REVIEW

by Prof. Dr. Hristo Stefanov Kiskinov, PhD,

Professor at Plovdiv University "Paisii Hilendarski" (PU),

Faculty of Mathematics and Informatics (FMI)

of a dissertation for the award of the educational and scientific degree **"Doctor" (PhD)**
by: Area of Higher Education 4. *Natural Sciences, Mathematics and Informatics*;
Professional Field 4.6. *Informatics and Computer Sciences;*
Doctoral Program *Informatics*.

Author: Nikolay Georgiev Handzhiyski.
Title: "An Iterative Parsing Algorithm with Application in the Profiling of Parsers".
Scientific Supervisor: Prof. Dr. Elena Petrova Somova.

1. General Presentation of the Procedure and the Presented Materials

By order No. RD-21-654 dated 21.03.2024 of the Rector of the Plovdiv University "Paisii Hilendarski" (PU), I have been appointed as a member of the scientific jury to ensure a procedure for the defense of a dissertation work on the topic "*Iterative Parsing Algorithm with Application in the Profiling of Parsers*" for the acquisition of the educational and scientific degree "doctor" (PhD) in the field of higher education 4. *Natural sciences, mathematics and informatics*, professional direction 4.5 *Informatics and Computer Sciences*, doctoral program *Informatics*. The author of the dissertation is Nikolay Georgiev Handzhiyski - PhD student in full-time study at the Department of Computer Informatics at the Faculty of Mathematics and Informatics at the Paisii Hilendarski University, with scientific supervisor Prof. Dr. Elena Petrova Somova. At the first non-attendance meeting of the scientific jury, held in the period 22.03.2024 - 23.03.2024 and reflected in Protocol No. 1, I am appointed to prepare a review.

The set of paper and electronic materials presented by the doctoral student Nikolay Georgiev Khandzhii is in full compliance with Article 36 (1) of the Regulations for the Development of the Academic Staff of the PU (RDASPU) and includes the following documents:

– a request to the Rector of the PU to disclose a procedure for the defense of a dissertation work;
– curriculum vitae in European format;
– protocol from the extended departmental council related to reporting the readiness to open the procedure and preliminary discussion of the dissertation work;
– dissertation work;
– abstract in Bulgarian and English;
– a list of scientific publications on the topic of the dissertation, including 6 titles;
– copies of scientific publications - 6 pcs.;
– declaration of originality and authenticity of the attached documents;
– references for compliance with the minimum national requirements.

The documents are very carefully prepared and I have no objections to them.

2. Brief Biographical Data of the Doctoral Student

Nikolay Georgiev Handzhiyski was born on 19.02.1995. In 2010, he obtained the bachelor's degree at the FMI of PU "Paisiy Hilendarski" majoring in "Informatics", and during his studies he was a student of the Cepus program at the Johannes Kepler University in Linz, Austria and under the Erasmus program at Koblenz-Landau University in Koblenz, Germany. In 2019, he obtained the "master's" degree again at "Paisii Hilendarski" PU, majoring in "Software Technologies". Since 2019, he has been a full-time doctoral student at the Department of Computer Informatics at FMI at PU.

From 2012 until now he has been working as a programmer in the company Software Systems - Bulgaria (headquartered in Plovdiv), and from 2018 until now as the manager of a company for software products in Offenburg, Germany, where he currently lives.

3. Relevance of the Topic and Appropriateness of the Set Goals and Tasks

Research related to translators (compilers and interpreters) and the mathematical grounds on which they are based (formal languages, generative grammars, abstract machines) is an evergreen topic in computer science. From the mid-1930s (1936), when the remarkable works on computability of Alan Turing, Emil Post and Alonso Church appeared, to the war years 1942-1945, when the first algorithmic programming language and compiler for it were created from Konrad Zuse, through the mid-1950s with Noam Chomsky's theoretical research on formal languages and their generating grammars, the rapid advance of electronic computing machines, and the creation of the first universal high-level programming languages FORTRAN and ALGOL and their compilers, up to the present day in this topic is constantly being worked on. That the topic continues to be extremely relevant today is easily established by following the publication activity of those working on this topic in the specialized journals.

The fact that the PhD student has obtained significant (in my opinion) and original results in a field that has been intensively worked on for nearly 90 years speaks to meaningful and original goals and tasks that the dissertation writer pursued in his dissertation.

4. Knowing the Problem

The overview of the scientific research on the problem, the selected literature and above all the obtained results speak unequivocally of deep knowledge in the researched area.

Moreover, I am firmly convinced that the PhD student knows the researched field very well. In addition to personal impressions (see point 12 in the review), there are also objective grounds for my conviction. It can easily be argued if we look closely at the overview chapter one of the dissertation. In it, for each discussed concept, the exact titles (and in most cases also the exact page in them) of the literature from where the presented information was taken are indicated. And nowhere are more than two titles cited at once. And all titles from the bibliography are cited... This clearly shows that the PhD student is familiar, and in detail, with each of the 191 sources cited in the dissertation.

5. Research Methodology

It is generally accepted that the standard methodology in research, leading to the achievement of specific results in the form of working software prototypes, is as follows: research of the area, identification of the problem, creation of a project model, synthesis of a suitable solution, implementation using known software tools, conducting experiments and analyzing the results obtained. However, I would venture to point out that in research leading to significant and original results, this is not exactly the case. Usually, after the initial study of the field and the selection of a problem to be solved, in the course of searching for its solution or its implementation, the researcher conducts additional research of the field, which often leads to the emergence of new ideas, respectively to a change of the problem to be solved etc. I.e. the process is, so to speak, iterative. In support of this hypothesis of mine is, for example, the fact that sometimes it is necessary to change the topic of the dissertation (in this case more than once). But in no case should this be taken as a shortcoming – on the contrary, it shows the sincere desire of the doctoral student to obtain significant and original results.

6. Characterization and Evaluation of the Dissertation Work

The dissertation "Iterative Parsing Algorithm with Application in Parser Profiling" is written in 179 pages and consists of an introduction, four chapters, a conclusion, a list of the author's publications on the topic, a list of noted citations, two appendices of 1 page each, a list of the literature

used and declaration of originality. The presentation also includes a list of used notations and abbreviations, 56 figures and 8 tables.

The two tables of the used terms from the literature and the terms introduced by the author, with their English translation and pointers for the source or the place of definition in the dissertation, respectively, are extremely useful (see point 11 of the review).

The introduction, in addition to formulating the goals and tasks, also contains a brief description of the structure of the dissertation work and the approval of the results.

The first chapter has an overview character, and in it are presented the main facts selected by the author from the theory of formal languages and abstract machines processing them, related to translation methods. The elements of translators performing lexical and syntactic analysis, as well as the generative grammars they use, are discussed. The chapter reviews finite state automata, stack automata, Turing machines, Markov algorithms, etc. The overview also includes well-known recognition and parsing algorithms based on context-free grammars.

The second chapter is devoted to the author's designed overall concept of data parsing, which he calls a parsing machine. In addition to the detailed description, it also contains formal definitions for the type of context-free grammars introduced by the author with additional components, called by him advanced grammars, as well as the phrase state machine introduced, which allows comparisons between classes of strings instead of comparisons between strings , facilitating the comparison of lexemes and phrases during parsing. Also defined are the formal languages that are defined by the new grammars with additional symbols. The parser machine shown contains different types of modules such as provider, scanner, lexer, parser, optimizer, builder and filter, all described in detail along with their functionalities. The chapter concludes with the definitions of different types of syntax trees and the construction commands based on which they are created.

The third chapter contains a detailed description of the algorithm created by the author, called Tunnel Parsing, which is built into the parser module of the parsing mashine and takes advantage of its added capabilities. The various objects that are created before parsing begins, based on a given advanced grammar, are described in detail. The list of these objects includes: execution stack, depth stack, replay stack, archive stacks, automata, reachable trees, conflicts, tunnels, routers, and control objects. The detailed description of the control objects contains the steps that the parser performs and effectively represents the pseudo-code of the Tunnel Parsing algorithm. The chapter contains an example of the defined objects that are used by the algorithm and concludes with an example implementation of the algorithm for a selected input string.

In the fourth chapter, a software tool specially created for the purpose of the thesis, called profiler, is presented, with the help of which a large number of context-free grammars and input data can be generated and experiments can be performed with parsing machines that are generated by different parser generators based on of these grammars. The chapter contains the description of a purpose-built template grammar language for imperative metaprogramming of grammars. The different grammars that define the scripts valid according to the templating language are shown. The chapter concludes with the interpretation of the results of four different experiments that were conducted using the profiler.

In the conclusion, the dissertation student summarized and systematized the main results, made a self-assessment of the contributions contained in the dissertation work and described the approbation of the obtained results. Specific prospects for further development of the dissertation topic are also formulated.

The list of used literature includes 191 (one hundred and ninety one) titles, two of which are in Bulgarian, two are in Russian, and the rest are in English.

The dissertation is written very competently, without spelling or stylistic errors. All concepts used or introduced are correctly defined, the exposition is strictly formalized, but at the same time understandable, and its flow is logical.

I do not detect "plagiarism" in the works of the author and the presented thesis in the sense of the Law on the Development of the Academic Staff in the Republic of Bulgaria.

7. Contributions and Significance of the Development for Science and Practice

I support the main contributions described by the PhD student in the current dissertation. Namely:

Scientific contributions:

- A conceptual model of a parsing machine is proposed, which is suitable for the implementation of various strategies, methods, approaches and algorithms, providing some additional possibilities compared to the existing ones;
- A new type of grammars, named by the author advanced grammars, with additional components (phrase symbols composed of rules and advanced symbols, according to the dissertation's terminology), with a structure close to context-free grammars in an augmented Backus-Naur form, are defined;
- A phrase machine model is proposed that pre-categorizes the different phrases in the parser grammar in order to speed up the analysis during parsing.

Scientific and applied contributions:

- The functionality of the introduced parsing machine was designed and described in detail;
- Designed and described in detail was the functionality of a new algorithm called Tunnel Parsing, parsing based on any non-left recursive advanced grammar, the parsing time being linear for some multivalued grammars and for any grammar that can be deduced from a deterministic automat with one stack (and if there is no recursion in the derived grammar, parsing can be performed with a constant amount of memory);
- A metaprogramming language for grammars was designed and created;
- A software tool called parser generator profiler was designed to allow performing tests (measuring the resources used during recognition and parsing) with parser machines created by different parser generators and compilers.

Applied contributions:

- A working prototype of a software tool has been developed - a profiler of parser generators (including a module for visualizing the results), allowing experimentation with a directly entered grammar or with a set of grammars programmatically derived from the tool;
- Experiments using the profiler were performed and analyzed.

In my opinion, the results obtained in the course of the dissertation research are significant and far exceed both in terms of quality and quantity the necessary required for the acquisition of the educational and scientific degree "doctor".

8. Assessment of the Dissertation's Publications

Част от резултатите от дисертационното изследване са представени в приложените 6 публикации на английски език, като всяка една от тях е индексирана в поне една от световноизвестните бази от данни с научна информация Web of Science (WoS) и SCOPUS. От тях 5 са в списания, индексирани в SCOPUS с SJR, а 4 са в списания, индексирани във WoS, като 2 от тях имат и Impact Factor (SJR) за съответната година. Те формират общо 168 т., което надхвърля над пет пъти минималните национални критерии по този показател, изискващи 30 т. Всичките 6 представени статии са в съавторство с научния си ръководител. Не е представен разделителен протокол, поради което считам, че участието в тях е равностойно. За мен, личният принос на кандидата в тях е несъмнен, като по-долу подробно съм описал мотивите за това мое становище.

Part of the results of the dissertation research are presented in the attached 6 articles in English, each of which is indexed in at least one of the world-knowed scientific information databases Web of Science (WoS) and SCOPUS. Of these, 5 are in journals indexed in SCOPUS with SJR, and 4 are in journals indexed in WoS, and 2 of them also have an Impact Factor (SJR) for the respective

year. They form a total of 168 points, which exceeds more as five times the minimum national criteria for this indicator, requiring 30 points. All 6 presented articles are co-authored with the scientific supervisor. No separation protocol has been presented, which is why I consider that participation in them is equal. For me, the personal contribution of the candidate in them is undoubted, and below (see point 9) I have described in detail the reasons for this opinion of mine.

One citation is also noted, something that is not mandatory according to the minimum national requirements, but makes a good impression.

9. Personal Participation of the PhD Student

For me, the candidate's personal contribution to the conducted dissertation research is unquestionable. Without wanting to belittle the significant contribution of the research supervisor (I hope Prof. Somova will forgive me for what I wrote), I believe that the leading author in the conducted research was the doctoral student. The grounds for this opinion of mine are both the doctoral student's perfect knowledge of the researched area (see point 4) and my personal impressions from our brief communication (see point 12).

10. Summary

The Summaries, written in Bulgarian and in English, have 32 pages, correspond to the requirements of RDASPU and contain the main results obtained in the dissertation work.

11. Critical Notes and Recommendations

I have no major critical notes. However, I will list a few not only non-essential, but also subjective.

– Some English terms are not well translated (in my opinion) into Bulgarian. For example, *powerset* should be translated as the *set of all subsets*. This is perhaps due to the small amount of literature in Bulgarian used by the PhD student. But the presence of a table with the terms used in the dissertation, their English prototype and the literary source, almost completely removes the problem.

– Some of the terms introduced in the dissertation do not sound/look good (especially in Bulgarian) to me. For example *advanced grammar*, *phrase symbol*, *advanced symbol* etc. Maybe *extended grammar* (*extended* instead of *advanced grammar*), *phrase-symbol* instead of *phrase symbol*, etc. would sound/look better. But here I must immediately point out that firstly - they are all strictly defined in the dissertation and secondly - that they are not easy to change, because they have been used that way in the already published articles. And here the presence of a table with the terms introduced in the dissertation, their English translation and the place of definition in the dissertation, makes this critical note nonessential.

– I would structure the overview chapter a little differently, but the current structure has its advantages (see point 4).

12. Personal Impressions

I have known the PhD student Nikolay Handzhiyski since his bachelor's student years, when I taught him discrete mathematics (16 years ago). Unfortunately, he did not show much interest in this field at the time. I gained actual personal impressions of him only around the preliminary acceptance of his thesis project, because I was included in the extended departmental council as a long-term lecturer in the discipline of Discrete Mathematics. I took the liberty of giving him some recommendations on the thesis, but only on layout and presentation, which he considered, accepted and followed through on, despite the short time frame for corrections. For my part, during our few brief discussions, I was able to personally convince myself of his encyclopedic knowledge of the researched field. I realized that during this years-long research the PhD student had written a huge amount of programming code and that he knew the field equally well both theoretically and practically. I was also impressed by his logical and critical thinking, but above all by his enthusiasm and hunger for creative research.

13. Recommendations for Future Use of the Dissertation Contributions and Results

A strong impression is made by the fact that the PhD student has correctly described six specific possibilities for continuing the research described in the dissertation work. There are certainly others. I hope to see corresponding articles describing their implementations soon. I would also recommend that in the future the places to publish become more and more prestigious, which will immediately affect the citation of the relevant articles.

Given the serious results achieved, as well as the large amount of theoretical knowledge and practical skills acquired by the PhD student during his research, I suggest that the PhD student consider publishing a monograph based on this dissertation work to be published by a serious publishing house.

I wish the PhD student to continue working in this interesting field and with the same enthusiasm.

CONCLUSION

The dissertation ***contains scientific, scientific-applied and applied results, which are an original contribution to the science*** and **meet all the requirements** of the Law on the Development of the Academic Staff in the Republic of Bulgaria (LDASRB), the Regulations for the Implementation of the LDASRB and the relevant Regulations of Plovdiv University "Paisii Hilendarski". I detect no plagiarism. The presented materials and dissertation results far exceed the minimum national requirements introduced by the Regulations for the Implementation of the LDASRB.

The dissertation work shows that the PhD student Nikolay Georgiev Handzhiyski possesses in-depth theoretical knowledge and professional skills in the scientific specialty Informatics, demonstrating qualities and skills for conducting research with obtaining original and significant scientific contributions.

Due to the above, I confidently give my **positive assessment** of the conducted research, presented by the above-reviewed PhD thesis, summary, achieved results and contributions, and ***I propose to the honorable scientific jury to award the educational and scientific degree "Doctor" (PhD)*** to Nikolay Georgiev Handzhiyski in the Area of Higher Education 4. *Natural sciences, Mathematics and Informatics*; Professional Field 4.6. *Informatics and Computer Sciences;* Doctoral Program *Informatics*.

14.05.2024 Scientific jury member: .....

Plovdiv Prof. Dr. Hristo Stefanov Kiskinov