

REVIEW

on a doctoral defense procedure featuring a PhD Thesis entitled:

**„AIDED DECISION MAKING FOR PUBLIC TRANSPORTATION OPTIMIZATIONS
USING BIG DATA“**

for awarding the PhD Degree

to

Candidate: **Georgi Kalinov Yosifov,**

In the Professional field 4.6 Informatics and Computer Sciences

Doctoral program: **„Software Technologies“ - Software Engineering,**

Faculty of Mathematics and Informatics (FMI),

Sofia University „St.. Kl. Ohridski“ (SU).

REVIEWER: Dr. Boris Blagovestov Shishkov, Assoc. Prof. in Context-Aware Information Systems at IMI – Bulgarian Academy of Sciences, Assoc. Prof. in Enterprise Information Systems at the University of Library Studies and Information Technologies.

I was elected a reviewer in the announced procedure on the 8th of November 2022, at a Scientific Jury meeting, summoned by Order № ПД-38-599 / 28.10.2022 of the Rector of SU. In my capacity of reviewer, I have received the documents (in electronic form) as included in the application package of Georgi Yosifov: the PhD Thesis, the Abstract (in both Bulgarian and English languages), a list of publications, copies of diplomas, and other material.

1. General characteristics of the PhD Thesis and the presented materials

The Candidate has published a PhD Thesis of 180 pages, entitled „Aided Decision Making for Public Transportation Optimizations Using Big Data” and organized in 6 chapters as follows: Introduction (6-15), Comparative analysis featuring the state-of-the-art research concerning data gathering methods and the determination the public transport traffic levels (16-30), Indexing the city traffic load, based on positioning data covering the public transport (31-91), Description of developed experiments handling software instruments (92-110), Predictability concerning the traffic load index (111-141), Conclusions and future work (142-146), , with 76 Figures presented as well as 17 tables and five appendices. The references list is featuring more than 100 items in Bulgarian language and English

languages. The contribution of this thesis is in several directions: (i) analytical classification methodology concerning the state-of-the art data gathering methods and the determination of traffic load levels (considering particularly the city of Sofia), as well as studies featuring corresponding analysis/forecast methods; (ii) algorithm for determination of traffic levels, validated by means of experiment scenarios; (iii) instruments, supporting the processing and management of data derived by the algorithm in carrying out the experiment scenarios.

2. About the Candidate

Having considered the curriculum vitae (that is just one page long), referring to the file "23.Автобиография.pdf", I was not able to get sufficient relevant information about the Candidate and I have no personal impressions on Georgi Yosifov since I have never met him. Still, it becomes clear from other documents submitted in his PhD application package that: (i) Georgi Kalinov Yosifov has graduated (Bachelor of Computer Science and Master of Informatics) from FMI - Sofia University „St. Kl. Ohridski“ (SU), scoring *very good*, while his Master Thesis has got an *excellent* grade; (ii) Afterwards he was appointed as an extramural PhD Student at SU - Doctoral program: „Software Technologies“ - Software Engineering; (iii) Georgi Yosifov has published 2 Springer (LNNS) Q4 publications indexed with SCOPUS as well as an ACM publication that is also indexed with SCOPUS; (iv) his SCOPUS h-index is 1.

3. Thorough analysis concerning the Author's scientific and application-oriented contributions as reflected in the PhD Thesis complemented by relevant publications

Enabling technologies concerning (wireless) telecommunications, hardware (featuring new miniature and multi-functional devices), and sensing, becoming widely accessible, have inspired (since 20 years ago today) the development and use of powerful computer-sensor networks capable of delivering “new generation” services to the end user, namely: such services that are not limited in the sense that user needs are considered static. On the contrary, many IT service providers have started considering the user within his/her context, adapting service provisioning accordingly, taking into account the user's location, time, and so on. Personal smart devices are capable of putting the user also in another role, namely: source of actual information. Further, sensors can be helpful in objectively establishing the user situation. Finally, computer networks and (wireless) telecommunications are effectively enabling the data flows concerning the service provider, the user and also relevant third parties. The DATA technologies developments, marking the past decade, based on an interdisciplinary and multi-disciplinary approach towards three classical scientific disciplines, namely: Mathematics, Statistics, and Informatics, but addressing the “tons” of data showering us nowadays (coming from cameras, Internet

“logs”, and so on, have opened new horizons: (i) We currently have not only the DATABASE perspective (and OLTP) but also a DATA WAREHOUSE perspective (and OLAP), considering “big” data that is not necessarily normalized, for the sake of conducting analyses and identifying trends; (ii) Classical Statistical instruments, such as Regression Analysis and Bayes Modeling are being put in a Computer Science perspective for the sake of enriching IT algorithmization in terms of probabilistic aspects; (iii) Machine Learning (and Artificial Intelligence) and Neural Networks are undercoming a new “boom” concerning training-data-related capabilities of effectively training IT systems. All this relates well to context-awareness (see above) since data (concerning the user situation), gathered by means of sensors (or in in another way), can be analyzed and used for predicting future situations. We have seen successful applications of this in tele-medicine, smart cities/homes, drone technology, and so on. Such aspects concern the functioning of a big contemporary city, such as Sofia – an EU capital and one of the big EU cities. Public transport appears to be a key issue as it concerns such cities and Sofia is especially characterized by multi-modality in this regard, with its citizens using bus servicing, tram servicing, trolley servicing, and metro servicing. In contrast to smaller towns, for example the Dutch city of Delft (where transport optimizations follow traditionally established transport “lines” and this works since overpopulation is not allowed and connectivity with adjacent places is adequately maintained), Sofia is challenged by population growth (that is fast in some periods), dynamic developments in some districts and lack of traditionally established communication “lines” (except for the central part of the city). Hence, public transport optimization in Sofia (and in other (European) big cities) is considered to be an actual challenge concerning the desired high-quality medical servicing of the population, the ease of move of people between their places of living and their places of work, the accessibility with regard to cultural and sports venues, entertainment, and so on.

As it can be seen from the PhD Thesis of Georgi Yosifov as well as from the abovementioned 3 publications, his research addresses this actual problem while his solution vision concerns issues discussed above, especially ones touching upon data analytics. This makes his work actual and relevant as it concerns problems that relate to current societal developments, in particular: developments that concern logistics, supply-chains, patient flows, and so on.

Considering the abovementioned thesis, I would present the (Sofia) public transport system (from the perspective of a desired flows optimization) as a CONTEXT-AWARE information system that has to adapt its “behavior” depending on situations (states) that concern USERS (people traveling using public transport), PUBLIC TRANSPORT VEHICLES, and PUBLIC TRANSPORT INFRASTRUCTURE. Relevant information in this regard can be gathered from personal smart devices (concerning the users), GPS positioning (concerning the vehicles), CCTV, and so on. Here I reckon that

the lack of explicit references to DATA FUSION algorithms is an **omission** because in my opinion they are of essential importance when considering big data gathered by diverse sources.

Predicting traffic-related situations (concerning the public transport) is an actual challenge considered in the thesis of Georgi Yosifov, as well as the possibility to analyze **historic (congestion-related) data** for the sake of identifying problematic “points” (such as “bottlenecks”) and proposing optimizations accordingly.

The PhD Thesis is mainly focused on the **data gathering** itself, „measures“ concerning **traffic load, software instruments for conducting experiments** and the abovementioned **predictions**.

The research goal is formulated as follows: “ **to support decision-making for the optimization of urban transport by determining, researching, and forecasting traffic load levels, using data,,**

With regard to the abovementioned goal, the scientific contribution concerning the PhD Thesis of Georgi Yosifov, is in the following directions:

- **A comparative state-of-the-art analysis has been conducted** (just on 15 pages), **focused on research concerning data gathering, relating this to traffic load (with respect to urban/public transport)**, see Chapter 2.
- A **model** has been proposed, concerning **computations featuring the traffic load index**, putting this in two perspectives: a city perspective, and a road segment perspective, inspired by real data gathered not only from Sofia but also from the Scottish city of Edinburg – see Chapter 3.
- **Requirements** have been identified, **modules** have been defined, and **classes** have been presented with regard to **software** that has been developed (using C#, in a .NET environment), fulfilling different decision-support aspects concerning the public/urban transport optimization, supported by big data and touching upon the gathering-related, processing-related, computation-related, and visualization-related issues that refer to corresponding experiments – see Chapter 4.
- **Simulation** (concerning schedules) has been specified and carried out, based on user-defined parameters.

As it concerns proof-of-principle / proof-of-concept with regard to what is proposed in the thesis, I consider the conducted experiments of just PARTIALLY relevant, since they are more into the goal achievement of the thesis, with no explicit PROOF DRIVE as it concerns the validity of what has been proposed in the thesis.

4. Results' approbation

As already mentioned, there are 3 articles submitted in the PhD application package of Georgi Yosifov; in all of them he is the first co-author. They are directly related to the content of the thesis:

- „Review of Urban Traffic Detection Approaches with Accent of Transportation in Sofia, Bulgaria” - Chapter 2.
- „Traffic Flow City Index Based on Public Transportation Vehicles Data” - Chapter 3, Chapter 4.
- „Predicting Traffic Indexes on Urban Roads based on Public Transportation Vehicle Data in Experimental Environment“ – Chapter 5.

As mentioned already, those publications, co-authored by Georgi Yosifov, are all SCOPUS-indexed - hence he meets the requirements of both ADASRB and SU. In most of those publications Georgi Kalinov Yosifov is a co-author (not the only author) – in those cases I assume equal contribution for each of the co-authors.

There is no plagiarism of the scientific works, proven accordingly following the legally established order.

5. Qualities of the abstract

In my opinion, the abstract meets in general the relevant requirements and correctly reflects the content of the thesis, even though, for example, what is presented in the second paragraph of the abstract (concerning “sensitive” data and GDPR) has no reflection in the thesis in the form of an explicit discussion. Further, I would like to note that I find the introductory part of the abstract insufficiently convincing since Georgi Kalinov Yosifov has not been good enough (in my opinion) in achieving an adequate alignment between the thesis goal (see Page 3 of the Abstract) and the identified problem (see Page 2). Next to that, a clear, thorough, and explicit statement is missing about the original scientific contribution of the thesis; on Page 43, the Candidate is listing conducted analyses, realized algorithms, and so on, but it remains insufficiently clear what makes those thing “original scientific contributions” and how they relate to the identified problem and the formulated goals. Taking this into account, I am in general satisfied with the content of the abstract.

6. Critical comments and recommendations

Georgi Yosifov's research is actual and interesting and his PhD Thesis is complemented by three SCOPUS-indexed publications. All this makes a good impression. Nevertheless, I have some critical

comments at the same time, that mostly concern the quality of the thesis, still stating my opinion that the main components (chapters) of the thesis are content-full and adequately related among each other:

- In my opinion, the thesis' Introduction is insufficiently convincing (similarly to the introductory part of the abstract, discussed above), since the content is not well-focused and the alignment concerning problem, goal, and contribution is not as good as it could be. I got a clear view on this because I firstly came through the abovementioned papers and afterwards I started reading the thesis. But I would not count just on the Introduction of the thesis (as it is currently written) to do the job.
- Here and there in the thesis the reader would encounter long, too wordy, and insufficiently clear sentences; I think that a careful proofread would lead to improvements (in terms of clarity and conciseness) of some paragraphs. Next to that, it is not always sufficiently clear whether a piece of content is original (from the Author of the thesis) or is reflection of what can be found in literature.
- The thesis lacks not only a SYSTEMATIC literature review concerning relevant disciplines, such as Statistics, Data Analytics, and Machine Learning, but also a thorough enough related work study, to justify the presented work, showing what has already been done by others (for solving the addressed problem), identifying “room for improvement” accordingly, and working in this direction.
- Georgi Kalinov Yosifov has not addressed (in the thesis) public values, such as privacy, accountability, transparency, and so on). In my opinion, this is an omission because any (technical) facilitations concerning urban public processes should be value- -sensitive).
- The Author has not shared his vision with regard to data fusion algorithms and their relevance as it concerns public-transport-related data.
- Even though Neural Networks are mentioned many times, they have not been thoroughly considered in the thesis, being just discussed at “high-level”; hence, it remains insufficiently clear how exactly what is proposed in the thesis relates to Neural Networks.

7. Conclusions

After I have familiarized myself with the PhD Thesis and publications, submitted for the doctoral defense procedure and on the basis of my analyzing their significance as well as the scientific and application-oriented contributions reflected in them, I **acknowledge** that the presented thesis and publications, and their corresponding quality and originality (as it concerns the results and achievements reflected in them) meet the requirements of ADASRB, the Regulation of Application

of the ADASRB, the corresponding SU regulations as it concerns the awarding of PhD degrees in the Professional field 4.6 Informatics and Computer Sciences, (Doctoral program "Computer Sciences"). In particular, the Candidate meets the min. national requirements in the professional field and there is no plagiarism proven as it concerns the submitted materials.

Hence, I suggest that the Scientific Jury awards Georgi Kalinov Yosifov with the educational and science degree of "Doctor" in the Professional field 4.6 Informatics and Computer Sciences.

20.01.2023

Reviewer:

(Assoc. Prof. Dr. Boris Shishkov)