

Este documento plantea una solución al problema de los comentarios multilínea basada en el uso de algoritmos vistos en la asignatura de Autómatas y Lenguajes Formales.

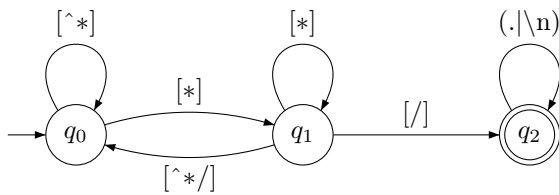
El problema consiste en buscar una expresión regular que describa el interior del comentario multilínea, ya que el comienzo y final del comentario se reduce a usar el prefijo `/*` y el sufijo `*/` en la expresión regular.

La descripción del interior del comentario se puede plantear como la descripción de las cadenas de caracteres que no contengan la subcadena `*/`, que es la marca de fin del comentario. Podemos seguir esta estrategia:

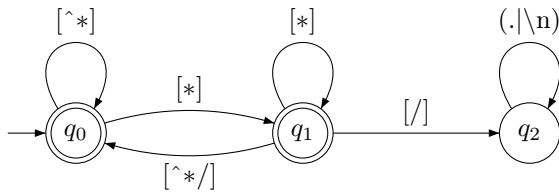
1. Pensar en el AFD completo que reconoce cadenas de caracteres que incluyen a la subcadena `*/`.
2. Complementar el AFD anterior.
3. Plantear el sistema de ecuaciones para obtener la expresión regular del estado inicial.

El resultado es lo que buscamos. Para plantear el AFD, podemos usar en las transiciones expresiones regulares que simbolicen el conjunto de caracteres que las etiquetan. Por ejemplo, para representar en una transición el carácter `*`, se puede usar: `[*]`. Y si hay que representar a todos los caracteres menos `*`, se puede poner así: `[^*]`. Por último, si se necesita representar a cualquier carácter, se puede usar `(.\n)`.

Comenzamos planteando el AFD completo que reconoce las cadenas que incluyen a la subcadena `*/`:



Ahora complementamos el AFD (es completo):



Y planteamos el sistema de ecuaciones:

$$\begin{aligned} q_0 &= [^*]q_0 \mid [*]q_1 \mid \lambda \\ q_1 &= [^*]q_0 \mid [*]q_1 \mid [/]q_2 \mid \lambda \\ q_2 &= (.\n)q_2 \end{aligned}$$

Si aplicamos el algoritmo para resolver el sistema, obtenemos:

$$\begin{aligned} q_2 &= (.\n)^* \emptyset = \emptyset \\ q_1 &= [^*]q_0 \mid [*]q_1 \mid \lambda = [*]^*([^*]q_0 \mid \lambda) \\ q_0 &= [^*]q_0 \mid [*]^*([^*]q_0 \mid \lambda) \mid \lambda = \\ &= [^*]q_0 \mid [*]^+[^*]q_0 \mid [*]^+ \mid \lambda = \\ &= ([^*] \mid [*]^+[^*])^*([*]^+ \mid \lambda) = \\ &= ([^*] \mid [*]^+[^*])^*[*]^+ \end{aligned}$$

Por tanto, podemos emplear esta expresión regular del comentario multilínea:

$$\begin{aligned} \text{comentario} &= "/*"([^*] \mid [*]^+[^*])^*[*]^+*"/" \\ &= "/*"([^*] \mid [*]^+[^*])^*[*]^+*"/" \end{aligned}$$