

TP – Modelos de RI



Universidad Nacional de Lujan

Materia: Recuperación de la Información

Profesores:

Gabriel Tolosa

Pablo Lavallen

Alumno: Leonardo Duville

- 1) B- Hay algunas diferencias con el archivo de documentos relevantes. Esto se debe a que con que un termino aparezca en el documento lo va a devolver en la consulta.
D- Se puede ver una pequeña mejora en la eficiencia, sobre todo en el modelo booleano, pero tampoco es en todas las queries. Con respecto al modelo vectorial, los documentos relevantes recuperados los resultados son los mismos.

4) Al utilizar TF*IDF para calcular los pesos de los términos y repetir las consultas, se aprecia una mejora en la precisión y relevancia de los resultados de recuperación de información, así como una mejor comprensión de la importancia de los términos en el proceso de búsqueda.

5)

QUERY	N = 10	N = 25	N = 50
software	0,99	1	0,99
hardware	0,99	1	1
english	0,99	1	0,99
spanish	0,99	1	0,99
italian	0,99	1	0,99

El coeficiente fue prácticamente uno en todas las queries con diferentes cantidades de documentos recuperados en los rankings, por lo que los rankings tienen una relación positiva y directa.

7)

Realice los coeficientes de correlación de las queries y tuvieron diversos resultados. La query 1, 2 y 3, obtuvieron correlación negativa, mientras que la query 4 y 5, la relación es positiva pero tampoco muy fuerte.

QUERY: software

```
Coeficiente de correlación de Spearman 5 elementos: 0.19999999999999998
Coeficiente de correlación de Spearman 10 elementos: -0.30909090909090909
```

QUERY: hardware

```
Coeficiente de correlación de Spearman 5 elementos: -0.09999999999999999
Coeficiente de correlación de Spearman 10 elementos: -0.01818181818181818
```

QUERY: english

```
Coeficiente de correlación de Spearman 5 elementos: -0.3
Coeficiente de correlación de Spearman 10 elementos: 0.24848484848484845
```

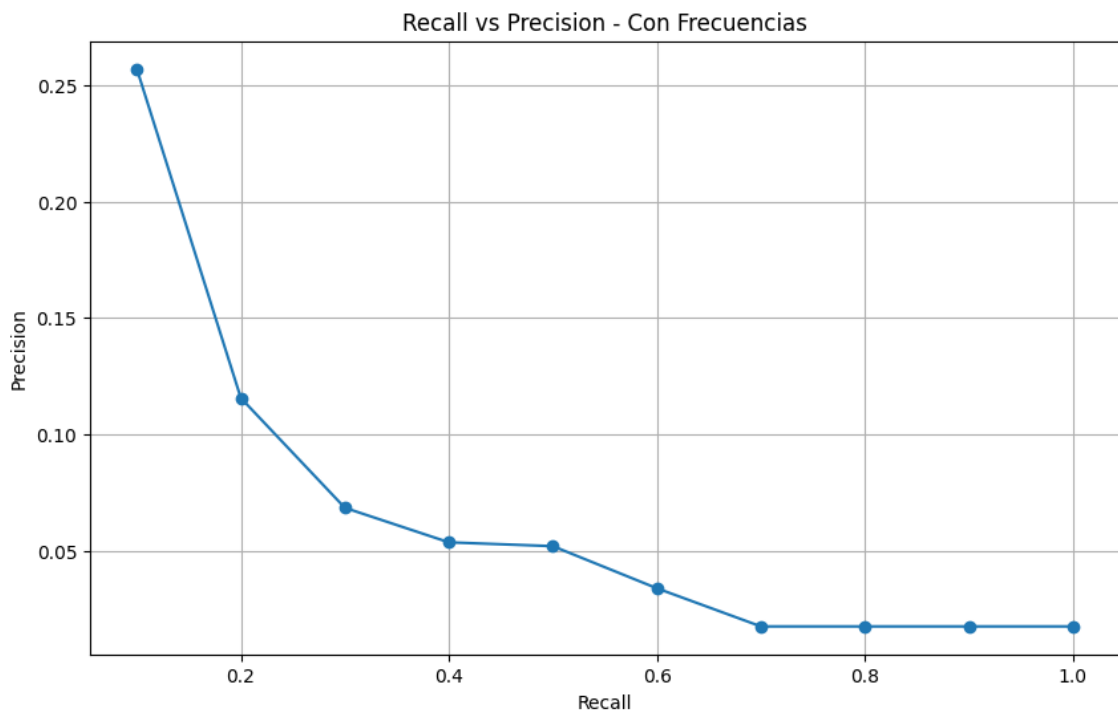
QUERY: spanish

```
Coeficiente de correlación de Spearman 5 elementos: 0.49999999999999994
Coeficiente de correlación de Spearman 10 elementos: 0.006060606060606061
```

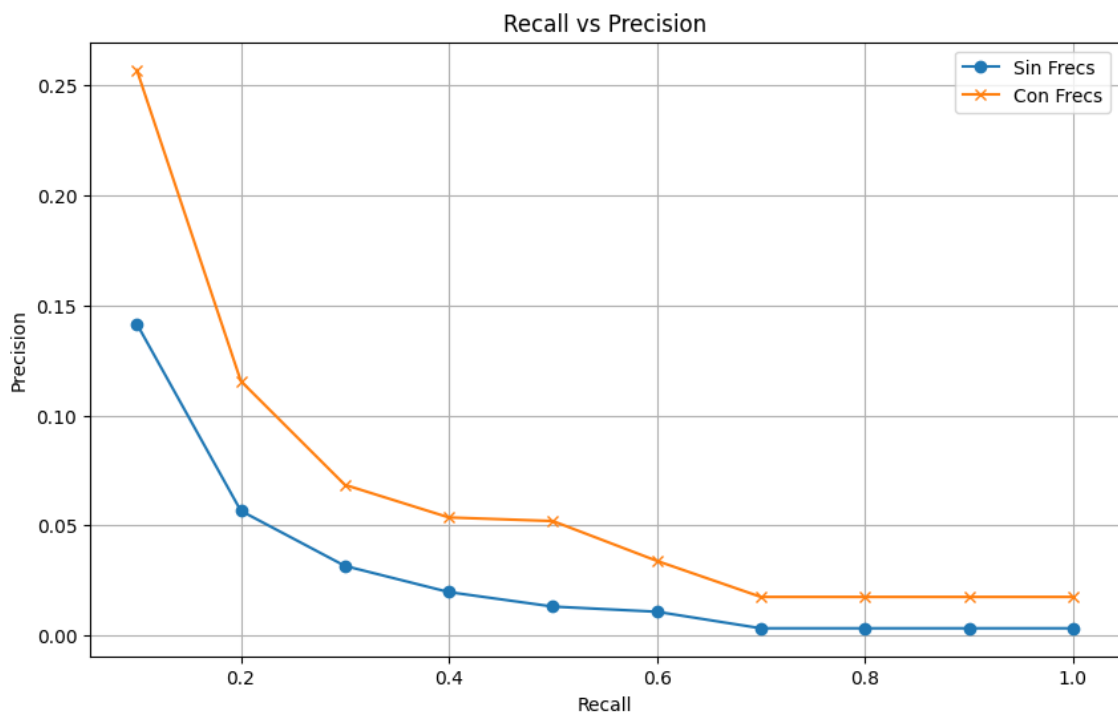
QUERY: italian

```
Coeficiente de correlación de Spearman 5 elementos: 0.49999999999999994
Coeficiente de correlación de Spearman 10 elementos: 0.2121212121212121
```

9) Precisión frente a recall interpolada (con frecuencias):

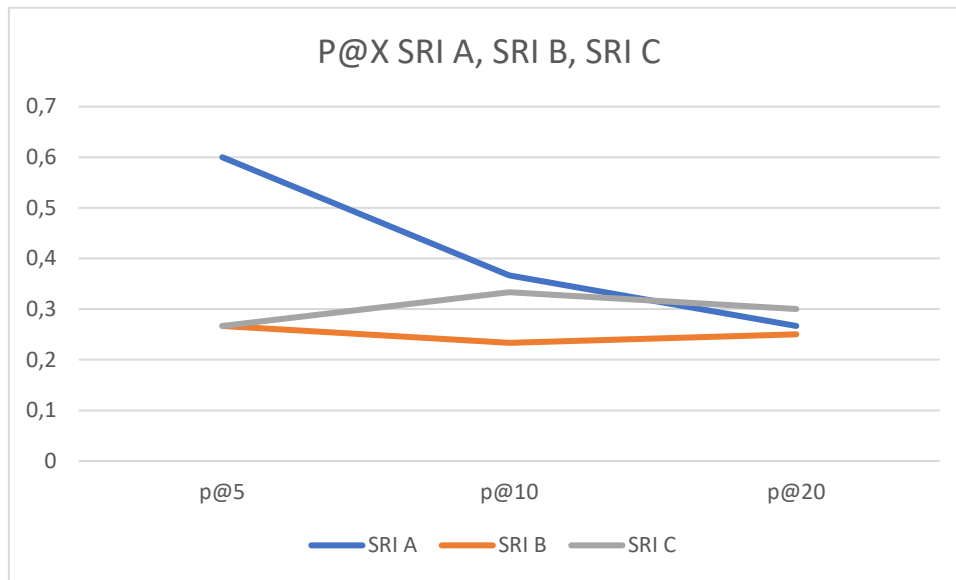


Cuando ejecuté el experimento sin tener en cuenta las frecuencias, obtuve una disminución en los 11 puntos estándar de R-P.



Esto quiere decir que Terrier puede obtener mejores resultados al tener en cuenta las frecuencias de los términos de la query. Al tener las frecuencias, Terrier puede identificar y recuperar documentos relevantes de forma más efectiva.

10)



El sistema A es que mayor precisión inicial posee. Sin embargo, su rendimiento disminuye a medida que avanza el intervalo de recall.

El sistema B y C se mantienen más estables, aunque su precisión no es la misma que el sistema A, estos pueden ser útiles en sistemas se valora la precisión en todos los rangos.

Escenario posible:

Una empresa de consultoría que necesita acceder rápidamente a información relevante sobre regulaciones gubernamentales para diferentes sectores industriales. Esta empresa utiliza sistemas de recuperación de información para buscar y analizar documentos relevantes que les ayuden a mantenerse actualizados con los cambios regulatorios y a ofrecer asesoramiento preciso a sus clientes.

En este caso usaría el sistema A, ya que es el que mayor precisión media posee y el que mejor recupera documentos relevantes en las primeras posiciones.