# 4. Corridas de ejemplo

Los resultados de las corridas de ejemplo están en el directorio del proyecto. Fueron identificadas tal y como se especificó en la guía de documentación (se lograron implementar todas las funcionalidades del proyecto para correr todas las consultas). Los resultados de las consultas se analizan en el siguiente apartado.

# 5. Análisis

A continuación se presenta el análisis de recall y precisión de las 3 consultas en modo vectorial y booleano.

## 5.1 Consulta 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Vectorial** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 3 | bzip2.1.txt | 0% 1/12 = 0.083 | 1/3 = 0.333 |
| 4 | bzcat.1 | 10% 2/12 = 0.166 | 2/4 = 0.5 |
| 5 | bunzip2.1.txt | 20% 3/12 = 0.25 | 3/5 = 0.6 |
| 9 | gunzip.1.txt | 30% 4/12 = 0.333 | 4/9 = 0.44 |
| 8 | zcat.1.txt | 40% 5/12 = 0.416 | 5/8 = 0.625 |
| 11 | gzip.1.txt | 50% 6/12 = 0.5 | 6/11 = 0.545 |
| 12 | znew.1.txt | 50% 7/12 = 0.583 | 7/12 = 0.583 |
| 13 | zgrep.1.txt | 60% 8/12 = 0.666 | 8/13 = 0.615 |
| 14 | zcmp.1.txt | 70% 9/12 = 0.75 | 9/14 = 0.642 |
| 15 | zdiff.1.txt | 80% 10/12 = 0.833 | 10/15 = 0.666 |
| 22 | zmore.1.txt | 90% 11/12 = 0.916 | 11/22 = 0.5 |
| 187 | gzexe.1.txt | 100% 12/12 = 1 | 12/187 = 0.06 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Booleano** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 1 | bunzip2.1.txt | 0% 1/12 = 0.083 | 1/3 = 0.33 |
| 5 | bzip2.1.txt | 10% 2/12 = 0.166 | 2/5 = 0.4 |
| 6 | bzcat.1.txt | 20% 3/12 = 0.25 | 3/6 = 0.5 |
| 8 | gunzip.1.txt | 30% 4/12 = 0.333 | 4/8 = 0.5 |
| 9 | zcat.1.txt | 40% 5/12 = 0.416 | 5/9 = 0.56 |
| 10 | gzip.1.txt | 50% 6/12 = 0.5 | 6/10 = 0.6 |
| 11 | zmore.1.txt | 50% 7/12 = 0.583 | 7/11 = 0.63 |
| 12 | zgrep.1.txt | 60% 8/12 = 0.666 | 8/12 = 0.66 |
| 15 | zcmp.1.txt | 70% 9/12 = 0.75 | 9/15 = 0.6 |
| 17 | zdiff.1.txt | 80% 10/12 = 0.833 | 10/17 = 0.58 |
| 41 | znew.1.txt | 90% 11/12 = 0.916 | 11/41 = 0.27 |

## 5.2 Consulta 2

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Vectorial** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 1 | ipc.5.txt | 0% 1/15 = 0.06 | 1/1 = 1 |
| 2 | socket.2.txt | 10% 2/15 = 0.133 | 2/2 = 1 |
| 3 | msgget.2.txt | 20% 3/15 = 0.2 | 3/3 = 1 |
| 4 | ipcs.8.txt | 20% 4/15 = 0.266 | 4/4 = 1 |
| 10 | msgctl.2.txt | 30% 5/15 = 0.333 | 5/10 = 0.5 |
| 26 | socket.7.txt | 40% 6/15 = 0.4 | 6/26 = 0.23 |
| 42 | msgrcv.2.txt | 40% 7/15 = 0.466 | 7/42 = 0.166 |
| 43 | msgop.2.txt | 50% 8/15 = 0.533 | 8/43 = 0.186 |
| 44 | msgsnd.2.txt | 60% 9/15 = 0.6 | 9/44 = 0.204 |
| 72 | recv.2.txt | 60% 10/15 = 0.666 | 10/72 = 0.138 |
| 73 | recvmsg.2.txt | 70% 11/15 =0.733 | 11/73 = 0.15 |
| 74 | recvfrom.2.txt | 80% 12/15 = 0.8 | 12/74 = 0.162 |
| 203 | tcp.7.txt | 80% 13/15 = 0.866 | 13/203 = 0.064 |
| 475 | sendto.2.txt | 90% 14/15 = 0.933 | 14/475 = 0.029 |
| 572 | sendmsg.2.txt | 100% 15/15 = 1 | 15/572 = 0.026 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Booleano** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 1 | msgget.2.txt | 0% 1/15 = 0.06 | 1/1 = 1 |
| 2 | ipc.5.txt | 10% 2/15 = 0.133 | ½ = 0.5 |
| 4 | socket.2.txt | 20% 4/15 = 0.266 | ¼ = 0.25 |
| 5 | socket.7.txt | 30% 5/15 = 0.333 | 1/5 = 0.2 |

## 5.3 Consulta 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Vectorial** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 1 | shmat.2.txt | 0% - 10% 1/7 = 0.142 | 1/1 = 1 |
| 2 | shmop.2.txt | 20% 2/7 = 0.285 | 2/2 = 1 |
| 3 | shmdt.2.txt | 30% - 40% 3/7 = 0.428 | 3/3 = 1 |
| 4 | shmget.2.txt | 50% 4/7 =0.571 | 4/4 = 1 |
| 5 | shmctl.2.txt | 60% - 70% 5/7 = 0.714 | 5/5 = 1 |
| 6 | ipcs.8.txt | 80% 6/7 = 0.857 | 6/6 = 1 |
| 7 | ipc.5.txt | 90% 7/7 = 1 | 7/7 = 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Modo Booleano** | | | |
| **Posición en escalafón** | **Documento relevante recuperado** | **Recall** | **Precisión** |
| 1 | shmdt.2.txt | 0% - 10% 1/7 = 0.142 | 1/1 = 1 |
| 5 | shmget.2.txt | 20% 2/7 = 0.285 | 1/5 = 0.2 |
| 6 | shmat.2.txt | 30% - 40% 3/7 = 0.428 | 1/6 = 0.16 |
| 7 | shmop.2.txt | 50% 4/7 =0.571 | 1/7 = 0.14 |
| 9 | shmctl.2.txt | 60% - 70% 5/7 = 0.714 | 1/9 = 0.11 |
| 11 | ipc.5.txt | 80% 6/7 = 0.857 | 1/11 = 0.09 |
| 21 | ipcs.8.txt | 90% 7/7 = 1 | 1/21 = 0.05 |

## 5.4 Resultados

En el archivo de Excel con las evaluaciones, se puede observar que el comportamiento de la consulta vectorial es irregular. En algunos casos crece, pero cuando decrece, lo hace de manera considerable.

La consulta booleana muestra un comportamiento muy irregular, a pesar de que mantiene un nivel de precisión entre el 20% y en 80%, sin embargo es muy bajo. En la consulta número 2 no se obtuvieron todos los archivos relevantes y por eso se da una caída muy rápida de la búsqueda booleana en el gráfico.

# 6. Comentarios finales

El proyecto logró cubrir todas las especificaciones de la primera tarea programada. Todas las funciones fueron implementadas. Sin embargo, para lograr la implementación de todas las tareas, se encontraron los siguientes problemas:

* En la línea de comandos no se logró ingresar algunos parámetros de la manera deseada. Por ejemplo, al ingresar el directorio que contiene la colección, este se debe ingresar como si fuera un path o ruta: D/man.es, así como el patrón de los documentos que se van a analizar. Estos se deben ingresar así: \.txt$ sin las comillas simples.
* Inicialmente, se implementaron las primeras versiones del proyecto en Perl para Ubuntu Linux. No obstante, hubo algunos problemas al reconocer las palabras separadas por una nueva línea y el carácter “-”. El problema era por la diferente codificación entre los archivos de la colección y los archivos de perl (.pl). Al final se optó por usar Perl para Windows y afortunadamente no se presentó de nuevo este inconveniente.

El programa no tiene limitaciones considerables, se puede considerar como limitaciones las soluciones a los problemas encontrados.