



# Moog!e!

Introduzca su búsqueda



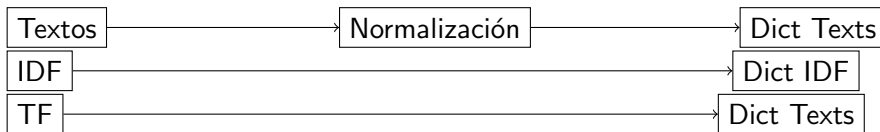
Buscar

Figure: Buscador

# ¿Qué es Moogle?

Moogle es un buscador de archivos .txt, el cual mediante una consulta debe entregar un grupo de nombres de textos ordenados por relevancia(máximo 3) mostrando una pequeña porción de los mismos donde salgan elementos de la query. Para determinar esta relevancia se utiliza el calculo del TF-IDF, así como también la función coseno.

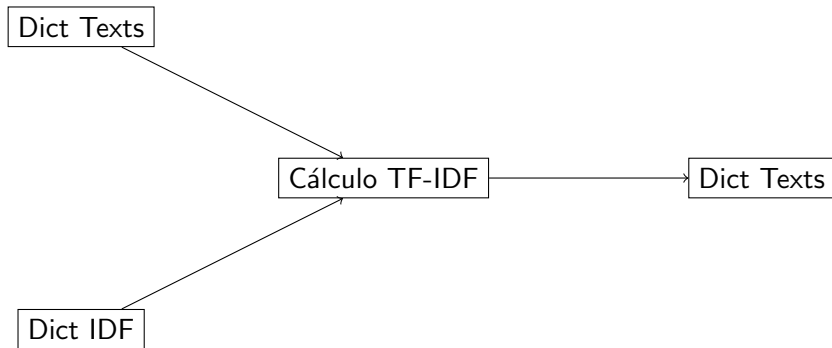
# Preprocesamiento



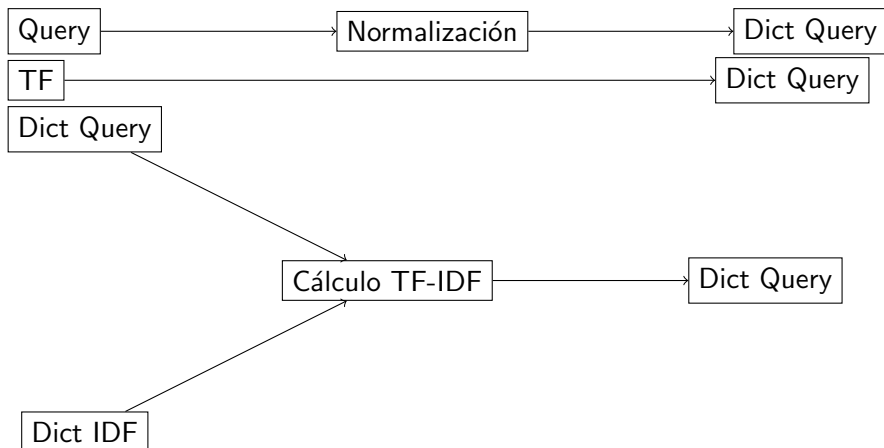
Normalización: Eliminación de mayúsculas y caracteres no alfanuméricos

Diccionario Texts: string, Diccionario de string, float

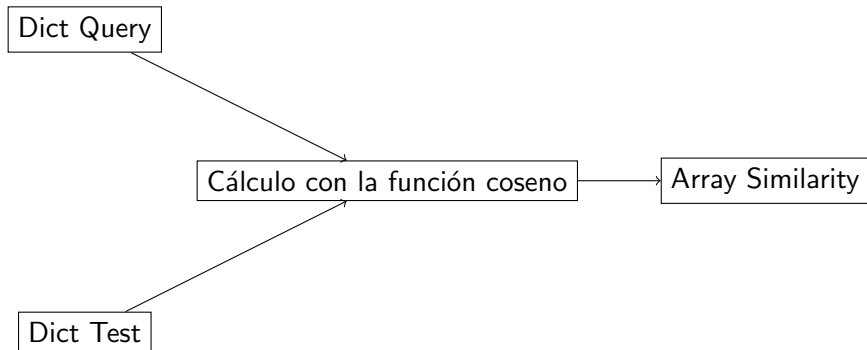
# Preprocesamiento



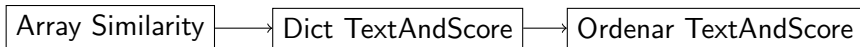
# Búsqueda



# Búsqueda



# Resultados



A partir del Diccionario TextAndScore se toman los textos más importantes y se muestra el título y un pequeño Snippet

# Si no coincide la búsqueda

¿Quisiste decir [Messi](#)?

- No se encontraron documetos :(

... ..

Figure: No hay coincidencias



# Si coincide la búsqueda



¿Quisiste decir [inercia](#)?

- **Adolf Hitler**

... discriminación orquestada desde arriba cada fase de radicalización era más intensa que la que la precedía de esta manera la inercia no se desvanecía nunca teorías sobre el origen de su antisemitismo desde su aparición en el mundo político surgieron ...

- **Conan Doyle, Arthur - El mundo perdido**

... con casco de corcho realmente usted llega ya al límite resopló el profesor excede lo que considero probable parálisis cerebral inercia mental resultaba demasiado absurdo para irritarme enojarse con un hombre como challenger era una pérdida de energía pues tendría ...

- **Einstein Albert - Este es mi pueblo**

... estudio de las curvas geodésicas deduciendo la ecuación general de las mismas siguió con la generalización del principio de la inercia y con el estudio del campo gravitatorio estableciendo la expresión general del mismo que en la primera aproximación contiene ...

Figure: Las 3 coincidencias mas importantes

# Función del ranking (Función Coseno)

$$sim(d, q) = \frac{\vec{d}_j * \vec{q}}{|\vec{d}_j| * |\vec{q}|} \quad (1)$$

$$sim(d, q) = \frac{\sum_{i=1}^n w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^n w_{i,j}^2} * \sqrt{\sum_{i=1}^n w_{i,q}^2}} \quad (2)$$

## Cálculo del peso en los documentos

$$TF - IDF_{i,j} = tf_{i,j} * idf_i \quad (3)$$

$$tf_{i,j} = \frac{freq_{i,j}}{\max_l freq_{l,j}} \quad (4)$$

$$idf_i = \log \frac{N}{n_i} \quad (5)$$