![ФОН logo]

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

To:
Saint Petersburg State University

Dr Ivan Luković, Full Professor
University of Belgrade
Faculty of Organizational Sciences
Chair of information Systems
E-mail: ivan.lukovic@fon.bg.ac.rs
Phone: +38163568749
ORCID: 0000-0003-1319-488X

Belgrade, 28-JUL-2022

Subject: An external reviewer report of a Ph.D. thesis

As a nominated external reviewer and a member of the dissertation council established for the defense of the dissertation prepared by **Ali Noaman Muhammad Aboalyazeed Muhammad** for the degree of a Candidate of Sciences / Doctor of Sciences in Technical Sciences, Scientific specialty 2.3.8, Informatics and Information Processes; dissertation title: **ANALYTICS OF TEXTUAL BIG DATA**, I submit here the following

REPORT
about the Ph.D. thesis: Analytics of Textual Big Data, proposed by Ali Noaman Muhammad Aboalyazeed Muhammad

The four main contributions of this Ph.D. thesis can be identified. The first, a data preprocessing model for textual data using a combination of NLP techniques has been proposed. The author justifies that a proposed model improves the quality of the resulting data by preserving the text's context. The second, the aspect term extraction model has been proposed. The model is based on clustering the word vectors generated using the pre-trained BERT model. Dimensionality reduction technique SOM has been employed to improve the quality of word clusters obtained using the K-Means++ clustering algorithm. The third, the author proposes a dynamic pruned n-gram model for recognizing the gender of the customers from their usernames. Finally, the author proposes an online product recommender system based on analyzing customer reviews utilizing aspect-based sentiment analysis.

The author clearly declares the theoretical and practical values of his Ph.D. thesis. As a theoretical contribution, he performed an analysis of the big data and analytics problems, approaches, and fundamental algorithms, and proposed a novel approach and method for application of NLP in the problems of gender classification and aspect term extraction. Finally, he proposed an online recommender system as a support of the proposed approach. The proposed method with the appropriate algorithms can be used in modern systems operating on data sciences, such as social networks, e-marketing, and real-time recommendations.

The practical value of the Ph.D. thesis has been demonstrated through the applications of the proposed algorithms in the selected case studies, and their embedding into the recommender system, also proposed in the thesis. The algorithms can be used in a way to meet the users' demands for fast data analysis and real-time result generation. The research validity of the

Јове Илића 154, 11000 Београд, Србија, Тел.: (011) 3950-800, Факс: (011) 2461-221
ПИБ: 100383934, Матични број: 07004044, Текући рачун: 840-1344666-69
Е пошта: dekanat@fon.bg.ac.rs; Посетите: www.fon.bg.ac.rs

33-06-738 от 01.08.2022

obtained results has been verified by using the appropriate metrics, mostly based on calculation of the precision, recall and F-score values, over the results of computational experiments.

After the introductory section, which gives a motivation for the research work, the Ph.D. report is organized into 5 chapters, followed by the concluding section. Chapter 1 presents a study of big data analytics and states the advantages of using big data solutions. Chapter 2 is about data gathering and preprocessing. Here, the author proposes his data preprocessing method, and presents an experiment of application of his method with the evaluation of the results being obtained. Next, in Chapter 3 the author discusses the aspect-based sentiment analysis approach, and presents his method for performing a task of aspect term extraction. Here, he also gives an experiment of application of his method with the evaluation of the results being obtained. Aspect-based sentiment analysis is a promising technique used widely to get insights of customers' reviews. Chapter 4 is about the problem of user gender identification, i.e. user gender recognition from theirs user names, often expressed in a leet language. The author proposes a multi-layered preprocessing framework to handle the gender identification problem, together with an experiment of application of his method with the evaluation of the results being obtained. In Chapter 5 it is presented a novel online product recommendation system, named SmartTips. It is based on analyzing customer reviews utilizing aspect-based sentiment analysis. It starts with capturing the user preferences using historical review data by extracting aspect terms and measuring the user sentiments towards each aspect separately. Next, the model searches the previously rated product sets for similar products to the current user preferences. Review data are used to rate various products using review text and users' votes as well. Finally, the recommendation engine generates a list of products that may be interesting to the user using a ranking scheme that applies a weight reflecting the relevance of each aspect in accordance with the user preferences.

The research results and contributions presented in the Ph.D. thesis are sound, well justified, and valuable. The thesis contains all necessary components, required for such kind of research work. Obviously, the author invested a lot of his efforts in researching related works, proposing novel approaches, methods, algorithms, and finally a recommender tool that unifies all of the components to a system with high capacities for its further practical applications. The author demonstrated a solid public visibility of the reported research results through the papers published in the two journals, as well as at several conferences.

Introduction is very well structured and brings valuable information about the main results and contributions of the author's Ph.D. thesis.

Also, I would like to point out to some minor concerns about the thesis. As I have noticed, the most of the author's publications related to the thesis results, are published in Russia. I know that the criteria at the refereed journals and conferences are strong. However, I believe that it is a must for the author to publish his results in a wider area, at some world-wide distinguished conferences or in journals.

From my point of view, the Research Goals in Introduction are not well elaborated. Namely, it is not quite clear, what research and professional values are to be reached by the proposed goals. Also, Research Main Objectives are formulated mostly as common research tasks, but not objectives. In Research Outcomes, it is not clear, for which potential stakeholders the research outcomes add some value. Who can benefit from the research outcomes, being declared in the thesis? It is even not discussed in Conclusion.

Methods and Methodology section is poorly elaborated, from my point of view. Namely, it is given on a very high and superficial level, without any concretization for the purpose of this Ph.D. thesis. In this form, it seems that the same text can be embedded into many other Ph.D. theses!

Chapter 1 gives a general overview of the topics of interest. However, it does not give an answer, how it is related to the author's research. For example, what are the precise research problems in the area of Aspect-Based Sentiment Analysis?

In Sections 3.2, 4.2, and 5.3, I notice a lack of some critical analysis of the related works for Aspect-Based Sentiment Analysis, Gender Identification, and Recommender Systems, in regard to the author's proposed research goals and outcomes. The author should discuss here with much more details, what are the findings and drivers identified in the literature that has a strong influence on his decisions about the research goals and outcomes, selected methodology, methods, etc.

Finally, from my point of view, Conclusion has to present some form of an elaborate of future research work (at least on two pages). The author should present here a general plan of his or someone's else future research, as by defending a Ph.D. thesis the author declares himself as a capable of being an advisor in some future research activities.

I can say that those my aforementioned findings do not significantly influence the evident main values and contributions reported in this Ph.D. thesis. By all facts given in this my report, I believe that the author did a great job with high potentials for developing future research results and publications, based on the presented Ph.D. thesis.

Conclusion

In conclusion, I declare that the dissertation with a title ANALYTICS OF TEXTUAL BIG DATA proposed by Ali Noaman Muhammad Aboalyazeed Muhammad meets all the requirements established by Saint-Petersburg State University for Doctor of Science degree, and Ali Noaman Muhammad Aboalyazeed Muhammad deserves the award of the degree of Doctor of Sciences in Technical Sciences, in scientific specialty 2.3.8, Informatics and Information Processes.

ИВАН ЛУКОВИЋ
2112965720014-2112965720014

Digitally signed by ИВАН ЛУКОВИЋ
2112965720014-2112965720014
Date: 2022.07.28 03:22:00 +02'00'

Ivan Luković