Lenguajes No Regulares

Problemas que los Autómatas No Resuelven.

Universidad de Cantabria

Esquema

Introducción

Todos los lenguajes no son regulares, simplemente hay que tener en cuenta que los lenguajes regulares son definidos por una expresión, cuando la intuición nos dice que se pueden definir lenguajes de una forma más compleja.

Lenguajes que no son Regulares

No hay ningún método que nos permita decidir si un lenguaje es regular o no, ya que depende de la descripción del lenguaje. Aunque si que tenemos diferentes herramientas que permiten probar que lenguajes específicos no son regulares.

Lenguajes que no son Regulares

Recordad que, usando las propiedades de las expresiones regulares, toda expresión regular se puede poner en **forma disyuntiva normal** y esto nos da una idea de como se puede generar palabras a partir de unas conocidas.

Lenguajes que no son Regulares

Recordad que en las expresiones regulares que generan lenguajes con infinitas palabras incluyen el operador estrella. Por lo que si tenemos una palabra, hay formas de generar infinitas nuevas.

Lenguajes que no son Regulares

Teorema (Pumping Lemma)

Sea L un lenguaje regular. Entonces, existe un número entero positivo $p \in \mathbb{N}$ *tal que para cada palabra* $\omega \in L$*, con* $|\omega| \geq p$ *existen* $x, y, z \in \Sigma^*$ *verificando los siguientes propiedades:*

- $y \neq \lambda$,
- $\omega = xyz$
- *Para todo* $\ell \in \mathbb{N}$*, las palabras* $xy^\ell z \in L$

UC UNIVERSIDAD DE CANTABRIA

Lenguajes que no son Regulares

Teorema (Stronger Pumping Lemma)

Sea L un lenguaje regular. Entonces, existe un número entero positivo $p \in \mathbb{N}$ tal que para cada palabra $\omega \in L$, con $|\omega| \geq p$ y cualesquiera $x, y, z \in \Sigma^*$ verificando los siguientes propiedades:

- $|y| \geq p$,
- $\omega = xyz$.

Se puede dividir y en $u, v, w \in \Sigma^*$ verificando las siguientes propiedades:

- $v \neq \lambda$,
- $y = uvw$
- *Para todo $\ell \in \mathbb{N}$, las palabras $xuv^\ell wz \in L$*

Lenguajes que no son Regulares

Probar que el siguiente lenguaje no es regular

$$L = \{0^n 1^n | n \in \mathbb{N}\}.$$

Idea del Método

- Asumir que es un lenguaje aceptado por un autómata.
- Tomar una palabra suficientemente larga α.
- Dividirla en tres partes x, y, z.
- Demostrar que cualquier división de la palabra y en tres partes no se puede bombear la parte del medio.

Prefijos

Definición (Prefijos)

Sea Σ un alfabeto finito y sea $L \subseteq \Sigma^*$ un lenguaje cualquiera. Definimos la siguiente relación de equivalencia sobre Σ^*: dados $x, y \in \Sigma^*$, $x \sim_L y$ si y solamente si:

$$\forall w \in \Sigma^*, xw \in L \Leftrightarrow yw \in L.$$

Myhill–Nerode

Teorema (Myhill–Nerode)

Si $L \subseteq \Sigma^*$ es un lenguaje aceptado por un autómata, entonces existen $x_1, \ldots, x_s \in \Sigma^*$ tal que cualquier otra palabra $\omega \in \Sigma^*$ es equivalente a alguna de las otras. El recíproco también es cierto.

Idea del Teorema

Estas palabras representaran los estados de un autómata muy especial que acepta el lenguaje. Las transiciones las definirá con que palabras sean equivalentes $ax_1, \ldots, ax_s$ para cada $a \in \Sigma$.

El Palíndromo

Los autómatas podrían haber sido buenos candidatos a procesos algorítmicos y uno estaría tentado a definir los problemas decisionales resolubles por un algoritmo como aquellos problemas que son funciones características de un lenguaje regular.

El Palíndromo

Un sencillo problema decisional *el Palíndromo* o, en buen catalán, el problema de la detección de los *“cap–i–cua”*, no será un lenguaje regular, como consecuencia del resultado de Myhill y Nerode anterior.

El Palíndromo

Dado un alfabeto finito Σ, y una palabra $\omega = x_1 \cdots x_n \in \Sigma^*$, denominamos el reverso de ω, ω^R a la palabra: $\omega^R = x_n \cdots x_1$. El lenguaje del Palíndromo es dado por

$$\mathcal{P} := \{x \in \Sigma^* \ : \ x^R = x\}.$$

Imposibilidad de Regularidad

Dada cualquier palabra $\omega \in \Sigma^*$ se tiene que $\omega\omega^R \in \mathcal{P}$. Claramente, pues

$$(\omega\omega^R)^R = (\omega^R)^R\omega^R = \omega\omega^R$$

Supongamos, entonces que $\mathcal{P}$ fuera un lenguaje regular y sea S un conjunto finito tal que $\Sigma^*/\sim_{\mathcal{P}} = \{[y] \ : \ y \in S\}$.

Imposibilidad de Regularidad

Sea $m := max\{| y | : \ y \in S\} + 1$, que existe por ser S un conjunto finito. Consideremos la palabra $x = 0^{[m]}1$ donde $0^{[m]}$ representa una lista de m 0's y los símbolos $\{0, 1\}$ se suponen dentro del alfabeto Σ. Ha de existir $y = y_1 \cdots y_r \in S$ tal que para todo $\omega \in \Sigma^*$

$$x\omega \in \mathcal{P} \Leftrightarrow y\omega \in \mathcal{P}$$

Imposibilidad de Regularidad

Ahora bien, tomando $x = 0^{[m]}1$, $xx^R \in \mathcal{P}$, luego $yx^R \in \mathcal{P}$. Pero esto implica

$$yx^R = (yx^R)^R = xy^R$$

Reescribamos esta identidad:

$$yx^R = y_1 \cdots y_r 10^{[m]} = 0^{[m]} 1 y_r \cdots y_1 = xy^R$$

Deducimos (dado que $r \leq m$ que $y_1 = \cdots = y_r = 0$) que $m + 1 = r + 1$, ya que la palabra solo tiene un 1.
Esto es una contradicción.

Otros ejemplos

Ejemplo

Los siguientes son también ejemplos de lenguajes no regulares:

- $\Sigma = \{0, 1\}$ *y el lenguaje L dado por la condición el número de* 1*'s es mayor que el número de* 0*'s.*
- *Para el mismo alfabeto el lenguaje:*

$$L := \{0^{[m]}1^{[m]} \ : \ m \in \mathbb{N}\}$$

- *Para el alfabeto* $\Sigma = \{0, 1, \ldots, 9\}$ *sea* $\pi \subseteq \Sigma^*$ *el lenguaje formado por las palabras que son prefijos de la expansión decimal de* $\pi \in \mathbb{R}$*, es decir:*

$$L := \{3, 31, 314, 3141, 31415, \ldots\}$$