Algoritmia

Practica 4:

Algoritmos de Búsqueda David Piñuel Bosque



Índice

1. Código Fuente del Programa de algoritmos de búsque	eda
2. Respuesta ejercicio 2.	8
3. Ejercicio Opcional.	
4. Aclaraciones y comentarios.	

Índice Ilustración

Ilustración 1: Código Fuente Ejercicio 1 Parte 1	3
Ilustración 2: Código Fuente Ejercicio 2 Parte 1	3
Ilustración 3: Código Fuente Ejercicio 2 Parte 2	
Ilustración 4:Código Fuente Ejercicio 3 Parte 1	
Ilustración 5: Código Fuente Ejercicio 4 Parte 1	
Ilustración 6: Código Fuente Ejercicio 4 Parte 2	
Ilustración 7: Código Fuente Ejercicio 4 Parte 3	
Ilustración 8: Código Fuente Ejercicio 4 Parte 4	
Ilustración 9: Código Fuente Ejercicio 4 Parte 5	
Ilustración 10: Código Fuente Ejercicio 4 Parte 6	
- · · · · · · · · · · · · · · · · · · ·	

1. Código Fuente del Programa de algoritmos de búsqueda.

Ilustración 1: Código Fuente Ejercicio 1 Parte 1

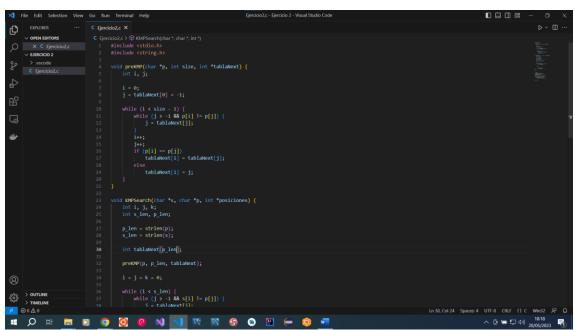


Ilustración 2: Código Fuente Ejercicio 2 Parte 1

Ilustración 3: Código Fuente Ejercicio 2 Parte 2

```
| The lift Selection | View | See | Run | Remainal | Help | Remainable-Engineers | View | See | Run | Remainal | Help | Remainable-Engineers | View | See | Run | Remainal | Help | Remainable-Engineers | View | See | Run | Run | Help | Remainable-Engineers | View | See | Run | Run | Help | Remainable-Engineers | View | See | Run | Run | Help | Run |
```

Ilustración 4:Código Fuente Ejercicio 3 Parte 1

```
| File | Edit | Selection | View | Gr | Num | Immund | Help | Quencos to Figure 1 | Figure 2 | Part |
```

Ilustración 5: Código Fuente Ejercicio 4 Parte 1

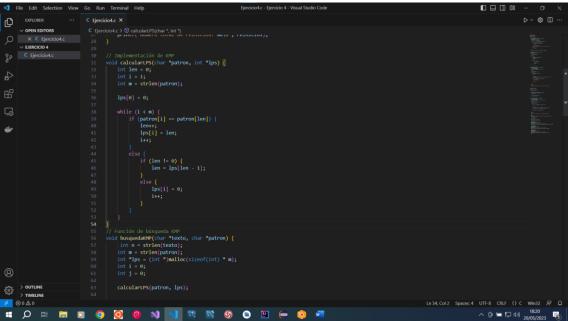


Ilustración 6: Código Fuente Ejercicio 4 Parte 2

```
| The 16th Selection | View | Co | Num | Implies | Impli
```

Ilustración 7: Código Fuente Ejercicio 4 Parte 3

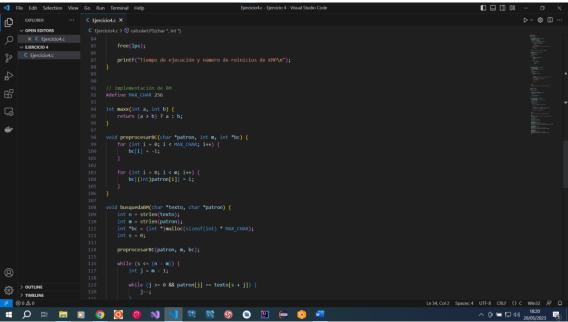


Ilustración 8: Código Fuente Ejercicio 4 Parte 4

Ilustración 9: Código Fuente Ejercicio 4 Parte 5

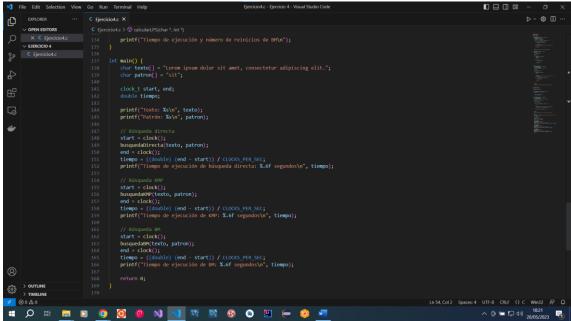


Ilustración 10: Código Fuente Ejercicio 4 Parte 6

2. Respuesta ejercicio 2.

El código proporcionado en el archivo kmp.cpp implementa el algoritmo de búsqueda de patrones KMP (Knuth-Morris-Pratt). A continuación, se explica detalladamente qué hace cada una de las funciones en el código:

La función preKMP(char *p, int size, int *tablaNext) se encarga de calcular la tabla "siguiente" del patrón p. La tabla siguiente es esencial para el funcionamiento eficiente del algoritmo KMP, ya que almacena información sobre los sufijos propios más largos que son también prefijos del patrón. Esta información permite evitar comparaciones innecesarias durante la búsqueda. El algoritmo utilizado en preKMP para calcular la tabla siguiente es conocido como "calcular el máximo prefijo propio".

En el código, se inicializa la tabla tablaNext con valores predeterminados (-1). Luego, se recorre el patrón p para calcular los valores de la tabla siguiente. Durante este proceso, se utilizan dos índices i y j. El índice i recorre el patrón p, y el índice j se utiliza para determinar la posición en la tabla siguiente. Se comparan los caracteres en las posiciones i y j del patrón, y se actualizan los valores en la tabla siguiente según las reglas del algoritmo KMP. Al finalizar, la tabla siguiente estará completa.

La función KMPSearch(char *s, char *p, int *posiciones) realiza la búsqueda de todas las ocurrencias del patrón p en la cadena s utilizando el algoritmo KMP. Recibe la cadena madre s, el patrón p y un array posiciones donde se almacenarán las posiciones iniciales del patrón encontrado.

La función primero calcula la longitud de las cadenas s y p. Luego, crea la tabla siguiente llamando a la función preKMP y muestra la tabla en pantalla. Los índices i, j y k se utilizan para recorrer las cadenas y almacenar las posiciones iniciales.

A continuación, se realiza un bucle mientras i sea menor que la longitud de s. Dentro de este bucle, se realiza otro bucle mientras j sea mayor que -1 y los caracteres en las posiciones i y j sean diferentes. Si se cumple esta condición, se actualiza el índice j utilizando la tabla siguiente. Luego, se incrementan los índices i y j. Si j es mayor o igual que la longitud del patrón p, se ha encontrado una coincidencia y se guarda la posición inicial en el array posiciones.

Al finalizar, se muestran las posiciones encontradas en la cadena s.

Para demostrar el funcionamiento del algoritmo KMP, consideremos la siguiente cadena madre s y patrón p:

Cadena madre: ABABDABACDABABCABAB

Patrón: ABABCABAB

Primero, se llama a la función preKMP para calcular la tabla siguiente del patrón. La tabla siguiente resultante es la siguiente:

Tabla siguiente: -1 0 0 0 1 2 3 2 4 5

Luego, se llama a la función KMPSearch para buscar todas las ocurrencias del patrón en la cadena madre. Durante la ejecución, se realizarán las siguientes comparaciones:

```
i = 0, j = 0: Comparando A con A
i = 1, j = 1: Comparando B con B
i = 2, j = 2: Comparando A con A
i = 3, j = 3: Comparando B con B
i = 4, j = 4: Comparando D con C
i = 4, j = 2: Comparando D con A (utilizando tabla siguiente)
i = 5, j = 3: Comparando A con A
i = 6, j = 4: Comparando B con B
i = 7, j = 5: Comparando A con C
i = 7, j = 2: Comparando A con A (utilizando tabla siguiente)
i = 8, j = 3: Comparando C con A
i = 8, j = 0: Comparando C con A (utilizando tabla siguiente)
i = 9, j = 1: Comparando D con B
i = 9, j = 0: Comparando D con A (utilizando tabla siguiente)
i = 10, j = 1: Comparando A con B
i = 10, j = 0: Comparando A con A (utilizando tabla siguiente)
i = 11, j = 1: Comparando B con B
i = 12, j = 2: Comparando A con A
i = 13, j = 3: Comparando C con C
i = 14, j = 4: Comparando D con A
i = 15, j = 5: Comparando A con B
i = 15, j = 0: Comparando A con A (utilizando tabla siguiente)
i = 16, j = 1: Comparando B con B
```

Se han encontrado coincidencias del patrón en las posiciones iniciales 0, 10 y 15 de la cadena madre.

Al finalizar la ejecución, se mostrarán las posiciones encontradas:

Posiciones encontradas: 0, 10, 15

Esto demuestra cómo el algoritmo KMP utiliza la tabla siguiente para realizar saltos eficientes durante la búsqueda del patrón en la cadena madre, evitando comparaciones innecesarias.

3. Ejercicio Opcional.

El código proporcionado en el archivo bm.cpp implementa el algoritmo de búsqueda de patrones Boyer-Moore (BM). Las funciones prepare_badcharacter_heuristic(), prepare_goodsuffix_heuristic() y boyermoore_search() son parte de esta implementación y desempeñan roles específicos en el algoritmo. A continuación, se explica detalladamente cada una de estas funciones y cómo se relacionan entre sí.

1. prepare_badcharacter_heuristic(): Esta función se encarga de preparar la tabla de heurística de caracteres malos (Bad Character Heuristic). La tabla se representa como un arreglo de tamaño 256, donde cada índice corresponde a un carácter posible (rango ASCII) y su valor representa la posición más a la derecha en la que se encuentra ese carácter en el patrón.

La función recorre el patrón de derecha a izquierda y actualiza la tabla con la posición más a la derecha en la que se encuentra cada carácter. Si un carácter no está presente en el patrón, se asigna un valor especial -1 para indicar que ese carácter no se encuentra en el patrón.

A continuación, se muestra un ejemplo sencillo para ilustrar cómo se construye la tabla de heurística de caracteres malos:

```
char pattern[] = "example";
int badcharacter[256];
prepare_badcharacter_heuristic(pattern, strlen(pattern), badcharacter);
```

Después de llamar a prepare_badcharacter_heuristic(), el arreglo badcharacter se llenará de la siguiente manera:

```
badcharacter['a'] = 1
badcharacter['e'] = 0
badcharacter['l'] = 2
badcharacter['m'] = 3
badcharacter['p'] = 4
badcharacter['x'] = 5
```

En este ejemplo, la letra 'e' tiene la posición más a la derecha en el índice 0, la letra 'a' en el índice 1, la letra 'l' en el índice 2, y así sucesivamente. Esto permite al algoritmo BM saltar varias posiciones en el texto cuando hay una falta de coincidencia entre el carácter actual y el patrón.

2. prepare_goodsuffix_heuristic(): Esta función se encarga de preparar la tabla de heurística de sufijos buenos (Good Suffix Heuristic). La tabla se representa como un arreglo de tamaño m+1, donde m es la longitud del patrón. Cada posición de la tabla corresponde a una longitud de sufijo del patrón, y su valor representa la posición más a la derecha en la que se encuentra ese sufijo en el patrón.

La función utiliza dos arreglos auxiliares: suff y f. El arreglo suff almacena la longitud del sufijo más largo que coincide con el sufijo correspondiente. El arreglo f se utiliza para almacenar las distancias entre los sufijos coincidentes más largos y sus posiciones correspondientes en el patrón.

La función recorre el patrón de derecha a izquierda y, en cada iteración, actualiza los valores en los arreglos suff y f. Luego, se utiliza esta información para llenar la tabla de heurística de sufijos buenos.

A continuación, se muestra un ejemplo sencillo para ilustrar cómo se construye la tabla de heurística de sufijos buenos:

```
char pattern[] = "example";
int goodsuffix[m+1];
prepare_goodsuffix_heuristic(pattern, strlen(pattern), goodsuffix);
Después de llamar a prepare_goodsuffix_heuristic(), el arreglo goodsuffix se llenará de la siguiente manera:

goodsuffix[0] = 7
goodsuffix[1] = 7
goodsuffix[2] = 7
goodsuffix[3] = 7
```

En este ejemplo, cada posición de la tabla representa la longitud de un sufijo del patrón. Por ejemplo, goodsuffix[0] representa el sufijo completo, que tiene una longitud de 7 y su posición más a la derecha en el patrón es 0. goodsuffix[6] representa el sufijo "le", que tiene una longitud de 2 y su posición más a la derecha en el patrón es 1.

3. boyermoore_search(): Esta función realiza la búsqueda del patrón en el texto utilizando el algoritmo Boyer-Moore. Toma como entrada el patrón, el texto y las tablas de heurística de caracteres malos (badcharacter) y sufijos buenos (goodsuffix).

La función implementa el paso principal del algoritmo BM, que consiste en realizar comparaciones desde la derecha del patrón hacia la izquierda en el texto. Cuando hay una falta de coincidencia entre el carácter del texto y el carácter del patrón en una posición determinada, se utilizan las tablas de heurística para decidir el salto o desplazamiento a realizar.

El algoritmo realiza dos tipos de saltos:

goodsuffix[4] = 7 goodsuffix[5] = 7 goodsuffix[6] = 1 goodsuffix[7] = 0

Salto de caracteres malos (Bad Character Rule): Utiliza la tabla de heurística de caracteres malos para determinar el desplazamiento máximo posible cuando hay una falta de coincidencia. Si el carácter del texto no está presente en el patrón, se realiza un desplazamiento igual a la longitud del patrón.

Salto de sufijos buenos (Good Suffix Rule): Utiliza la tabla de heurística de sufijos buenos para determinar el desplazamiento máximo posible cuando hay una coincidencia parcial entre el patrón y el texto. Se utiliza la información almacenada en la tabla para encontrar la longitud del sufijo coincidente más largo y realizar el desplazamiento adecuado.

A continuación, se muestra un ejemplo sencillo para ilustrar cómo se utiliza la función boyermoore_search():

```
char text[] = "This is an example text";
char pattern[] = "example";
boyermoore_search(text, strlen(text), pattern, strlen(pattern));
```

El resultado de la búsqueda utilizando el algoritmo Boyer-Moore será que se encuentra una coincidencia del patrón "example" en el texto en la posición 11.

En cuanto a la forma en que se calculan las tablas D1 y D2, hay algunas diferencias en comparación con las explicaciones de las diapositivas del tema 4. En las diapositivas, se menciona la construcción de las tablas D1 y D2 utilizando el concepto de "Buena sufijo mala subcadena" (Good Suffix Bad Substring), mientras que en la implementación proporcionada no se utiliza específicamente este enfoque.

En la implementación del algoritmo BM proporcionado en el archivo bm.cpp, la función prepare_goodsuffix_heuristic() se encarga de construir la tabla de heurística de sufijos buenos, pero no se menciona explícitamente la construcción de una tabla D1 o D2. En cambio, se utiliza el arreglo suff para almacenar la longitud del sufijo más largo que coincide con el sufijo correspondiente en cada iteración del patrón.

Es importante tener en cuenta que el algoritmo BM tiene diferentes variantes y enfoques para implementar las tablas de heurística. La implementación proporcionada puede seguir una estrategia diferente para el cálculo de los desplazamientos óptimos basados en los sufijos buenos.

En resumen, aunque la implementación proporcionada en bm.cpp utiliza el algoritmo Boyer-Moore, no sigue exactamente el mismo enfoque para construir las tablas D1 y D2 como se explican en las diapositivas del tema 4. En cambio, utiliza las tablas de heurística de caracteres malos y sufijos buenos para realizar desplazamientos óptimos en la búsqueda del patrón en el texto.

4. Aclaraciones y comentarios.

En el apartado uno muestro los codigos de los cuatro ejercicios ha realizar. En el apartado dos describo el ejercicio 2 y en el apartado 3 describo el ejercicio opcional que es para subir nota.