

TP2: Modelos de Recuperación de Información (y evaluación)

Alumno: Gallozo, Luis Angel 156308

1. Utilizando la colección provista por el equipo docente, cuya estructura es la siguiente:

vocabulary.txt → [id termino, idf, termino]

documentVectors.txt → [id doc, lista(id terminos)]

queries.txt → [id query, lista(id terminos)]

relevants.txt → [id query, listarelevantes (id doc)]

informationNeeds.txt → [id in, texto libre]

a) Calcule los conjuntos de respuestas usando el modelo booleano y el modelo vectorial (asuma en todos los casos $TF = 1$).

Se unieron los términos de cada query con un “or”. Además no se ordenaron los resultados, para no hacer un ranking, esto es debido a que es modelos booleano no nos serviría para hacerlo.

Conjunto de respuestas del modelo booleano:

- **Query 1:** 5, 15, 17, 25, 29, 32, 36
- **Query 2:** 1, 4, 6, 8, 9, 10, 14, 18, 19, 21, 29, 32, 36, 37, 38
- **Query 3:** 2, 8, 10, 11, 14, 18, 19, 20, 22, 32, 33, 34
- **Query 4:** 7, 8, 10, 13, 18, 27, 30, 32, 35
- **Query 5:** 8, 10, 18, 32

Para el modelo vectorial, podemos crear un ranking por relevancia. Para esto se utilizó el método de similitud por coseno:

- **Query 1:** 36, 17, 25, 29, 15, 32, 5
- **Query 2:** 8, 32, 10, 18, 4, 21, 9, 38, 37, 6, 14, 19, 1, 29, 36
- **Query 3:** 32, 22, 8, 10, 34, 18, 14, 11, 20, 33, 19, 2
- **Query 4:** 8, 32, 10, 18, 35, 27, 30, 7, 13
- **Query 5:** 8, 32, 10, 18

b) Compare los resultados contra los relevantes y trate de explicar las diferencias.

Comparación Booleana:

Comparación Query 1:

Recuperados: **5, 15, 17, 25, 29, 32, 36**

Relevantes: **5, 15, 25, 36**

Se recuperaron todos los documentos relevantes, aunque los documentos 17, 29 y 32 no son relevantes, lo más probable es que contenían términos de la query.

Comparación Query 2:

Recuperados: **1, 4, 6, 8, 9, 10, 14, 18, 19, 21, 29, 32, 36, 37, 38**

Relevantes: **1, 4, 6, 9, 14, 17, 19, 21, 23, 29, 36, 37, 38, 147, 195, 196**

No se recuperaron todos los documentos relevantes. Ocurre debido a que los documentos (17, 23, 147, 195 y 196) no contienen los términos de la query. Además surge que se recuperaron documentos (8,10,18 y 32) que no son relevantes debido a que contienen 1 o más términos.

Comparación Query 3:

Recuperados: **2, 8, 10, 11, 14, 18, 19, 20, 22, 32, 33, 34**

Relevantes: **2, 3, 11, 12, 13, 16, 20, 22, 24, 26, 28, 33, 34**

No se recuperaron todos los relevantes. Como primera observación, vemos que los documentos (8,10,18 y 32) fueron recuperados a pesar de no ser relevantes, esto es debido al término "PATIENTS". Luego como segunda observación, pasa lo mismo con el término "DIET" agregando los documentos no relevantes: 14,19,32.

Comparación Query 4:

Recuperados: **7, 8, 10, 13, 18, 27, 30, 32, 35**

Relevantes: **7, 27, 30, 31, 35**

Se recuperan todos los relevantes menos el 31. Vemos que sucede otra vez la recuperación de documentos no relevantes debido al término "PATIENTS". Además el término SWEATING no aparece en ninguno de los documentos.

Comparación Query 5:

Recuperados: **8, 32, 10, 18**

Relevantes: **8, 10, 18, 32**

En esta query se da el caso que todos los términos están en todos los documentos recuperados, además que los documentos recuperados resultan ser todos los relevantes.

Comparación vectorial:

Luego de calcular los scores de los archivos relevantes de la colección, basándonos en la similitud coseno y se realizó un ranking de los documentos por cada query.

Comparación Query 1:

Recuperados: **36**, 17, **25**, 29, **15**, 32, **5**

Relevantes: **5, 15, 25, 36**

Se observa la aparición del documento 17 como segundo en el ranking de recuperados, esto es debido a que contiene al término 117 y tiene un IDF de 5.2, lo cual hace que suba hasta esa posición, quizás si tuviésemos frecuencias (TF) mayor a 1, podríamos rankear mejor los documentos. Además vemos que ocurre lo mismo con los documentos 29 y 32, aparecen en los recuperados debido a que contienen al término 191 VITAMINA-A.

Comparación Query 2:

Recuperados: 8, 32, 10, 18, **4, 21, 9, 38, 37, 6, 14, 19, 1, 29, 36**

Relevantes: **1, 4, 6, 9, 14, 17, 19, 21, 23, 29, 36, 37, 38**, 147, 195, 196

En la comparación booleana se habló de los docs 8,32,10,18 que a pesar de no ser relevantes aparecen por tener el término PATIENTS, aún así aparecen en esas posiciones debido al IDF alto que posee el término.

Comparación Query 3:

Recuperados: 32, **22**, 8, 10, **34**, 18, 14, **11, 20, 33**, 19, **2**

Relevantes: **2, 3, 11, 12, 13, 16, 20, 22, 24, 26, 28, 33, 34**

Comparación Query 4:

Recuperados: 8, 32, 10, 18, **35, 27, 30, 7**, 13

Relevantes: **7, 27, 30, 31, 35**

Comparación Query 5:

Recuperados: **8, 32, 10, 18**

Relevantes: **8, 10, 18, 32**

Como se puede observar en la mayoría de los recuperados, algunos relevantes son recuperados pero debido a uno o más términos se recuperan no relevantes y además surgen ocasiones en donde estos documentos tienen un ranking alto por el IDF de los términos que maneja. Entonces terminamos teniendo un ranking muy disperso y poco preciso al realizar el ranking, quitando claro al último query.

En el caso de la Query 3, vemos que se omitió utilizar el término “ENZYMES” para la elaboración de la querys, de igual forma no se contaba con algún documento que lo contuviera.

c) Usando las necesidades de información reescriba los 5 queries y repita la operación.

d) Indique si pudo mejorar la eficiencia a partir de las nuevas consultas.

Query 1:

Antes:

Query 1: (ENERGY-METABOLISM, LIVER-CIRRHOSIS,VITAMINA-A)

Query 1: (72, 117, 191)

Nueva:

Query 1: (ENERGY-METABOLISM, VITAMINA-A)

Query 1: (72, 191)

Se decidió que era mejor opción eliminar el término LIVER-CIRRHOSIS de la query, debido a que este no estaría sumando documentos relevantes.

Resultados:

Recuperados: **5, 15, 25, 29, 32, 36**

Relevantes: **5, 15, 25, 36**

Query 2:

Antes:

Query 2: (PATIENTS, VITAMIN-E, VITAMIN-E-DEFICIENCY)

Query 2: (147, 195, 196)

Nueva:

Query 2: (VITAMIN-E, VITAMIN-E-DEFICIENCY)

Query 2: (195, 196)

Se decidió que era mejor opción eliminar el término PATIENTS de la query, debido a que este no estaría sumando documentos relevantes. Se analizó si podríamos agregar algún término como por ejemplo "VITAMINS" pero solo aportaba documentos no relevantes. Además se pensó en agregar términos que estuvieran en los documentos relevantes no recuperados pero no tiene sentido ya que no están representados en la necesidad de información.

Resultados:

Recuperados: **4, 21, 9, 38, 37, 6, 14, 19, 1, 29, 36**

Relevantes: **1, 4, 6, 9, 14, 17, 19, 21, 23, 29, 36, 37, 38, 147, 195, 196**

Query 3:

Antes:

Query 3: (DIET, DIET-THERAPY, PANCREATIC-EXTRACTS, PANCREATIC-JUICE, PATIENTS)

Query 3: (55, 56, 141, 142, 147)

Nueva:

Query 3: (DIET, DIET-THERAPY, PANCREATIC-EXTRACTS, PANCREATIC-JUICE)

Query 3: (55, 56, 141, 142)

Se decidió que era mejor opción eliminar el término PATIENTS de la query, debido a que este no estaría sumando documentos relevantes. Se pensó en agregar el término PANCREAS, pero al revisar los docs donde está ubicado, encontramos que son el 8, 13, 24, 27 y 35, de los cuales el 13 y 24 son relevantes, si bien los otros 3 docs no son relevantes se podría plantear la query:

Query 3: (DIET, DIET-THERAPY, PANCREAS, PANCREATIC-EXTRACTS, PANCREATIC-JUICE)

Query 3: (55, 56, 139, 141, 142)

Resultados:

Recuperados: **22, 34, 14, 11, 20, 33, 19, 2**

Relevantes: **2, 3, 11, 12, 13, 16, 20, 22, 24, 26, 28, 33, 34**

Query 4:

Antes:

Query 4: (PATIENTS, SWEAT, SWEATING)

Query 4: (147, 179, 180)

Nueva:

Query 4: (SWEAT)

Query 4: (179)

Se decidió que era mejor opción eliminar el término SWEATING que no tenía documentos que la contuvieran, además de eliminar también PATIENTS que solo aportaba documentos no relevantes.

Resultados:

Recuperados: **35, 27, 30, 7, 13**

Relevantes: **7, 27, 30, 31, 35**

Query 5:

Antes:

Query 5: (PATIENTS, TASTE, TASTE-DISORDERS)

Query 5: (147, 182, 184)

Variaciones:

1) Query 5: (TASTE, TASTE-DISORDERS)

Query 5:(182, 184)

2) Query 5: (TASTE-DISORDERS)

Query 5:(184)

3) Query 5: (TASTE)

Query 5:(182)

Esta query, tuvo resultados excelentes, ya que se recuperaron todos los relevantes, pero también se observó que todos los términos tenían aparición en todos los documentos relevantes, por lo que cada una de las variaciones, serviría. Cabe aclarar que el término PATIENTS tiene como condición especial estar relacionado con cada documento relevante pero es simple casualidad, ya que no tendría una fuerte relación con la necesidad de información.

Resultados:

Recuperados: **8, 32, 10, 18**

Relevantes: **8, 10, 18, 32**

Cabe aclarar que en cada query se pensó en agregar el término CYSTIC-FIBROSIS (51) o CF, pero no se realizó debido a que no aporta valor debido a su IDF de 0.

2. Dados los siguientes documentos, arme la matriz termino-documento (TD)

- Doc 1 = {El software libre ha tenido un papel fundamental en el crecimiento de Internet. Además, Internet ha favorecido la comunicación entre los desarrolladores de software.}
- Doc 2 = {La mayor riqueza que tiene un país es la cultura, eso lo hace más libre.}
- Doc 3 = {La producción de software es fundamental para nuestro país, como así también lo es la producción de tecnología de hardware y comunicación}
- Doc 4 = {La cultura del software libre está en crecimiento. Es fundamental que nuestro país incorpore software libre en el estado.}

¿Qué documentos se recuperan en cada caso para las siguientes consultas booleanas? (Muestre mediante operaciones con conjuntos como se resuelven las consultas)

a) (not software) or (pais and fundamental)

(not software) = {d2}

(pais) = {d2,d3,d4}

(fundamental) = {d1,d3,d4}

(pais and fundamental) = {d2,d3,d4} \cap {d1,d3,d4} = {d3,d4}

R: = {d2} \cup {d3,d4} = {d2,d3,d4}

b) produccion and (cultura or libre)

(produccion) = {d3}

(cultura) = {d2,d4}

(libre) = {d1,d2,d4}

(cultura or libre) = {d2,d4} \cup {d1,d2,d4} = {d1,d2,d4}

R: {d3} \cap {d1,d2,d4} = { \emptyset }

c) fundamental or libre or pais

(fundamental) = {d1,d3,d4}

(libre) = {d1,d2,d4}

(pais) = {d2,d3,d4}

$R:\{d1,d3,d4\} \cup \{d1,d2,d4\} \cup \{d2,d3,d4\} = \{d1,d2,d3,d4\}$

3. Utilizando los documentos del ejercicio anterior arme la matriz TD pero calculando w_{ij} como la frecuencia del i-esimo termino en el j-esimo documento. Calcule el ranking para la siguientes consultas utilizando como metrica el producto escalar y luego repita con la metrica del coseno.

a) software

b) pais libre

c) produccion software pais

| TF - Producto Escalar | | | | | |
|-----------------------|-------|---------------|-------|-----------------------------|-------|
| a) Software | | b) Pais Libre | | c) Produccion Software Pais | |
| Ranking | Score | Ranking | Score | Ranking | Score |
| d1 | 2 | d4 | 3 | d3 | 4 |
| d4 | 2 | d2 | 2 | d4 | 3 |
| d3 | 1 | d3 | 1 | d1 | 2 |
| d2 | 0 | d1 | 1 | d2 | 1 |

| TF - Coseno | | | | | |
|-------------|--------------|---------------|--------------|-----------------------------|--------------|
| a) Software | | b) Pais Libre | | c) Produccion Software Pais | |
| Ranking | Score | Ranking | Score | Ranking | Score |
| d4 | 0,5547001962 | d2 | 0,632455532 | d3 | 0,7302967433 |
| d1 | 0,5163977795 | d4 | 0,5883484054 | d4 | 0,4803844614 |
| d3 | 0,316227766 | d3 | 0,2236067977 | d1 | 0,298142397 |
| d2 | 0 | d1 | 0,1825741858 | d2 | 0,2581988897 |

En las querys a) y b), podemos apreciar que si bien el ranking es similar, gracias a la métrica del coseno podemos definir con más precisión los scores de cada documento, logrando esas pequeñas diferencias en las posiciones.

Específicamente en la query a) tuvimos que solo con el producto escalar con TF no podemos ser capaces de determinar cual documento es más relevante que otro (en el caso del d1 y d4), ambos tienen score 2 debido al TF de 2 para el término.

Al probar el ranking usando el coseno, vemos que la diferencia entre los scores es más detallada, obtenemos un ranking más preciso. Esto ocurre debido a que tenemos en

cuenta la longitud normalizada del query y del documento. El documento 1 obtuvo una penalización en su score debido a que es más grande en tamaño que todos los demás documentos, y además, se penaliza levemente al score del documento 4 debido a que es el segundo más grande.

Así como en el ejemplo anterior, sucede en las demás queries.

4. Rearme la matriz del ejercicio anterior pero calcule los pesos de acuerdo a TF * IDF. Repita todas las consultas (por ambas métricas). ¿Puede obtener alguna conclusión?

| TFxIDF - Producto Escalar | | | | | |
|---------------------------|---------------|---------------|---------------|-----------------------------|---------------|
| a) Software | | b) Pais Libre | | c) Produccion Software Pais | |
| Ranking | Score | Ranking | Score | Ranking | Score |
| d1 | 0,03121937581 | d4 | 0,04682906371 | d3 | 0,7561718421 |
| d4 | 0,03121937581 | d2 | 0,03121937581 | d4 | 0,04682906372 |
| d3 | 0,01560968791 | d1 | 0,0156096879 | d1 | 0,03121937581 |
| d2 | 0 | d3 | 0,0156096879 | d2 | 0,01560968791 |

En la primera query a), tenemos que 2 documentos(d1 y d4) tienen el mismo score, esto sucede debido a que ambos tienen el mismo término y la misma frecuencia. Por lo que seguimos sin poder determinar un ranking preciso.

En la segunda query b), sucede que los scores son similares (d1 y d3), ya que los términos “Pais” y “Libre” tienen la misma frecuencia e IDF. Por lo que usar TFxIDF da una aproximación a un score más preciso, pero no lo suficiente cuando tenemos frecuencia e IDF's similares.

La query c) mantiene resultados iguales que en TF-Producto Escalar.

| TFxIDF - Coseno | | | | | |
|-----------------|---------------|---------------|---------------|-----------------------------|---------------|
| a) Software | | b) Pais Libre | | c) Produccion Software Pais | |
| Ranking | Score | Ranking | Score | Ranking | Score |
| d4 | 0,2987009834 | d4 | 0,3168202363 | d3 | 0,7925334618 |
| d1 | 0,1490052933 | d2 | 0,1920100747 | d4 | 0,08921640843 |
| d3 | 0,08216266207 | d3 | 0,05809777551 | d1 | 0,02967006661 |
| d2 | 0 | d1 | 0,05268132668 | d2 | 0,02703496697 |

Cambios:

- Query a) ranking más preciso para d4 y d1.
- Query b) ranking más preciso para d3 y d1.
- Query c) sin cambios en el ranking.

Comparándolo con los resultados del ranking de Coseno usando solo TF, vemos que cambia el ranking en ocasiones, esto es debido a que se está teniendo en cuenta la penalización de los documentos de mayor y menor longitud. Debido a la utilización combinada de TFxIDF y Coseno, podemos determinar una relevancia más precisa, ya que tenemos mejor valoración de términos comunes y las longitudes de los archivos.

5. Utilizando Terrier indexe la colección Wiki-Small. Tome 5 necesidades de información y – de forma manual – derive una consulta (query). Para cada una, pruebe la recuperación por los modelos basados en TF * IDF y BM25. ¿Cómo se comportan los rankings? Calcule el coeficiente de correlación para los primeros 10, 25 y 50 resultados. ¿Qué conclusiones obtiene?

Necesidades de información:

- INFORMATION RETRIEVAL
- MACHINE LEARNING
- CLOUD COMPUTING
- INFORMATION SECURITY
- ARTIFICIAL INTELLIGENCE

| | Query 1 | Query 2 | Query 3 | Query 4 | Query 5 |
|-------------|---------|--------------------|-------------------|--------------------|-------------------|
| Primeros 10 | 1 | 0,9999618003771690 | 0,999651758136061 | 1 | 0,999999142093939 |
| Primeros 25 | 1 | 0,9999284380377080 | 0,999472275920751 | 0,9999999999999999 | 0,999996809766000 |
| Primeros 50 | 1 | 0,9999141029793130 | 0,999547371646568 | 1 | 0,999996248886504 |

Como se ve en los resultados, los coeficientes son muy cercanos a 1, lo cual indica que ambos rankings están altamente correlacionados. Es decir que, los rankings son casi similares (al menos en esta prueba). Solo en hubo 2 lugares del ranking (del 1 al 50) que fue diferente:

```
Diferencias en rankings: Docs id y sus scores
Pos - Doc_1 <> Doc_2 - score_Doc_1 <> score_Doc_2
-----
[29] 494 2929 7.443157638061469 4.0950731033062375
[30] 2929 494 7.4430118666611325 4.081767178196611
```

Como ya se dijo, todos los rankings fueron casi similares salvo por estas 2 posiciones, por lo que se concluye que la relación es positiva y directa, además de que al menos en esta prueba no podemos decir o determinar cuál de los 2 modelos de recuperación es más efectivo o no para las necesidades de información elegidas (debido a que no se cuenta con juicios de relevancia).

El script y la planilla excel, serán adjuntados en la entrega.

6. Escriba un pequeño programa que lea un directorio con documentos de texto y arme una estructura de datos en memoria para soportar la recuperación. Luego, debe permitir ingresar un query y devolver un ranking de los documentos relevantes utilizando el modelo vectorial. Se debe soportar la ponderación de los términos de la consulta. Implemente las versiones sugeridas en MIR [1].

El programa se basa en 2 archivos:

- programa_ri.py: Se encarga de mostrar el menu, 3 opciones:
 - 1) Procesar Corpus.
 - 2) Realizar Query
 - 3) Salir
 - Al final la query se genera un archivo ranking.txt
- motor_ri.py: Tiene toda la lógica de:
 - Tokenización
 - Cálculo de TF-IDF de cada Documento
 - Calcula la similitud entre los documentos y la query, utilizando el método del coseno.
 - Genera los rankings.

7. Indexe la colección del ejercicio 5 con su software. Ejecute las consultas y compare los resultados con los obtenidos con Terrier. ¿Son consistentes?

Según la comparación de los primeros 100 resultados de terrier y mi script, obtenemos que la correlación general es positiva y moderada, no muy fuerte en algunos casos, como se muestra en la siguiente imagen.

Correlación Q1 :0.6785124267222156

Correlación Q2 :0.6538535893631239

Correlación Q3 :0.5886652762669767

Correlación Q4 :0.7466581773801598

Correlación Q5 :0.5041458623344943

Por lo que se considera que mi script no es muy robusto en: tokenización, normalización, stemming y cálculo de similitudes. Todo esto comparado con la complejidad que maneja Terrier en las características ya mencionadas.

8. Se requiere evaluar la performance en la recuperación de un sistema. Para una consulta q1, dicho sistema entregó la siguiente salida.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| R | N | N | R | R | N | N | N | N | R | N | N | N | R | N |

Los documentos identificados como R son los relevantes, mientras que las N's corresponden a documentos no relevantes a q1. Suponga – además – que existen en el corpus otros 6 documentos relevantes a q1 que el sistema no recuperó. A partir de esta salida calcule las siguientes medidas:

- Recall y Precision para cada posición j**
- Precision promedio**
- Precision al 50% de Recall**
- Precision interpolada al 50% de Recall**
- Precision-R**

Finalmente, realice las gráficas interpoladas y sin interpolar. Luego, intérprete brevemente los resultados y brinde una explicación.

| a) | | |
|----|--------|-----------|
| | Recall | Precisión |
| R | 0,09 | 1,00 |
| N | 0,09 | 0,50 |
| N | 0,09 | 0,33 |
| R | 0,18 | 0,50 |
| R | 0,27 | 0,60 |
| N | 0,27 | 0,50 |
| N | 0,27 | 0,43 |
| N | 0,27 | 0,38 |
| N | 0,27 | 0,33 |
| R | 0,36 | 0,40 |
| N | 0,36 | 0,36 |
| N | 0,36 | 0,33 |
| N | 0,36 | 0,31 |
| R | 0,45 | 0,36 |
| N | 0,45 | 0,33 |

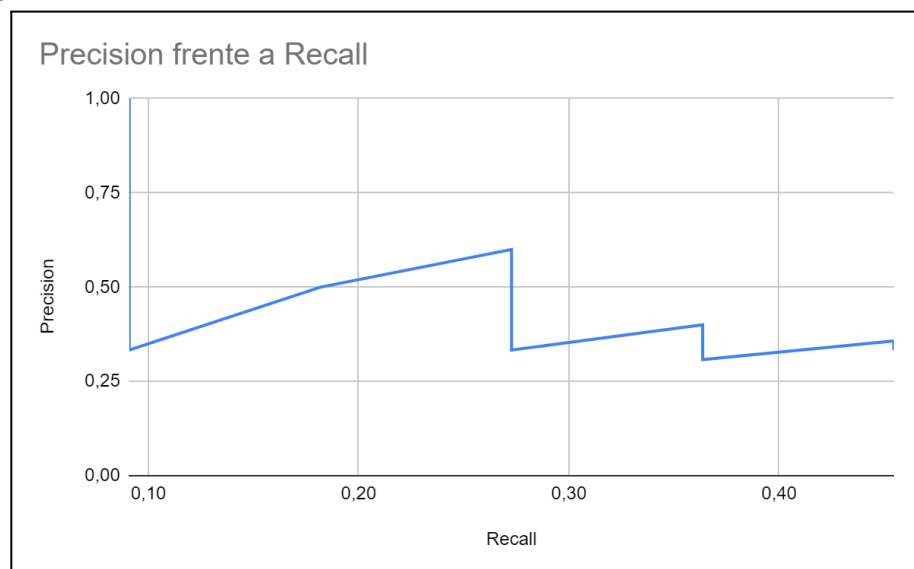
b) Precisión promedio: 0,25974026

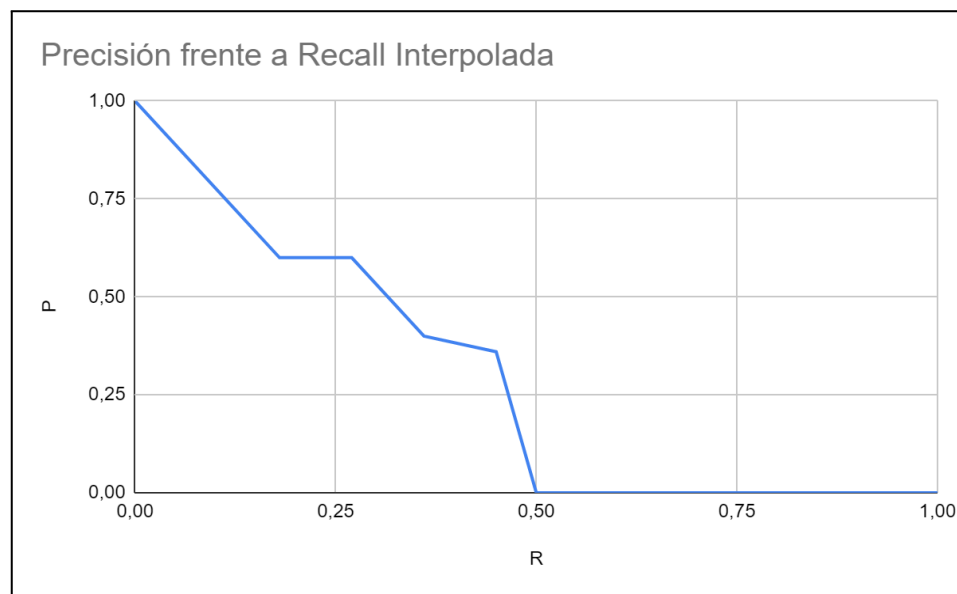
c) No se logra alcanzar el 50% del recall.

d) Precisión interpolada al 50 % de Recall = 0

e) R-Precision: 0,36 (sería la precisión en la posición 11, ya que los relevantes son 11)

P@5: 0,60; P@10: 0,40





Como se puede observar, la precisión se mantiene constante en 1 para el primer punto y luego disminuye a medida que aumenta el recall. Esto quiere decir que al aumentar el recall se recuperan más documentos relevantes pero también más irrelevantes, reduciendo así la precisión.

En conclusión se observa una relación inversa entre el recall y la precisión. Indicando el gran desafío de hallar un equilibrio entre la recuperación de documentos relevantes y minimizar la recuperación de documentos irrelevantes.

9. Utilizando la colección de prueba CISI y Terrier se debe realizar la evaluación del sistema. Para ello, es necesario construir un índice con los documentos de la colección y luego ejecutar las consultas, las cuales se deben armar a partir de los términos que considere de las necesidades de información. Los resultados deben ser comparados contra los juicios de relevancia de la colección utilizando el software trecval. Realizar el análisis y escribir un reporte indicando los resultados obtenidos, junto con la gráfica de R-P en los 11 puntos standard. Realice dos experimentos: en el primero, no considere la frecuencia de los términos en el query mientras que en el segundo lo debe tener en cuenta.

En primer lugar, preparé la colección CISI para transformarla al formato trec:

- Preprocesamiento del archivo de documentos(CISI.ALL)
- Preprocesamiento del archivo de relevantes(CISI.REL).
- Preprocesamiento del archivo queries (CISI.QRY). Se realizó tokenizado y luego extracción de términos.

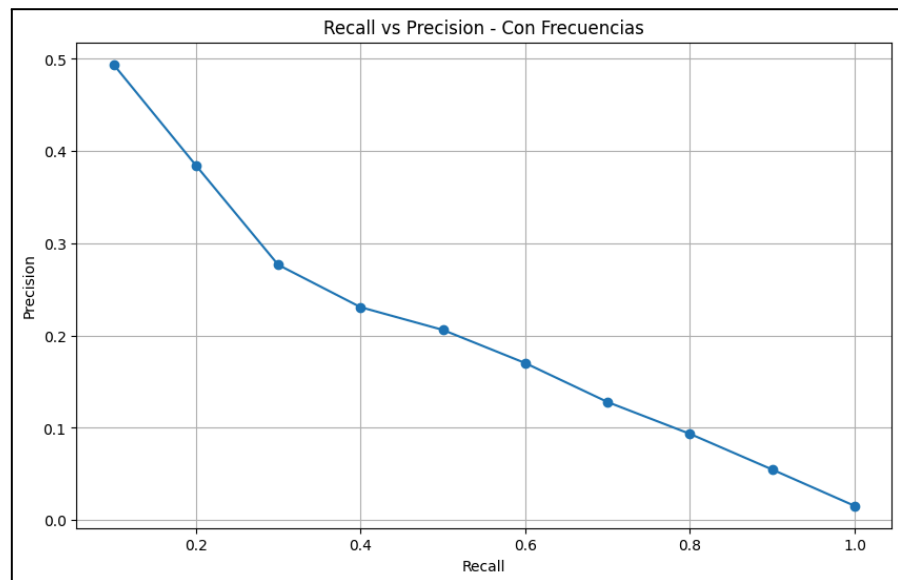
Se generaron dos archivos de queries, uno que tiene términos repetidos en la query, y otro que sólo tiene términos unívocos.

Luego se indexó la colección con Terrier, se ejecutaron las queries utilizando el modelo TF_IDF, y se realizó la evaluación de resultados, obteniendo:

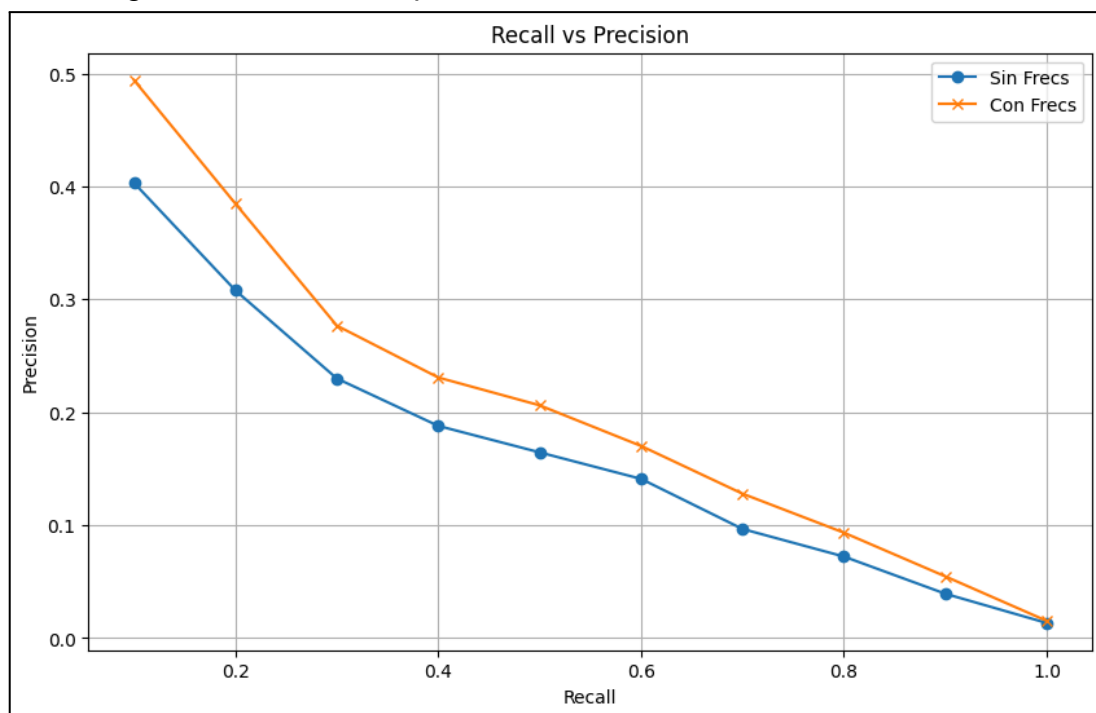
- Se recuperaron 65535 documentos
- En total, eran relevantes 3114 pero se recuperaron 2673 de estos.

Con Frecuencias: precisión frente a recall interpolada:

| | Recall | Precision |
|---|--------|-----------|
| 0 | 0.1 | 0.494026 |
| 1 | 0.2 | 0.384637 |
| 2 | 0.3 | 0.276709 |
| 3 | 0.4 | 0.230806 |
| 4 | 0.5 | 0.206150 |
| 5 | 0.6 | 0.170184 |
| 6 | 0.7 | 0.127931 |
| 7 | 0.8 | 0.093429 |
| 8 | 0.9 | 0.054718 |
| 9 | 1.0 | 0.015327 |



Al realizar el experimento quitando los términos repetidos en las queries, se obtuvo una disminución significativa en los 11 puntos standard de R-P:



Esto sugiere que al tener en cuenta la frecuencia de los términos en la query, Terrier puede obtener mejores resultados en términos de precisión. La frecuencia de los términos puede proporcionar información adicional sobre la importancia de esos términos para la query, lo que ayuda a Terrier a identificar y recuperar documentos relevantes de manera más efectiva.

10. Dadas las salidas de tres sistemas de recuperacion de informacion para 3 consultas cualquiera y los juicios de relevancia creados por asesores humanos, calcule para cada sistema:

a) La precision media

b) La precision media a intervalos de Recall de 20%

c) P@5, P@10, P@20

Luego, exponga un escenario posible y medidas complementarias para decidir que sistema utilizar.

a)

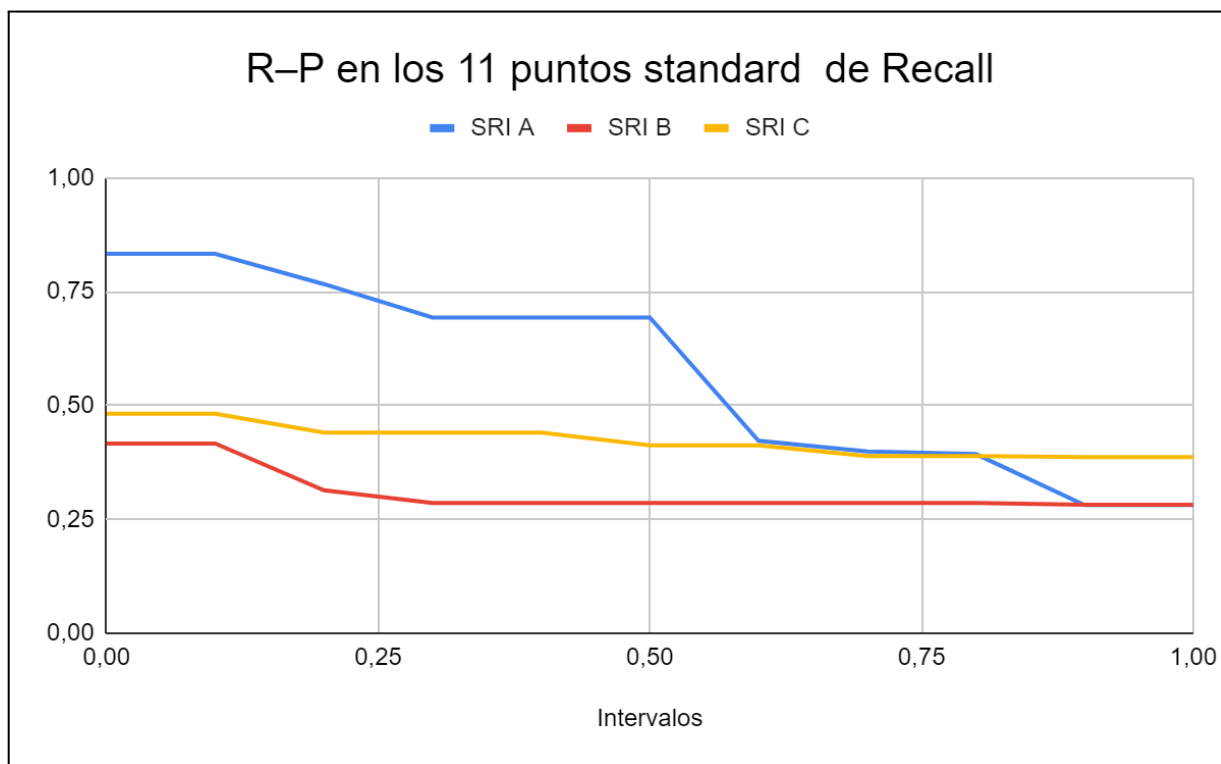
| SRI A | | | |
|------------------------|--------------|--------------|-------------|
| | Query 1 | Query 2 | Query 3 |
| Average Precision | 0,724025974 | 0,6178571429 | 0,328488082 |
| Mean Average Precision | 0,5567903996 | | |

| SRI B | | | |
|------------------------|--------------|--------------|--------------|
| | Query 1 | Query 2 | Query 3 |
| Average Precision | 0,2291353383 | 0,2875505275 | 0,3486185154 |
| Mean Average Precision | 0,2884347937 | | |

| SRI C | | | |
|------------------------|--------------|--------------|--------------|
| | Query 1 | Query 2 | Query 3 |
| Average Precision | 0,4952380952 | 0,3320516612 | 0,2918911289 |
| Mean Average Precision | 0,3730602951 | | |

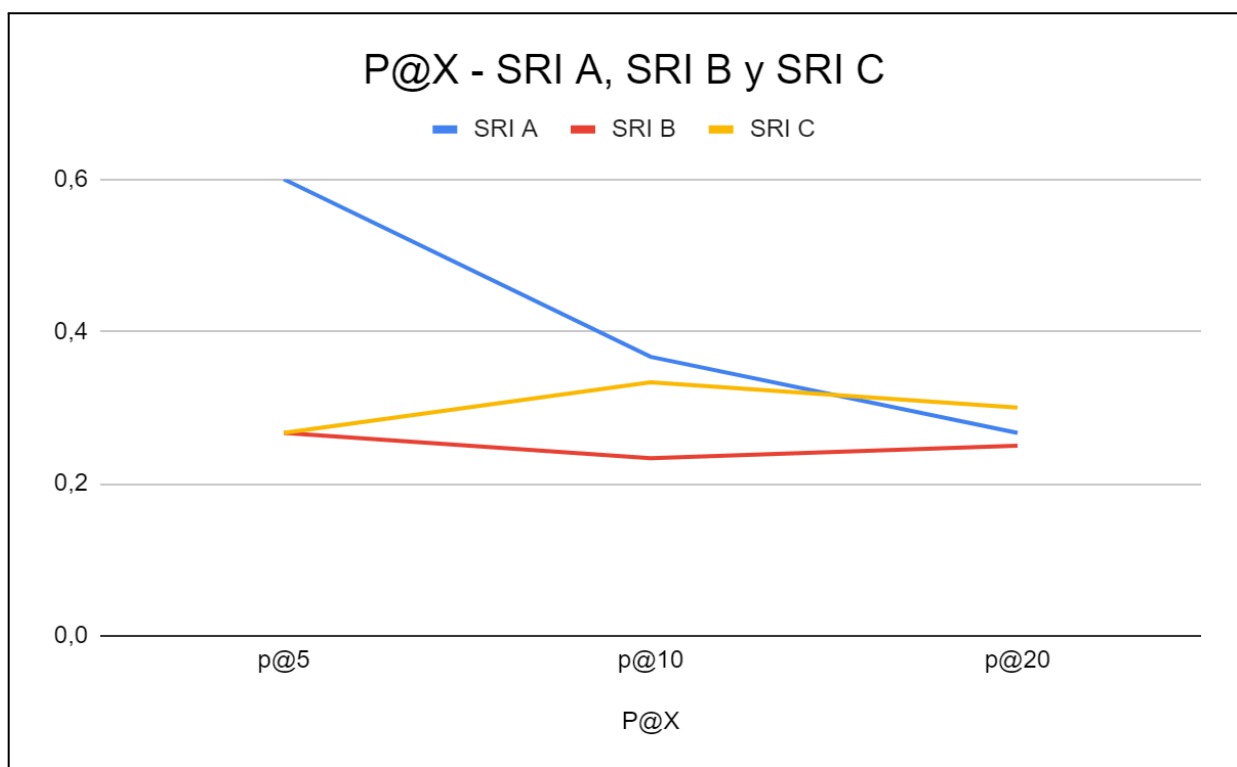
b)

| Intervalos | SRI A | SRI B | SRI C |
|------------|--------------|--------------|--------------|
| 0 | 0,8333333333 | 0,4166666667 | 0,4821428571 |
| 0,1 | 0,8333333333 | 0,4166666667 | 0,4821428571 |
| 0,2 | 0,7666666667 | 0,3138888889 | 0,4404761905 |
| 0,3 | 0,6933333333 | 0,2858187135 | 0,4404761905 |
| 0,4 | 0,6933333333 | 0,2858187135 | 0,4404761905 |
| 0,5 | 0,6933333333 | 0,2858187135 | 0,4126984127 |
| 0,6 | 0,4226984127 | 0,2858187135 | 0,4126984127 |
| 0,7 | 0,3988888889 | 0,2858187135 | 0,3888888889 |
| 0,8 | 0,3933333333 | 0,2858187135 | 0,3888888889 |
| 0,9 | 0,2814646465 | 0,2820048309 | 0,3866959064 |
| 1 | 0,2814646465 | 0,2820048309 | 0,3866959064 |



c)

| | Promedio en P@X | | |
|------|-----------------|--------------|--------------|
| P@X | SRI A | SRI B | SRI C |
| p@5 | 0,6 | 0,2666666667 | 0,2666666667 |
| p@10 | 0,3666666667 | 0,2333333333 | 0,3333333333 |
| p@20 | 0,2666666667 | 0,25 | 0,3 |



Es crucial considerar no sólo la precisión en los primeros resultados, sino también cómo se mantiene esta precisión a medida que avanzamos en el ranking de resultados. Si bien el sistema A exhibe una alta precisión inicial, su rendimiento disminuye significativamente a medida que aumenta el intervalo de recall. Esto puede ser problemático en escenarios donde la relevancia de los resultados más profundos es crucial.

Por otro lado, aunque el sistema C no presenta la misma precisión que el sistema A en los primeros resultados, su rendimiento se mantiene estable a lo largo de todos los intervalos de recall. Esto sugiere que el sistema C puede ser más confiable y consistente en aplicaciones donde se valora la precisión en todo el rango de resultados.

El sistema B, aunque no tan impresionante como el A en los primeros resultados, muestra un rendimiento constante y competitivo en comparación con el sistema C.

Uso del Sistema A:

El sistema A puede ser útil para obtener resultados muy relevantes en escenarios de navegación web, debido a que se requiere que los primeros resultados sean lo más relevante posible a la necesidad de información.

Uso del Sistema C:

Por otro lado, el sistema C puede ser más confiable en situaciones de investigación, donde la precisión constante es crucial. Aunque puede no destacar tanto en los primeros resultados, mantiene una precisión estable a lo largo de todos los intervalos de búsqueda.