Lenguajes (gramáticas y autómatas)

Elvira Mayordomo

Universidad de Zaragoza

19 de septiembre de 2014

Definiciones básicas

- Alfabeto
- Cadena o palabra
- Lenguaje

Alfabeto

- Un **alfabeto** es un conjunto finito
- Un **símbolo o letra** es un elemento del alfabeto

Nuestros ejemplos suelen ser pequeños:
$\Sigma = \{a, b\}$
$\Sigma = \{0, 1\}$
Pero pueden ser de otro tipo:
$\Sigma = \{Rojo, Negro, Azul, Verde\}$
$\Sigma = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$
$\Sigma = \{el, la, los, las, un, una, unos, unas\}$

Cadena

- Una **cadena** es una secuencia finita de letras

Por ejemplo, si el alfabeto es $\Sigma = \{a, b\}$, algunas palabras
$abbaaa$
$bbbbbbbbb$

- La **longitud** de una cadena es el número de letras

$w = a_1 \ldots a_n$, la longitud de w es $|w| = n$
La cadena de longitud 0 o cadena vacía es ϵ
¡Cuidado!, también se usa $|A|$ para denotar el número de elementos del conjunto A

Cadena: prefijo y sufijo

- **Prefijo** de una cadena: Caracteres consecutivos de la cadena empezando por el principio
- **Sufijo** de una cadena: Caracteres consecutivos de la cadena terminando al final

Prefijos de abb: ϵ, a, ab, abb
Sufijos de abb: ϵ, b, bb, abb

Lenguaje

- Un lenguaje es un conjunto de cadenas

Por ejemplo, para $\Sigma = \{a, b\}$.
$L = \{aa, bbbb, \epsilon\}$
$A = \{\epsilon\}$
$B = \{w \mid |w| = 3\}$
$C = \{a^n b^n \mid n \in \mathbb{N}\}$

Operaciones con cadenas: concatenación

$v = a_1 \dots a_n$
$w = b_1 \dots b_m$

- La concatenación es $vw = a_1 \dots a_n b_1 \dots b_m$
- Si hace falta se pone un punto $v \cdot w$

$aaab \cdot bbb = aaabbbb$
$baba \cdot \epsilon = baba$

- La longitud de la concatenación es la suma de las longitudes

$$|uv| = |u| + |v|$$

Operaciones con cadenas: Potencia

- $w^n = \underbrace{w \dots w}_{n}$
- Por definición $w^0 = \epsilon$

Ejemplo: $w = babb$, $w^0 = \epsilon$, $w^2 = babbbabb$

Operaciones con cadenas: Reverso

- $w^R = a_n \dots a_1$

$(abbb)^R = bbba$

Operaciones con lenguajes: Reverso

- Dado un lenguaje L, $L^R = \left\{ w^R \mid w \in L \right\}$.

Ejemplos: $L = \{aa, ba, bba\}$, $L = \{w \mid w \text{ termina en } b\}$.

Operaciones con lenguajes: Concatenación

- Dados dos lenguajes A, B,

$$A \cdot B = \{uv \mid u \in A, v \in B\}$$

Ejemplos: $A = \{b, ba\}$, $B = \{a\}$

$A = \{w \mid w \text{ empieza en } a\}$. $B = \{w \mid w \text{ termina en } b\}$.
Propuesto: Demostrar quién es $A \cdot B$.

Operaciones con lenguajes: Potencia

- $A^n = \underbrace{A \dots A}_{n}$
- De otra forma, $A^n = \{w_1 \dots w_n \mid w_i \in A\}$

Operaciones con lenguajes: estrella de Kleene

- $L^* = \bigcup_{n=0}^{\infty} L^n = L^0 \cup L^1 \cup L^2 \dots$
- Σ^* es el lenguaje formado por todas las cadenas

Operaciones con lenguajes: $+$

- $L^+ = \bigcup_{n=1}^{\infty} L^n = L^1 \cup L^2 \dots$

Propuesto: Averiguar cuándo $\epsilon \in L^+$.

Operaciones con lenguajes: unión, intersección, diferencia de conjuntos

- Son las operaciones que ya conocemos de los conjuntos.

Definición

Definición

Un conjunto A es *contable* si existe una función inyectiva $f : A \to \mathbb{N}$.

- Intuitivamente, un conjunto contable **se puede contar** ya que corresponde a un subconjunto de los naturales.

Cómo demostrarlo

Una primera propiedad básica es que se puede demostrar que un conjunto es contable a partir de otro contable.

Theorem

Si B es contable y existe una función inyectiva $g : A \to B$ entonces A es contable.

Ejercicios

- Demostrar que, dado Σ un alfabeto, Σ^* es contable.

Ejercicios

- En informática nos interesa especialmente "codificar" conjuntos, es decir, representar cada elemento como una cadena. Veamos que todo conjunto que se puede codificar es contable.
- Demostrar que si existe una función inyectiva $g : A \to \Sigma^*$ entonces A es contable.

¿Algún conjunto no contable?

Más adelante veremos algún ejemplo de conjunto no contable (por ejemplo los reales). ¿Se te ocurre alguna forma de codificar TODOS los números reales?

Volviendo a lenguajes ... ¿por qué nos interesan los lenguajes?

- Principalmente buscaremos programas que nos digan si una cadena está en un lenguaje (es decir, que reconocen el lenguaje)
  Ejemplo. $L = \{a^n b^n\}$
- Queremos saber cómo de difícil es un programa que reconoce un lenguaje
- ¿Para qué?

Los lenguajes humanos

Inglés, Español, Cantonés, ...
¿Qué frases de las siguientes son ciertas, con sentido, gramaticalmente correctas?

- vgrlum qp#dn aoiuiui brubrubrubru 3jc6r
- perro deberes comió mis. Mi
- El hierro es más denso que el glup.
- Two is less than three.
- El libro de este curso tiene 10 páginas.

Los lenguajes de programación

C, Java, Python, Prolog, Pascal, ...
¿Cuándo es un programa:

- sintácticamente correcto?
- compilable?
- sin errores de ejecución fatales?
- sin bloqueos o bucles infinitos?
- una implementación correcta de su espacificación?

Un lenguaje es un problema decisional

Un **problema decisional** es el que tiene sólo 2 respuestas posibles:

- Dado un número natural n, ¿es n primo?
- ¿Este programa es compilable?
- ¿Esta frase f es correcta en este idioma I?

Cada problema decisional corresponde a reconocer un lenguaje
Los problemas decisionales son los más simples de analizar, y su análisis nos ayuda a reconocer problemas generales

Lenguajes, gramáticas y autómatas

- Estudiaremos máquinas que reconozcan los distintos lenguajes (cómo de difícil es un programa que reconoce si una cadena está en un lenguaje)
  Autómatas finitos, autómatas de pila, máquinas de Turing, ordenadores
- Estudiaremos generadores de las cadenas de un lenguaje
  Expresiones regulares, gramáticas

Las dos preguntas importantes

- ¿Qué puede y no puede hacer un ordenador?
- ¿Cuál es el mecanismo más sencillo para resolver cada problema?

La Teoría de la Computación busca la respuesta a esas dos preguntas.

Bibliografía

- Sipser (2a edición), páginas 13 y 14 (en sección 0.2) .
- Kelley, capítulo 1.