimira Ivanova. The five main problems of the "Artificial Intelligence". International Conference „Knowledge Management and Competitive Intelligence", Vol 8, Kharkov, Ukraine, 2012, p. 6-7.

[Markov et al, 2007] Kr. Markov, Kr. Ivanova, I. Mitov. Basic Structure of the General Information Theory. Int. Journal "Information Theories and Applications", Vol.14/2007, No.:1, pp.5-19.

[Markov et al, 2009] Kr. Markov, Kr. Ivanova, I. Mitov. Theory of Infos. Int. Book Series "Information Science & Computing" – Book No: 13. Intelligent Information and Engineering Systems, Sofia, 2009, pp.9-16.

[Markov et al, 2012] Naidenova X. et al. "Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems". Release Date: July, 2012. p. 367; Chapter 7: Kr. Markov, K. Vanhoof, I. Mitov, B. Depaire, Kr. B. Ivanova, V. Velychko, V. Gladun. "Intelligent Data Processing Based on Multi-dimensional Numbered Memory Structures". IGI Global, USA DOI: 10.4018/978-1-4666-1900-5.ch007, pp. 156 – 184.

[Markov et al, 2014] Kr. Markov, Kr. Ivanova, K. Vanhoof, B. Depaire, V. Velychko, J. Castellanos, L. Aslanyan, St. Karastanev. Storing Big Data Using Natural Language Addressing. In: N. Lyutov (ed.), Int. Sc. Conference "Informatics in the Scientific Knowledge", VFU, Varna, Bulgaria. 2014. ISSN: 1313-4345. pp. 147-164.

[Markov et al, 2015] Krassimir Markov, Krassimira Ivanova, Koen Vanhoof, Vitalii Velychko, Juan Castellanos, „Natural Language Addressing", ITHEA® Hasselt, Kyiv, Madrid, Sofia, IBS ISC No.: 33, 2015, ISBN: 978-954-16-0070-2 (printed), ISBN: 978-954-16-0071-9 (online), 315 p.

[Markov, 1984] Markov Kr. A Multi-domain Access Method.//Proceedings of the International Conference on Computer Based Scientific Research. Plovdiv, 1984. pp. 558-563.

[Markov, 2004] Markov, K. Multi-domain information model. Int. J. Information Theories and Applications, 11/4, 2004, pp. 303-308.

[Mitov et al, 2010] I. Mitov, Kr. Markov, Kr. Ivanova. The Intelligence. Plenary paper. Third International Scientific Conference “Informatics in the Scientific Knowledge”. University Publishing House, VFU “Chernorizets Hrabar”, 2010. ISSN: 1313-4345. pp. 7-13

[Palagin et al, 2011] Palagin A.V., Krivii S.L., Petrenko N.G., “Ontological methods and instruments for processing domain knowledge”, (А. В. Палагин, С. Л. Крывый, Н. Г. Петренко. Онтологические методы и средства обработки предметных знаний: монография/Луганск: изд-во ВНУ им. В. Даля, 2011. – 300 с.), (in Russian)

Authors' Information

Krassimira Ivanova – *Assoc. prof. Dr.; University of Telecommunications and Posts, Sofia, Bulgaria; Institute of Mathematics and Informatics, BAS, Bulgaria; e-mail: krasy78@mail.bg;*

Major Fields of Scientific Research: *Software Engineering, Business Informatics, Data Mining, Multidimensional multi-layer data structures in self-structured systems*

Stefan Karastanev – *Assist. prof.; Institute of Mechanics, BAS, Bulgaria;*

e-mail: stefan@imbm.bas.bg

Major Fields of Scientific Research: *Software Engineering, Data Processing and Mining, Data structures in information systems.*