

Tabla Informativa

Cuatrimestre	Agosto – Diciembre del 2022
Cuatrimestre y grupo	7B
Asignatura	Lenguajes y Autómatas
Corte	1
Actividad	LA.C1. A1 Caso de aplicación
Fecha de asignación	
Fecha de entrega	<i>Lunes, 3 de octubre del 2022</i>

Matricula	Nombre
191229	Maza Ricardo Ernesto

Contenido

Tabla de ilustración:	2
Introducción:	3
Modelado:	4
Desarrollo:	6
1. Método lectura de datos:	6
2. Método rescate de posibles correos del texto:	6
3. Método validaciones:	6
4. Autómata AFND:	7
5. método mostrar correos electrónicos:	7
Resultados:.....	8
Bibliografía:.....	11

Tabla de ilustración:

Ilustración 1 Autómata.....	5
Ilustración 2 ventana para leer la URL	8
Ilustración 3 vista con la url.....	8
Ilustración 4 mensaje de aviso	8
Ilustración 5 mensaje de aviso de protección	8
Ilustración 6 muestra de correos validos	9
Ilustración 7 más de un correo.....	9

Introducción:

En la historia o disciplina de la informática, se logran a saber tres tipos de autómatas que reconocen distintos tipos de lenguajes los cuales son:

- Los autómatas finitos.
- Los autómatas a pila.
- Las máquinas de Turing.

El autómata es la primera máquina con lenguaje, se conoce como un calculador lógico cuyo juego de instrucciones se orienta hacia los sistemas de evaluación secuencial.

El propósito de este documento es mostrar como se realizo un autómata con el fin de rescatar correos electrónicos de paginas web, los correos a rescatar son:

- @gmail.
- @hotmail.
- @outlook.

Mas adelante se definirá que dominós estaremos buscando para afirmar que se un correo valido.

Modelado:

Digamos que entramos una pagina web y queremos recatar lo que son los correos electrónicos, que sean: @Gmail, @hotmail, @outlook.

Pero también debemos saber que los correos se manejan con dominios después del (.) entonces tenemos que ver que tipo de dominio manejan, tanto como Gmail, Outlook y Hotmail, manejan diferentes, pero antes de esto se acompaña con un nombre de usuario, en la regla de estos correos un usuario puede estar compuesto por letras del abecedario, números, punto, guion bajo y guion ([a-z], [0-9] y [._-]) como ejemplo estos correos:

Gmail: dominios validos (.com, .mx)

Ejemplo: (usuario.08@gmail.com y usuario@gmail.mx)

Outlook: dominios validos (.com, .mx, .es, .net y .live)

Ejemplo: (usu_u-ari.o@outlook.com, usu02-ario@outlook.mx,
us89_uario@outlook.es,
us20-uario@outlook.net y usuario@outlook.live)

Hotmail: dominios validos (.com, .mx, .es, .net y .live)

Ejemplo: (usu-_ario@hotmail.com, usu-.ario@hotmail.mx,
usu02.ario@hotmail.es, u.56suario@hotmail.net y usu-02ario@hotmail.live).

Desarrollo:

1. Método lectura de datos:

En este método tenemos la parte donde mandamos la url de la pagina que queremos extraer el texto, para poder recuperar los correos electrónicos que existan en ella, lo primero es pasarlo por un filtro donde limpia el archivo quitando los scripts y el estilo del HTML, para luego extraer el puro texto, así después todo se escribe en un texto.

2. Método rescate de posibles correos del texto:

Lo que realizamos es leer todo el contenido del texto que extrajimos para luego rescatar los posibles correos electrónicos que puedan existir para luego mandarlo en un arreglo, ahí es donde almacenaremos los tipos de correos que existan en el texto, lo rescatamos validando que exista un @ en la parte del texto entonces puede ser el correo.

3. Método validaciones:

Usamos un try-except para poder validar si la pagina esta protegida, nos encontramos casos donde las paginas no nos deja acceder y extraer información desde el HTML y este evita que el programa se detenga por ese error.

Por otra parte, al igual usamos una validación donde vemos si no hay correos que se hayan encontrado validos no se muestra la tabla de contenido mas que solo un mensaje.

4. Autómata AFND:

Lo primero es mandar abrir nuestro archivo de texto que contiene la información del autómata que no se servirá para validar correos electrónicos, para esto se extrae la información, limpiamos todo lo que nos trae para luego pasarlo en su variables específicas, luego de esto pasamos a hacer un recorrido con un for, donde mandamos a validar los correos que rescatamos hace unos momentos, ahí es donde definimos todo, primero se pasa la cadena, para seguir y validar si lo que contiene la cadena esta dentro de nuestro alfabeto, en dado caso de estar, va seguir con el recorrido para esto es donde vamos validando cada estado he ir verificando cada punto del recorrido si existe en nuestra tabla de transición va ir actualizando los estados nuevos que vayan entrando, luego de pasar tales procesos también se agrega un elemento a la lista de los estados actuales para poder ir verificando valor por valor y si el recorrido dela cadena llega a un estado final valido, luego de ir haciendo los recorridos vamos limpiando los estados y cambiándolo al nuevo estado que vaya llegando para a si finalizar y ver si llega aun estado final valido, en caso de serlo se va a una validación donde dice que si es una cadena aceptada se vuelve verdadero y se guarda y en dado caso de no serlo simplemente se elimina. En caso de tener correos validos este se escribe en un archivo de texto (TXT).

5. método mostrar correos electrónicos:

para esto validamos si en el archivo de texto (TXT) encontramos que hay uno o mas de un correo se muestra una tabla con los correos encontrados, en dado caso que no se muestra un mensaje donde nos dice que no se encontraron correos validos que acepte nuestro autómata.

Resultados:

Primero ingresamos la URL de la página que queremos ver si encuentra correos válidos:

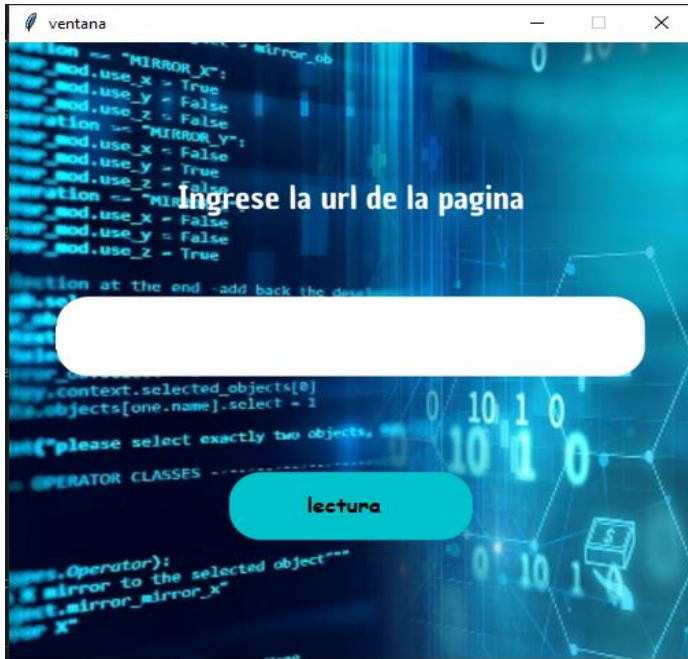


Ilustración 3 ventana para leer la URL

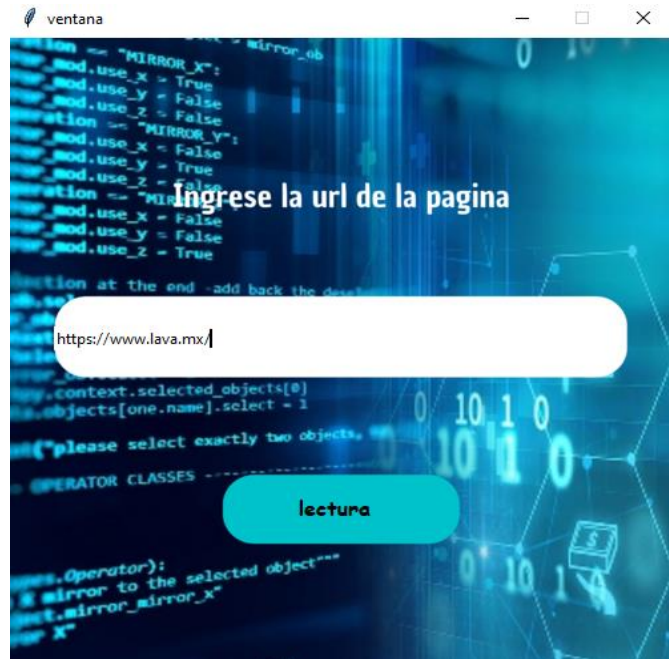


Ilustración 2 vista con la url

En este caso la pagina no contiene correos que sean válidos para nuestro autómata en este caso nos manda un mensaje:

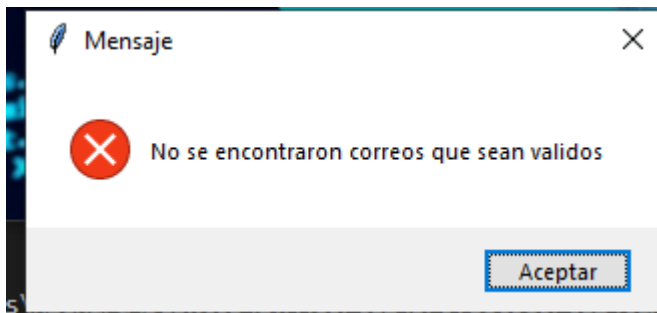


Ilustración 4 mensaje de aviso

En caso de encontrar una pagina con seguridad nos manda este mensaje:

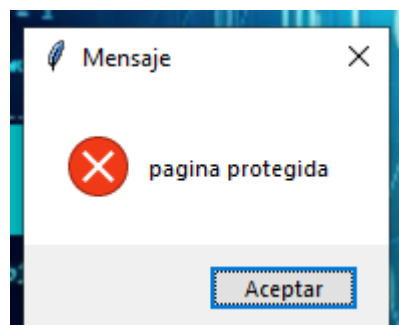


Ilustración 5 mensaje de aviso de protección

Ya al tener un resultado con correos que si aceten nuestro autómeta se nos muestra la tabla de contenido:

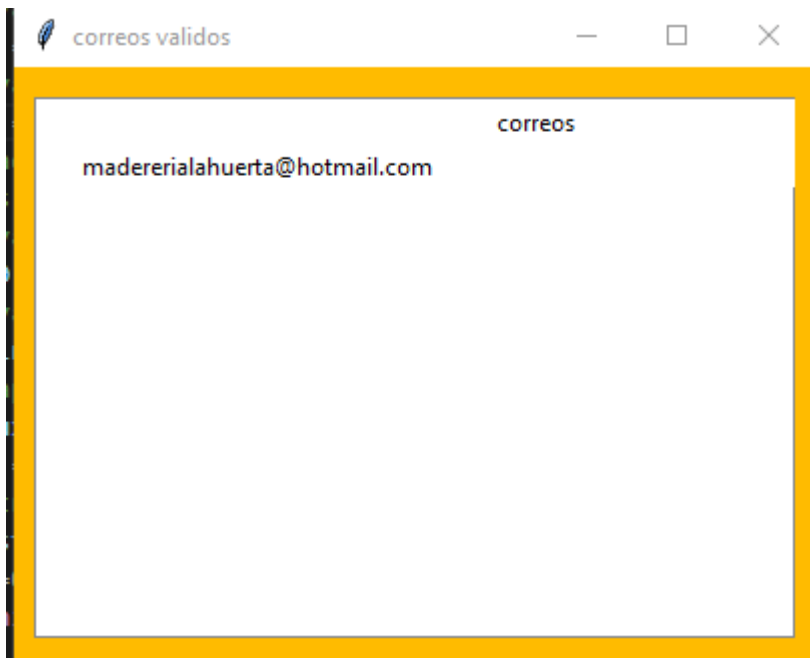


Ilustración 6 muestra de correos validos

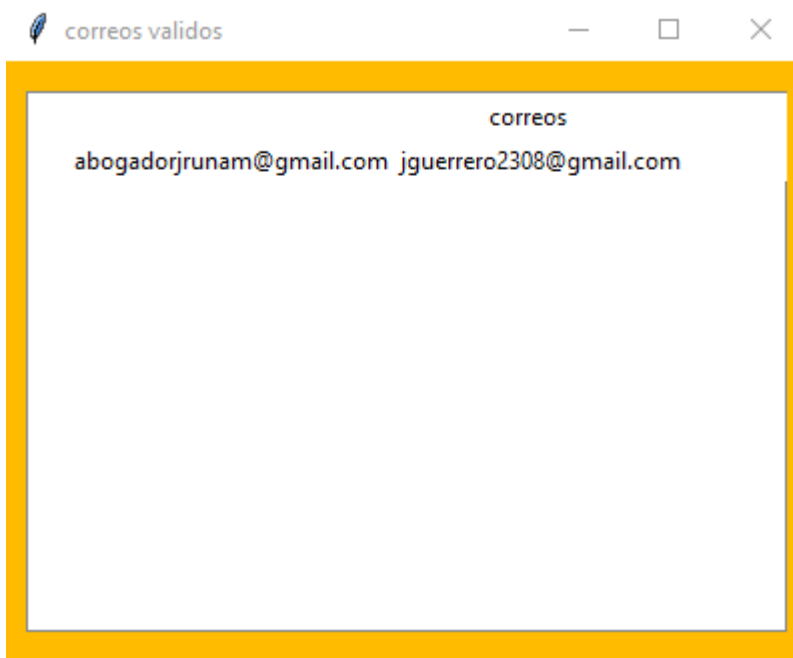


Ilustración 7 más de un correo

Así también vemos como quedan los archivos de textos (TXT) y también vemos como es la página web:

texto: Bloc de notas

Archivo Edición Formato Ver Ayuda

Contact - Innovaciones Jurídicas

Ir al contenido

Inicio

Nosotros

Servicios

Contacto

Main Menu

Inicio

Nosotros

Servicios

Contacto

Contacto

Raúl Jimenez

EMAIL

abogadorjrunam@gmail.com

55-3972-6928

Jorge Guerrero

EMAIL

jguerrero2308@gmail.com

55-1686-3532

Inicio

Nosotros

Servicios

Contacto

Innovaciones Jurídicas ©

correosrescatados: Bloc de notas

Archivo Edición Formato Ver Ayuda

abogadorjrunam@gmail.com jguerrero2308@gmail.com

Ilustración 8 correos encontrados.

Ilustración 10 extrac

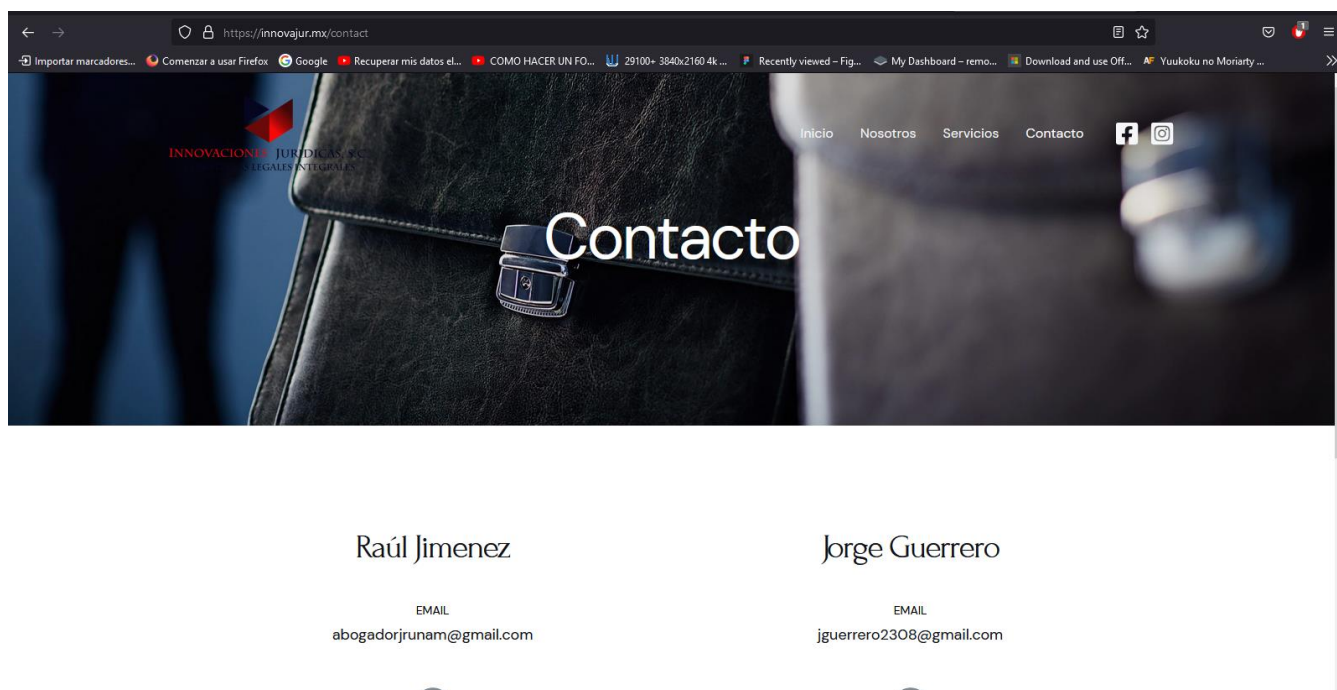


Ilustración 9 página web

Bibliografía:

Wolfram, Stephen (2002). «Starting From Randomness». A New Kind of Science. Wolfram Media. p. 958. Consultado el 31 de marzo de 2010.

Saltar a: a b Rabin, Michael O.; Scott, Dana (1959). «Finite automata and their decision problems». IBM Journal of Research and Development (en ingles) (IBM Corp. Riverton, NJ, USA) 3 (2): 114-125. ISSN 0018-8646. Consultado el 5 de abril de 2010.

Wolfram, Stephen (2002). A New Kind of Science (en ingles). Wólffram Media. p. 893. Consultado el 31 de marzo de 2010.