



# Proyecto Coordinado ALT-SAR

Asier Serrano Aramburu, Javier Maroto Llamas, Rodrigo Juanes Busolo



# Índice de Presentación

1. División del trabajo
2. Tareas realizadas
3. Medición de tiempos

# 1. División del Trabajo

## Javier:

- Damerau-Levenshtein restringida e intermedia.
- Damerau-Levenshtein restringida con y sin umbral.
- Damerau-Levenshtein restringida e intermedia con trie-cadena.
- Modificar el Indexador y el Recuperador de SAR para permitir la búsqueda aproximada de cadenas.

## Asier:

- Levenshtein con y sin umbral, y trie (modo backwards).
- Clase trie.
- Solo para Levenshtein, probar cota optimista (tarea 3).

## Rodrigo:

- Método suggest con threshold.
- Cotas inferiores y superiores de los algoritmos de cálculo de distancia de edición, para evitar ejecuciones no deseadas.
- Damerau-Levenshtein intermedia con y sin umbral.

# Tarea 1

- Distancia de levenshtein sin optimización espacial.
- Distancia de levenshtein con optimización espacial.
- Distancia de Damerau-levenshtein restringida.
- Distancia de Damerau-levenshtein intermedia.

## Tarea 2

- Distancia de levenshtein con threshold.
- Distancia de Damerau-levenshtein restringida con threshold y distancia de Damerau-levenshtein intermedia con threshold.

# Tarea 3

- Método suggest sin trie.
- Método suggest con trie.
- Cotas optimistas para levenshtein.

## Tarea 4

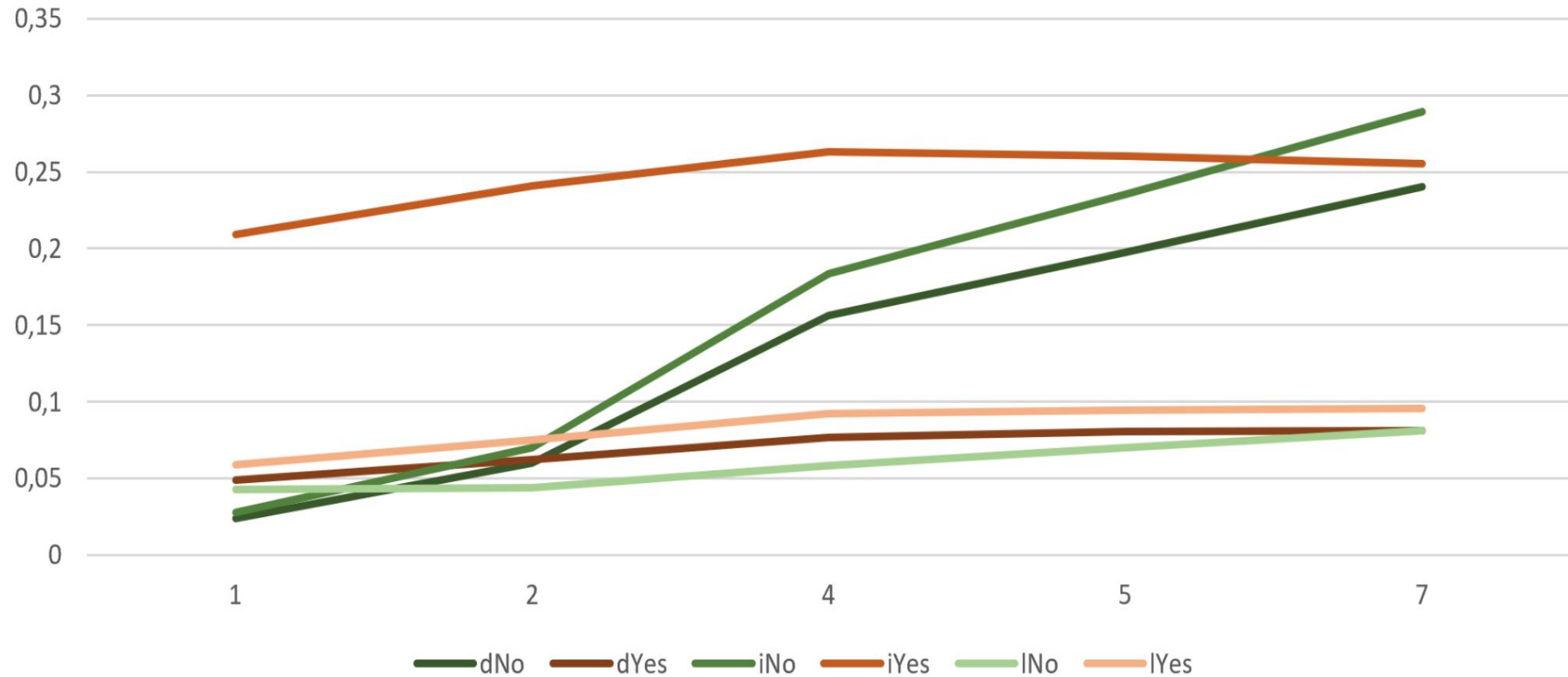
- Distancia de Damerau restringida cadena-trie con threshold.
- Distancia de levenshtein restringida cadena-trie con threshold.
- Distancia intermedia cadena-trie con threshold.

# Tarea 5 y Medición de Tiempos

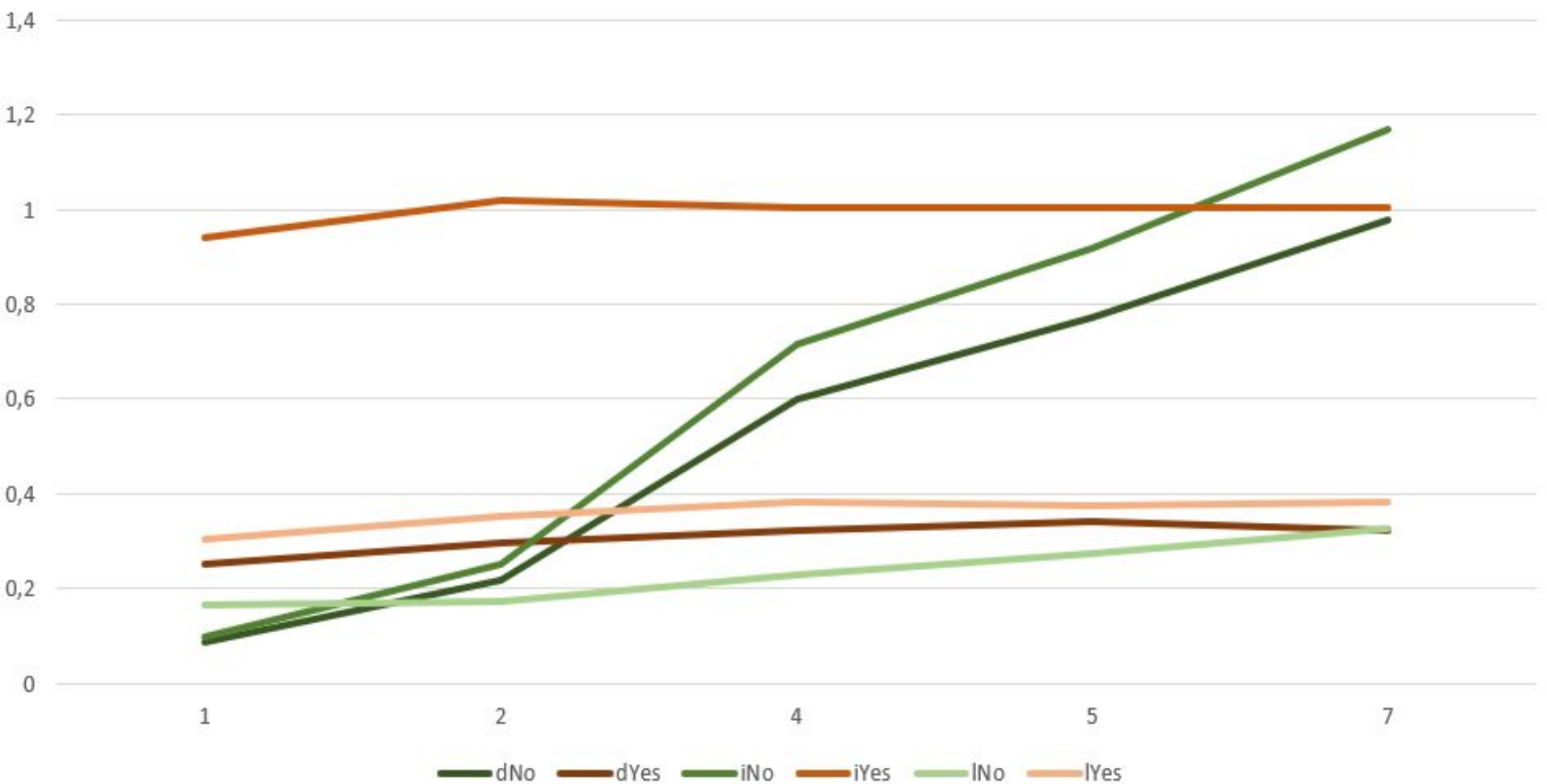
- Main en *spellsuggest.py*



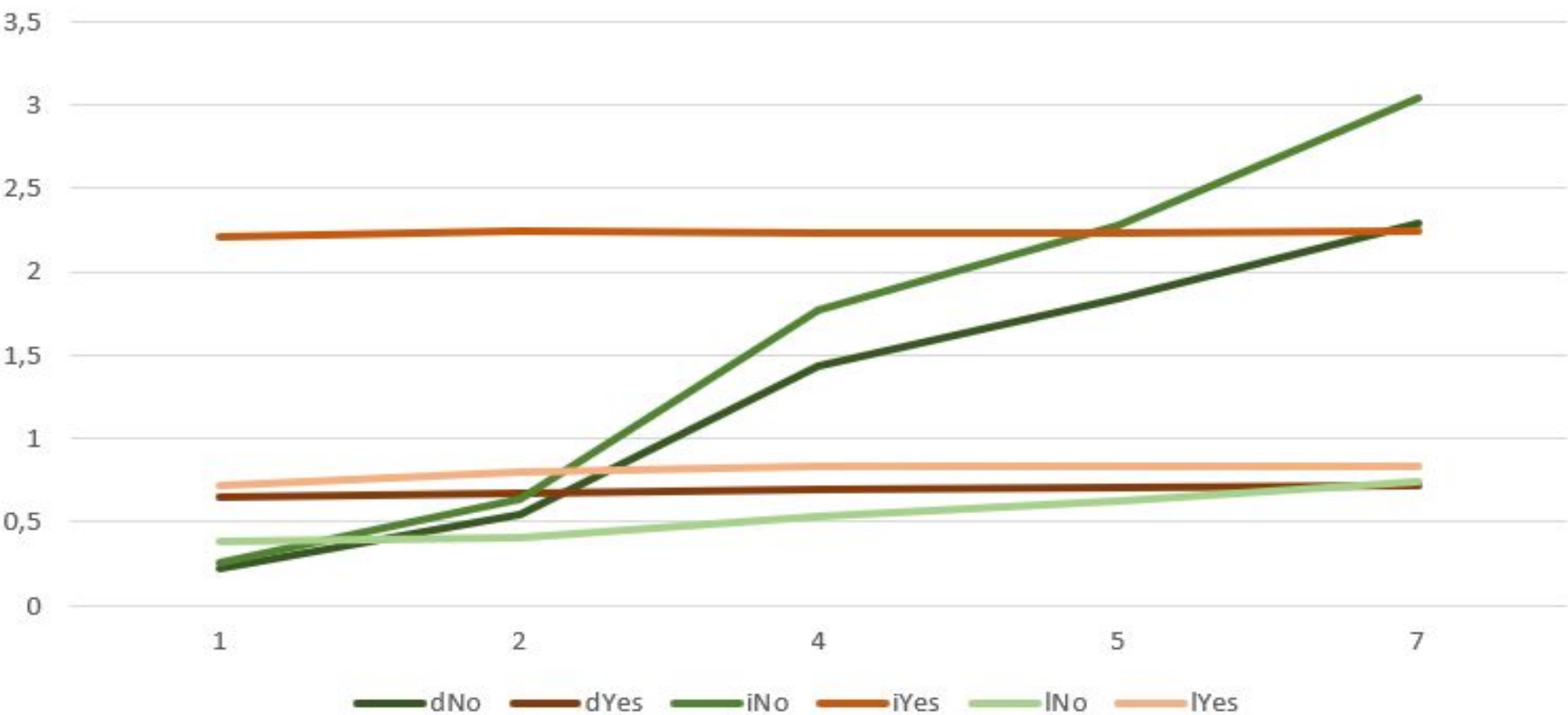
# Tiempos medios todas la palabras, talla 2500 (en segundos)



Tiempos medios todas las palabras, talla 10000 (en segundos)



# Tiempos medios todas las palabras, talla 35000 (en segundos)



# Tarea 6: Indexador y Buscador + demo

```
>python SAR_Indexer.py 2015 index_20151.bin -G
```

```
-----  
Number of indexed days: 285  
-----
```

```
Number of indexed news: 803  
-----
```

```
TOKENS:
```

```
    # of tokens in 'article': 44684  
-----
```

```
Positional queries are NOT allowed  
-----
```

```
Time indexing: 1.22s.
```

```
Time saving: 0.16s.
```

```
>python SAR_Searcher.py index_20151.bin -Q "alexanderx" -Z "3" -Tr False -B "levenshtein"  
18
```

```
Query: alexanderx
```

```
Number of results: 18
```

```
>python SAR_Searcher.py index_20151.bin -Q "alexanderx" -Z "3" -Tr True -B "levenshtein"  
18
```

```
Query: alexanderx
```

```
Number of results: 18
```

```
>python SAR_Searcher.py index_20151.bin -Q "alexanderx" -Z "3" -Tr False -B "intermediate"  
18
```

```
Query: alexanderx
```

```
Number of results: 18
```

```
>python SAR_Searcher.py index_20151.bin -Q "alexanderx" -Z "3" -Tr False -B "restricted"  
18
```

```
Query: alexanderx
```

```
Number of results: 18
```