

Instituto Politécnico Nacional

ESCUELA SUPERIOR DE COMPUTACIÓN

BUSCADOR DE PALABRAS

Práctica 4

Alumno:

Alcántara Covarrubias Erik

Profesor:

Juarez Martinez Genaro

Grupo: 4CM6

1 Introducción

Un autómata finito no determinístico es aquel que puede estar en varios estados a la vez, al igual que un autómata finito determinístico tiene un conjunto finito de estados, un conjunto finito de símbolos de entrada, un estado inicial y un conjunto de estados finales. También tiene funciones de transición.

Un autómata finito no determinístico se representa esencialmente como un autómata finito determinístico:

$$A = \{Q, \Sigma, \delta, q_0, F\} \quad (1)$$

donde:

- Q es un conjunto finito de estados.
- Σ es un conjunto finito de símbolos de entrada.
- q_0 , un elemento de Q , es el estado inicial.
- F , un subconjunto de Q , es el conjunto de estados finales (o de aceptación).
- δ la función de transición, es una función que toma como argumentos un estado de Q y un símbolo de entrada de Σ y devuelve un subconjunto de Q . Observe que la única diferencia entre un AFN y un AFD se encuentra en el tipo de valor que devuelve δ : un conjunto de estados en el caso de un AFN y un único estado en el caso de un AFD.

2 Marco teórico

Primero denominaremos palabras clave a las palabras que buscaremos, en este caso es se procede a diseñar un autómata finito no determinista que va a indicar, mediante un estado final, que ha encontrado una de las palabras clave. El texto de un documento se introduce carácter a carácter en este AFN, el cual reconoce a continuación las apariciones de las palabras clave en dicho texto. Existe una forma simple para que un AFN reconozca un conjunto de palabras clave.

- Hay un estado inicial con una transición a sí mismo para cada uno de los símbolos de entrada, por ejemplo, todos los caracteres ASCII imprimibles si estamos examinando texto. Intuitivamente, el estado inicial representa una “conjetura” de que todavía no hemos detectado una de las palabras clave, incluso aunque hayamos encontrado algunas de las letras de una de esas palabras.
- Para cada palabra clave $a_1a_2\dots a_k$, existen k estados, por ejemplo, q_1, q_2, \dots, q_k . Existe una transición desde el estado inicial a q_1 para el símbolo a_1 , una transición desde q_1 a q_2 para el símbolo a_2 , etc. El estado q_k es un estado de aceptación e indica que se ha encontrado la palabra clave $a_1a_2\dots a_k$.

Ya teniendo el AFN, las reglas para poder construir el AFD desde el AFN son los siguientes:

- Si q_0 es el estado inicial del AFN, entonces $\{q_0\}$ es uno de los estados del AFD.
- Suponemos que p es uno de los estados del AFN y se llega a él desde el estado inicial siguiendo un camino cuyos símbolos son $a_1a_2\dots a_m$. Luego uno de los estados del AFD es el conjunto de estados del AFN constituidos por:
 - q_0 .
 - p .
 - Cualquier otro estado del AFN al que se pueda llegar desde q_0 siguiendo un camino cuyas etiquetas sean un sufijo de $a_1a_2\dots a_m$ es decir, cualquier secuencia de símbolos de la forma $a_ja_{j+1}\dots a_m$.

3 Explicación del problema e implementación del algoritmo

3.1 Problema a resolver

Programar el autómata finito determinístico que reconozca las palabras: web, webpage, website, webmaster, ebay, page, site

1. Diseñar el NFA.
2. Realizar la conversión a DFA mostrando todo el proceso a través de los subconjuntos y tablas.
3. El programa deberá de leer un archivo de texto, podría ser de una página web.
4. El autómata deberá de identificar cada palabra reservada con el DFA, contarlas e indicar dónde las encontró.
5. En un archivo imprimir la evaluación del autómata por cada carácter que lea y cambio de estado, es decir, toda la historia del proceso.
6. En otro archivo enumerar, contar y anotar donde están las palabras encontradas.
7. Tener una opción para ver el autómata, es decir, hay que graficarlo.

3.2 Implementación

```
import graphviz
import tempfile
from os.path import exists
import urllib.request
from inscriptis import get_text

def Grafo():
    """
    It creates a graphviz graph with the nodes and edges that correspond to the words in
    """
    Grafo = graphviz.Digraph('Diccionario')
    Grafo.attr(rankdir='LR')
    Grafo.attr('node', shape='circle')
    Grafo.node('Q0')
    Grafo.attr('node', shape='plaintext')
    Grafo.edge('', 'Q0', label='Inicio')

    Grafo.attr('node', shape='circle')
    Grafo.edge('Q0', 'Q1', label='w')
    Grafo.edge('Q1', 'Q2', label='e')
    Grafo.edge('Q2', 'Q3', label='b')
    Grafo.edge('Q3', 'Q4', label='p')
    Grafo.edge('Q3', 'Q5', label='s')
    Grafo.edge('Q3', 'Q6', label='m')
    Grafo.edge('Q4', 'Q7', label='a')
    Grafo.edge('Q7', 'Q8', label='g')
    Grafo.edge('Q5', 'Q10', label='i')
    Grafo.edge('Q10', 'Q11', label='t')
    Grafo.edge('Q6', 'Q13', label='a')
    Grafo.edge('Q13', 'Q14', label='s')
    Grafo.edge('Q14', 'Q15', label='t')
    Grafo.edge('Q15', 'Q16', label='e')
    Grafo.edge('Q0', 'Q18', label='e')
    Grafo.edge('Q18', 'Q19', label='b')
```

```

Grafo.edge('Q19', 'Q20', label='a')
Grafo.edge('Q0', 'Q22', label='p')
Grafo.edge('Q22', 'Q23', label='a')
Grafo.edge('Q23', 'Q24', label='g')
Grafo.edge('Q0', 'Q26', label='s')
Grafo.edge('Q26', 'Q27', label='i')
Grafo.edge('Q27', 'Q28', label='t')

Grafo.attr('node', shape='doublecircle')
Grafo.edge('Q8', 'Q9', label='e')
Grafo.edge('Q11', 'Q12', label='e')
Grafo.edge('Q16', 'Q17', label='r')
Grafo.edge('Q20', 'Q21', label='y')
Grafo.edge('Q24', 'Q25', label='e')
Grafo.edge('Q28', 'Q29', label='e')
Grafo.view(tempfile.mktemp())

def DIAGRAMA():
    """
    It creates a graph, and if it fails, it prints "Error"
    """
    try:
        Grafo()
    except Exception:
        print("Error")

def Escritura(lista):
    """
    It takes a list of lists of integers, and writes a file with the number of integers
    and the integers themselves.

    :param lista: a list of lists, where each list contains the positions of the words in
    """
    diccionario = ["web", "website", "webpage", "webmaster", "ebay", "site"]
    Pos = open("Concurrencia.txt", "w+")
    index = 0
    for i in diccionario:
        Pos.write("\n"+i+"'s encontrados: ")
        Pos.write(str(len(lista[index])))
        Pos.write("\n*En las posiciones:*")
        Pos.write(', '.join(map(str, lista[index])))
        index += 1
    Pos.close()

def DES(direccion, flag):
    """
    It takes a string and returns a list of lists of integers

    :param direccion: The address of the file or the URL
    :param flag: 1 for reading from a file, 2 for reading from a URL
    :return: a list of lists.
    """
    if (flag == 1):
        archivo = open(direccion, 'r')

```

```

    palabras = archivo.read()

elif (flag == 2):
    html = urllib.request.urlopen(direccion).read().decode('utf-8')
    palabras = get_text(html)

else:
    print("Error, contacte al programador")
    return None

web = [ ]
webpage = [ ]
website = [ ]
webmaster = [ ]
ebay = [ ]
page = [ ]
site = [ ]
listas = [ ]

contador = 0
estado = 1
his = open("Historia.txt", "w")

for letra in palabras:
    contador = contador + 1
    his.write("f( Q"+str(estado)+","+letra+"))-> ")

    if(estado == 1):
        if(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 19
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 2):
        if(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 3
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 3):
        if(letra == "w"):

```

```

        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "b"):
        web.append(contador-3)
        estado = 4
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    elif(letra == "b"):
        estado = 20
    else:
        estado = 1
    continue

if(estado == 4):
    if(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 5
    elif(letra == "s"):
        estado = 6
    elif(letra == "m"):
        estado = 7
    elif(letra == "a"):
        estado = 21
    else:
        estado = 1
    continue

if(estado == 5):
    if(letra == "a"):
        estado = 8
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 6):
    if(letra == "i"):
        estado = 11
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):

```

```

        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 7):
    if(letra == "a"):
        estado = 14
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 8):
    if(letra == "g"):
        estado = 9
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 9):
    if(letra == "e"):
        webpage.append(contador-7)
        estado = 10
    elif(letra == "w"):
        estado = 2
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 10):
    if(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19

```

```

        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        elif(letra == "b"):
            estado = 20
        else:
            estado = 1
        continue

    if(estado == 11):
        if(letra == "t"):
            estado = 12
        elif(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 19
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 12):
        if(letra == "e"):
            website.append(contador-7)
            estado = 10
        elif(letra == "w"):
            estado = 2
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 14):
        if(letra == "s"):
            estado = 15
        elif(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 19
        elif(letra == "p"):
            estado = 23
        else:
            estado = 1
        continue

    if(estado == 15):
        if(letra == "t"):
            estado = 16
        elif(letra == "w"):

```



```

        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 16):
    if(letra == "e"):
        estado = 17
    elif(letra == "w"):
        estado = 2
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 17):
    if(letra == "r"):
        webmaster.append(contador-9)
        estado = 10
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    elif(letra == "b"):
        estado = 20
    else:
        estado = 1
    continue

if(estado == 19):
    if(letra == "b"):
        estado = 20
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

```

```

if(estado == 20):
    if(letra == "a"):
        estado = 21
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 21):
    if(letra == "y"):
        ebay.append(contador-4)
        estado = 10
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 23):
    if(letra == "a"):
        estado = 24
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23
    elif(letra == "s"):
        estado = 27
    else:
        estado = 1
    continue

if(estado == 24):
    if(letra == "g"):
        estado = 25
    elif(letra == "w"):
        estado = 2
    elif(letra == "e"):
        estado = 19
    elif(letra == "p"):
        estado = 23

```

```

        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 25):
        if(letra == "e"):
            page.append(contador-4)
            estado = 10
        elif(letra == "w"):
            estado = 2
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 27):
        if(letra == "i"):
            estado = 28
        elif(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 19
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 28):
        if(letra == "t"):
            estado = 29
        elif(letra == "w"):
            estado = 2
        elif(letra == "e"):
            estado = 19
        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    if(estado == 29):
        if(letra == "e"):
            site.append(contador-4)
            estado = 10
        elif(letra == "w"):
            estado = 2

```

```

        elif(letra == "p"):
            estado = 23
        elif(letra == "s"):
            estado = 27
        else:
            estado = 1
        continue

    listas.append(web)
    listas.append(website)
    listas.append(webpage)
    listas.append(webmaster)
    listas.append(ebay)
    listas.append(site)

    Escritura(listas)
    his.close()
    if(flag == 1):
        archivo.close()

if __name__ == "__main__":

    while(True):
        print("****Diccionario****")
        print("1.- Cargar el archivo")
        print("2.- Grafo")
        print("3.- Salir")
        OP = int(input("Seleccione una opcion: "))
        if(OP == 1):
            direccion = str(input('Escribe el nombre del archivo o el link (incluyendo www o '))
            if(exists(direccion)):
                DES(direccion, 1)
            elif(direccion[:3] == "www" or direccion[:4] == "http"):
                print("Selecciono web, espere en lo que analizamos la pagina")
                DES(direccion, 2)
            else:
                print("El archivo especificado no existe")
        elif(OP == 2):
            DIAGRAMA()

        elif(OP == 3):
            print("Adios!")
            break

        else:
            print("Esa opcion no es valida , Adios!")
            break

```

4 Pruebas de escritorio y sus grafos

5 Conclusión

6 Bibliografía

- Introduction to Automata Theory, Languages, and Computation John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman Pearson/Addison Wesley, 2nd edition
- Software Foundation, P. (2022, 13 enero). 3.10.2 Documentation. Documentación de Python. Recuperado 15 de febrero de 2022, de <https://docs.python.org/es/3/>