TREND ANALYSIS OF SOFTWARE ENGINEERING SCIENTIFIC PUBLICATIONS

TENGKU NADZIRAH BINTI
TENGKU MAJANI

BACHELOR OF COMPUTER SCIENCE
(SOFTWARE ENGINEERING)

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis, and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science (Software Engineering).

PROFESSOR DR. VITALIY MEZHUYEV
PROFESSOR
FACULTY OF COMPUTER SYSTEMS
& SOFTWARE ENGINEERING
UNIVERSITI MALAYSIA PAHANG
LEBUHRAYA TUN RAZAK, 26300 GAMBANG, KUANTAN
TEL: 09-549 2168 FAX: 09-549 2144

(Supervisor's Signature)

Full Name : PROF. DR. VITALIY MEZHUYEV

Position : PROFESSOR

Date : 04 JANUARY 2019

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Student's Signature)

Full Name : TENGKU NADZIRAH BINTI TENGKU MAJANI

ID Number : CB15019

Date : 04 JANUARY 2019

TREND ANALYSIS OF SOFTWARE ENGINEERING SCIENTIFIC PUBLICATIONS

TENGKU NADZIRAH BINTI
TENGKU MAJANI

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science (Software Engineering)

Faculty of Computer Systems & Software Engineering
UNIVERSITI MALAYSIA PAHANG

JANUARY 2019

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Vitaliy Mezhuyev who always guided me and giving me motivation, comments, positivity vibes and encouragement to perform and express my capabilities throughout this research field. I also would like to thank for his time spent to correcting my mistakes and proofreading. This research will not able to be completed on time without his moral support and guidance.

Apart from that, this research could not have been possible without the participant and assistance of my family and friends who helped me a lot and supporting me mentally and physically during my final year project. Their contributions are sincerely appreciated.

Lastly, I would like to thank you to everyone who contributes and support me directly or indirectly in this research to reinforce my basic knowledge and theories during my studies in Universiti Malaysia Pahang. of your document. You can use these galleries to insert tables, headers, footers, lists, cover pages, and other document building blocks. When you create pictures, charts, or diagrams, they also coordinate with the current look of your document.

ABSTRAK

Salah satu masalah yang paling penting untuk seorang penyelidik ialah bagaimana mengenal pasti trend dan topik semasa dalam penerbitan saintifik. Terdapat kepelbagaian kata kunci tertentu dalam jurnal dan kertas saintifik yang telah diterbitkan. Kajian ini bertujuan untuk menganalisis kaedah dan teknik yang sedia ada untuk meramalkan trend penerbitan saintifik. Untuk mendapatkan penyelesaian yang sesuai untuk masalah ini, kerja ini mencadangkan pelaksanaan algoritma untuk analisis trend dalam literatur sains Kejuruteraan Perisian. Anggaran trend masa depan dilakukan dengan membangun algoritma algoritma regresi linear algoritma dalam Matlab.

Dengan data yang dikumpul dari pangkalan data *Science Direct* utama dengan kata kunci Kejuruteraan Perisian, pengesahan algoritma yang dicadangkan pada kajian kes telah dilakukan. Pelaksanaan algoritma ini akan membantu para pengguna melihat tren dan mengenali corak dalam penerbitan saintifik Kejuruteraan Perisian pada masa akan datang. Dalam Matlab, kajian ini memberi tumpuan kepada kod Matlab regresi linear untuk mencapai hasil keinginan.

ABSTRACT

One of the most important problems for a researcher is how to identify the current trends and topics in the scientific publications. There is a diversity of specific keywords in the scientific journals and papers that had been published. The study aimed to analyse the existing methods and technique to predict trends of scientific publications. To get the suitable solution to the problem, this work proposes an implementation of the algorithm for trend analysis in Software Engineering scientific literature. The estimation of the future trend is done by development of the algorithm linear regression algorithm in Matlab.

With data collected from the main Science Direct database with of Software Engineering keywords, validation of the proposed algorithms on the case study was done. The implementation of the algorithm will help the users to see the trends and recognize patterns in Software Engineering scientific publications in future. In Matlab, the study focused on linear regression Matlab code to achieve the desire results.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARIMA	Auto-Regressive Integrated Moving Averages
CRF	Conditional Random Fields
CS	Computational Science
HMM	Hidden Markov Model
HPC	High-Performance Computing
ICCS	International Conference on Computational Science
LDA	Latent Dirichlet Allocation
RSME	Root Mean Squared Error
SLR	Systematic Literature Reviews

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

The role of journal papers, conference proceedings and other publications is gradually increasing, especially in scientific fields. (Larsen & von Ins, 2010). Scientific publication is a central piece of the mechanism that makes science advance. The rapid change of the trends of scientific publications is pushed by the availability of many tolls that did not exist before. (Bisquert, 2016).

Software Engineering fields tends to have a huge amount of the scientific journals and papers. With the rapid growing amount of publications, sometimes a user (a student or researcher at the beginning of a carrier) might have lost a track of the data and would like to predict the scientific trends in future.

There are many approaches could be used for analysing the trends such as manual systematic literature reviews (SLR) and automated text mining. (Wahono, 2015) An SLR is known as a process of identifying, assessing and interpreting all available research evidence with the purpose to provide answers for specific research questions. (Nie & Sun, 2017) Text mining techniques area is a subset of data mining that aims to extract knowledge from semi-structured textual data. Text mining is one of the methods of quantitative trend analyses and therefor by using the method of Latent Dirichlet Allocation (LDA), a generation model that could possibly finds all the hidden topics in document. (Hwang & Lee, 2018)

Next session of research aims to analyse the existent methods for trends analyses in scientific publications. It will result in a new method allowing to define a trend by the set of related to Software Engineering keywords. The well-organized dataset of scientific publications is a necessary condition for high efficiency of the scientific effort.

1.2 PROBLEM STATEMENT

The dramatic growth of the number of scientific publications in Software Engineering is a challenge for a student or a beginner researcher (Bornmann & Mutz, 2015). The most crucial part is how to identify the current trends in scientific research. Existing methods and algorithms of trend analysis were not applied in Software Engineering domain. This research will develop algorithm for trend analysis which will use linear regression.

1.3 OBJECTIVES

i. To analyse the existing methods and technique to predict trends of scientific publications.

ii. To adopt an algorithm of trending topic analysis in Software Engineering scientific literature and implement it in Matlab

iii. To validate proposed approach on the case study with data collected from main Science Direct databases for Software Engineering domain.

1.4 SCOPE

i. The study is considered Science Direct database for Software Engineering domain.
ii. The approach is based on the allocation of Software Engineering keywords applied to find amount of scientific publications.
iii. The algorithm for trend analyses of the scientific publications is developed with Matlab.
iv. Algorithm uses a linear and polynomial regression method.
v. Users of the proposed methods are students and researchers at the beginning of a carrier.

1.5 SIGNIFICANCE

Help the users, students and researchers at the beginning of a carrier to see the future trends and recognize patterns in Software Engineering scientific publications.

1.6 THESIS ORGANISATION

This thesis consists of five chapters which are introduction, literature review, methodology, result and discussion and conclusion. PSM 1 thesis covers chapter 1 until chapter 3, while chapter 4-5 will be discussing in PSM 2.

Chapter 1 gives introduction to the project, formulates problem statement, objectives and scope of the project. Chapter 2 gives literature review of the project. This chapter compares the existing approaches. Chapter 3 focuses on the methodology to develop an approach for trend analyses of the scientific publications to suit research objectives and scope. Chapter 4 discusses the findings of the project and the implementation of the algorithm. Lastly, chapter 5 concludes the research findings and presents limitations of research

REFERENCES

Abuhay, T., & Bochenina, K. (2018). Analysis of Publication Activity of Computational Science Society in 2001-2017 Using Topic Modeling and Graph Theory. Retrieved from https://www.researchgate.net/publication/324255534_Analysis_of_Publication_Activity_of_Computational_Science_Society_in_2001-2017_Using_Topic_Modeling_and_Graph_Theory

Bisquert, J. (2016). Trends of Scientific Publication. Journal of Physical Chemistry Letters, 7(9), 1703. https://doi.org/10.1021/acs.jpclett.6b00846

Buckland, M., & Gey, F. (1994). The Relationship between Recall and Precision. Journal Of The American Society For Information Science, 12-19. doi: 10.1002/(SICI)1097-4571(199401

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal Of The Association For Information Science And Technology, 66(11), 2215-2222. doi: 10.1002/asi.23329

Erduran, S., Ozdem, Y., & Park, J.-Y. (2015). Research trends on argumentation in science education: a journal content analysis from 1998–2014. International Journal of STEM Education, 2(1), 5. https://doi.org/10.1186/s40594-015-0020-1

Hwang, M., & Lee, K. (2018). A Method of Trend Analysis using Latent Dirichlet Allocation, (July). https://doi.org/10.14257/ijca.2018.11.5.15

Isenberg, P., Isenberg, T., Sedlmair, M., Chen, J., & Möller, T. (2017). Visualization as Seen through its Research Paper Keywords. IEEE Transactions on Visualization and Computer Graphics, 23(1), 771–780. https://doi.org/10.1109/TVCG.2016.2598827

Kanoun, K., & Laprie, J. (1991). The Role of Trend Analysis in Software Development and Validation. IFAC Proceedings Volumes, 24(13), 169-174. doi: 10.1016/s1474-6670(17)51385-3

Kivikunnas, S. (2001). Overview of Process Trend Analysis Methods and Applications. Retrieved from https://www.researchgate.net/publication/2907853_Overview_of_Process_Trend_Analysis_Methods_and_Applications

Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics, 84(3), 575–603. https://doi.org/10.1007/s11192-010-0202-z

Lei, L., & Liu, D. (2018). The research trends and contributions of System's publications over the past four decades (1973–2017): A bibliometric analysis. System, 80, 1-13. doi: 10.1016/j.system.2018.10.003

Limitations of bibliometrics | Measuring research impact | Library | University of Leeds. (2018). Retrieved from https://library.leeds.ac.uk/info/1406/researcher_support/17/measuring_research_impact/2

Mathew, G., Agrawal, A., Menzies, T., & Ieee, S. M. (n.d.). Finding Trends in Software Research, 45(1), 1–12.

Nie, B., & Sun, S. (2017). Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research. Applied Sciences, 7(4), 401. https://doi.org/10.3390/app7040401

Nigatie, G. (2018). ScienceDirect ScienceDirect Towards Predicting Trend of Scientific Research Topics using Topic Towards Predicting Trend of Modeling Scientific Research Topics using Modeling. Procedia Computer Science, 136, 304–310. https://doi.org/10.1016/j.procs.2018.08.284

Paper, C. (2014). Human Language Technology Challenges for Computer Science and Linguistics, 8387(November). https://doi.org/10.1007/978-3-319-14120-6

Powers, D. (2011). Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. Journal of Machine Learning Technologies, 2(1), 1–24. https://doi.org/10.1.1.214.9232

Sehra, S., Brar, Y., Kaur, N., & Sehra, S. (2017). Research patterns and trends in software effort estimation. Information And Software Technology, 91, 1-21. doi: 10.1016/j.infsof.2017.06.002

Wahono, R. S. (2015). A Systematic Literature Review of Software Defect Prediction : Research Trends , Datasets , Methods and Frameworks. Journal of Software Engineering, 1(1), 1–16.