

Расстояние Дамерау — Левенштейна

Материал из Википедии — свободной энциклопедии

Расстояние Дамерау — Левенштейна (названо в честь учёных Фредерика Дамерау и Владимира Левенштейна) — это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Является модификацией расстояния Левенштейна: к операциям вставки, удаления и замены символов, определенных в расстоянии Левенштейна добавлена операция транспозиции (перестановки) символов.

Алгоритм

Расстояние Дамерау — Левенштейна между двумя строками ***a*** и ***b*** определяется функцией $d_{a,b}(|a|, |b|)$ как:

$$d_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} d_{a,b}(i-1, j) + 1 \\ d_{a,b}(i, j-1) + 1 \\ d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{if } i, j > 1 \text{ and } a_i = b_{j-1} \text{ and } a_{i-1} = b_j \\ \min \begin{cases} d_{a,b}(i-1, j) + 1 \\ d_{a,b}(i, j-1) + 1 \\ d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise,} \end{cases}$$

где $1_{(a_i \neq b_j)}$ это индикаторная функция, равная нулю при $a_i = b_j$ и 1 в противном случае.

Каждый рекурсивный вызов соответствует одному из случаев:

- $d_{a,b}(i-1, j) + 1$ соответствует удалению символа (из *a* в *b*),
- $d_{a,b}(i, j-1) + 1$ соответствует вставке (из *a* в *b*),
- $d_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)}$ соответствие или несоответствие, в зависимости от совпадения символов,
- $d_{a,b}(i-2, j-2) + 1$ в случае перестановки двух последовательных символов.

Реализации

- На языке программирования Perl в виде модуля Text::Levenshtein::Damerau
- На языке программирования PIPgSQL
- Дополнительный модуль для MySQL

См. также

- Расстояние Левенштейна
-

Источник — «https://ru.wikipedia.org/w/index.php?title=Расстояние_Дамеру_—_Левенштейна&oldid=80173793»

- Последнее изменение этой страницы: 20:35, 13 августа 2016.
- Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.
Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.