December 7, 2021

Abstract

1

Algorithm 1: identify Row Context

Input: r_i, $Backgrd(T_i){=}T_1, T_2, \ldots, T_n$ and similarity threshold θ_r

Output: $con(r_i)$

1 $con(r_i) = \Phi$;

2 **for** $j = 1; j \leq n; j \neq i$ **do**

3 $\quad$ float $maxSim = 0$;

4 $\quad$ $r^{maxSim} = null$;

5 $\quad$ **while** *not end of T_j* **do**

6 $\quad\quad$ compute $\text{Jaro}(r_i, r_m)(r_m \in T_j)$;

7 $\quad\quad$ **if** $(Jaro(r_i, r_m) \geq \theta_r) \wedge (Jaro(r_i, r_m) \geq r^{maxSim})$ **then**

8 $\quad\quad\quad$ replace r^{maxSim} with r_m;

9 $\quad$ $con(r_i) = con(r_i) \cup r^{maxSim}$;

10 return $con(r_i)$;