In order to reflect these differences in attribute values, we introduce a modified version of the Damerau-Levenshstein distance, that not only reflects the difference between two process instances, but also the attribute values. We achieve this by introducing a cost function $cost_{a_i,b_j}$, which applies to a normed vector-space[1]. Concretely, we formulate the modified Damerau-Levenshstein distance as shown in Equation 1. For the remainder, we refer to this edit-distance as Semi-strucured Damerau-Levenshtein distance (SSDLD).

$$d_{a,b}(i,j) = \min \begin{cases} d_{a,b}(i-1,j) + cost(\mathbf{0}, b_j) & \text{if } i > 0 \\ d_{a,b}(i,j-1) + cost(a_i, \mathbf{0}) & \text{if } j > 0 \\ d_{a,b}(i-1,j-1) + cost(a_i, b_j) & \text{if } i,j > 0 \\ & \&\ \overline{a}_i = \overline{b}_j \\ d_{a,b}(i-1,j-1) + cost(a_i, \mathbf{0}) + cost(\mathbf{0}, b_j) & \text{if } i,j > 0 \\ & \&\ \overline{a}_i \neq \overline{b}_j \\ d_{a,b}(i-2,j-2) + cost(a_i, b_{j-1}) + cost(a_{i-1}, b_j) & \text{if } i,j > 1 \\ & \&\ \overline{a}_i = \overline{b}_{j-1} \\ & \&\ \overline{a}_{i-1} = \overline{b}_j \\ 0 & \&\ i = j = 0 \end{cases} \tag{1}$$

Here, $d_{a,b}(i,j)$ is the recursive form of the Damerau-Levenshtein-Distance. a and b are sequences and i and j specific elements of the sequence. $cost(a,b)$ is a cost function which takes the attribute values of a and b into account. The first two terms correspond to a deletion and an insertion from a to b. The idea is to compute the maximal cost for that the wrongfully deleted or inserted event. The third term adds the difference between two events with identical activities $\overline{a}_i$ and $\overline{b}_j$. As mentioned earlier, two events that refer to the same activity can still be different due to event attributes. The distance between the event attributes determines *how* different these events are. The fourth term handles the substitution of two events. Here, we compute the substitution cost as the sum of an insertion and a deletion. The fifth term computes the cost after transposing both events. This cost is similar to term 3 only that we now consider the differences between both events after they were aligned. The last term relates to the stopping criterion of the recursive formulation of the Damerau-Levenshstein distance.

[1] A normed vector-space is a vector space, in which all vectors have the same dimensionality. For instance, if all vectors have three dimensions, we can call the vector-space *normed*.