**1.- DISTANCIA LEVENSHTEIN**

La distancia Levenshtein determina la diferencia entre dos secuencias de caracteres. Su interés se ha puesto de manifiesto en la corrección ortográfica, en la detección de plagio y en el análisis de secuencias de ADN, entre otros. La distancia calcula el número mínimo de ediciones necesarias (inserciones, eliminaciones o sustituciones de caracteres) para transformar una cadena en otra. Así, por ejemplo, la distancia entre los caracteres «kitten» y «sitting» es 3 y la distancia normalizada es 3/7 (7=max {6,7} la longitud máxima de los caracteres),

|  |  |  |
| --- | --- | --- |
| kitten |  |  |
| sitten | k→s | 1 |
| sittin | e→i | 2 |
| sitting | Añadir g | 3 |
|  |  | distancia=3 |

La distancia de Levenshtein se calcula mediante un algoritmo dinámico que compara dos cadenas de texto carácter a carácter. Comienza con una matriz que representa todas las combinaciones posibles de caracteres entre las dos cadenas y luego calcula el número mínimo de ediciones necesarias para transformar una cadena en otra.

Algoritmo dinámico de Levenshtein

* Se crea una matriz con la longitud de las palabras más uno 8x7 (sitting, 7 posiciones; kitten, 6 posiciones
* Texto

  Descripción generada automáticamenteSe llena la matriz utilizando un algoritmo de programación dinámica

Texto

Descripción generada automáticamente

siendo aᵢ el carácter de la posición i en la cadena a y bⱼ el carácter de la posición j en la cadena b

* La celda inferior derecha de la matriz representa la distancia de Levenshtein entre las dos palabras. En este caso, la distancia es 3.

Pseudocódigo, código y explicaciones en páginas web:

<https://es.wikipedia.org/wiki/Distancia_de_Levenshtein>

<https://www.glc.us.es/~jalonso/exercitium/04-oct-23/>

<https://medium.com/@ethannam/understanding-the-levenshtein-distance-equation-for-beginners-c4285a5604f0>

<https://www.analyticslane.com/2020/06/17/la-distancia-de-levenshtein/>

<https://github.com/alu0100723304/DistanciaLevenshtein>

**2. AJUSTE BILBAO A LA DISTANCIA**

Las distancias IRI, IPI, Levenshtein evalúan las analogías y diferencias entre poblaciones caracterizadas por una única respuesta. Cuando en una misma población tenemos varias respuestas, resulta de interés aplicar el ajuste Bilbao,

Imagen que contiene Diagrama

Descripción generada automáticamente

**Ejemplo**. Para dos poblaciones, A y B, en las que tenemos 3 respuestas a un mismo concepto/ítem A={aaa, bcd, ecd} y B={aaa, ecd, baa}, el ajuste Bilbao referido a la distancia de Levenshtein es,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dL** | **aaa** | **bcd** | **ecd** | **min** |
| **aaa** | 0 | 1 | 1 | **0** |
| **ecd** | 1 | 1/3 | 0 | **0** |
| **baa** | 1/3 | 2/3 | 1 | **1/3** |
| **min** | **0** | **1/3** | **0** |  |