

Universidad Tecnológica Nacional - Facultad Regional Buenos Aires Sintaxis y Semántica de los Lenguajes

2022

Trabajo Práctico Grupal N° 2

Tema: AUTOMATAS

Curso: K2006

Profesora: Roxana Leituz

Fecha Estipulada de Entrega: 18 de Septiembre

Grupo N° 5

Alumno	Legajo
Beretta Chiara Sofía	204.619-2
Gribnicow Irina Pérez	178.165-0
Iglesias Zoe	203.734-8
Pangaro Lucas	164.142-6
Romano Mateo Agustín	204.018-9
Santucho Gianlucca	204.050-5

	Fecha	Nota	Observaciones	Fecha Devol.	Firma Docente
Entrega	18/09				
Corrección					

Ejercicio 1: Dada una cadena que contenga varios números que pueden ser decimales, octales o hexadecimales, con o sin signo para el caso de los decimales, separados por el carácter '&', reconocer los tres grupos de constantes enteras, indicando si hubo un error léxico, en caso de ser correcto contar la cantidad de cada grupo. Debe diagramar y entregar el o los autómatas utilizados y las matrices de transición. La cadena debe ingresar por línea de comando o por archivo.

Para desarrollar este ejercicio antes tenemos que saber qué forma tienen los datos mencionados.

Las constantes hexadecimales comienzan con cero seguidas de una equis (mayúsculo o minúscula). Posteriormente puede continuar con un numero (del cero al nueve) o una letra (de la 'a' a la 'f') tanto mayúscula como minúscula.

Algunos ejemplos válidos para este tipo de constante son: 0X123, 0x123, 0xAb, 0X0, 0Xa9f.

Las constantes octales tienen la particularidad que comienzan siempre con 0 y luego tienen un número del cero al siete.

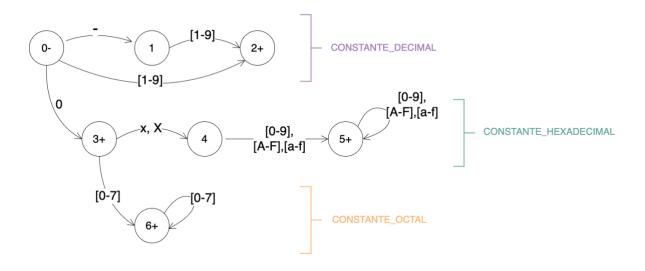
Algunos ejemplos son: 012, 0034, 076.

Como mención especial consideramos que el 0 pertenece a las constantes octales, aunque también podría incluirse dentro de las decimales. El único motivo de esto es simplificar el código, cuando el usuario ingresa el 0 podemos decir que el numero será un octal y quedaron en ese estado preguntando por los siguientes, salvo que continue con 'x', en ese caso será un hexadecimal (siempre y cuando el resto de caracteres también cumplan la condición).

Finalmente, las constantes decimales son todos los numero que se pueden formar con los caracteres del cero al nueve, también incluye los negativos, es decir los que comienzan con '-'. Pero no reconoce los que tienen como primer carácter numérico un 0. Es decir 03, -03.

Algunos ejemplos de caracteres decimales validos son: 123, -45, 7, -9, 10.

En representación de todo lo anterior se realizó el siguiente autómata con su respectiva tabla.



	-	0	1	2	3	4	5	6	7	8	9	a, A	b, B	c, C	d, D	e, E	f, F	x, X
0-	1	3	2	2	2	2	2	2	2	2	2	7	7	7	7	7	7	7
1	7	2	2	2	2	2	2	2	2	2	2	7	7	7	7	7	7	7
2+	7	2	2	2	2	2	2	2	2	2	2	7	7	7	7	7	7	7
3+	7	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	4	4
4	7	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7
5+	7	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7
6+	7	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7

Algunos detalles sobre el código...

Se comenzó realizando la declaración de nombres de constantes con el tipo de dato enum, esto nos permite una mayor expresividad en el código. (Estos nombres se usarán más adelante en un switch).

La función main es muy sencilla y se utiliza para llamar al resto de funciones que harán el trabajo duro del programa; sin embargo, es en este momento que el usuario elige como ingresar la cadena deseada. Puede optar por un ingreso en la línea de comandos (realizado con un scanf) o bien mediante un archivo, para esta alternativa además se le pedirá el nombre del archivo y con este se llama a una función.

```
grintf("Escriba una opcion: ");
scanf ("%d", &menu);
system("clear");

switch (menu){
    case 1:
        printf("Ingrese la cadena de caracteres numericos (Hexadecimales, Ocatales o Decimales): ");
scanf ("%s", vector_palabras);
break;
case 2:
    printf("Ingrese el nombre del archivo (ej: entrada.txt): ");
scanf ("%s", archivo);
ingreso_por_archivo(archivo, vector_palabras);
break;
default:
    printf ("No ingreso una opcion valida \n");
}
```

Cualquiera sea la opción elegida se guarda la cadena en el vector_palabras que luego es enviado a la función separar y reconocer.

```
separar_y_reconocer(vector_palabras, f_salida); // PASO EL ARRAY CON LAS PALABRAS A RECONOCER printf("Archivo creado exitosamente. Para visiualizarlo escriba: ´cat salida.txt´\n");
```

Dentro de esta función se crean variables del tipo int que serán pasadas por referencia a una nueva función reconocer, de esta forma reconocer puede "devolver" la cantidad de constataste de cada tipo.

Pero antes de ello se tiene que obtener la sub-cadena de char que representa cada número. Como dice el enunciado cada una de estas sub-cadenas se encuentra separa por el carácter '&'; por lo tanto, con la utilización de la función strtok y pasándole como argumento el puntero al vector y el centinela (&) se obtiene cada cadena a analizar.

La función reconocer es la más compleja de todas, es la encargada de recorrer cada sub-cadena y decir a que tipo de constante pertenece.

Para esto comienza con un while (1) que se utiliza para que cicle indefinidamente.

Dentro de este ciclo tenemos un switch que utiliza como estados los nombres definidos al principio con ayuda de enum.

En cada uno de estos estados procura reconocer un carácter y dependiendo de su contenido cambia el estado.

Para comenzar tenemos el estado inicio en el que se lee el primer carácter de la sub-cadena. Si este carácter es un 0 es posible que el número leído sea un octal o un hexadecimal. Es por esto que el estado cambia a OCTAL_O_HEXA.

```
case INICIO:
    if(valor[i] == '0')
        estado = OCTAL_O_HEXA;
    else if(isdigit(valor[i]))
        estado = CONSTANTE_DECIMAL;
    else if (valor[i] == '-'){
        estado = DECIMAL_NEGATIVO;
    }
    else
        estado = DESCONOCIDO;
    i++;
    break;
```

Si el carácter que se reconoció no fue un '0'

esta la posibilidad que sea un decimal asique con la función isdigit se pregunta si dicho carácter es un número, si lo es cambia el estado a CONSTATNE DECIMAL.

(isdigit también reconoce al '0' como un número, por lo que si el carácter fuese '0' también se validaría esta condición. Sin embargo, como los programas en C corren línea por línea al cumplirse la condición anterior ya no entraría en esta).

Finalmente tenemos dos opciones más, si el carácter leído no es un número, puede ser que sea un menos '-', carácter valido para los decimales, por lo que se envía a un estado llamado DECIMAL_NEGATIVO. Y si no es ni un numero ni '-' el estado cambia a DESCONOCIDO.

Si el estado llega a desconocido se imprime que la cadena no fue reconocida y termina el programa

```
case DESCONOCIDO:
    fprintf(f_salida, "%-20s\tNO RECONOCIDA\n", valor);
    return;
```

Continuando en orden y recordando si el carácter era un '0' se enviaba al estado OCTAL_O_HEXA.

Es en este estado en que se pregunta por el carácter siguiente. Si es '\0' quiere decir que la cadena termino y el numero era 0 (octal), se guarda en un archivo y retorna rompiendo el ciclo while.

Si no es '\0' seguimos preguntando si es una equis mayúscula o minúscula, si lo es todo indica que el número es un hexadecimal por lo que el estado cambia a CONSTANTE_HEXADECIMAL. En cambio, si lo que se lee es un numero entre el 0 y el 7 es un octal (CONSTANTE_OCTAL).

```
case OCTAL_0_HEXA:
   if(valor[i] == '\0'){
        (*cant_oct)++;
        fprintf(f_salida, "%-20s\t %d OCTAL\n", valor, *cant_oct);
        return;
}
else if(valor[i] == 'x' || valor[i] == 'X')
        estado = CONSTANTE_HEXADECIMAL2;
else if(isdigit(valor[i]) && valor[i] - '0' <= 7)
        estado = CONSTANTE_OCTAL;
else
        estado = DESCONOCIDO;
i++;
break;</pre>
```

Otro estado especial es el DECIMAL_NEGATIVO a este entramos cunado el primer carácter de la sub-cadena era un '-', ahora preguntamos si el carácter que le sigue es un digito salvo el 0 (porque no tendría sentido tener -0) si es un digito, el estado cambia a CONSTANTE_DECIMAL

Para los siguientes estados es siempre lo mismo, CONSTANTE_OCTAL, CONSTANTE_DECIMAL y CONSTANTE_HEXADECIMAL, realizan tres preguntas. Si estoy en ese estado y **Ilega el carácter '\0**' quiere decir que la sub-cadena termino, por lo que se guarda en un archivo de salida el valor de la constante, el tipo y su cantidad, retornando y saliendo del while (1).

Si el carácter leído no es '\0' pregunto si se sigue cumpliendo la condición que sea octal, decimal o Hexa. Si la cumple, el estado sigue siendo el mismo y en la próxima vuelta del ciclo pregunto por el carácter siguiente.

Tercero y por último si no cumple con ninguna de las anteriores quiere decir que la cadena NO ES RECONCOIDA, el estado cambia a DESCONOCIDO.

```
case CONSTANTE_HEXADECIMAL2:
    if(valor[i] == '\0'){
        (*cant_hexa)++;
        fprintf(f_salida, "%-20s\t %d HEXADECIMAL\n", valor, *cant_hexa);
        return;
    }
    else if(isxdigit(valor[i]))
        estado = CONSTANTE_HEXADECIMAL2;
    else
        estado = DESCONOCIDO;
    i++;
    break;
```

```
case CONSTANTE_DECIMAL:
    if(valor[i] == '\0'){
        (*cant_deci)++;
        fprintf(f_salida, "%-20s\t %d DECIMAL\n", valor, *cant_deci);
        return;
}
else if(isdigit(valor[i]))
        estado = CONSTANTE_DECIMAL;
else
        estado = DESCONOCIDO;
i++;
break;
```

```
case CONSTANTE_OCTAL:
    if(valor[i] == '\0'){
        (*cant_oct)++;
        fprintf(f_salida, "%-20s\t %d OCTAL\n", valor, *cant_oct);
        return;
    }
    else if(isdigit(valor[i]) && valor[i] - '0' <= 7)
        estado = CONSTANTE_OCTAL;
    else
        estado = DESCONOCIDO;
    i++;
    break;</pre>
```

Por último, podemos mencionar un estado extra de los Hexadecimales, a este estado entra luego de haber leído 0x y se asegura que posteriormente de ese inicio de cadena continue con por lo menos otro valor. De esta manera no permite las cadenas "0x" o "0X" como hexadecimal.

Como se puede ver, para la resolución de este problema se tomó la decisión de cambiar el estado en cada uno de los caracteres y que se retorne inmediatamente el error, esto permite que si tengo una cadena que contine un error al principio no continue procesándola. Por ejemplo, la cadena "0x^1234ab4578f" deja de ser procesada en el tercer carácter ya que entra al estado DESCONOCIDO y allí se hace un return que rompe el while (1).

Ejemplo del programa corriendo con ingreso desde archivo

```
entrada.txt
 0xFF&127&0159&0xaBb1&0Xx&0&010&09&127ABACB&07&-10
                                     Ejercicio_1 — -zsh — 80×24
           Ingrese el nombre del archivo (ej: entrada.txt): entrada.txt
           Archivo creado exitosamente. Para visiualizarlo escriba: 'cat salida.txt'
           [lpangaro@MacBook-Air-de-Lucas Ejercicio_1 % cat salida.txt
                                  Cant
                                         Tipo
           0xFF
                                    1 HEXADECIMAL
           127
                                    1 DECIMAL
           0159
                                   NO RECONOCIDA
                                    2 HEXADECIMAL
           0xaBb1
                                   NO RECONOCIDA
           0Xx
           0
                                    1 OCTAL
           010
                                    2 OCTAL
           09
                                   NO RECONOCIDA
           127ABACB
                                   NO RECONOCIDA
           07
                                    3 OCTAL
           -10
                                    2 DECIMAL
ECIMAL_N lpangaro@MacBook-Air-de-Lucas Ejercicio_1 % ▮
(valor[]
    estac
se if(is
  estado
```

Ejercicio 2: Debe realizar una función que reciba un carácter numérico y retorne un número entero.

Las computadoras solo entienden números, por lo que cuando se escriben caracteres (char) estos tienen una representación numérica basada en ASCII.

Esto quiere decir que por ejemplo al ingresar el carácter 'a' la computadora lo que recibe es 97 y es en un segundo momento, cuando se realiza la salida del flujo de datos por pantalla, en que se le asigna la letra.

De igual manera sucede con los números. Una cosa es el **CARACTER** '1' y otra el **NUMERO** 1. El carácter '3' se almacena en una variable de tipo char, que al imprimirla en pantalla se ve como 3. Sin embargo, esta variable char no es más que un número, específicamente el carácter '3' tiene un valor de 51 en la tabla ASCII.

En este programa de prueba podemos ver como una misma variable de tipo char puede imprimir tanto su "valor interno" como la representación del mismo.

```
#include<stdio.h>

#include<stdio.h

#inclu
```

Utilizando el mismo ejemplo, se le asigna a caracter_num el carácter 3. Al hacer un printf con "%c" la terminal muestra el carácter '3' mientras que al imprimirlo con

```
Ipangaro@MacBook-Air-de-Lucas Desktop % ./prueba
Ingrese un CARACTER NUMERICO: 3
El caracter numerico '3'
Se representa con el numero 51
Ipangaro@MacBook-Air-de-Lucas Desktop %
```

"%d" podemos ver el valor que tiene dicho carácter internamente (51).

Guiándonos con la siguiente tabla sabemos que si el carácter '0' tiene un valor de 48 para transformarlo en un entero simplemente debemos restar 48 al valor del char.

Además de tener en cuenta esta propiedad de los char y su representación, para la resolución de este ejercicio tenemos que saber algo elemental de las matemáticas, esto es el **valor posicional.**

En un número, por ejemplo, el 234 sabemos que el 4 representa las unidades, el 3 las decenas y el 2 las centenas.

Si descomponemos este número obtendremos 2 x 100 + 3 x 10 + 4 x 1.

TABLA ASCII DE NUMEROS '0' = 48 '1' = 49 '2' = 50 '3' = 51 '4' = 52 '5' = 53 '6' = 54 '7' = 55 '8' = 56 '9' = 57 elcodigoascii.com.ar

Ahora sí, algunos detalles del código:

```
#define diferenciaASCII 48 ==
48

int my_atoi(char* string){

int largo, i, numero = 0;
int VP = 1;  // Valos posicional 1, 10, 100, 1000 ...

largo = strlen(string);

for (i = (largo-1) ; i >= 0 ; i--) { // uso largo-1 porque el indice del vector comienza en 0 numero += (string[i] - diferenciaASCII) * VP;
VP *= 10; // Se utiliza para calcular unidades, decenas, centenas ...
}

return numero;
}
```

Podemos destacar dos partes de este algoritmo, una es la resta al carácter numérico char para obtener el valor int. Se utilizo un #define para dotar de mayor expresividad al código.

Y la otra mención importante es sobre el valor posicional. Al crear el ciclo se comenzó de atrás hacia delante, de esta forma el primer número se multiplica por 1 (unidades), el segundo por 10 (decenas), el tercero por 100 (centenas). En cada vuelta la variable VP se multiplica por 10 y el numero calculado finalmente se retorna al main.

Ejercicio 3: Ingresar una cadena que represente una operación simple con enteros decimales y obtener su resultado, se debe operar con +,_ y *. Ejemplo = 3+4*7+3-5 = 29. Debe poder operar con cualquier número de operandos y operadores respetando la precedencia de los operadores aritméticos.

La cadena ingresada debe ser validada previamente preferentemente reutilizando las funciones del ejercicio 1.

Para poder realizar la operación los caracteres deben convertirse a números utilizando la función 2.

La cadena debe ingresar por línea de comando o por archivo.

Antes de comenzar con la resolución del ejercicio se explicará el razonamiento utilizado.

Se parte de un vector de caracteres que denominaremos cadena. La misma tiene el siguiente formato:

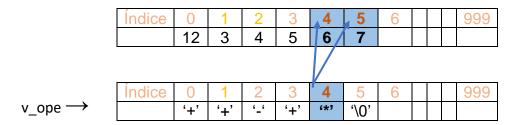
char cadena [1000];

Una función que se desarrolló se encarga de diferenciar los números de los operadores y separarlos en distintos vectores. Para ello utilizamos v_num y v_ope.

int v_num [1000] v_num
$$\rightarrow$$
 12 3 4 5 6 7 | char v_ope[1000] v_ope \rightarrow '+' '+' '-' '+' '*' '\0'

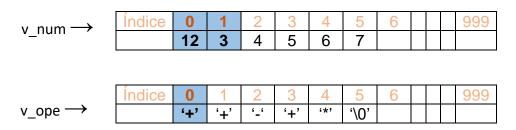
Ahora bien, si vemos en detalle podemos apreciar cierta relación entre los índices de ambos vectores. Por ejemplo, el operador '*' se encuentra en la posición 4 de v_ope. Lo que debería hacer es operar el 6 con el 7; que tienen el índice 4 y 5 respectivamente en v_num.

 $v_num \rightarrow 11$



Ahora el signo menos '-' opera con los números 4 y 5. El índice del '-' en v_ope es (2) y del 4 y 5 en v_num el (2) y (3). Es decir, cada operador se relaciona con dos números, y los índices de esos números son: el mismo que el del operador y el siguiente.

Más ejemplos:



Para sumar el 12 y el 3 los índices serian:

En el vector de operadores: 0 del '+' En el vector de números: 0 del 12

1 del 3

$v_num \longrightarrow$	Índice	0	1	2	3	4	5	6		999
		12	3	4	5	6	7			



Para sumar el 3 y el 4 los índices serian:

En el vector de operadores: 1 del '+'

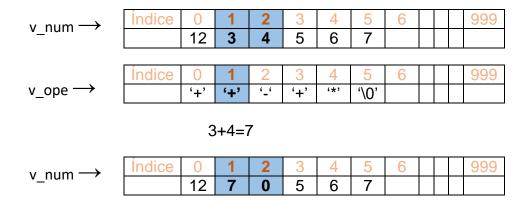
En el vector de números: 1 del 3

2 del 4

En todos los casos coincide el índice del operador con los incides de los números que debe operar (el mismo y el siguiente).

Ahora bien, cuando se realiza una operación podemos decir que los números se consumen. Dentro de nuestra función vamos realizando semi-resultados de la

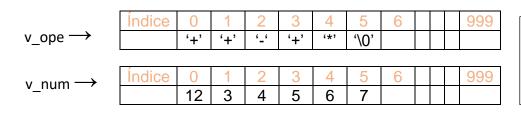
cadena ingresada. Cada uno de estos semi-resultados, o resultados parciales, se guarda dentro del vector de números. ¿En que posición? En la del índice de la operación, es decir, en el índice del primer digito que estoy operando; y para no repetir números ponemos un 0 en la posición del otro de los números. Por ejemplo



Una vez que se realizan todas las operaciones del vector v_ope nos queda v_num con todos los resultados parciales; Lo único que queda por realizar es la suma de todos ellos.

Pero antes de ver en detalle el código vamos a mostrar un poco más de este funcionamiento.

"12+3+4-5+6*7"



Vamos a realizar el producto. Para ello vemos el índice del carácter '*'. (4) y multiplicamos el número que se encuentre en la posición (4) con el de la siguiente (5). En este caso 6*7

	Índice	0	1	2	3	4	5	6		999
v_num \longrightarrow		12	3	4	5	6*7 = 42	0			

Se resuelve la operación (6*7). Su resultado se guarda en donde antes estaba el 6 (primero termino) y en el lugar del 7 (segundo término) se coloca un 0 para no volver a operar con dicho valor

Ahora realizando el proceso con

Índice	0	1	2	3	4	5	6		999
	12+3								
	=	0	4	5	42	0			
	15								

continuamos mismo las sumas...

	Índice	0	1	2	3	4	5	6		999
v_num \rightarrow		15	0+4 = 4	0	5	42	0			
	Índice	0	1	2	3	4	5	6		999
v_num \longrightarrow		15	4	0-5 = -5	0	42	0			
	Índice	0	1	2	3	4	5	6		999
v_num →		15	4	-5	0+42 = 42	_	0			
	Índice	0	1	2	3	4	5	6		999
$v_num \longrightarrow$	maice	15	4	-5	42	0	0	0		333

Ahora recorriendo este vector y sumando cada uno de sus elementos obtenemos el resultado de la operación escrita en la cadena!

Luego de esa explicación comencemos con algunas menciones especiales sobre el código.

Al igual que en el punto 1 se tiene un main bastante breve donde se da a elegir si se quiere ingresar la cadena por línea de comandos o por un archivo.

```
switch (menu){
case 1:
    printf("Ingrese la cadena de caracteres numericos (Hexadecimales, Ocatales o Decimales): ");
scanf ("%s", cadena);
break;

case 2:
    printf("Ingrese el nombre del archivo (ej: entrada.txt): ");
scanf ("%s", name_file);
ingreso_por_archivo(name_file, cadena);
break;

default:
    printf ("No ingreso una opcion valida \n");
return 1;
}
```

Posteriormente se valida la cadena ingresada, para ello se realizó una función muy simple, pero con un detalle especial.

```
int validar_vector(char* cadena) {

int i = 0;

while (cadena[i] != '\0') {
    if(!isdigit(cadena[i])) { // isdigit devuelve un valor no nulo si es un digito
    i++;
    }
    else if (isoperator(cadena[i]) && !isdigit(cadena[i - 1]) && !isdigit(cadena[i + 1])) {
        // SI el caracter leido es un operador tambien verifico que en la posicion siguiente y anterior haya un numero
    i++;
    }
    else
    return 1;
}

return 0;
```

Se recorre la cadena preguntando si el carácter leído es un digito o un operador. Pero si es un operador debe estar ante- y pre-cedido por un número. Esto hace que la cadena " *1+2 ", " 1+2* " o " 1++2 " no sean reconocidas.

La función isoperator también fue desarrollada y únicamente recibe un char, pregunta si es '+', '-', o '*' y devuelve 1 o 0.

Una vez que la cadena está validada se separan los números de los operadores en v_num y v_ope. El prototipo de la función que realiza esto es

No hay mucho que explicar sobre esta función más que se le pasa por referencia la variable largo para saber la cantidad de elementos que tiene el vector de números.

```
if (validar_vector(cadena)){

separar_numeros_de_operadores(cadena, v_ope, v_num, &largo);

resultado = calcular(v_ope, v_num, largo);

printf("El Resultado es: %d \n", resultado);
}
```

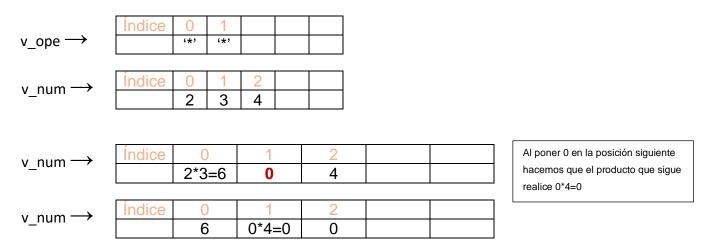
Una vez separados los números de los operadores se pasan ambos vectores a la función calcular. Esta devuelve el resultado de la cadena ingresada anteriormente.

Sebe comenzar preguntando por el operador "*" para respetar la precedencia de operadores.

Y el otro detalle importante es que se recorre de atrás hacia delante. ¿por qué?, Porque como vimos anteriormente cuando se resuelve una operación guardamos el valor del resultado en el índice del primer número y colocamos un cero en el índice siguiente. Si tenemos una multiplicación sucesiva y se recorre de izquierda a derecha el siguiente producto se realizaría con valor 0. Por ejemplo:

$$2*3*4 = 24$$

Si recorremos de izquierda a derecha



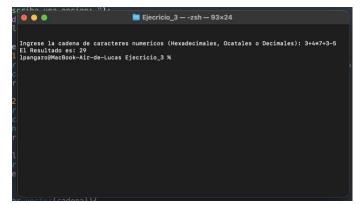
Mientras que si lo recorremos de derecha a izquierda:

	Índice	0	1		
$v_{ope} \rightarrow$		(*)	(*)		
v_num \longrightarrow	Índice	0	1	2	
		2	3	4	
	•				
$v_num \longrightarrow$	Índice	0	1	2	
v_nam ,		2	3*4=12	0	
$v_num \rightarrow$	Índice	0	1	2	
v_nam ,		2*12=24	0	0	

Por esta razón el bucle del "*" se realiza de atrás para delante.

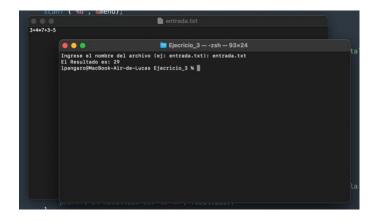
A continuación, se analizan los siguientes operadores ('+', '-'). Si es un más se hace la suma del número y el siguiente; y se "limpia" esta última posición. Si es un menos se realiza la resta ya también se "limpia" el valor siguiente.

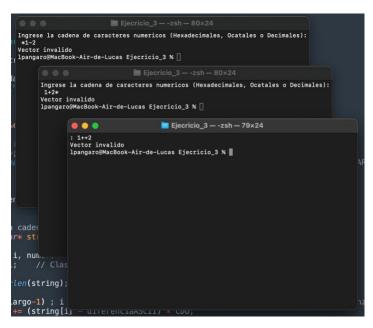
Finalmente, un for recorre el vector de enteros sumando los resultados y se retorna dicho valor.



Programa corriendo con ingreso por línea de comando.

Mismo programa pero corrido con un archivo .txt como entrada





Prueba de validación del vector:

1er caso: operador al comienzo de la cadena.

2do caso: Operador al final de la cadena.

3er caso: Operador repetido.