


**Certifico que este trabajo es INDIVIDUAL y que lo he completado respetando las normas de ética académica de la materia. Entiendo que si violo estas normas que seré reportado a las autoridades de la Facultad y estaré sujeto a sanciones que incluyen hasta expulsión de la Facultad.**

**Firma:**  **Aclaración:** Agustin Arguello **Fecha:** 11/05/2025

# Algoritmos y Estructura de Datos 2

## Proyecto 3 - Buscador de Palabras

**Proyecto:** <https://github.com/agusquartz/AYED2-P3-TextSearchEngine/tree/main>

**Autor:** Agustin Argüello Bernal

### Como se hizo

Carga archivos de texto, analiza las palabras y almacena sus posiciones en una tabla hash, donde cada palabra clave apunta a una lista de ocurrencias que indica en qué documentos y posiciones aparece.

### Compilación en Debian Linux:

```
gcc -g -o EJECUTABLE InvertedIndex.c HashTable.c ArrayList/arraylist.c  
Occurrence/occurrence.c FileManager.c Main.c -Ipath/to/headers
```

### Uso:

./EJECUTABLE <archivo>

```
172.31.163.149 [windows-lenovo] agusquartz Proyecto3 gcc -g -o MAIN InvertedIndex.c HashTable.c ArrayList/arraylist.c  
t.c Occurrence/occurrence.c FileManager.c Main.c -Ipath/to/headers  
172.31.163.149 [windows-lenovo] agusquartz Proyecto3 ./MAIN "DonQuijote.txt"  
Cargando archivo id=0...  
===== Word/s Search Engine =====  
exit() para salir  
Ingrese la/s palabra/s a buscar (separadas por comas): escribano,Consejo  
Buscando la/s palabra/s: escribano,Consejo  
Resultados encontrados (1):  
Documento 0: líneas 42 a 43  
--- Contenido aproximado: ---  
Yo, Juan Gallo de Andrada, escribano de Cámara del Rey nuestro señor, de  
los que residen en su Consejo, certifico y doy fe que, habiendo visto por  
-----  
----- Word/s Search Engine -----  
exit() para salir  
Ingrese la/s palabra/s a buscar (separadas por comas): exit()  
Buscando la/s palabra/s: exit()  
Saliendo...  
172.31.163.149 [windows-lenovo] agusquartz Proyecto3
```

## Logros

El programa puede leer y almacenar correctamente el contenido de un file y sus posiciones  
Se pueden realizar búsquedas de dos palabras

```
Clave en índice 32960: randas
Clave en índice 32961: disuene
Clave en índice 32962: have
Clave en índice 32964: mostraba
Clave en índice 32968: falsos
Clave en índice 32969: tuviste
Clave en índice 32971: suya
Clave en índice 32972: reliquia
Clave en índice 32973: reverend
Clave en índice 32974: competencias
Clave en índice 32975: ofrecen
Clave en índice 32976: adalid
Clave en índice 32977: altaneros
Clave en índice 32979: viuda
Clave en índice 32980: ofrecer
Clave en índice 32981: perseguir
Clave en índice 32982: remediaron
Clave en índice 32983: perdonad
Clave en índice 32984: ofreces
Clave en índice 32985: suyo
Clave en índice 32986: invisibles
Clave en índice 32988: orbaneja
Clave en índice 32989: sentimiento
Clave en índice 32991: ahorrarse
Clave en índice 32992: grabando
Clave en índice 32993: viudo
Clave en índice 32994: cohetes
Clave en índice 32995: antifaz
Clave en índice 32996: macho
Clave en índice 32997: perdonar
Clave en índice 32998: aire
Clave en índice 32999: tibulo
Clave en índice 33000: xlvii
Clave en índice 33002: monstruos
Clave en índice 33003: allegados
Clave en índice 33004: amasa
Clave en índice 33008: nube
Clave en índice 33009: inaccurate
Clave en índice 33010: paveses
Clave en índice 33014: rceles
Clave en índice 33015: cortesana
Clave en índice 33017: ndolas
Clave en índice 33018: ficiesen
Clave en índice 33019: perpe
Clave en índice 33020: arrugado
Clave en índice 33021: perdona
```

## Dificultades y “No” Logros

No estoy seguro si puede realizar búsquedas de una sola palabra o más de dos palabras.

Actualmente no puede leer varios files a la vez, solo uno, algun error en el manejo de la inserción en el file InvertedIndex puede estar causando eso, me quedé sin tiempo.

Puede que tenga bugs si usas caracteres latinos como los que llevan acentos, me asegure de que la inserción y búsqueda haga un “toLowerCase” solamente, para manejar todo en minúsculas.