

## INTISARI

### PENGARUH PENAMBAHAN FITUR *TOPICS* DALAM PERHITUNGAN NILAI KEMIRIPAN REPOSITORY PADA GITHUB

Oleh

Syafina Nurul Aida

14/364133/PA/15910

Pengembangan sistem deteksi kemiripan repository GitHub dilakukan untuk mendeteksi repository yang memiliki kemiripan fungsi satu sama lain dengan menghitung kemiripan fitur repository. Dalam penelitian sebelumnya, fitur *readme files*, *stargazers*, dan waktu pemberian *star* digunakan untuk perhitungan kemiripan repository. Namun, diperlukan pengembangan lebih lanjut dalam proses perhitungan kemiripan repository. Ditambahkan fitur *topics* untuk menghitung kemiripan repository GitHub. Pada penelitian ini, dilakukan perhitungan besarnya pengaruh fitur *topics* sebagai tambahan dalam perhitungan kemiripan repository.

Dilakukan perbandingan nilai *success rate*, *confidence*, dan presisi dari evaluasi output antara perhitungan kemiripan repository dengan tiga fitur (*readme files*, *stargazers*, dan waktu pemberian *star*) dan dengan empat fitur (tiga fitur ditambah fitur *topics*) untuk mengetahui pengaruh penambahan fitur *topics*. Sebanyak 501 data repository digunakan sebagai basis data dan sebanyak 20 repository diantaranya digunakan sebagai kueri untuk menghasilkan output lima repository teratas yang memiliki nilai kemiripan paling tinggi dibanding kueri.

Pada proses evaluasi hasil pengujian, dilibatkan sebanyak empat orang penilai yang memberikan penilaian pada hasil output perhitungan kemiripan dengan nilai kesepakatan *Kappa* sebesar 0,47. Nilai evaluasi akhir dari semua penilai kemudian ditentukan menggunakan skala *Likert*. Perhitungan kemiripan repository dengan penambahan fitur *topics* mendapatkan nilai *success rate*, *confidence*, dan *presisi* lebih tinggi jika dibandingkan dengan perhitungan kemiripan repository tanpa fitur *topics* dengan nilai *success rate* pada T=4 sebesar 55%, nilai *success rate* pada T=5 sebesar 25%, *mean confidence* sebesar 2,35, median *confidence* sebesar 2,00, *mean presisi* sebesar 0,16, dan median presisi sebesar 0,20.

**Kata Kunci:** Perhitungan Nilai Kemiripan, GitHub, Fitur Topik, TF-IDF, *Fleiss Kappa*, Skala *Likert*

## ABSTRACT

### *THE INFLUENCE OF A TOPICS FEATURE ADDITION IN CALCULATING THE SIMILARITY VALUE OF REPOSITORY ON GITHUB*

By

Syafina Nurul Aida  
14/364133/PA/15910

The development of repository similarity detection system was conducted to detect similarity between repositories by calculating features similarity. In the research that has been done, readme files, stargazers, and timestamp of star addition can be used for repository similarity calculations. However, further development is needed in the process of repository similarity calculation. Topics feature was added to calculating the similarity of repositories. In this study, the influence of topics features to the repository similarity calculation was calculated.

In this study, we conduct a comparison of success rate, confidence, and precision value of final evaluation score from the output between two systems, i.e. system with three features (readme files, stargazers, and timestamp) and one with four features (with topics feature) to determine the influence of topics feature. The number of repository data use as database was 501. From the database there were 20 repositories ID used as query and for each of the query in the similarity calculation there were five first highest value of similarity selected as an output.

The evaluation process involved four raters who gave an assessment of the output system with Kappa value was 0,47. The final evaluation score was calculated using Likert scale. System with topics feature addition obtained higher success rate, confidence, and precision values than system with only 3 features (without topics feature), with success rate obtained at T=4 was 55%, success rate at T=5 was 25%, mean confidence was 2,35, median confidence was 2,00, mean precision was 2,00, and median precision was 0,20.

**Keyword:** Similarity Value Measurement, TF-IDF, Topic Feature, GitHub, Fleiss Kappa, Likert Scale