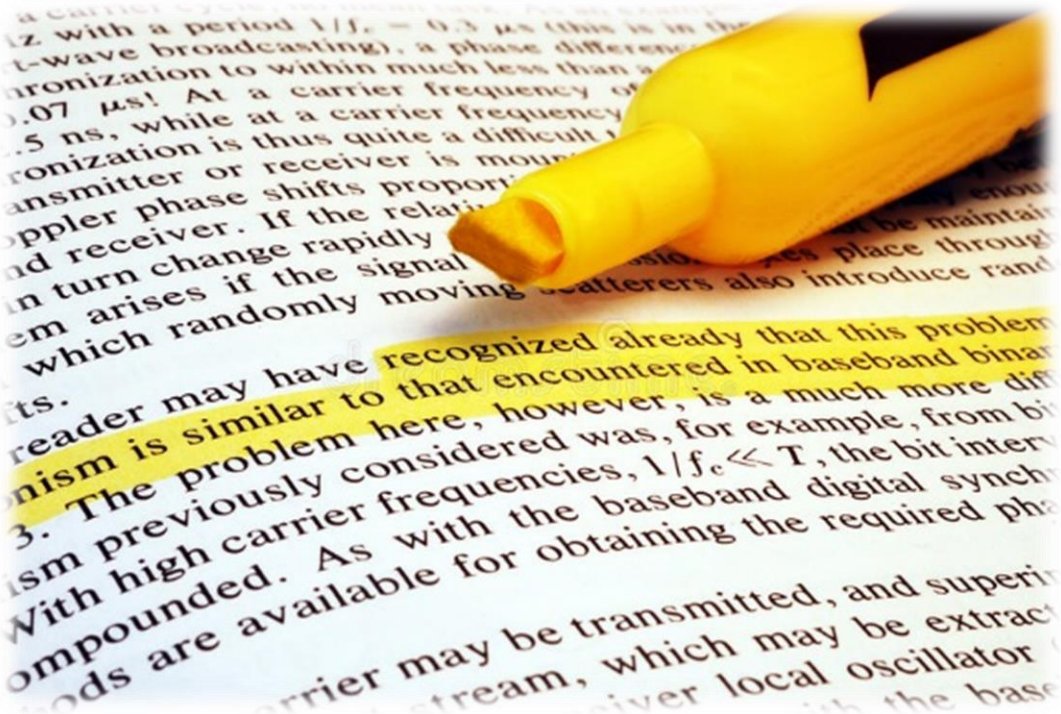
UTEM

UNIVERSIDAD TECNOLÓGICA METROPOLITANA

Proyecto Semestral:
Analizador de textos con el algoritmo de Porter



Daniel Vásquez Robles.
Mayo 2019.

Universidad Tecnológica Metropolitana.
Depto. Ingeniería en Informática.
Teoría de Autómatas.
Profesor Ricardo Corbinaud.

Tabla de Contenidos

Capítulo 1 Introducción e información general 1
¿Qué es el algoritmo de Porter? 1
Text-Mining 2
Autómatas 3
Expresiones regulares 3
Funciones relevantes 4
Comparación de resultados obtenidos por el algoritmo 5
Lista de referencias 7
Apéndice 8
Pronombres: 8
Articulos: 42
Prefijos: 46
Sufijos: 54

Lista de tablas

Tabla 1. Comparación de resultados obtenidos 6

Capítulo 1

Introducción e información general

¿Qué es el algoritmo de Porter?

El algoritmo de derivación de Porter es un proceso para eliminar las terminaciones morfológicas e inflexionales más comunes de las palabras en inglés, sin embargo, para efectos de la realización de este proyecto se aplicará a las palabras en español.

El documento original del algoritmo de derivación se escribió en 1979 en el laboratorio de computación, Cambridge (Inglaterra), como parte de un proyecto de RI (recuperación de información) más grande, el cual sería publicado en 1980 con el nombre de *"An algorithm for suffix stripping"*.

Originalmente el algoritmo fue escrito en BCPL, un lenguaje que alguna vez fue popular, pero que ahora está extinguido. Durante los primeros años posteriores a su publicación, se distribuyó en su versión BCPL, a través de una cinta de papel perforada. Pronto comenzaron a aparecer versiones en otros lenguajes, y para 1999 se estaba utilizando, citando y adaptando ampliamente.

Remover los sufijos de las palabras automáticamente significa una operación que resulta especialmente útil en el campo de la recuperación de información. Podemos decir que un documento está formado por un vector largo de términos, los cuales pueden poseer una raíz en común, por ejemplo:

Conectar Conectado Conectando Conexión Conexiones

Para este ejemplo podemos dar cuenta que si a estas palabras se les extrae el sufijo la raíz en común de estas palabras es “conec-”, de esta forma el algoritmo realiza un conteo de la frecuencia de las diferentes raíces obtenidas y con esta información podremos deducir de que trata el documento analizado por este algoritmo.

Algoritmos basados en este método conocido como Stemming (reducir una palabra a su raíz o stem) ayudan en sistemas de recuperación de información ya que aumentan el *recall* que es una medida sobre el número de documentos que se pueden encontrar con una consulta. Por ejemplo: Al obtener el stem “bibliotec- “, encontrara documentos para “biblioteca” y “bibliotecario” ya que ambas palabras poseen dicho stem en común.

El motor de búsqueda *Muscat* proviene de una investigación realizada por Porter en la Universidad de Cambridge y fue comercializado en 1984 por Cambridge CD Publishing; posteriormente se vendió a MAID, que se convirtió en la Corporación Dialog.

Text-Mining

Consiste en el proceso de analizar colecciones de materiales textuales con el fin de capturar conceptos y temas clave, descubrir relaciones y tendencias ocultas sin requerir que conozca las palabras o términos precisos que los autores han usado para expresar esos conceptos. Se suele confundir el text-mining con la recuperación de información, sin embargo, tienen un objetivo diferente ya que el text-mining se centra en la extracción y gestión de contenido de calidad, la terminología y las relaciones contenidas en la información.

En el text-mining destacan tres técnicas. La primera de ellas es la extracción de términos, la cual identifica los términos clave y entidades lógicas. Luego la extracción de información, consiste en identificar las relaciones básicas entre los términos extraídos del texto. Por último, el análisis relacional permite formar modelos complejos, que permiten dar una idea de las relaciones entre varias entidades.

Al utilizar text-mining se pueden analizar grandes volúmenes de información de manera más rápida y eficiente en comparación a la codificación manual, ya que una vez establecidos y validados los parámetros, el proceso es automatizado. Además, facilita el análisis de datos no estructurados permitiendo así filtrar, buscar y hacer referencias cruzadas de manera más sencilla.

Autómatas

En el primer avance de proyecto se trabajó modelando autómatas que permitieran satisfacer los requerimientos exigidos. Como resultado se obtuvieron los diferentes autómatas para reconocer las diferentes categorías de palabras ya sean los sufijos, prefijos, artículos y pronombres. Cabe destacar que el modelo de autómata es el mismo tratándose de la misma categoría de palabra, es decir, se puede aplicar el autómata tanto para un sufijo como para otro sufijo diferente, lo mismo aplica para cada categoría anteriormente mencionada. Todas las figuras correspondientes a los autómatas están ubicadas en el apéndice.

Expresiones regulares

Posteriormente se trabajó en el segundo avance de proyecto, esta vez modelando expresiones regulares capaces de identificar los sufijos, prefijos, pronombres o artículos

presentes en un texto. En el caso de los sufijos se modelo por ejemplo que las palabras que contenían sufijos primero venia una serie de combinaciones de letras y luego le seguía el sufijo en cuestión. Para los prefijos, caso contrario, la palabra iniciaba con el prefijo y le seguía una combinación de letras.

Previo análisis del texto se utilizaron otras expresiones regulares para normalizar el documento, es decir, se transformó el texto solo a minúsculas, se eliminó los símbolos, tildes y números. De esta manera el algoritmo podrá cumplir su función.

Funciones relevantes

Finalmente, luego de trabajar con los avances, se sintetizo las ideas en un código JAVA, que como resultado de analizar un documento .txt crea un "Diccionario.txt" que contiene los 10 stem (o raíces) más frecuentes. Además, se observó que las expresiones regulares (regex) obtenidos en el avance dos, encapsulaban la palabra en su totalidad por lo que de hacer esto, eliminaría la palabra completa generando resultados erróneos, es por esto que fue necesario re-definir los regex a utilizar por el código final.

A continuación, destacare algunas de las funciones del código que ayudaron a que el proyecto cumpliera con su objetivo:

La función "Tildes" recibe un tipo de dato string que contiene una palabra obtenida del texto por el algoritmo y reemplaza los caracteres de vocales que presenten tildes por su formato sin tilde, esta es una de las funciones que ayuda a normalizar el texto como se mencionó anteriormente.

```
public String Tildes(String s){
   s = s.replaceAll("[áàä]","a");
```

```
    s = s.replaceAll("[éèë]","e");
    s = s.replaceAll("[íìï]","i");
    //s = s.replaceAll("ñ","n"); //En caso de querer sustituir las "ñ" por "n".
    s = s.replaceAll("[óòö]","o");
    s = s.replaceAll("[úùü]","u");
    s = s.replaceAll("[ýÿ]","y");
    return s;
}
```

La siguiente función “Articulos” recibe un tipo de dato string que contiene una palabra obtenida del texto, verifica si es un artículo especificado por medio de la expresión regular, y en caso de que así sea, lo elimina. Esto lo realizamos por que los artículos se repiten mucho a lo largo de los textos y de no eliminarlos en los resultados aparecerían entro de los diez primeros stem más frecuentes entorpeciendo nuestro resultado final ya que los artículos no nos sugieren de que trata el texto.

```
public String Articulos(String s){
    s = s.replaceAll("\\b[en|el|l[o|a](s)?]\\b","");
    return s;
}
```

Comparación de resultados obtenidos por el algoritmo

A continuación, mostrare los resultados obtenidos por el algoritmo aplicando primero la eliminación de los sufijos y prefijos de las palabras, para luego comparar los resultados obtenidos solo aplicando la eliminación de sufijos.

Tabla 1. Comparación de resultados obtenidos.

Sufijos y Prefijos	***Solo Sufijos***
'sley' se ha repetido 315 veces.	'presley' se ha repetido 315 veces.
'elvis' se ha repetido 95 veces.	'como' se ha repetido 150 veces.
'cant' se ha repetido 79 veces.	'elvis' se ha repetido 95 veces.
'habia' se ha repetido 59 veces.	'cant' se ha repetido 79 veces.
'musica' se ha repetido 53 veces.	'habia' se ha repetido 59 veces.
'spues' se ha repetido 46 veces.	'musica' se ha repetido 53 veces.
'rock' se ha repetido 46 veces.	'años' se ha repetido 50 veces.
'memphis' se ha repetido 41 veces.	'despues' se ha repetido 46 veces.
'cuando' se ha repetido 37 veces.	'rock' se ha repetido 46 veces.
'parker' se ha repetido 37 veces.	'memphis' se ha repetido 41 veces.
'canciones' se ha repetido 30 veces.	'cuando' se ha repetido 37 veces.
'lbum' se ha repetido 30 veces.	'parker' se ha repetido 37 veces.
'roll' se ha repetido 28 veces.	'desde' se ha repetido 34 veces.
'cada' se ha repetido 28 veces.	'canciones' se ha repetido 30 veces.
'tras' se ha repetido 28 veces.	'album' se ha repetido 30 veces.
'untry' se ha repetido 27 veces.	'roll' se ha repetido 28 veces.
'greso' se ha repetido 27 veces.	'country' se ha repetido 27 veces.
'donde' se ha repetido 26 veces.	'estilo' se ha repetido 26 veces.
'blues' se ha repetido 26 veces.	'donde' se ha repetido 26 veces.
'gran' se ha repetido 26 veces.	'blues' se ha repetido 26 veces.

Como conclusión luego de comparar los resultados obtenidos en la tabla, los stem resultantes presentan pequeñas diferencias debido a que se aplicó un leve cambio al "filtro de palabras" por así decirlo, sin embargo, al solamente eliminar los sufijos la información obtenida resulta más fácil de deducir, ya que, por ejemplo, el stem "Presley" resulta más fácil de comprender, nos dice más en definitiva que el stem "sley", de modo que al observar el ligero cambio en los resultados da a pensar que el algoritmo funciona en ambos casos pero para facilitar la deducción y comprensión de los resultados, es más practico solamente eliminar los sufijos.

Lista de referencias

Martin F. Porter (2006). Recuperado de https://tartarus.org/martin/PorterStemmer/index.html

Porter, M. F. (1980). *"An algorithm for suffix stripping".* Recuperado de https://www.cs.odu.edu/~jbollen/IR04/readings/readings5.pdf

Universidad de Alcalá (2018). Recuperado de https://www.master-data-scientist.com/en-que-consiste-el-text-mining/

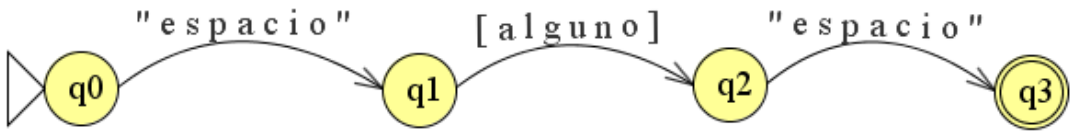
Rochina, Paula (2017). *¿Qué es y cuáles son las aplicaciones del text-minig?*. Recuperado de https://revistadigital.inesem.es/informatica-y-tics/text-mining/

Apéndice

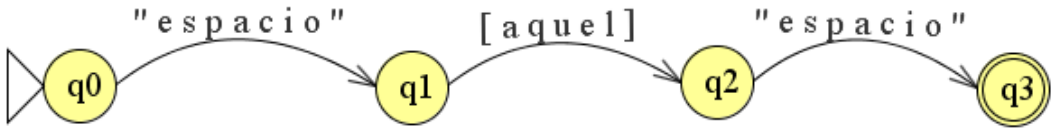
Pronombres:

El regex correspondiente es: ^[pronombres]$


"espacio"
[algunas]
"espacio"
q0
q1
q2
q3


"espacio"
[alguno]
"espacio"
q0
q1
q2
q3


"espacio"
[algunos]
"espacio"
q0
q1
q2
q3


"espacio"
[aquel]
"espacio"
q0
q1
q2
q3


"espacio"
[aquella]
"espacio"
q0
q1
q2
q3


"espacio"
[aquellas]
"espacio"
q0
q1
q2
q3


"espacio"
[aquello]
"espacio"
q0
q1
q2
q3


"espacio"
[aquellos]
"espacio"
q0
q1
q2
q3

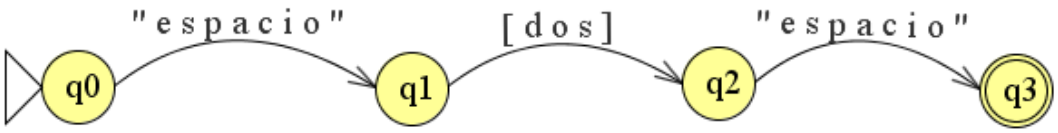

"espacio"
[bastante]
"espacio"
q0
q1
q2
q3


"espacio"
[bastantes]
"espacio"
q0
q1
q2
q3


"espacio"
[conmigo]
"espacio"
q0
q1
q2
q3


"espacio"
[consigo]
"espacio"
q0
q1
q2
q3


"espacio"
[cuatro]
"espacio"
q0
q1
q2
q3


"espacio"
[dos]
"espacio"
q0
q1
q2
q3


"espacio"
[el]
"espacio"
q0
q1
q2
q3


"espacio"
[ella]
"espacio"
q0
q1
q2
q3


"espacio"
[ellas]
"espacio"
q0
q1
q2
q3


"espacio"
[ello]
"espacio"
q0
q1
q2
q3


"espacio"
[ellos]
"espacio"
q0
q1
q2
q3


"espacio"
[esa]
"espacio"
q0
q1
q2
q3


"espacio"
[esas]
"espacio"
q0
q1
q2
q3


"espacio"
[eso]
"espacio"
q0
q1
q2
q3


"espacio"
[esos]
"espacio"
q0
q1
q2
q3


"espacio"
[esta]
"espacio"
q0
q1
q2
q3

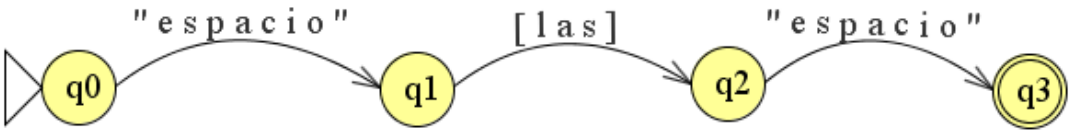

"espacio"
[estas]
"espacio"
q0
q1
q2
q3

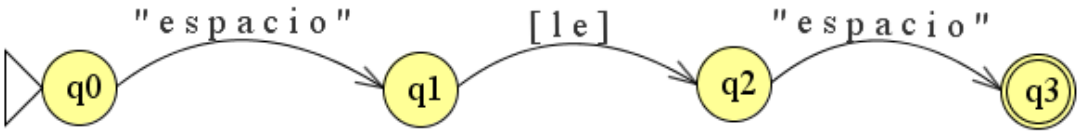

"espacio"
[este]
"espacio"
q0
q1
q2
q3


"espacio"
[estes]
"espacio"
q0
q1
q2
q3

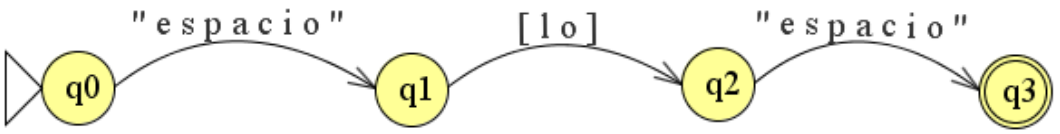

"espacio"
[esto]
"espacio"
q0
q1
q2
q3

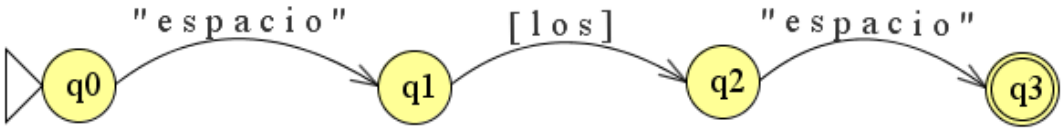

"espacio"
[la]
"espacio"
q0
q1
q2
q3

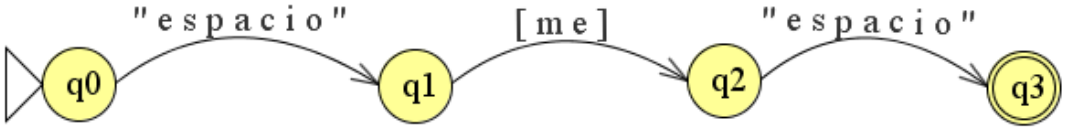

"espacio"
[las]
"espacio"
q0
q1
q2
q3

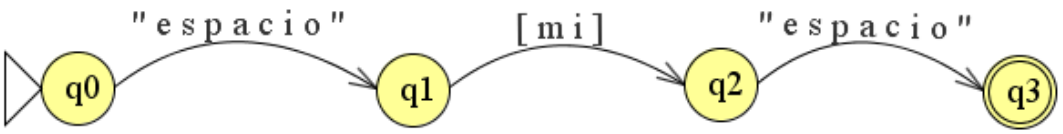

"espacio"
[le]
"espacio"
q0
q1
q2
q3

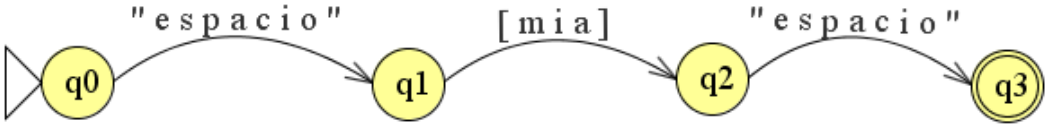

"espacio"
[les]
"espacio"
q0
q1
q2
q3

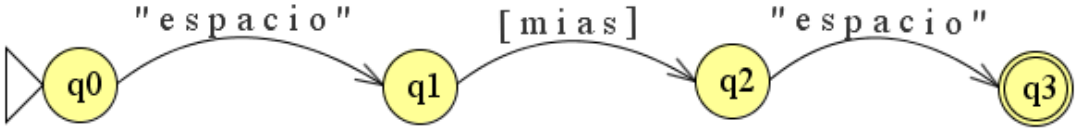

"espacio"
[lo]
"espacio"
q0
q1
q2
q3

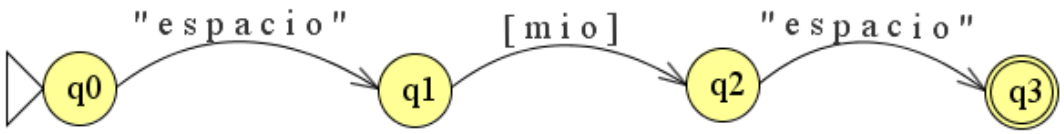

"espacio"
[los]
"espacio"
q0
q1
q2
q3

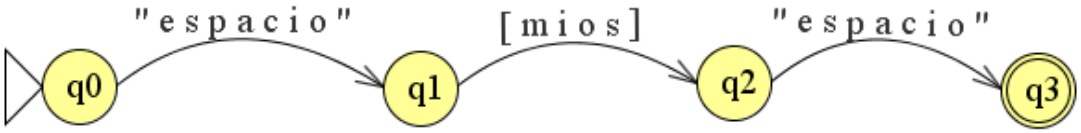

"espacio"
[me]
"espacio"
q0
q1
q2
q3

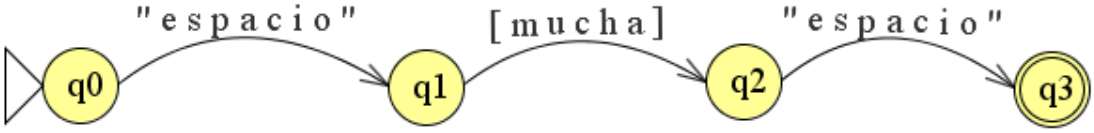

"espacio"
[mi]
"espacio"
q0
q1
q2
q3

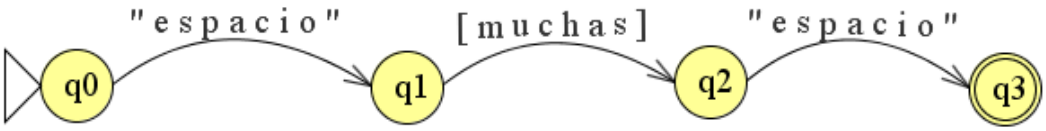

"espacio"
[mia]
"espacio"
q0
q1
q2
q3

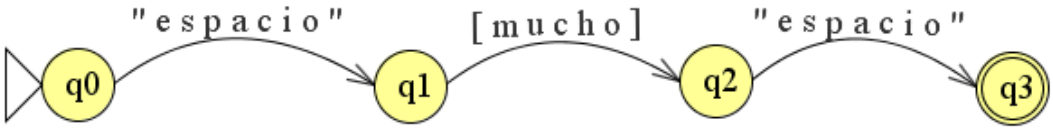

"espacio"
[mias]
"espacio"
q0
q1
q2
q3

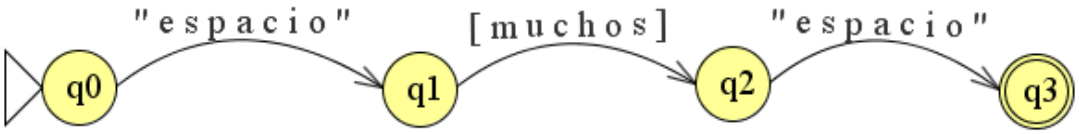

"espacio"
[mio]
"espacio"
q0
q1
q2
q3


"espacio"
[mios]
"espacio"
q0
q1
q2
q3


"espacio"
[mucha]
"espacio"
q0
q1
q2
q3


"espacio"
[muchas]
"espacio"
q0
q1
q2
q3


"espacio"
[mucho]
"espacio"
q0
q1
q2
q3

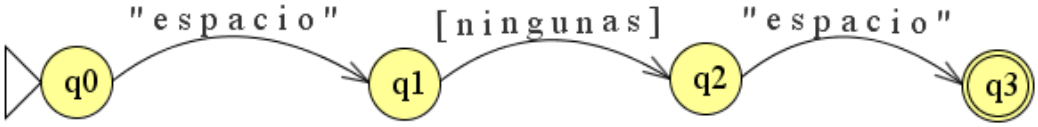

"espacio"
[muchos]
"espacio"
q0
q1
q2
q3


"espacio"
[nada]
"espacio"
q0
q1
q2
q3

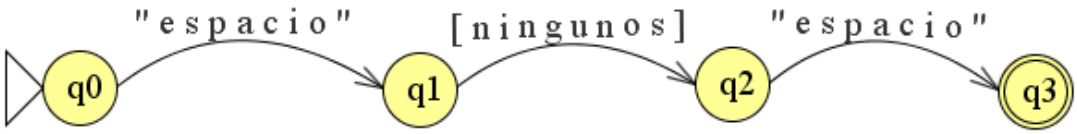

"espacio"
[nadie]
"espacio"
q0
q1
q2
q3


"espacio"
[ningun]
"espacio"
q0
q1
q2
q3


"espacio"
[ninguna]
"espacio"
q0
q1
q2
q3


"espacio"
[ningunas]
"espacio"
q0
q1
q2
q3


"espacio"
[ninguno]
"espacio"
q0
q1
q2
q3


"espacio"
[ningunos]
"espacio"
q0
q1
q2
q3


"espacio"
[nosotras]
"espacio"
q0
q1
q2
q3


"espacio"
[nosotros]
"espacio"
q0
q1
q2
q3


"espacio"
[nuestra]
"espacio"
q0
q1
q2
q3


"espacio"
[nuestras]
"espacio"
q0
q1
q2
q3


"espacio"
[nuestro]
"espacio"
q0
q1
q2
q3

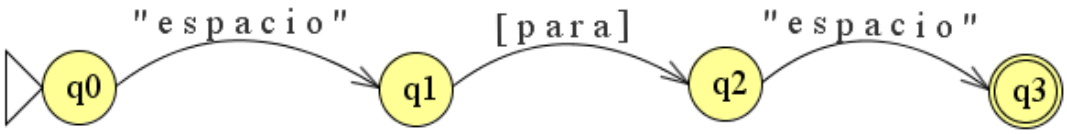

"espacio"
[nuestros]
"espacio"
q0
q1
q2
q3


"espacio"
[otra]
"espacio"
q0
q1
q2
q3


"espacio"
[otras]
"espacio"
q0
q1
q2
q3


"espacio"
[otro]
"espacio"
q0
q1
q2
q3


"espacio"
[otros]
"espacio"
q0
q1
q2
q3

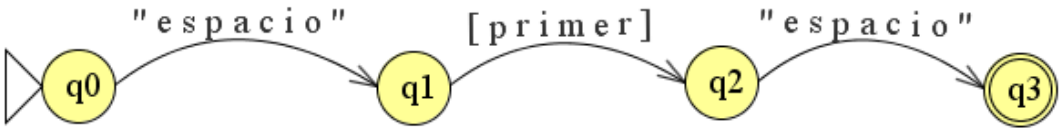

"espacio"
[para]
"espacio"
q0
q1
q2
q3


"espacio"
[poca]
"espacio"
q0
q1
q2
q3

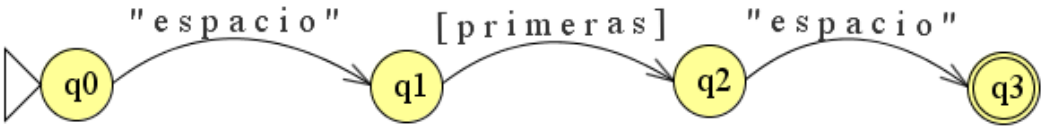

"espacio"
[pocas]
"espacio"
q0
q1
q2
q3


"espacio"
[poco]
"espacio"
q0
q1
q2
q3

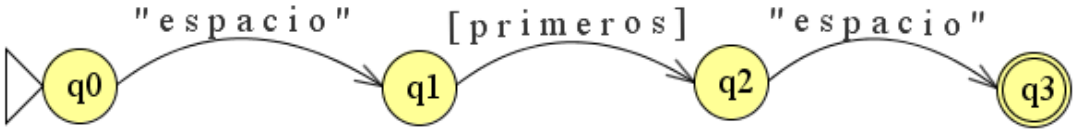

"espacio"
[pocos]
"espacio"
q0
q1
q2
q3


"espacio"
[primer]
"espacio"
q0
q1
q2
q3

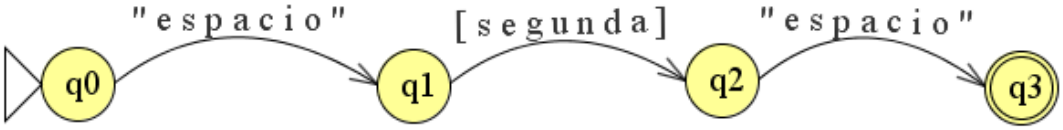

"espacio"
[primera]
"espacio"
q0
q1
q2
q3

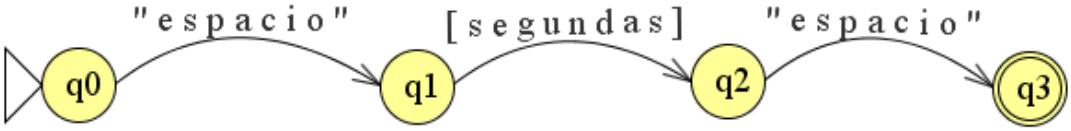

"espacio"
[primeras]
"espacio"
q0
q1
q2
q3

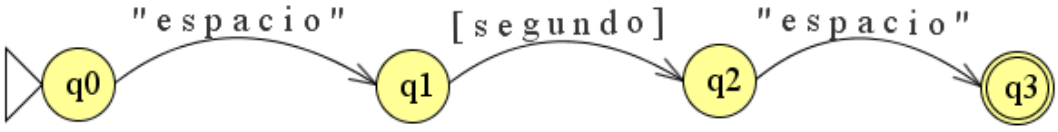

"espacio"
[primero]
"espacio"
q0
q1
q2
q3


"espacio"
[primeros]
"espacio"
q0
q1
q2
q3

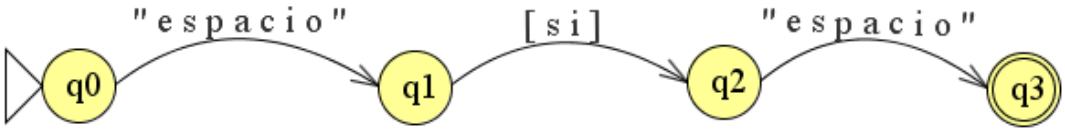

"espacio"
[se]
"espacio"
q0
q1
q2
q3


"espacio"
[segunda]
"espacio"
q0
q1
q2
q3

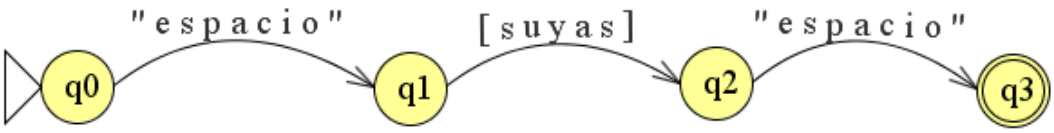

"espacio"
[segundas]
"espacio"
q0
q1
q2
q3


"espacio"
[segundo]
"espacio"
q0
q1
q2
q3

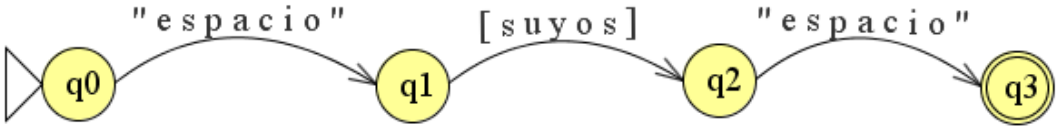

"espacio"
[segundos]
"espacio"
q0
q1
q2
q3

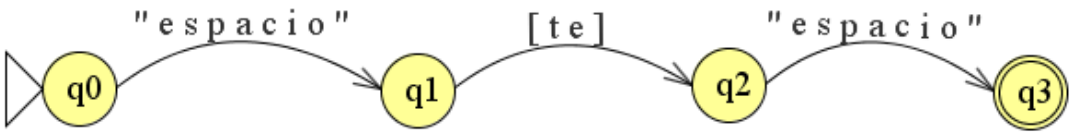

"espacio"
[si]
"espacio"
q0
q1
q2
q3


"espacio"
[suya]
"espacio"
q0
q1
q2
q3


"espacio"
[suyas]
"espacio"
q0
q1
q2
q3


"espacio"
[suyo]
"espacio"
q0
q1
q2
q3


"espacio"
[suyos]
"espacio"
q0
q1
q2
q3


"espacio"
[te]
"espacio"
q0
q1
q2
q3


"espacio"
[tercer]
"espacio"
q0
q1
q2
q3


"espacio"
[tercera]
"espacio"
q0
q1
q2
q3


"espacio"
[terceras]
"espacio"
q0
q1
q2
q3


"espacio"
[tercero]
"espacio"
q0
q1
q2
q3


"espacio"
[terceros]
"espacio"
q0
q1
q2
q3


"espacio"
[ti]
"espacio"
q0
q1
q2
q3


"espacio"
[toda]
"espacio"
q0
q1
q2
q3


"espacio"
[todas]
"espacio"
q0
q1
q2
q3


"espacio"
[todo]
"espacio"
q0
q1
q2
q3

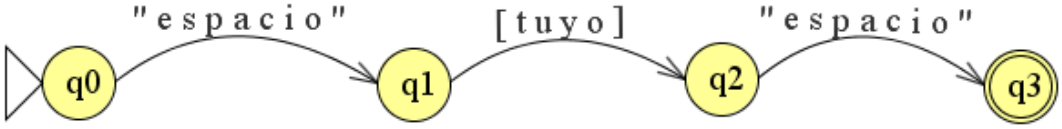

"espacio"
[todos]
"espacio"
q0
q1
q2
q3

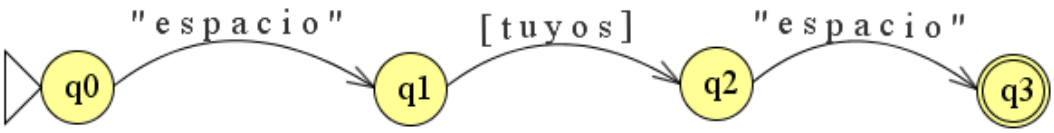

"espacio"
[tres]
"espacio"
q0
q1
q2
q3

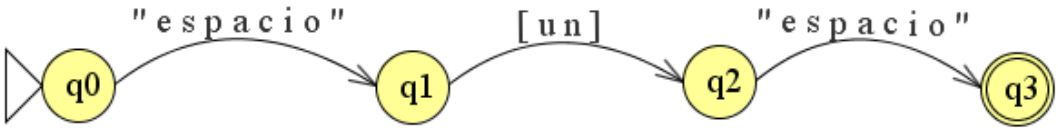

"espacio"
[tu]
"espacio"
q0
q1
q2
q3

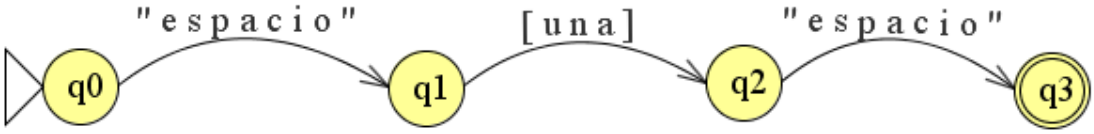

"espacio"
[tuya]
"espacio"
q0
q1
q2
q3


"espacio"
[tuyas]
"espacio"
q0
q1
q2
q3

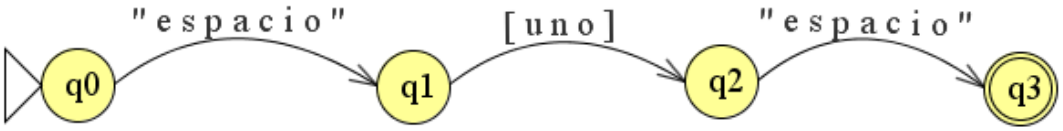

"espacio"
[tuyo]
"espacio"
q0
q1
q2
q3


"espacio"
[tuyos]
"espacio"
q0
q1
q2
q3


"espacio"
[un]
"espacio"
q0
q1
q2
q3


"espacio"
[una]
"espacio"
q0
q1
q2
q3


"espacio"
[unas]
"espacio"
q0
q1
q2
q3


"espacio"
[uno]
"espacio"
q0
q1
q2
q3


"espacio"
[unos]
"espacio"
q0
q1
q2
q3


"espacio"
[varias]
"espacio"
q0
q1
q2
q3


"espacio"
[varios]
"espacio"
q0
q1
q2
q3


"espacio"
[vosotras]
"espacio"
q0
q1
q2
q3

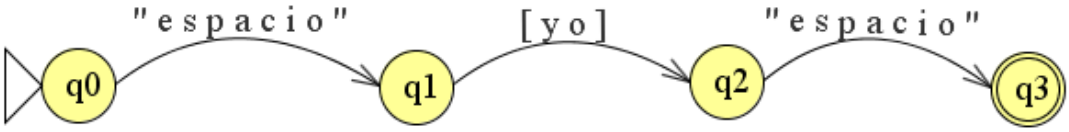

"espacio"
[vosotros]
"espacio"
q0
q1
q2
q3


"espacio"
[vuestra]
"espacio"
q0
q1
q2
q3


"espacio"
[vuestras]
"espacio"
q0
q1
q2
q3

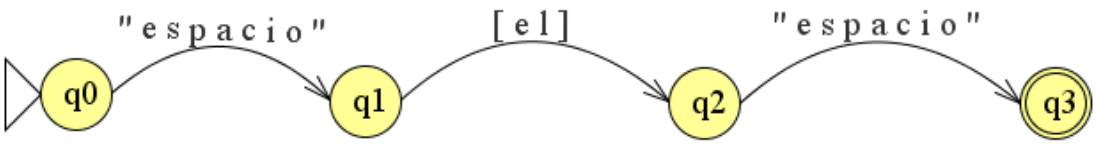

"espacio"
[vuestro]
"espacio"
q0
q1
q2
q3

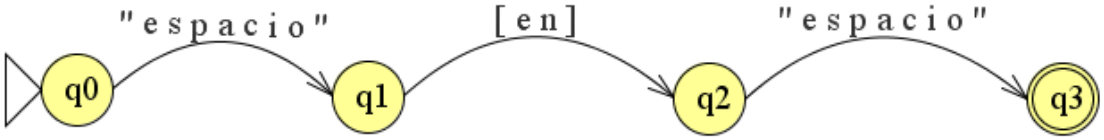

"espacio"
[vuestros]
"espacio"
q0
q1
q2
q3

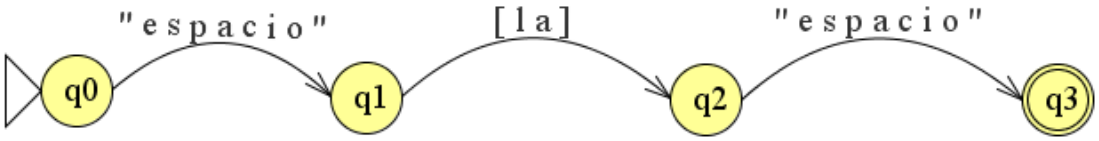

"espacio"
[yo]
"espacio"
q0
q1
q2
q3

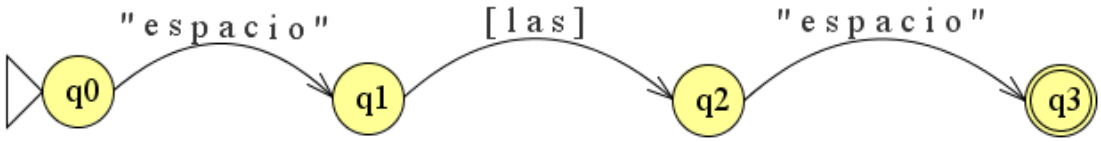
Articulos:

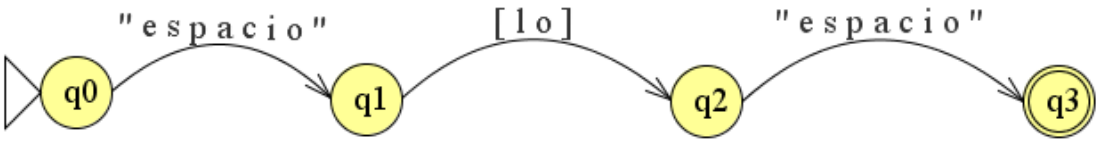
El regex correspondiente es: ^[articulo]$

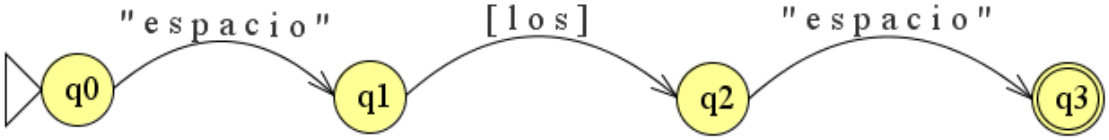


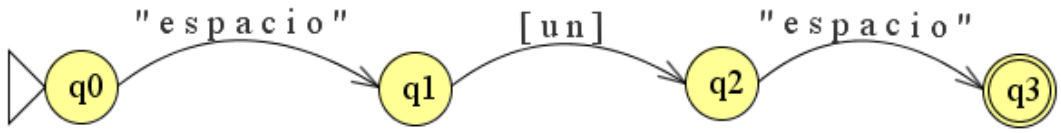


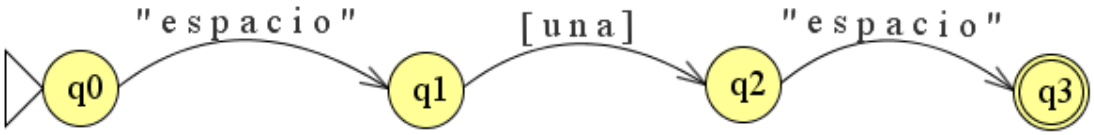



"espacio"
[las]
"espacio"
q0
q1
q2
q3


"espacio"
[lo]
"espacio"
q0
q1
q2
q3


"espacio"
[los]
"espacio"
q0
q1
q2
q3


"espacio"
[un]
"espacio"
q0
q1
q2
q3

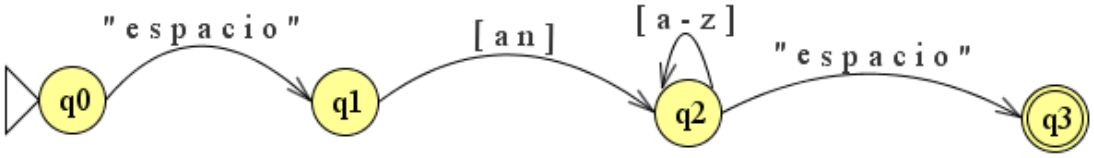

"espacio"
[una]
"espacio"
q0
q1
q2
q3

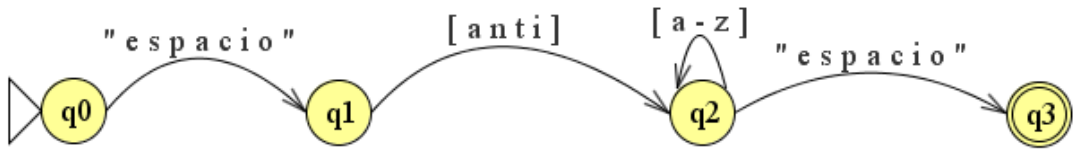

"espacio"
[unas]
"espacio"
q0
q1
q2
q3

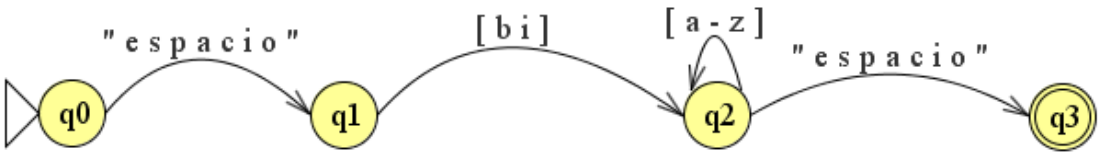

"espacio"
[unos]
"espacio"
q0
q1
q2
q3

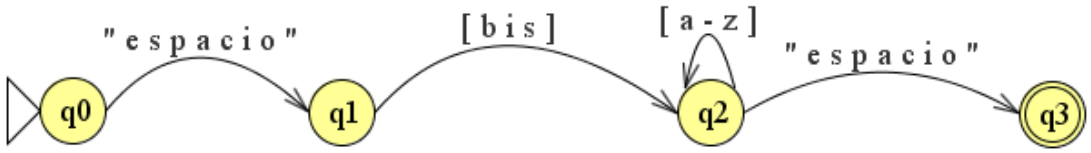
Prefijos:

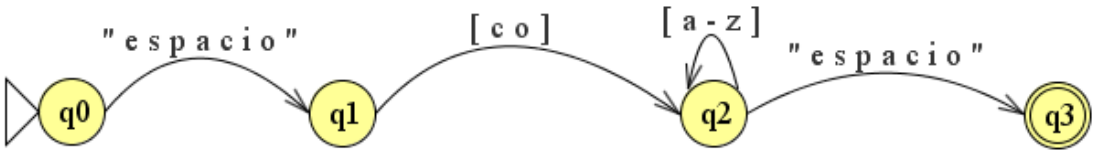
El regex correspondiente es: ^prefijo

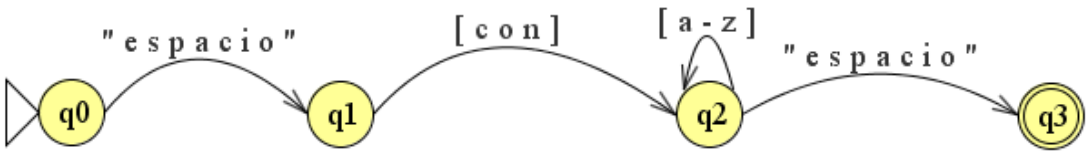


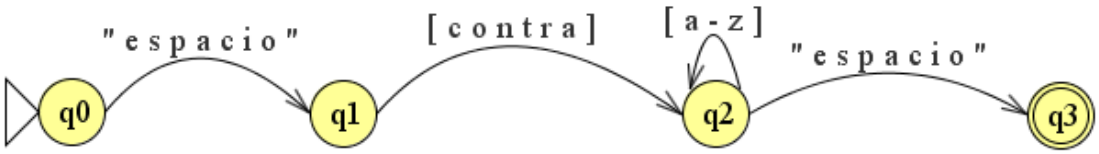

"espacio"
[anti]
[a-z]
"espacio"
q0
q1
q2
q3

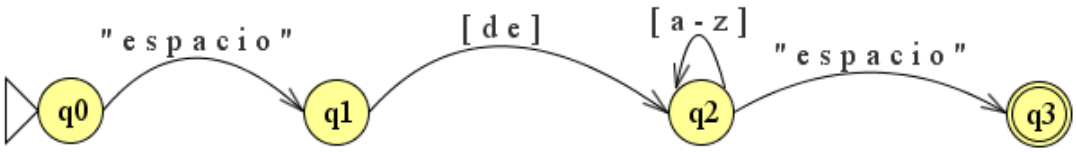

"espacio"
[bi]
[a-z]
"espacio"
q0
q1
q2
q3

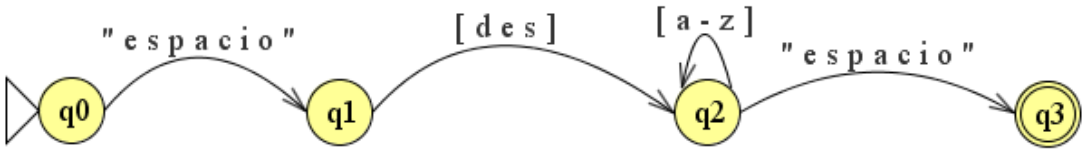

"espacio"
[bis]
[a-z]
"espacio"
q0
q1
q2
q3

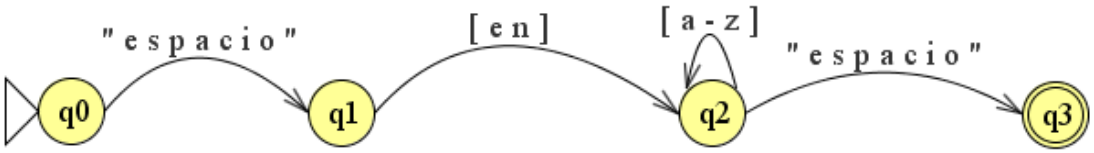

"espacio"
[co]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[con]
[a-z]
"espacio"
q0
q1
q2
q3

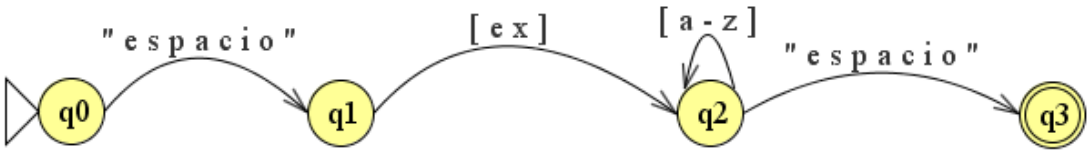

"espacio"
[contra]
[a-z]
"espacio"
q0
q1
q2
q3

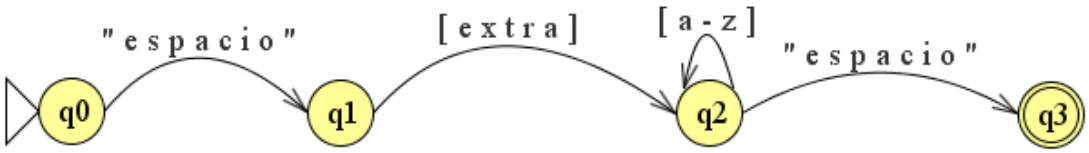

"espacio"
[de]
[a-z]
"espacio"
q0
q1
q2
q3


" e s p a c i o "
[d e s]
[a - z]
" e s p a c i o "
q0
q1
q2
q3

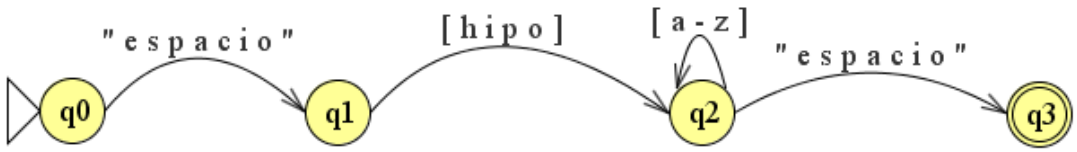

" e s p a c i o "
[e n]
[a - z]
" e s p a c i o "
q0
q1
q2
q3

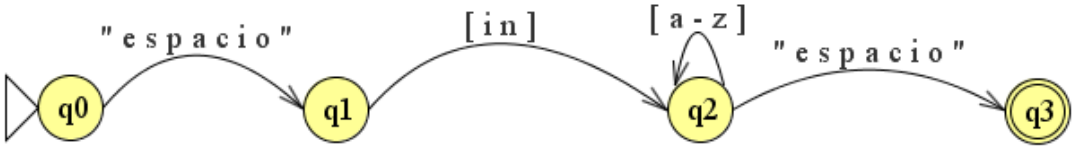

" e s p a c i o "
[e n t r e]
[a - z]
" e s p a c i o "
q0
q1
q2
q3


"espacio"
[ex]
[a-z]
"espacio"
q0
q1
q2
q3

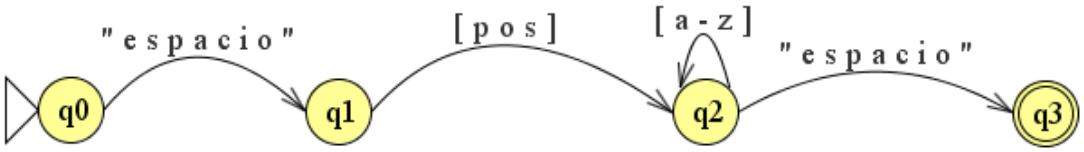

"espacio"
[extra]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[hiper]
[a-z]
"espacio"
q0
q1
q2
q3

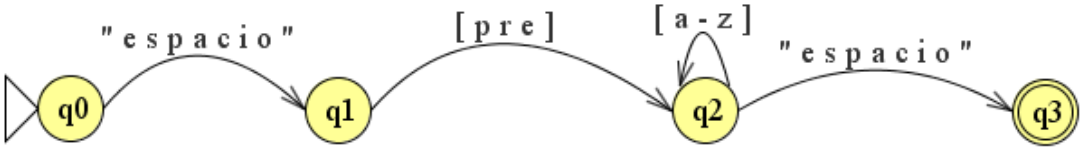

"espacio"
[hipo]
[a-z]
"espacio"
q0
q1
q2
q3

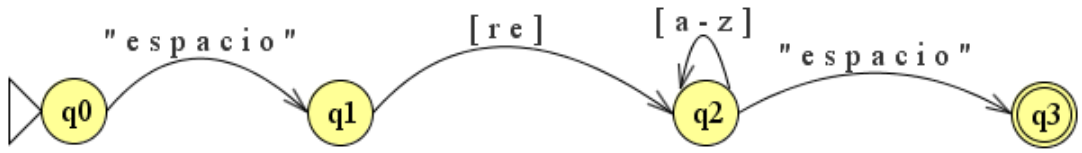

"espacio"
[in]
[a-z]
"espacio"
q0
q1
q2
q3

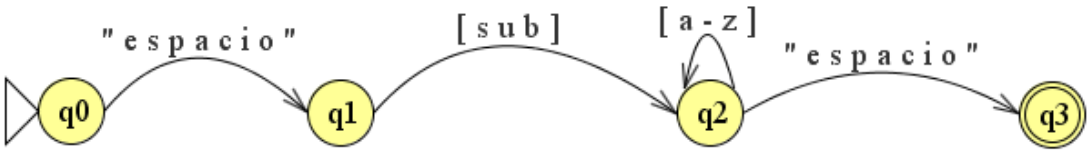

"espacio"
[inter]
[a-z]
"espacio"
q0
q1
q2
q3

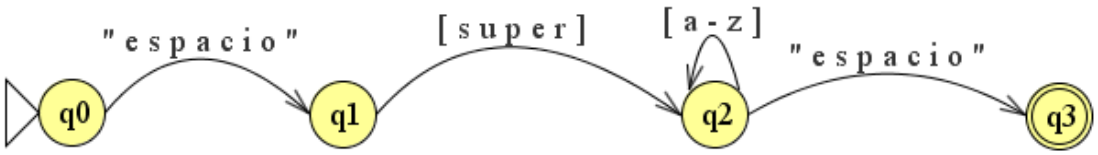

"espacio"
[pos]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[post]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[pre]
[a-z]
"espacio"
q0
q1
q2
q3

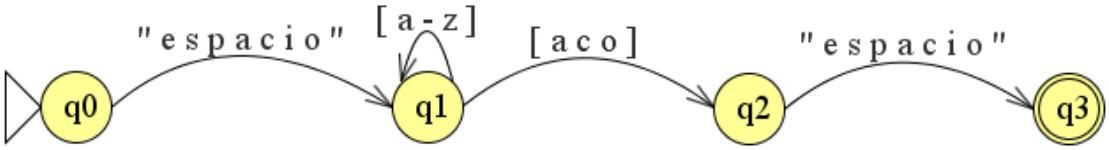

"espacio"
[re]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[sub]
[a-z]
"espacio"
q0
q1
q2
q3


"espacio"
[super]
[a-z]
"espacio"
q0
q1
q2
q3

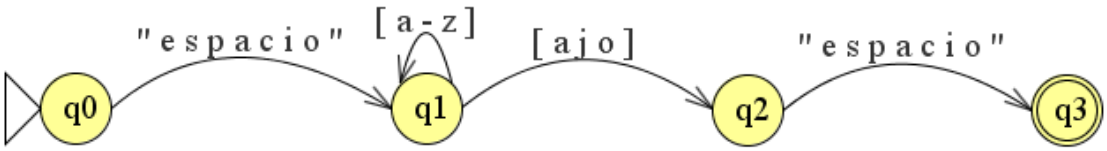
Sufijos:

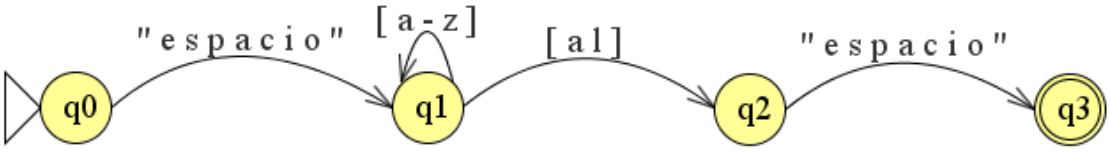
El regex correspondiente es: sufijo$



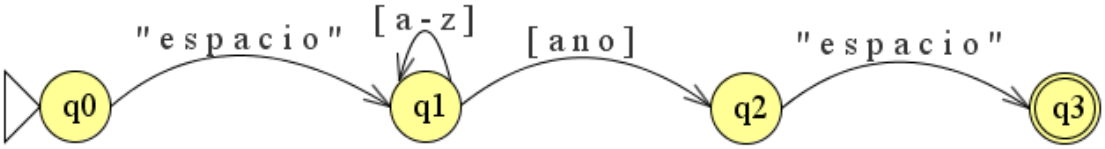

"espacio"
[a-z]
[ado]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[aje]
"espacio"
q0
q1
q2
q3

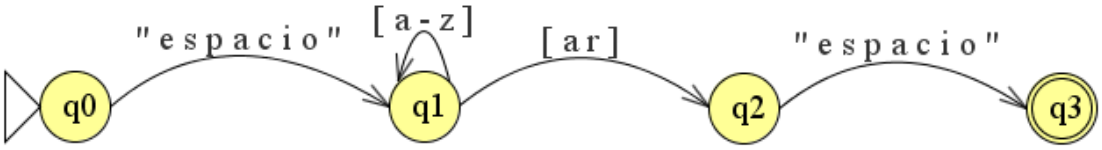

"espacio"
[a-z]
[ajo]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[al]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ancia]
"espacio"
q0
q1
q2
q3

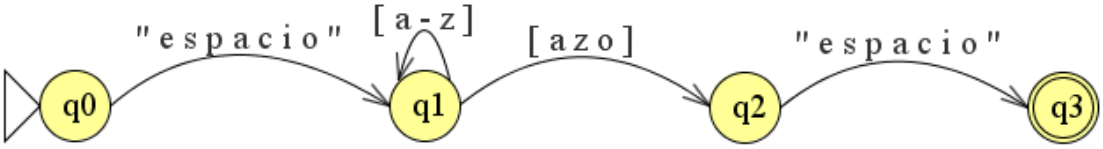

"espacio"
[a-z]
[ano]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ante]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ar]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ario]
"espacio"
q0
q1
q2
q3

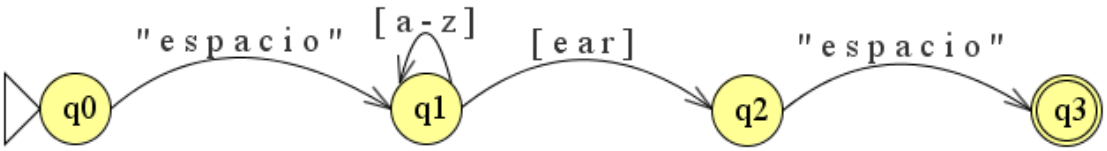

"espacio"
[a-z]
[astro]
"espacio"
q0
q1
q2
q3

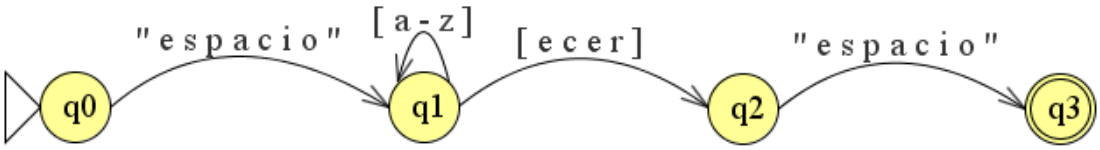

"espacio"
[a-z]
[azo]
"espacio"
q0
q1
q2
q3

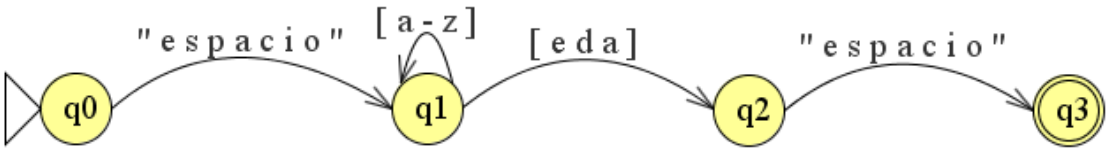

"espacio"
[a-z]
[cion]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[dad]
"espacio"
q0
q1
q2
q3

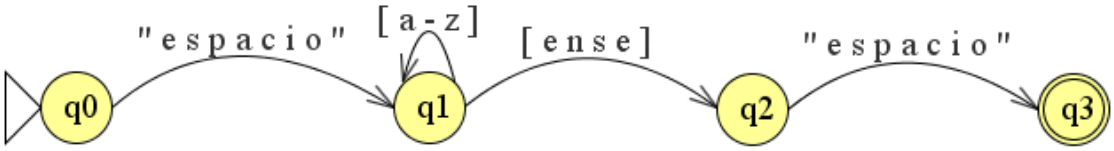

"espacio"
[a-z]
[dor]
"espacio"
q0
q1
q2
q3

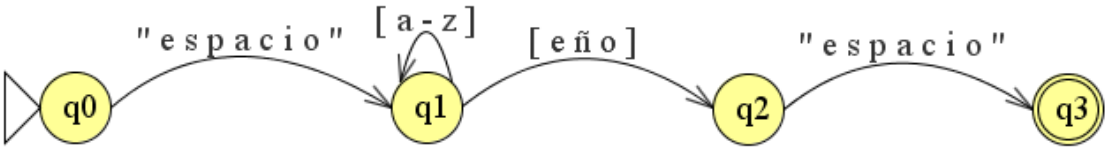

"espacio"
[a-z]
[ear]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ecer]
"espacio"
q0
q1
q2
q3

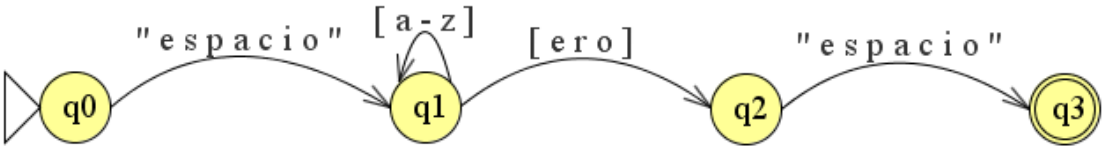

"espacio"
[a-z]
[eda]
"espacio"
q0
q1
q2
q3

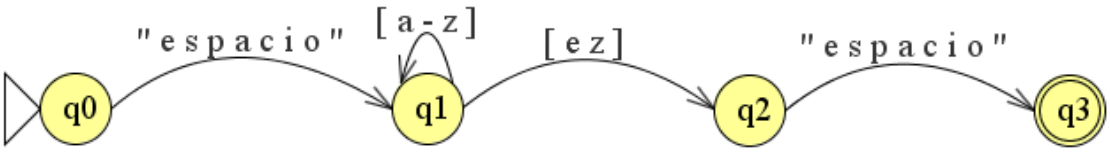

"espacio"
[a-z]
[encia]
"espacio"
q0
q1
q2
q3

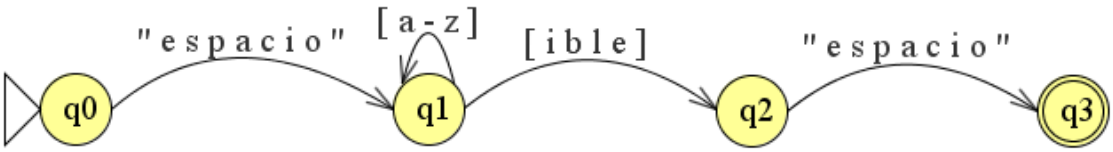

"espacio"
[a-z]
[ense]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[eño]
"espacio"
q0
q1
q2
q3

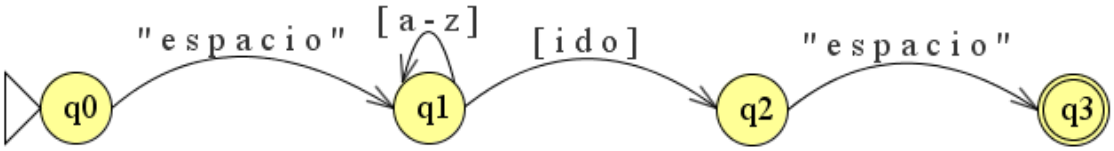

"espacio"
[a-z]
[eria]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ero]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ez]
"espacio"
q0
q1
q2
q3

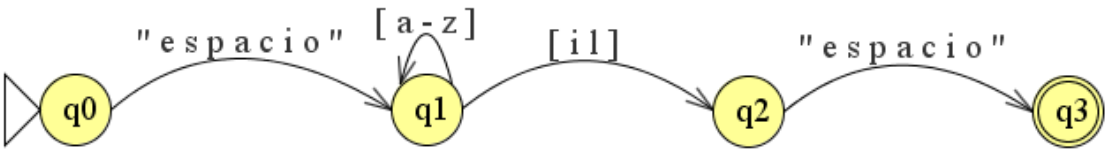

"espacio"
[a-z]
[ible]
"espacio"
q0
q1
q2
q3

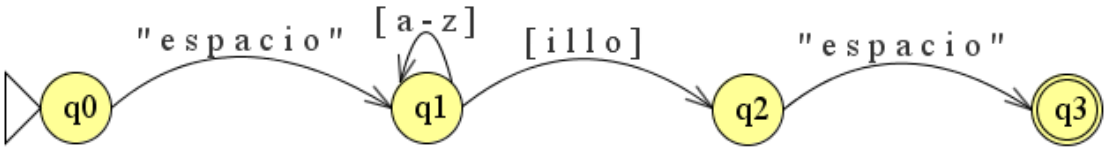

"espacio"
[a-z]
[ico]
"espacio"
q0
q1
q2
q3

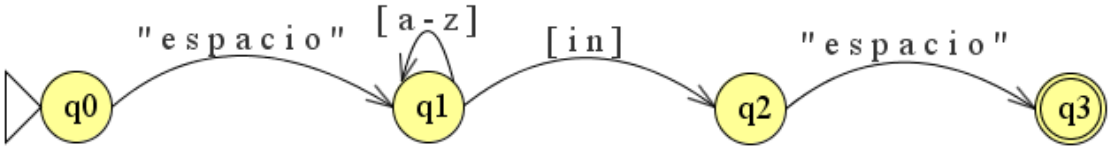

"espacio"
[a-z]
[ido]
"espacio"
q0
q1
q2
q3

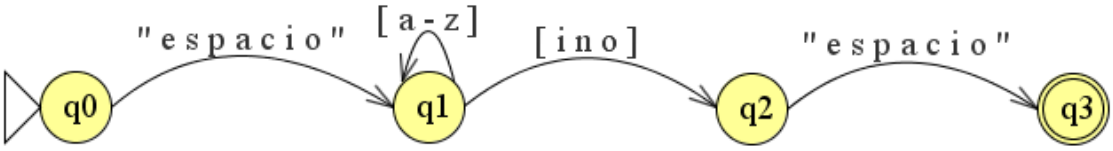

"espacio"
[a-z]
[iento]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ificar]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[il]
"espacio"
q0
q1
q2
q3

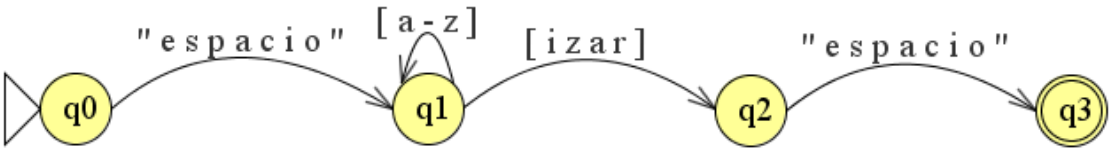

"espacio"
[a-z]
[illo]
"espacio"
q0
q1
q2
q3

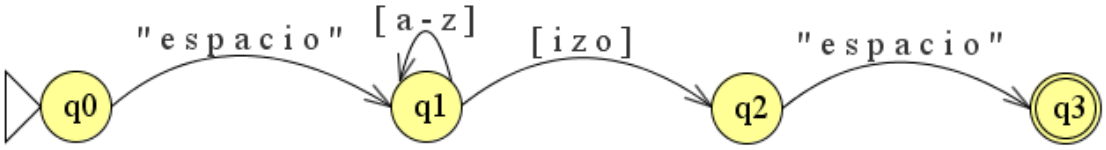

"espacio"
[a-z]
[in]
"espacio"
q0
q1
q2
q3

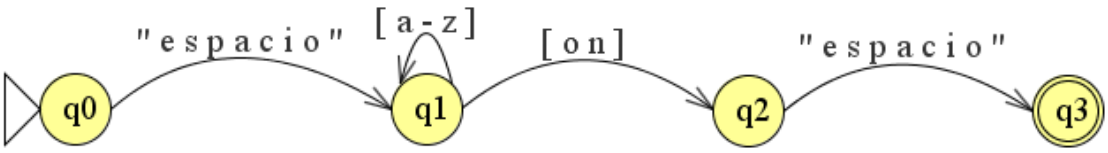

"espacio"
[a-z]
[ino]
"espacio"
q0
q1
q2
q3

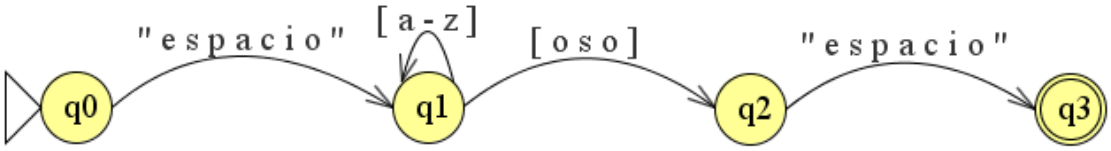

"espacio"
[a-z]
[ista]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ito]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[izar]
"espacio"
q0
q1
q2
q3

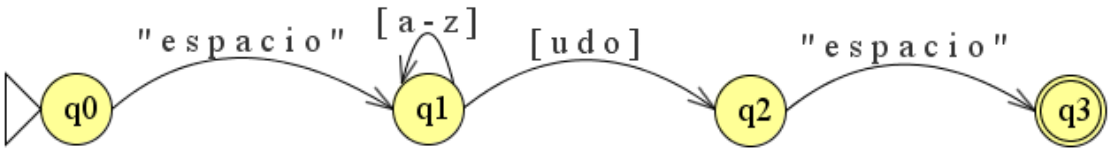

"espacio"
[a-z]
[izo]
"espacio"
q0
q1
q2
q3

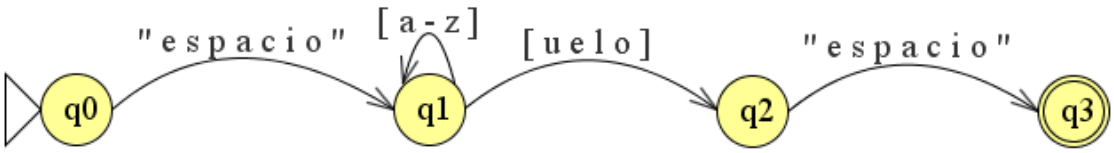

"espacio"
[a-z]
[on]
"espacio"
q0
q1
q2
q3

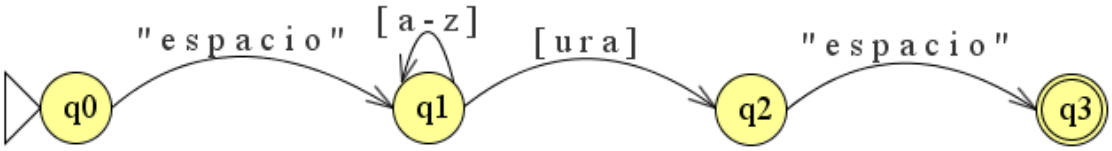

"espacio"
[a-z]
[oso]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ote]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ucho]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[udo]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[uelo]
"espacio"
q0
q1
q2
q3


"espacio"
[a-z]
[ura]
"espacio"
q0
q1
q2
q3