Universidad Simón Bolívar
Jormar Martín Arellano Gonzalez 0537840
Manuel Gómez Chacon 0535456

Proyecto I

Expresiones regulares E1 , E2 y E3 que corresponden, respectivamente, al reconocimiento de la palabra clave “as”, la palabra clave “array “y de un identificador de variables.

E1: as
E2: array
E3: (a + b + ... + z + A + B + ... + Z)(a + b + ... + z + A + B + ... + Z + 0 + 1 + ... + 9 + _)*

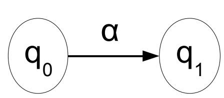
Vemos que la expresión regular más compleja es la del identificador. Vamos a identificar algunos conjuntos de interés.

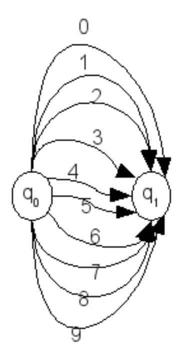
Σ = ASCII
Sea $\bar{A} \subseteq \Sigma$, tal que $\bar{A}$ = {a,b,c, ... ,z,A,B,C, ... , Z}, es decir, solo contiene las leras.
Sea $\bar{E} \subseteq \Sigma$, tal que $\bar{E}$ = {0,1,2,3,4,5,6,7,8,9}, es decir, solo contiene los numeros.

En el futuro, podremos referirnos a $\bar{A} \subseteq \Sigma$ como el conjunto LETRA, y a $\bar{E} \subseteq \Sigma$ como el conjunto NUMERO.

Se puede ver que si $\alpha \in \bar{A}$, entonces $\alpha \in \Sigma$. Mismo análisis para cualquier $\beta \in \bar{E}$.
Por simplicidad, usaremos una notación simplificada en los autómatas; si usamos una letra literal, que perenece a Σ, nos estaremos refiriendo precisamente a esa letra, mientras que si usamos un caracter especial, nos estaremos refiriendo a un conjunto de letras, dependiendo del conjunto del que hablemos.
Por ejemplo:

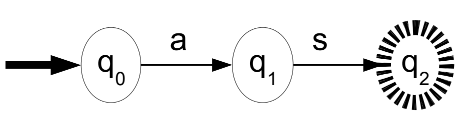
Si $\alpha \in$ NUMERO, enonces los dos automatas de abajo son equivalentes:



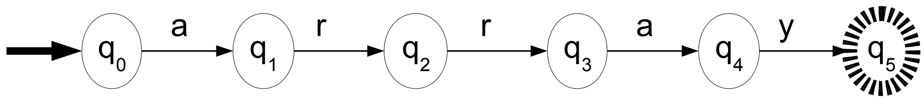


A continuación presentaremos los diagramas de transición de los autómatas finitos, M1 M2 y M3, que reconocen, respectivamente los lenguajes E1, E2 y E3. Trabajaremos sobre:

M1:

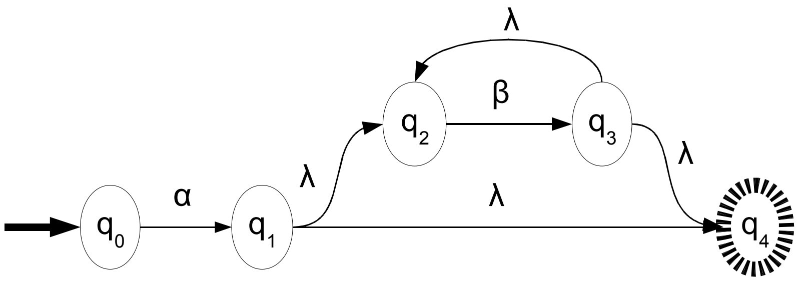


M2:



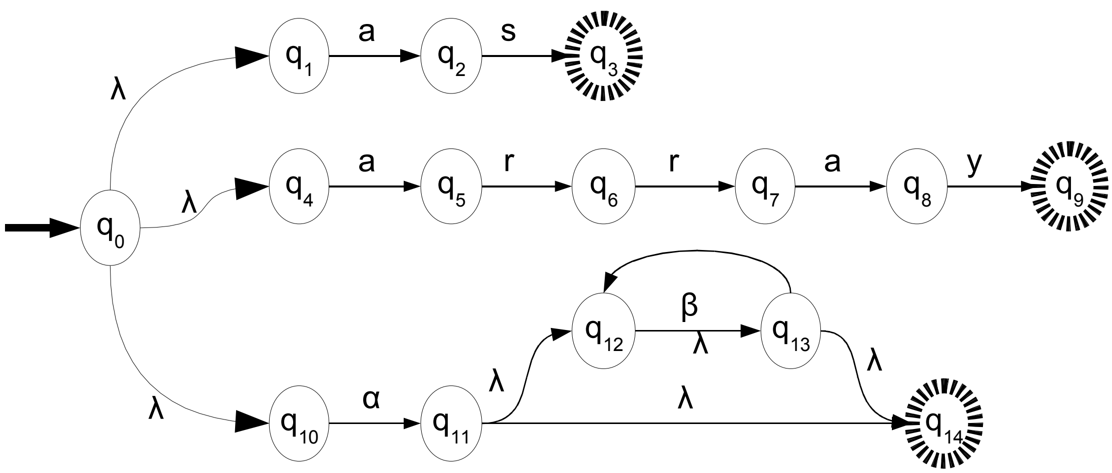
Para α ϵ LETRA, β ϵ (LETRA ∪ NUMERO ∪ {_})

M3:



Estos tres autómatas son no-determinísticos, obtenidos como resultado de aplicar el algoritmo de conversión de ER a AFND. Si quisiéramos construir un autómata M que reconociera los lenguajes de M1, M2 y M3, es decir, L(M) = L(M1) ∪ L(M2) ∪ L(M3), bastaría con agregar λ-transiciones de la siguiente forma:

M:



A la hora de implementar un analizador lexicográfico, es importante que el autómata M presentado reconozca a que lenguaje pertenece la palabra identificada. Hay que notar que, “as” y “array” son palabras reservadas, y también son identificadores validos, por lo que “as” pertenece a L(M1) y L(M3), mientras que “array” pertenece a L(M2) y L(M3).

Este conflicto se puede resolver aplicando alguna distinción y/o prioridad a los estados finales de la maquina M. Dicho en otras palabras, **los estados finales son diferentes entre si.** Para esta primera maquina M podríamos resolver la prioridad como sigue:

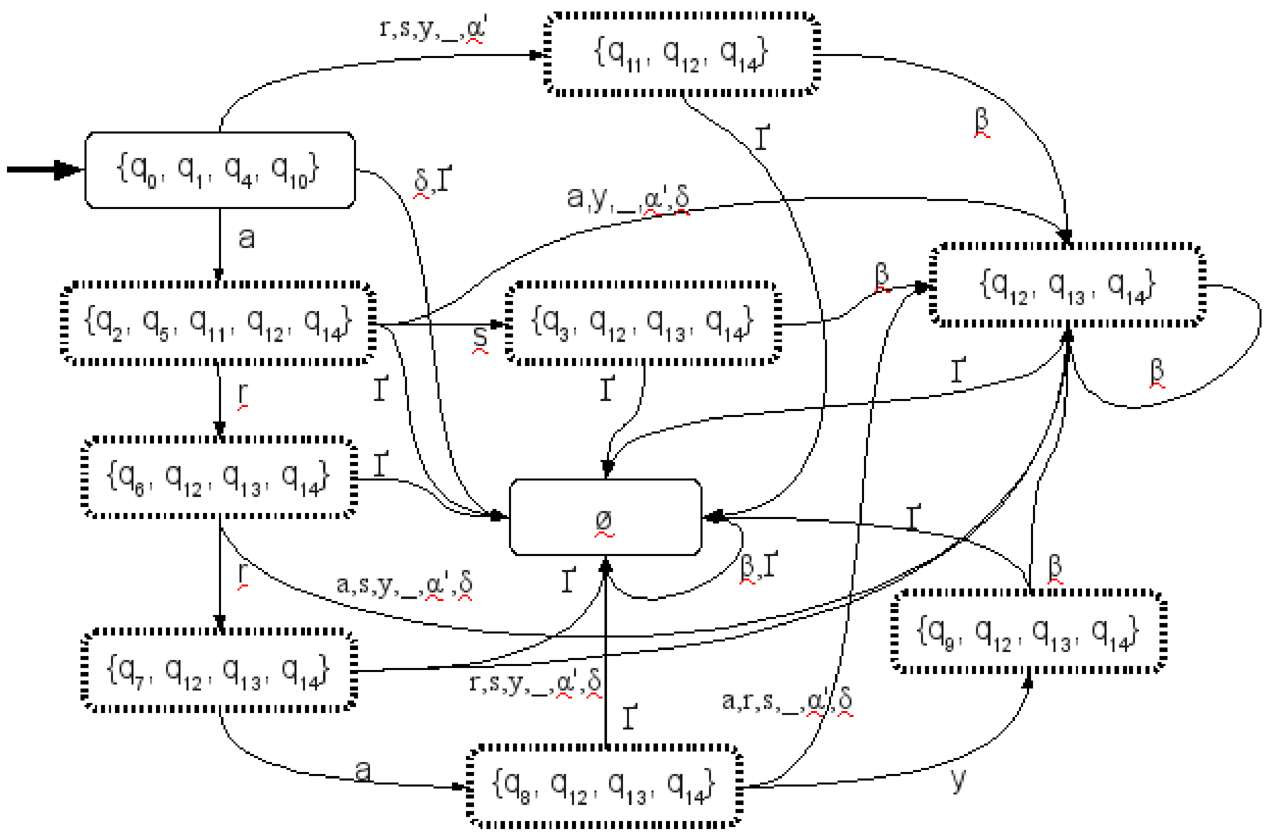
Para dos estados finales q_i y q_p cualesquiera, q_i es más relevante que q_p si y solo si i<p. Si al reconocer una palabra se llega a dos estados finales distintos, se tomara únicamente el **estado final más relevante.** Los lenguajes reconocidos serían:

q_3 indica que la palabra pertenece a L(E1)

q_9 indica que la palabra pertenece a L(E2)

q_{14} indica que la palabra pertenece a L(E3)

A continuación presentaremos el autómata finito determinístico mínimo para M, pero primero debemos convertirlo en AFD, usando el algoritmo repasado en la practica, resultando:



Con:

- α є LETRA
- δ є NUMERO
- β є (LETRA U NUMERO U {_})
- α' є (LETRA – {a, r, s, y}), es decir, α' puede ser cualquier letra, acepto la “a”, “r”, “s” y “y”.
- Ґ є Σ - (LETRA U NUMERO U {_}), es decir, Ґ es cualquier caracter de Σ que no sea letra o numero o el '_' (underscore).

Los estados finales son todos aquellos que tienen el borde punteado (todos, exepto el estado inicial y el estado vacio). En este autómata, llámese M', debemos mantener la prioridad de estados finales de la misma forma, quedando asi:

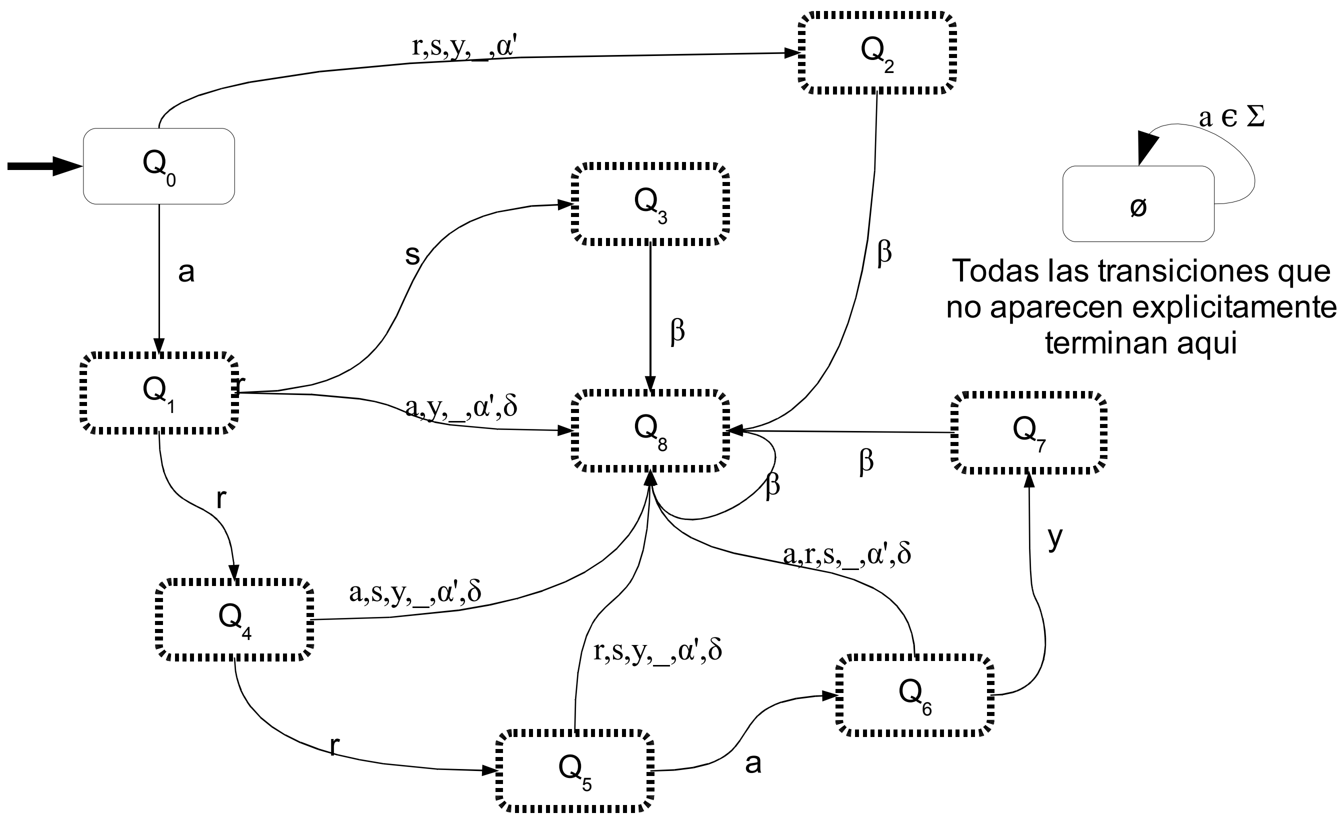
$\{q_3, q_{12}, q_{13}, q_{14}\}$ Tiene precedencia 3, y reconoce el lenguaje L(E1).

$\{q_9, q_{12}, q_{13}, q_{14}\}$ Tiene precedencia 2, y reconoce el lenguaje L(E2).

Todos los estados finales restantes tiene precedencia 1, y reconoce el lenguaje L(E3).

Siempre que se tengan dos posibles estados finales, tomaremos aquel que tenga mayor precedencia, y asi determinaremos a que lenguaje pertenece la palabra.

Para mayor claridad, vamos a reescribir al AFD M', pero sin las transiciones con el estado vacio (asumiremos que si no se dice algo, caemos en el estado vacio), y renombrando los estados:



Con:

- α є LETRA
- δ є NUMERO
- β є (LETRA ∪ NUMERO ∪ {_})
- α' є (LETRA – {a, r, s, y}), es decir, α' puede ser cualquier letra, acepto la "a", "r", "s" y "y".

En este nuevo grafo (recordemos que es un ADF, previamente definido, solo que no dibujamos las transiciones que caen en el estado vacio), llamémoslo M'_Q, nos permite visualizar mejor lo que esta pasando, y asi podemos aplicar el algoritmo de minimización más cómodamente. Vamos a mostrar las clases de equivalencia resultantes al aplicar el algorimo.

Ξ_0 = { {Q_0, ø}, {Q_3}, {Q_7}, {Q_1 , Q_2 ,Q_4 ,Q_5 ,Q_6 , Q_8} }

Recordemos que los estados finales son **distinguibles**, lo que significa que no todos son equivalentes: Q_3 reconoce el lengueje L(E1), Q_7 reconoce el lengueje L(E2) y Q_1 , Q_2 ,Q_4 ,Q_5 ,Q_6 , Q_8 reconocen el lengueje L(E3).

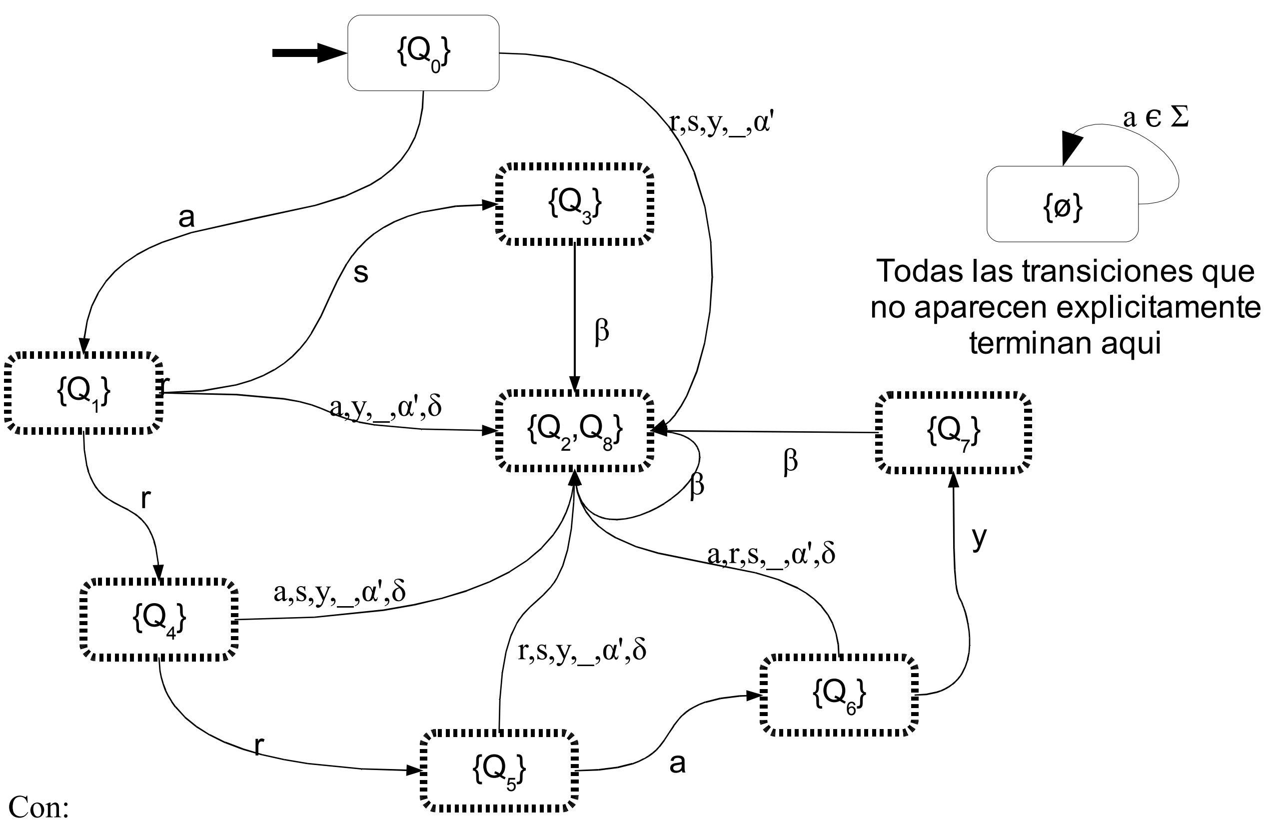
Ξ_1 = { {ø}, {Q_0}, {Q_1}, {Q_3}, {Q_6}, {Q_7}, {Q_2 ,Q_4 ,Q_5 ,Q_8} }

Ξ_2 = { {ø}, {Q_0}, {Q_1}, {Q_3}, {Q_5}, {Q_6}, {Q_7}, {Q_2 ,Q_4 , Q_8} }

Ξ_3 = { {ø}, {Q_0}, {Q_1}, {Q_3}, {Q_4}, {Q_5}, {Q_6}, {Q_7}, {Q_2 , Q_8} }

Ξ_4 = { {ø}, {Q_0}, {Q_1}, {Q_3}, {Q_4}, {Q_5}, {Q_6}, {Q_7}, {Q_2 , Q_8} } = Ξ_3

Finalmente, vemos reducidos solo dos estados, quedando el autómata de la siguiente manera:



Con:

- α є LETRA
- δ є NUMERO
- β є (LETRA ∪ NUMERO ∪ {_})
- α' є (LETRA – {a, r, s, y}), es decir, α' puede ser cualquier letra, excepto la "a", "r", "s" y "y".

Recordemos, una vez más, que se trata del AFD mínimo, solo que, por simplicidad visual, no colocamos las transiciones que terminan en el estado {ø}.

Para terminar, estos son los estados y los lenguajes asociados a ellos:
$\{Q_3\}$ Reconoce el lenguaje L(E1).
$\{Q_7\}$ Reconoce el lenguaje L(E2).
$\{Q_1\}$, $\{Q_4\}$, $\{Q_5\}$, $\{Q_6\}$, $\{Q_2 , Q_8\}$ Reconocen el lenguaje L(E3).