

Distância de Edição

Distância de edição: forma de medir a distância entre 2 strings, pela contagem do número mínimo de operações necessárias para transformar uma string na outra. Existem diferentes definições de distância de edição.

Distância de Levenshtein

A métrica mais comum de distância de edição.

Métrica de Similaridade

Métrica que mede similaridade entre strings

Operações Mínimas

Número mínimo de operações para transformar uma palavra em outra

Operações Permitidas

- Inserção
- Deleção
- Substituição

Usos da Distância de Levenshtein



Corretores Ortográficos



Alinhamento de Sequências de DNA



Busca e Recuperação de Dados



Processamento de Linguagem Natural

Onde ajuda a encontrar palavras corretas ou a comparar textos

Exemplo Prático

"casa" → "casa" (0 operações)

"casa" → "caso" (1 operação: substituir 'a' por 'o')

"casa" → "causa" (1 operação: inserir 'u')

"casa" → "cas" (1 operação: deletar 'a')

Como Funciona o Algoritmo?

O algoritmo implementa a distância de Levenshtein de forma dinâmica:

DISTÂNCIA 0:

Verifica se a palavra está no vocabulário. Se estiver, retorna a própria palavra: "casa" → "casa" (0 operações)

DISTÂNCIA 1:

Se não estiver, ele gera todas as palavras possíveis com uma operação (inserção, substituição, exclusão), verifica quais delas existem no vocabulário, forma uma lista com elas e as retorna.

Exemplo: "Ols"

Todas as combinações possíveis com distância de Levenshtein 1 da palavra "ols", divididas pelas três operações.

1. Deleção (Remoção de 1 caractere)

- ls (removendo 'o')
- os (removendo 'l')
- ol (removendo 's')

(Total: 3 combinações)

2. Substituição (Troca de 1 caractere por outro do alfabeto a-z)

- Substituindo 'o': als, bls, cls, dls, els, fls, gls, hls, ils, jls, kls, lls, mls, nls, pls, qls, rls, sls, tls, uls, vls, wls, xls, yls, zls.
- Substituindo 'l': oas, obs, ocs, ods, oes, ofs, ogs, ohs, ois, ojs, oks, **oms**, ons, oos, ops, oqs, ors, oss, ots, ous, ovs, ows, oxs, oys, ozs.
- Substituindo 's': **ola**, olb, olc, old, **ole**, olf, olg, olh, oli, olj, olk, oll, olm, oln, olo, olp, olq, olr, olt, olu, olv, olw, olx, oly, olz.

(Total: 75 combinações, assumindo um alfabeto de 26 letras e que a letra original não é contada)

Exemplo: "Ols" (Continuação)

3. Inserção (Adição de 1 caractere do alfabeto a-z)

- No início (_ols): aols, bols, cols, dols, eols, fols, **gols**, hols, iols, jols, kols, lols, **mols**, nols, ools, pols, qols, rols, sols, tols, uols, vols, wols, xols, yols, zols.
- Entre 'o' e 'l' (o_ls): oals, obls, ocls, odls, oels, ofls, ogls, ohls, oils, ojls, okls, olls, omls, onls, ools, opls, oqls, orls, osls, otls, ouls, ovls, owls, oxls, oyls, ozls.
- Entre 'l' e 's' (ol_s): olas, olbs, olcs, olds, oles, olfs, olgs, olhs, olis, oljs, olks, olls, olms, olns, olos, olps, olqs, olrs, olss, olts, olus, olvs, olws, olxs, olys, olzs.
- No final (ols_): olsa, olsb, olsc, olsd, olse, olsf, olsg, olsh, olsi, olsj, olsk, olsl, olsm, olsn, olso, olsp, olsq, olsr, olss, olst, olsu, olsv, olsw, olsx, olsy, olzs.

(Total: 104 combinações, assumindo um alfabeto de 26 letras) Somando todas as operações, existem 182 combinações únicas com distância de Levenshtein 1 da palavra "ols".

Como Funciona o Algoritmo?

DISTÂNCIA 2:

Se nenhuma dessas palavras estiver na lista, o algoritmo forma todas as palavras possíveis com 2 edições (as quais aumentam de forma exponencial) e busca cada uma no vocabulário, criando uma lista com elas e as retornando.

Seleção por Frequência

MAIOR FREQUÊNCIA:

Entre as palavras de menor distância encontradas, ele então seleciona a que for mais frequente.

Caso ainda não encontre nenhuma das palavras formadas no vocabulário, é feita uma busca por força bruta em todo o vocabulário para encontrar a palavra .