Computer Science Journal of Moldova, vol.**18**, no.3(54), 2010

Abstracts of Doctor Habilitatus Thesis

Title: Study of language-theoretic computational paradigms inspired by biology
Author: Serghei Verlan
University: Université Paris Est – Creteil Val de Marne
Date of defence: 22 October 2010

Understanding the surrounding matter, the laws and the functioning of the universe, this is the main challenge of the humankind, appeared at the dawn of times. Significant progress was made during last centuries, especially in the field of physics, which ultimately permitted the engineering of new devices and the development of new technologies. The mathematics is in tight relation with all these discoveries, as it provided a formal basis for these achievements.

The challenges of the actual science more and more reside in the field of biology. This is quite natural as the fundamental questions about life: what distinguishes living and non-living matter and what is the organization of a living thing, are considered from the very old times. Recent developments in most areas of science permit to approach above questions in a methodological way giving a hope for substantial breakthrough.

Still the biology is an experimental science that accumulates a large collection of knowledge and that needs the help of other disciplines in order to explain and correlate the gathered information. The mathematics is the primary tool sewing this goal, for example, with differential equations in systems biology or statistical analysis for populations evolution.

The computer science also plays an important role in biological investigations. Large-scale databases, data mining, sequence analysis and protein structure prediction – these are several tasks that would never be possible without the arrival of the computers, computational theory and corresponding algorithms. The molecular biology highlighted the key role of DNA and of related mechanisms for the information processing, which is crucial for the understanding of the functioning of living organisms. The computer science traditionally deals with sequences, so it was natural to use the provided tools for the sequence-related research, forming the field of bioinformatics.

We note that computer science is specialized on the discrete representation of the universe and on the representation of information flow processes. Many key concepts from the field where borrowed from the biology. This is why it fits well to the biological modeling providing discrete models in contrast to a modeling with differential equations, which are continuous. This property stimulated the investigation of (computational) models and operations issued from biology. In this case real biological phenomena are abstracted and a discrete system based on the involved operations and functioning is proposed.

There are two motivations for the investigation of such models. The first motivation is very clear – since the model represents an abstraction of a biological phenomena, it is possible to describe the last one in precise terms and provide simulations, estimations and predictions. A close relationship with target biological system permits to express it in a short and clear manner giving the hope to extract additional knowledge about the model. This approach, very nice in theory, encounters important difficulties while faced to the practice. Usually, simple models are too abstract for obtaining nontrivial relations for the initial biological system, while more complete models encounter rapidly the combinatorial explosion of their size, which makes them very difficult to analyze. The success of the approach relies on a deep knowledge of the target phenomena that permits to find a good balance between abstraction and adequateness.

The second motivation comes from the fact that most successful applications of computer science are based on ideas borrowed from the biology. During the evolution of living organisms, the nature provided solutions for many complicated problems encountered and it is very fruitful to adapt these solutions to different problems. Ant colony optimization, cellular automata, neural networks, evolutionary algorithms are some examples of such an approach.

This thesis is centered around two main topics: insertion-deletion systems and P systems. The first part does a systematical study of the operations of insertion and deletion. While it is possible to consider them as biologically inspired operations, we perform in the thesis a pure theoretical investigation in terms of the theory of formal languages. Thus, our research provides new classes of formal grammars which extend the Chomsky hierarchy by introducing new levels. We introduce a new proof method and show a series of computational-completeness and non-completeness results and discuss possibilities for the extension of the computational power for the latter case.

The research on P systems follows a different motivation. Primary, they represent a general framework for distributed multiset rewriting, which captures important processes in cell biology. In contrast to traditional approaches in systems biology, P systems provide a discrete framework for the representation of molecular interactions[1] that focuses on the structure of the system and on the identification of its components. The field of application for P systems is not limited to systems biology, the topic incorporates concepts from cell biology and aims to propagate them to all fields of the computer science.

The first part of our work targets the core of the P systems framework – its formal definition, which, surprisingly, was not always clear; moreover the introduction of new concepts like the minimal parallelism lead to different interpretations by different authors. We provide a single formal definition, which covers most of the classes of P systems with static structure known up to now.

The second part of Chapter 3 deals with one of most important models in the area of P systems – P systems using only communication, i.e., the objects present in the system cannot evolve, but can only be moved from one

---

[1] We remark that other discrete models like Petri nets or process algebra are also used, each of them having advantages and limitations.

compartment to another. We generalise the idea of co-transporters from the cell biology and end up with an interesting computational model based on the synchronization of signals. It is worth to note that the obtained model generalises all previous communication- only models of P systems.

The last part of Chapter 3 deals with two concrete problems. The first problem, the problem of the synchronization on a tree, is the generalization of the firing squad synchronization problem for cellular automata, where a linear communication structure is replaced by a hierarchical one. The second problem, the construction of universal P systems of small size, is closely related to the problem of the construction of small size universal Turing machines, which is a fundamental problem of the computer science.

The last part of the thesis presents an exploratory topic, where we performed all the stages for the construction of a new model – we start with a biological phenomena, give its formalization and start theoretical investigations. The closeness of the obtained model to the initial biological subject permit us to affirm that results that are obtained theoretically, can be verified in vitro.