

WikiRacer

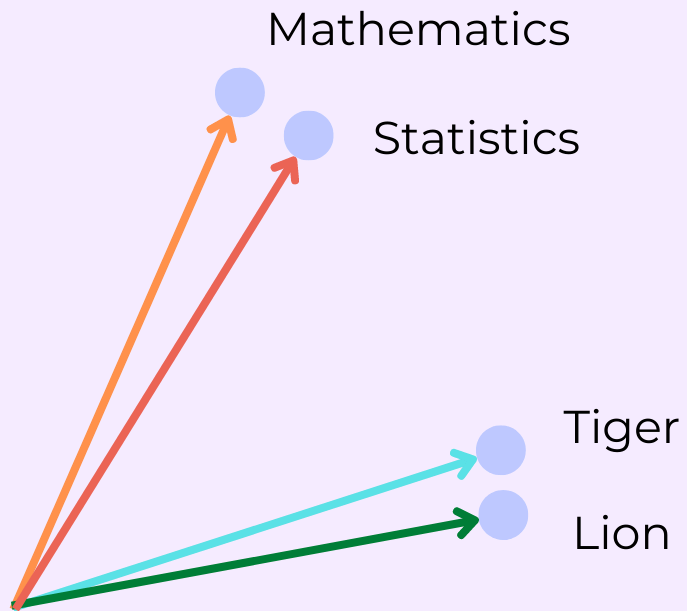
Programación avanzada

José Ocampo

Ideas fallidas

Word embeddings

Word Embeddings



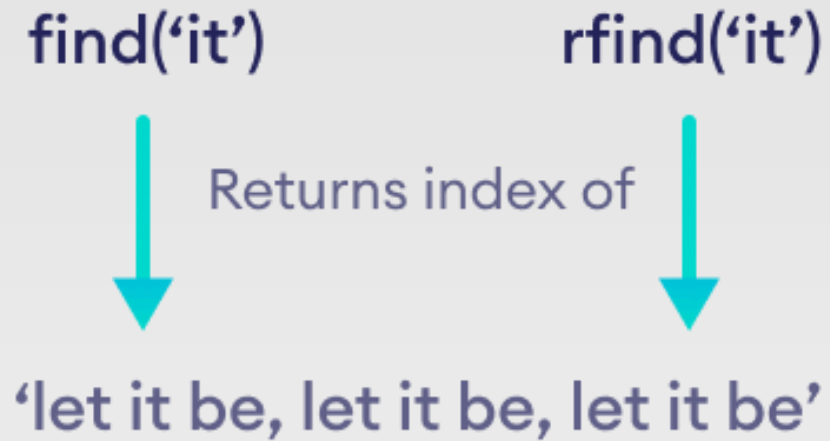
"Mathematics" is similar to "Statistics"

"Tiger" is similar to "Lion"

Word embeddings

- Rendimiento terrible
- No terminó el test `test_wikiracer_1` en menos de 5 minutos
- Similaridad semántica < String similarity (hipótesis)

Substrings y superstrings



Substrings y superstrings

- Las coincidencias exactas son muy difíciles de encontrar
- Las que son fáciles de encontrar quitan en lugar de añadir (Ej. `/wiki/A`)

Levenshtein Distance

“

The minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other.

”

Levenshtein Distance

rain
sain
shin
shine

(a)

shine
rhine
raine
rain

(b)

shine
tshine
trhine
traine
train

(c)



Substitution



Insertion



Deletion

Levenshtein Distance

- Falla en todos los tests complejos
- Al principio pensé que esta era la clave del éxito, pero mientras hacía la presentación me di cuenta que no.

Función de costo

Implementación

- Dar prioridad máxima a una página si está presente en los links del objetivo
- Eso es todo! (para estos tests)

```
def costFn(current_link, neighbor_link, goal_links, goal):  
    if neighbor_link in goal_links:  
        return 0  
  
    # Esta métrica por si misma falla los tests  
    # En general, return a, con  $a > 0$ , da el mismo resultado  
    return Levenshtein.distance(neighbor_link, goal)
```

Ejemplo: Jesus -> Kobe_Bryant

```
current_distance=0, current_link='/wiki/Jesus'  
current_distance=0, current_link='/wiki/Eucharist'  
current_distance=0, current_link='/wiki/ISBN_(identifier)'  
current_distance=0, current_link='/wiki/ISSN_(identifier)'  
current_distance=0, current_link='/wiki/Main_Page'  
current_distance=0, current_link='/wiki/United_States'  
current_distance=0, current_link='/wiki/Academy_Awards'  
current_distance=0, current_link='/wiki/90th_Academy_Awards'  
  
# Path  
['/wiki/Jesus', '/wiki/ISBN_(identifier)', '/wiki/United_States',  
 '/wiki/Academy_Awards', '/wiki/90th_Academy_Awards', '/wiki/Kobe_Bryant']
```

Resultados

Métricas generales

- 56.56s para correr todos los tests (sin caché)
- Entre 1 y 9 requests para llegar al objetivo
- Longitud del camino: entre 2 y 6

Resultados delglosados

Inicio	Objetivo	Requests	Longitud del camino
United_Nations	Brazil	1	2
Michael_Jordan	Kobe_Bryant	1	2
Michael_Jordan	Brazil	2	3
Adolf_Hitler	United_Nations	2	3
Jesus	Adolf_Hitler	3	4
Michael_Jordan	United_Nations	3	4
Adolf_Hitler	Michael_Jordan	3	4

Inicio	Objetivo	Requests	Longitud del camino
Adolf_Hitler	Brazil	3	4
Jesus	United_Nations	3	3
Adolf_Hitler	Kobe_Bryant	4	4
United_Nations	Kobe_Bryant	4	4
Jesus	Brazil	7	5
Jesus	Kobe_Bryant	8	6
Jesus	Michael_Jordan	9	6
Kobe_Bryant	Brazil	9	4