

Lenguajes de programación - T08: Implementación de la distancia Damerau-Levenshtein en el lenguaje R

Jorge Aurelio Morales Manrique

C.C. 1010075711

jomorales@unal.edu.co

Universidad Nacional de Colombia

Abril 13 de 2021

Distancia propuesta entre lenguajes de programación

Se propone una función de distancia entre dos lenguajes de programación como la distancia promedio entre cada una de las distancias obtenidas entre pares de palabras de ambos vocabularios. Expresado en términos matemáticos se tiene:

$$\begin{aligned} f: W \times W &\rightarrow R \\ (A, B) &\mapsto \mu(D). \end{aligned}$$

Donde A y B son los vocabularios de los lenguajes de programación, mientras que D es el conjunto de distancias obtenidas para cada par distinto de palabras entre los lenguajes. La función de distancia para dos palabras a y b se define de la siguiente manera:

$$d(a, b) = \text{abs}(|a| - |b|) + \sum_{i=1}^{\min(|a|, |b|)} \text{cmp}(a(i), b(i))$$

Donde $|x|$ representa la longitud del lenguaje, $x(i)$ el i -ésimo caracter de la palabra x y $\text{cmp}(w, y)$ una función que retorna 1 en caso de que los caracteres w e y sean diferentes, 0 en el otro caso.