# FINDINGS ON SIMILARITY COEFFICIENTS TO IDENTIFY REFACTORING OPPORTUNITIES

**This briefing reports scientific evidence on the proposal of better similarity coefficients to identify code refactoring opportunities.**

## FINDINGS

The results presented in this briefing were obtained through the analysis of the accuracy of the proposed similarity coefficients along with other 18 coefficients in 101 software systems. This analysis considered similarities values for Move Class, Move Method and Extract Method refactoring operations.

- The research results show that it is indeed possible to apply optimization algorithms to adapt the weights of the formula of an existing coefficient to generate new coefficients with a higher accuracy.

- As the research objective, three new similarity coefficients were proposed. **PTMC** for Move Class operations, **PTMM** for Move Method operations and **PTEM** for Extract Method operations.

**Findings about the major 18 previously existing coefficients in literature analyzed:**

- Considering the analyzed coefficients and target systems, PSC was the coefficient with the best precision for Move Class operations, in which it achieved an average accuracy of 55.03% in the first attempt, 69.20% in the second, and 76.36% in the third.

- Sokal and Sneath 2 reached the best precision among the coefficients for Move Method operations. It achieved an average accuracy of, respectively, 40.79%, 51.86% and 58.58% for the first, second, and third attempts.

- For Extract Method operations, which consider only one attempt, Russell and Rao presented an accuracy of 86.58%, the best precision among the previously existing coefficients.

**Findings about the coefficients proposed in the research:**

- It can be observed, through the analysis of the results, that the proposed coefficient **PTMC** was superior to the previously best coefficient for Move Class operations (PSC), achieving an average accuracy of 58.20% in the first attempt, 70.82% in the second, and 77.22% in the third, and having therefore an accuracy, respectively, 3.17%, 1.62%, and 0.86% higher.

- For Move Method operations, the proposed coefficient **PTMM** obtained an average accuracy of 51.36% in the first attempt, 60.93% in the second ,and 66.41% in the third. Therefore, it was superior to Sokal and Sneath 2 (coefficient with the previous highest precision), achieving an accuracy, respectively, 10.57%, 9.07% and 7.83% higher.

- The proposed coefficient **PTEM** reached an average accuracy of 86.88%, and therefore 0.3% higher than the Russell and Rao, previously best coefficient for Extract Method operations. And although such difference in the accuracy is relatively small, PTEM was not superior in only five of the 101 analyzed systems.

### *Who is this briefing for?*

Software engineering practitioners who want to find more suitable similarity coefficients to identify code refactoring opportunities based on scientific evidence.

### *Where the findings come from?*

All findings of this briefing were extracted from the empirical study conducted by Arthur F. Pinto and Ricardo Terra.

### *What is included in this briefing?*

The main findings of the original empirical study related to the proposed coefficients.

### *What is not included in this briefing?*

Additional information not supported by the findings of the original empirical study as well as descriptions about the applied methodology and details of the developed tool features.

**ORIGINAL RESEARCH REFERENCE**

Arthur F. Pinto and Ricardo Terra. Better Similarity Coefficients to Identify Refactoring Opportunities. In Proceedings of SBCARS, 2017.