Dear Phuong,

Thank you for submitting your work to the ESEM 2020 Technical Track.

Congratulations! We are pleased to inform you that your submission titled

TopFilter: An Approach to Recommend Relevant GitHub Topics

has been accepted as a full paper.

This year the ESEM Technical Track received 142 submissions, with full paper uploads for 123. After a desk-rejection of seven papers, we ended up with a set of 116 viable papers. At least three reviewers reviewed each of these papers. We accepted 26 full papers based on these reviews and on the discussions between the reviewers to reach consensus (acceptance rate of 21%).

The reviews for your paper are included at the end of this email. Please take these reviews into account when preparing your final, Camera-Ready Copy. Do your best to address the feedback, paying particular attention to the points mentioned in the meta-review. It is crucial to proofread your submission to improve its quality and fix any language issues.

Your Camera-Ready Copy must be prepared according to the instructions specified at:

https://eseiw2020.di.uniba.it/esem_conf/how-to-submit/

In particular, please verify if the formatting of your paper adheres to the ACM proceedings template for conferences.

Camera-Ready Copy. We need your final Camera-Ready Copy of your paper by ** Thursday, July 27 **. By July 13, the Proceedings Chair will publish the author kit at http://eseiw2020.di.uniba.it/esem_conf/author-kit/ with the instructions on how to submit the Camera-Ready Version of your paper (please wait for the instructions and DO NOT submit it to EasyChair). As part of this process, you will also need to provide a copyright form. Therefore, please read the instructions carefully immediately after they arrive.

Page Limits. The 12-page (10 main + 2 references) limit stated in the Call for Papers is strict. Please make sure to respect the page limit since we will not be able to make any exceptions.

List of Authors. The list of authors (names, emails, affiliations, order) is not allowed to be changed after notification. If a correction is needed (e.g., because the author name was misspelled), the track chairs need to approve the change.

Author Registration Policy. At least one of the paper's authors must register to the conference by the Camera-Ready Copy deadline **Monday, July 27 **. Note that each accepted contribution must have a minimum of one author registered by this date. The registration policy for the conference is per paper, rather than per author; i.e., if you have more than one paper accepted to the conference, then you will need to register for each accepted paper. We will also contact you with further details on our dissemination strategy, which we are planning to allow authors to increase the visibility of their accepted research papers. If you have any questions regarding the Author Registration Policy, please contact the General Chairs.

Deadline, Page-Limit, and Author Registration Policy Enforcement. We emphasize that the deadline for submitting your Camera-Ready Copy and registering for the conference is strict. We cannot guarantee the inclusion of your work in the conference proceedings and program if the stated deadlines, instructions, and policies are not respected.

Open-Science Initiative. ESEM’s open-science policy supports transparency, reproducibility, and replicability. The Open-Science Initiative Chair will contact you about opening up your data and replication packages as well as making a pre-print version of your paper available in an open-access archive. More information is available at http://eseiw2020.di.uniba.it/esem_conf/open-science/

Congratulations on the acceptance of your paper!

Best Regards,

Marcos Kalinowski and Federica Sarro
ESEM 2020 Program Chairs

SUBMISSION: 108
TITLE: TopFilter: An Approach to Recommend Relevant GitHub Topics

------------------------- METAREVIEW ------------------------
This paper proposes a new approach TopFilter to recommend relevant topics for repositories. When the proposed approach is combined with MNB network, it performs better than the state-of-the-art approach. This paper is well written and it provides a replication package. This paper can be further improved. For example, the motivation could be strengthened, especially why repositories already with some topics need the recommendation of new topics.

----------------------- REVIEW 1 ---------------------
SUBMISSION: 108
TITLE: TopFilter: An Approach to Recommend Relevant GitHub Topics
AUTHORS: Juri Di Rocco, Davide Di Ruscio, Claudio Di Sipio, Phuong Nguyen and Riccardo Rubei

----------- Paper Summary -----------
An approach to recommend GitHub topics
----------- Strengths and Weaknesses -----------
+ important topic
+ well written up
+ useful to practitioners and researchers
+ nice approach

- motivation could be strengthened
- more illustrations of actual topic generation would be valuable
----------- Detailed evaluation -----------
Comprehensively describing the topic of projects is an important area. Properly described projects make those projects much more useful and accessible to both developers, users and researchers. So I like the topic of the paper, but I do think that the importance of the work could be much more fully presented in the Introduction. Currently the motivation is not fully explained and the work could be motivated in a more compelling way. I would also very much like to see the RQs early in the paper as it would be useful to see the specific direction of the work from the beginning.

The paper would be much improved with some examples of topic generations. The analysis is all at the statistical level (rightly so) but actually showing some examples of the difference in project labels before and afterwards would be really illuminating, perhaps with a narrative around how the generation of these labels could be useful to specific circumstances and more information on situations when ‘good’ labels or generated versus ‘bad’ labels.

The methods are well described, the results are interesting and potentially useful and the paper is well written up and I enjoyed reading it.

----------------------- REVIEW 2 --------------------
SUBMISSION: 108
TITLE: TopFilter: An Approach to Recommend Relevant GitHub Topics
AUTHORS: Juri Di Rocco, Davide Di Ruscio, Claudio Di Sipio, Phuong Nguyen and Riccardo Rubei

----------- Paper Summary -----------
Developers uses topics to annotate repositories in GitHub. This paper builds a project-topic matrix, and designs an approach TopFilter to recommend topics for projects. Results show that TopFilter achieves good performance in recommending topics. Furthermore, better performance can be achieved by combining TopFilter with another method MNBN.
----------- Strengths and Weaknesses -----------
Weaknesses
====================================

This paper uses collaborative filtering technique to design the approach TopFilter. A project must have some initial topics for collaborative filtering. Then this paper writes that ‘Given an input project p, and an initial set of related topics decided by the developer’. If the developer gives some initial topics of a project, the develop may directly give all related topics. Why is necessary to recommend topics for projects which already have some topics? In subsection 3.2, this paper mentions that ‘TopFilter recommends topics by iterating over refining steps: once we select some topics from the recommended ones, TopFilter can discover new topics using the selected ones as new input.’ It is unclear why it is necessary to iteratively recommends topics.

In subsection 4.2, this paper mentions that ‘we removed irrelevant topics to reduce probable noise during the prediction phase.’ ‘we improve the overall quality of recommendation by removing “bad” terms.’ How do authors determine irrelevant topics or bad topics? How do authors ensure the correctness of removing topics? If some relevant topics are wrongly removed, it may affect the performance evaluation.
----------- Detailed evaluation -----------
In abstract, this paper mentions that ‘labeling GitHub repositories should be carefully conducted to avoid adverse effects on project popularity and reachability.’ What are potential adverse effects on popularity and reachability?

In subsection 3.1.3, the equation (2) is hard to understand. What does equation (2) work? This paper mentions that ‘r_p and r_q are the mean of the ratings of p and q, respectively’ What are ratings of projects? How do authors obtain ratings of projects?

Table 3 describes statistics of datasets. It is unclear how datasets are collected. Though datasets are from the previous work, authors would better describe data collection.

In subsection 5.1, the value of τ is always considered as half of the number of topics already assigned to the project under analysis. How does the value of τ affect recommendation performance?

----------------------- REVIEW 3 --------------------
SUBMISSION: 108
TITLE: TopFilter: An Approach to Recommend Relevant GitHub Topics
AUTHORS: Juri Di Rocco, Davide Di Ruscio, Claudio Di Sipio, Phuong Nguyen and Riccardo Rubei

----------- Paper Summary -----------
Summary

In order to make repositories accessible and exposed to a broader range of users, GitHub repositories are categorized into some topics. TopFilter introduced in this paper to assist developers for selecting a better topic. Using a graph-based approach, TopFilter recommend missed topics for to categorize repositories in a GitHub. For evaluation of TopFilter, authors respond to two research questions to (i) evaluate the impact of configuration on the performance and (ii) the extent that the accuracy that can be added to MNBN state of the art method. As the result authors conclude that their proposed method can improve the performance of the state of the art research.

----------- Strengths and Weaknesses -----------

In Favor

+ The method is replicable and the paper is accompanied by a replication package (both tool and the dataset)
+ The similarity calculator for offering new topics is sound and novel in this context
+ There is a clear comparison with the state of the art research using the same data as former research to evaluate.
+ the evaluation process is rigor and sound.

Against

- The applicability and usefulness of the provided topics for the develoeprs are not clear.
- The motivation of the paper is not clear.
- The paper is not empirically rigor and is mostly rely on number crunching results.
- Toward the mid and end of the paper the paper is actually is and sound like improvement of the existing method and hence indicate the lack of novelty

----------- Detailed evaluation -----------

Major Concerns

The paper is motivated with the hope that the provided recommendations the popularity of the repositories would be improved

Figure 2 is very high level and does not provide detailed view on the methods used better to refine and add details in order to clarify the input, output and techniques

In order to evaluate the tool fairly, authors better perform an evaluation with real world developers. One very relevant paper on the github repositories and for labeling commit message used this method which is also very relevant to the current study:

[R] Nayebi, Maleknaz, Shaikh Jeeshan Kabeer, Guenther Ruhe, Chris Carlson, and Francis Chew. "Hybrid Labels Are the New Measure!." IEEE Software 35, no. 1 (2017): 54-57.

Please add legend to Figure 9.

Minor

“as initially proposed by its authors in their paper [8].” – name the authors in the format X et al.