Alfabetos, palabras y lenguajes

Enric Sesa i Nogueras

P03/75015/01047

Universitat Oberta de Catalunya

Índice

Introducción

Si nos adentramos en el corazón de la informática teórica, sin duda descubriremos los lenguajes formales. Y es que, en el fondo, cualquier problema decisional se puede reducir al problema de la pertenencia de una palabra a un lenguaje. Y muchos problemas que no tienen un aspecto decisional también se pueden reducir o relacionar, con más o menos dificultad, con problemas que tienen que ver con reconocedores y/o generadores de lenguajes.

Un problema decisional...

... es aquel que se plantea en términos de una pregunta que sólo puede tener dos respuestas: cierto o falso.

Sin embargo, no es necesario ir al corazón de la informática teórica para descubrir los lenguajes formales y sus operaciones. Los podemos encontrar en herramientas tan informáticamente cotidianas como los lenguajes de programación y sus compiladores, en la utilidad de búsqueda de palabras de cualquier editor de textos moderno, en los intérpretes de comandos de la mayoría de los sistemas operativos y en algunas de las utilidades de muchos entornos basados en UNIX.

En este módulo didáctico estudiaremos algunos de los conceptos más básicos de la teoría de lenguajes formales. No se trata de una materia que tenga gran interés por sí misma, pero constituye el fundamento, indispensable, de un edificio rico y variado que será parcialmente revelado en otros módulos.

El estudio de una materia eminentemente teórica puede comportar algunas dificultades en los momentos iniciales. Por esta razón, encontraréis cada concepto ilustrado con algún ejemplo. No paséis de página hasta haberlos entendido todos y cada uno.

Objetivos

Los materiales didácticos de este módulo contienen las herramientas fundamentales para que el estudiante alcance los siguientes objetivos:

1. Familiarizarse con los conceptos de alfabeto, palabra y lenguaje.

2. Entender el significado de los operadores habitualmente aplicados a las palabras y conocer sus propiedades más destacadas.

3. Entender el significado de los operadores habitualmente aplicados a los lenguajes y conocer sus propiedades más destacadas.

4. Conocer la existencia de lenguajes distinguidos, como $\emptyset$, Σ^* o Σ^+, y su significado.

5. Entender la naturaleza conjuntista de los lenguajes y ser capaz de razonar, en casos simples, sobre la pertenencia o no de una palabra a un lenguaje.

6. Saber interpretar la definición formal de un lenguaje.

1. Elementos básicos: alfabetos, palabras y lenguajes

1.1. Alfabetos: conjuntos de símbolos

Un **alfabeto** es un conjunto finito y no vacío cuyos elementos se denominan **símbolos**.

Para designarlo se utilizan letras mayúsculas del alfabeto griego (especialmente, Σ y Γ).

Algunos ejemplos de alfabetos son los siguientes:

- $\{0, 1\}$, el alfabeto binario.
- $\{a, b, c, ..., x, y, z\}$, el alfabeto latino de las letras minúsculas.
- $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, el alfabeto de los dígitos.
- $\{\wedge, \rightarrow, \neg, \vee, (,), P, Q, R, ...\}$, el alfabeto de la lógica de enunciados.

Alfabeto con dos símbolos

A pesar de que cotidianamente se utilizan alfabetos de más de dos símbolos (las letras, los diez dígitos, etc.), cualquier alfabeto con dos símbolos es suficiente para representar todo tipo de información.

Conviene señalar que en este módulo didáctico, en la mayoría de los casos usaremos el alfabeto $\{a, b\}$ y, de manera esporádica, usaremos el alfabeto $\{0, 1\}$. !

1.2. Palabras: secuencias de símbolos

Una **palabra** es una secuencia finita de símbolos de un alfabeto.

Cuando se quiere dejar claro que los símbolos que se utilizan son los de un alfabeto determinado, se habla de **palabras sobre un alfabeto**.

A continuación mostraremos algunos ejemplos de palabras sobre un alfabeto. Son los siguientes:

a) Si se considera el alfabeto $\{a, b\}$, las palabras siguientes son sobre este alfabeto: *aba, bab, bbbbaaaabbbaaabbaaba, a* (una palabra de un único símbolo) y *b* (otra palabra de un único símbolo).

b) Como ejemplos de palabras sobre $\{0, 1\}$, podemos considerar *0, 1, 00, 01,10,* y *11.*

c) Como ejemplos de palabras sobre $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, podemos considerar *1, 12, 23, 34, 45* y *980.*

1.2.1. La palabra sin símbolos: λ

Hay una palabra que es una secuencia vacía de símbolos: se representa con la letra griega λ y, a menudo, se la llama **palabra vacía**.

Dado que no tiene ningún símbolo, λ es una palabra sobre cualquier alfabeto.

1.2.2. Subpalabras, prefijos y sufijos

Las subsecuencias de símbolos consecutivos de una palabra reciben el nombre de **subpalabras** de esta palabra. A menudo se usan las palabras **factor** o **infijo**.

Por ejemplo, la palabra *bba* contiene el siguiente conjunto de subpalabras (o factores): {λ, *a* , *b, bb, ba, bba*}.

Observad que λ (la palabra vacía) y *bba* (la palabra entera) son consideradas subpalabras de la palabra *bba*. Estas subpalabras se llaman **subpalabras impropias**. Todas las subpalabras distintas de λ y de la propia palabra se denominan **subpalabras propias.**

Las subpalabras del principio de una palabra se llaman **prefijos**, y las subpalabras del final se llaman **sufijos**. La palabra vacía (λ) y la palabra entera se consideran prefijos y sufijos –impropios– de cualquier palabra.

Por ejemplo, los prefijos de la palabra *bbaab* son {λ, *b, bb, bba, bbaa, bbaab*} y sus prefijos propios son {*b, bb, bba, bbaa*}. De esta misma palabra, los sufijos son {λ, *b, ab, aab, baab, bbaab*} y el conjunto de sufijos propios es {*b, ab, aab, baab*}.

1.3. Lenguajes: conjuntos de palabras

Un **lenguaje** es un conjunto de palabras sobre un alfabeto determinado.

Para designarlos se usa la letra *L*, con subíndices, si es necesario, y otras letras mayúsculas del alfabeto latino.

Algunos ejemplos de lenguajes sobre el alfabeto $\{a, b\}$ son los siguientes:

- $L_1 = \{a, aa, aaa, aaaa\}$.
- $L_2 = \{a, b, aa, ab, ba, bb\}$.
- $L_3 = \{aabb\}$ (lenguaje con una sola palabra).
- $L_4 = \{\lambda\}$ (lenguaje con una sola palabra).
- $L_5 = \{\} = \varnothing$ (lenguaje sin ninguna palabra)*.

* No es lo mismo un lenguaje sin palabras que un lenguaje que sólo contiene la palabra sin símbolos.

Un lenguaje no debe ser, necesariamente, finito. Los siguientes son ejemplos de lenguajes con un número infinito de palabras:

- Lenguaje de todas las palabras sobre $\Sigma = \{0, 1\}$ que comienzan por el símbolo *1*.
- Lenguaje de todas las palabras sobre $\Sigma = \{a, b\}$ que contienen tantas *a* como *b*.
- El lenguaje de todas las palabras sobre $\Sigma = \{a, b, c\}$ (el conjunto de todas las palabras que se pueden construir utilizando el alfabeto $\{a, b, c\}$).

2. Operaciones sobre palabras

2.1. La concatenación: una palabra tras otra

Concatenar dos palabras significa construir una palabra nueva añadiendo los símbolos de la segunda tras los símbolos de la primera.

El operador de concatenación de palabras se representa con el símbolo $\cdot$.

Algunos ejemplos de concatenación de palabras son los siguientes:

- $aaa \cdot bbb = aaabbb$.
- $abba \cdot baabbb = abbabaabbb$.
- $aba \cdot \lambda = aba$.

2.1.1. Propiedades de la concatenación de palabras

Las propiedades de la concatenación de palabras que deberemos tener presentes son las siguientes: !

Elisión

Es muy frecuente que el símbolo $\cdot$ se elida, como pasa con el producto de números. De esta manera, si w_1 y w_2 son dos palabras cualquiera, entonces $w_1w_2 = w_1 \cdot w_2$.

1) La concatenación no tiene la propiedad conmutativa porque, en general, $w_1w_2 \neq w_2w_1$.

2) La concatenación tiene la propiedad asociativa, ya que $(w_1w_2)w_3 = w_1(w_2w_3)$. Esto permite escribir $w_1w_2w_3$ sin que haya ningún tipo de ambigüedad.

3) La palabra vacía, λ, es el elemento neutro de la concatenación, ya que $w\lambda = \lambda w = w$.

2.1.2. Concatenación de una palabra consigo misma

La concatenación de una palabra consigo misma se suele representar usando la notación exponencial, de manera que $w^2 = ww$, $w^3 = www$ etc. Notad que $w^0 = \lambda$.

2.1.3. Concatenación de una palabra y un símbolo

La concatenación de una palabra y un símbolo se denota de la misma manera que la concatenación de dos palabras. Así, si a es un símbolo y w es una palabra, entonces aw es la palabra resultante de añadir el símbolo a a la izquierda de la palabra w.

2.2. La longitud: el número de símbolos de una palabra

Con la notación $|w|$ se designa la **longitud** (número de símbolos) de la palabra w.

Será interesante que recordéis los dos puntos siguientes:

1) $|w_1 \cdot w_2| = |w_1| + |w_2|$.
2) $|w| = 0 \Leftrightarrow w = \lambda$.

Ejemplos de longitud

- $|010| = 3$.
- $|a| = 1$.
- $|\lambda| = 0$.

2.3. El número de ocurrencias de un símbolo

La notación $|w|_x$ denota el número de apariciones (**número de ocurrencias**) del símbolo x en la palabra w.

Como ejemplos de número de ocurrencias, considerad los siguientes:

- $|ababb|_a = 2$.
- $|aaab|_c = 0$.

2.4. La inversión: invertir una palabra

La **inversión** es una operación que consiste en escribir al revés una palabra dada. La palabra resultante de la inversión se denomina **inversa.** Si w es una palabra cualquiera, entonces w^R denota su inversa.

Ejemplos de inversión:

- $(ab)^R = ba$.
- $(010011)^R = 110010$.
- $\lambda^R = \lambda$.
- $(abba)^R = abba$.

La inversión puede ser definida muy fácilmente de manera recursiva:

$$w^R = \begin{cases} \lambda; & \text{si } w = \lambda, \\ a_n(a_1 \ldots a_{n-1})^R; & \text{si } w = a_1 \ldots a_n (a_i \in \Sigma) \end{cases}$$

Fijaos que, en general, se cumple que $(w_1 \cdot w_2)^R \neq w_1^R \cdot w_2^R$. Pero en cambio, $(w_1 \cdot w_2)^R = w_2^R \cdot w_1^R$.

2.4.1. Palíndromos

Cuando una palabra es igual que su inversa ($w = w^R$), se dice que es un **palíndromo**.

Las siguientes reglas definen, recursivamente, qué es un palíndromo:

1) λ es un palíndromo.

2) Las palabras formadas por un solo símbolo son palíndromos.

3) Si a es un símbolo cualquiera y w es un palíndromo, entonces tenemos que la palabra $awa = (a \cdot w \cdot a)$ también es un palíndromo.

4) Cualquier palíndromo se obtiene de la aplicación de las tres reglas anteriores.

3. Operaciones sobre lenguajes

3.1. Operaciones conjuntistas

Dado que los lenguajes son conjuntos, todas las **operaciones conjuntistas** habituales son aplicables a los lenguajes. Estas operaciones son las siguientes:

1) La unión ($\cup$).

2) La intersección ($\cap$).

3) La complementación (c).

4) La diferencia (–).

Recordad que algunas de las **propiedades básicas** de estas operaciones son las que exponemos a continuación:

a) $(L^c)^c = L$.

b) $(L_1 \cup L_2)^c = L_1^c \cap L_2^c$.

c) $(L_1 \cap L_2)^c = L_1^c \cup L_2^c$.

d) $L_1 - L_2 = L_1 \cap L_2^c$.

3.2. La concatenación

La **concatenación de dos lenguajes**, L_1 y L_2, es otro lenguaje que contiene todas las palabras que se pueden construir concatenando una palabra de L_1 con una palabra de L_2. El operador de la concatenación de lenguajes se representa, también, con el símbolo ·. Formalmente, tenemos que:

$$L_1 \cdot L_2 = \{x \cdot y \mid x \in L_1 \wedge y \in L_2\}.$$

Algunos ejemplos de concatenación de lenguajes son los siguientes:

- Si $L_1 = \{a, aa\}$ y $L_2 = \{b, bb\}$, entonces $L_1L_2 = \{ab, abb, aab, aabb\}$.
- Si $L_1 = \{aa, ab, \lambda\}$ y $L_2 = \{a, ba\}$, entonces $L_1L_2 = \{aaa, aaba, aba, abba, a, ba\}$.
- Si $L_1 = \{\lambda\}$ y $L_2 = \{abbb, babb\}$, entonces $L_1L_2 = \{abbb, babb\}$.
- Si $L_1 = \{b, ba, abb\}$ y $L_2 = \{ba, ab\}$, entonces $L_1L_2 = \{bba, bab, baba, baab, abba, abbab\}$.

En la concatenación de lenguajes...

... es habitual la elisión del símbolo •, al igual que en la concatenación de palabras.

3.2.1. Propiedades de la concatenación de lenguajes

Es importante que tengáis presentes las siguientes propiedades:

1) La concatenación de lenguajes no es conmutativa, ya que en general:

$$L_1L_2 \neq L_2L_1.$$

2) La concatenación de lenguajes es asociativa, porque $(L_1L_2)L_3 = L_1(L_2L_3)$. Esta propiedad permite escribir $L_1L_2L_3$, sin que esto represente ninguna ambigüedad.

3) El elemento neutro de la concatenación de lenguajes es $\{\lambda\}$, el lenguaje que sólo contiene la palabra vacía, ya que $L\{\lambda\} = \{\lambda\}L = L$.

4) La concatenación de lenguajes es distributiva respecto de la unión, ya que:

a) $(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3$.
b) $L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3$.

5) La concatenación de lenguajes no es distributiva respecto de la intersección, porque, en general, se cumple que:

a) $(L_1 \cap L_2)L_3 \neq L_1L_3 \cap L_2L_3$.
b) $L_1(L_2 \cap L_3) \neq L_1L_2 \cap L_1L_3$.

No obstante, sí se dan las siguientes inclusiones:

a) $(L_1 \cap L_2)L_3 \subset L_1L_3 \cap L_2L_3$.
b) $L_1(L_2 \cap L_3) \subset L_1L_2 \cap L_1L_3$.

Distinción entre λ y $\varnothing$

No deben confundirse $\{\lambda\}$ (el lenguaje que sólo contiene la palabra vacía) con $\varnothing$ (el lenguaje que no contiene ninguna palabra). Notad que $L\{\lambda\} = \{\lambda\}L = L$, mientras que $L\varnothing = \varnothing L = \varnothing$.

A menudo, $\varnothing$ se denomina *lenguaje vacío*.

3.2.2. Concatenación de un lenguaje consigo mismo

La concatenación de un lenguaje consigo mismo también puede representarse utilizando la notación exponencial, de manera que L^2 corresponda a LL, L^3 a LLL, etc.

Notad que $\forall L\ (L^0 = \{\lambda\})$ y que, por consiguiente, $\varnothing^0 = \{\lambda\}$.

3.2.3. Concatenación de una palabra y un lenguaje

Las concatenaciones de la forma $\{w\}L$ y $L\{w\}$ se pueden abreviar como wL y Lw, y denotan el resultado de añadir una palabra w delante y detrás de cada una de las palabras de L.

Las concatenaciones de la forma aL y La, en las que a es un símbolo, también son aceptables, y denotan el resultado de añadir el símbolo a delante y detrás de cada palabra de L.

3.3. La inversión

La **operación de inversión** también puede extenderse a los lenguajes. Así, si L es un lenguaje, L^R representa su inverso, que no es más que el lenguaje formado por los inversos de las palabras de L. Es decir:

$$L^R = \{w^R \mid w \in L\}.$$

3.4. Cierres

Existen dos tipos de cierres: el cierre de Kleene y el cierre positivo de Kleene.

3.4.1. El cierre de Kleene (la estrella de Kleene): todas las concatenaciones y λ

El **cierre de Kleene de un lenguaje** es la unión de todas las potencias de este lenguaje. También se denomina **estrella de Kleene**. Si L es un lenguaje, su cierre de Kleene se representa mediante L^*, y se define formalmente como:

$$L^* = \bigcup_{n \geq 0} L^n = L^0 \cup L^1 \cup L^2 \ldots$$

Pese a la aparente complejidad de la definición, el concepto es simple: el cierre de Kleene de un lenguaje es la unión de {λ} con el conjunto de todas las palabras que se pueden formar concatenando entre sí palabras de este mismo lenguaje.

Por ejemplo, si L= {*a, ba*}, entonces:

- $L^0 = \{\lambda\}$.
- $L^1 = L = \{a, ba\}$.
- $L^2 = \{aa, aba, baa, baba\}$.
- $L^3 = \{aaa, aaba, abaa, ababa, baaa, baaba, babaa, bababa\}$.
- $L^4 = \ldots$

Así, $L^* = \{\lambda\} \cup \{a, ba\} \cup \{aa, aba, baa, baba\} \cup \{aaa, aaba, abaa, ababa, baaa, baaba, babaa, bababa\} \cup \ldots$

Podemos considerar otros ejemplos, como los que presentamos a continuación:

- Si $L = \{b\}$, entonces $L^* = \{\lambda, b, bb, bbb, bbbb, bbbbb, \ldots\}$.
- Si $L = \{aa\}$, entonces $L^* = \{\lambda, aa, aaaa, aaaaaa, aaaaaaaa, aaaaaaaaaa, \ldots\}$.

- Si $L = \{a, bb\}$, entonces $L^* = \{\lambda, a, bb, aa, bbbb, abb, bba, aaa, bbbbbb, abba, abbbb, aabb, ...\}$.

Excepto para $L = \{\lambda\}$ y $L = \varnothing$,...

... L^* es siempre un lenguaje con un número infinito de palabras.

Es importante que recordéis que $\forall L\ (\lambda \in L^*)$. Esto es así porque tenemos que $\forall L\ (L^0 = \{\lambda\})$.

3.4.2. El cierre positivo de Kleene: todas las concatenaciones posibles

El **cierre positivo de un lenguaje *L*** se representa por L^+ y se define formalmente como:

$$L^+ = \bigcup_{n>0} L^n = L^1 \cup L^2 \cup L^3 ...$$

Notad que L^* y L^+ sólo difieren en la presencia de L^0 en la definición del primero.

Algunos ejemplos de cierres positivos de un lenguaje son los que mostramos a continuación:

- Si $L = \{b\}$, entonces $L^+ = \{b, bb, bbb, bbbb, bbbbb, ...\}$.
- Si $L = \{aa\}$, entonces $L^+ = \{aa, aaaa, aaaaaa, aaaaaaaa, aaaaaaaaaa, aaaaaaaaaaaa, ...\}$.
- Si $L = \{a, bb\}$, entonces $L^+ = \{a, bb, aa, bbbb, abb, bba, aaa, bbbbbb, abba, abbbb, aabb, ...\}$.
- Si $L = \{\lambda, ab\}$, entonces $L^+ = \{\lambda, ab, abab, ababab, ...\}$.

Es importante darse cuenta de las tres propiedades siguientes:

1) $\forall L\ (L^+ \subset L^*)$, propiedad fácilmente verificable a partir de las definiciones de L^+ y de L^*, ya que, $L^1 \cup L^2 \cup L^3 ... \subset L^0 \cup L^1 \cup L^2 \cup L^3 ...$

2) $\forall L\ (\lambda \in L^+ \Leftrightarrow \lambda \in L)$.

3) $\forall L\ (L^+ = L^* \Leftrightarrow \lambda \in L)$, es decir, L^+ y L^* coinciden si, y sólo si, $\lambda \in L$. Si $\lambda \notin L$, entonces $\lambda \notin L^+$, pero $\lambda \in L^*$.

3.4.3. Cierres de un alfabeto: Σ^* y Σ^+

Dado que cualquier alfabeto puede ser considerado un lenguaje formado por palabras de un solo símbolo (longitud 1), las operaciones de cierre de Kleene y de cierre positivo de Kleene también son aplicables a los alfabetos.

Σ^* representa el conjunto (lenguaje) de todas las palabras sobre Σ.

Σ^+ representa el conjunto (lenguaje) de todas las palabras sobre Σ de longitud no nula ($\Sigma^* = \Sigma^+ \cup \{\lambda\}$).

Si nos restringimos a un alfabeto Σ determinado, entonces Σ^* es el **conjunto universal** (conjunto de todas las palabras), de manera que:

1) $(\Sigma^*)^c = \varnothing$ y $\varnothing^c = \Sigma^*$, es decir, el complementario de Σ^* es $\varnothing$ y el complementario de $\varnothing$ es Σ^*.

2) $\forall L$ sobre Σ ($L \subseteq \Sigma^*$), es decir, cualquier lenguaje sobre Σ es un subconjunto de Σ^*.

4. Definición de lenguajes

A menudo nos encontramos con la necesidad de especificar lenguajes más o menos de manera formal. A menudo se dispone de una definición informal dada en lenguaje natural que debe convertirse en una definición cuanto más formal mejor.

Las definiciones formales más habituales son las que utilizan la siguiente notación:

$$L = \{\underbrace{w \in \Sigma^*}_{\substack{\text{Palabras}\\ \text{de } \Sigma^*}} \mid \underbrace{\text{condición/nes de pertenencia}}_{\substack{\text{Condiciones que deben}\\ \text{cumplir para ser palabras de } L}}\}.$$

De hecho, dado que toda palabra sobre Σ pertenece a Σ^*, no es estrictamente necesario indicar esta circunstancia, de manera que la notación $L = \{w \mid$ condición/es de pertenencia$\}$ también es totalmente aceptable.

Como ejemplos consideramos las siguientes definiciones informales y sus contrapartidas formales:

- Palabras con el doble de símbolos *a* que de símbolos *b*. La definición formal sería la siguiente: $L = \{w \mid |w|_a = 2|w|_b\}$.
- Palabras que empiezan con el símbolo *a* y que acaban con el sufijo *aba*. Para este lenguaje, tendríamos la siguiente definición: $L = \{w \mid \exists x \in \Sigma^*(w = axaba)\}$. En este caso, también sería totalmente aceptable $L = \{a\}\ \Sigma^*\ \{aba\}$.
- Palabras de longitud impar que acaban con el símbolo *a*. La definición formal de este lenguaje sería: $L = \{w \mid (|w| = 2n + 1,\ n \geq 0) \wedge \exists x \in \Sigma^*\ (w = xa)\}$.
- Palabras que tienen un prefijo palíndromo de cinco símbolos. En este caso, la definición formal sería la siguiente: $L = \{w \mid \exists x, y \in \Sigma^*\ (w = xy \wedge |x| = 5 \wedge x = x^R)\}$.
- Palabras que tienen un prefijo propio que es igual que un sufijo propio. Para este lenguaje, tendríamos la siguiente definición: $L = \{w \mid \exists x, y, z \in \Sigma^*(w = xyz \wedge x = z \wedge x \neq w \wedge x \neq \lambda)\}$. También sería correcta una definición como la siguiente: $L = \{w \mid \exists x, y \in \Sigma^*\ (w = xyx \wedge x \neq \lambda)\}$.

En algunos casos, el hecho de dar una definición de lenguaje especificado informalmente es más complejo que dar una del complementario.

Por ejemplo, en el caso de palabras que no contienen la subpalabra *aba*, es mucho más fácil dar una definición formal del lenguaje complementario: el de las palabras que si contienen la subpalabra *aba*.

La definición de L^c sería $L^c = \{w \mid \exists x, y \in \Sigma^* (w = xabay)\}$. Sin embargo, si conviene, la negación de la condición de pertenencia permite definir L de la siguiente forma: $L = \{w \mid \forall x, y \in \Sigma^* (w \neq xabay)\}$.

También sería totalmente aceptable definir L^c como $L^c = \Sigma^* \{aba\}\Sigma^*$, y, consiguientemente, definir L como $L = (\Sigma^* \{aba\}\Sigma^*)^c$.

Resumen

En este módulo didáctico hemos visto los siguientes aspectos:

1) Hemos empezado definiendo los conceptos ***alfabeto*** (conjunto finito de símbolos) y ***palabra*** (secuencia de símbolos). También nos hemos interesado por una palabra especial, λ, que tiene la particularidad de no poseer ningún símbolo. Finalmente, hemos llegado al concepto ***lenguaje*** (conjunto de palabras).

2) En el segundo apartado hemos presentado las operaciones que se aplican más habitualmente a las palabras: **concatenación** (añadir una palabra tras otra), **longitud** (número de símbolos), **número de ocurrencias de un símbolo** (cuántas veces aparece un símbolo determinado) e **inversión** (inversión de una palabra). Se ha dedicado una especial atención a la concatenación, mostrando sus propiedades más importantes: asociatividad, no-conmutatividad y existencia de elemento neutro (λ).

3) El tercer apartado se ha dedicado a las operaciones que se aplican habitualmente sobre lenguajes: las **operaciones conjuntistas** (unión, intersección y complementación), la **concatenación** y la **inversión**. Al igual que en el caso de las palabras, se ha dedicado especial atención a la concatenación y a sus propiedades más destacadas: asociatividad, no-conmutatividad, existencia de elemento neutro ($\{\lambda\}$), distributividad respecto de la unión y no-distributividad respecto de la intersección. La aplicación reiterada de la operación de la concatenación de un lenguaje consigo mismo y la unión de todos los resultados posibles nos han conducido a la definición de cierres de un lenguaje: el **cierre de Kleene** (la estrella de Kleene) y el **cierre positivo de Kleene**.

Hemos acabado este apartado definiendo el **lenguaje universal** sobre un alfabeto determinado con el cierre de Kleene de este alfabeto (Σ^*).

4) La manera de definir los lenguajes formalmente ha sido el tema del cuarto y último apartado.

Ejercicios de autoevaluación

1. Sea $\Sigma = \{a, b\}$. Explicad por qué es imposible que exista una palabra $w \in \Sigma^*$ tal que verifique la igualdad $aw = wb$.

2. Sea $L = \{a, aa, ba\}$. Proponed, de forma razonada, un ejemplo de palabra que pertenezca a cada uno de los lenguajes que se dan a continuación:
a) L^cL.
b) LL^R.
c) $(L^RL) \cap L$.
d) $(L^R)^c$.
e) $(L^*)^c$.
f) $(L^c)^*$.

3. A pesar de que en general, la concatenación de lenguajes no es conmutativa, es posible encontrar parejas L_1 y L_2 de lenguajes tales que $L_1L_2 = L_2L_1$. Proponed de forma justificada tres ejemplos con $L_1 \neq L_2$ (buscad alguno con lenguajes que tengan un número infinito de palabras).

4. ¿Es cierto que $\Sigma^*L = L\Sigma^*$, para cualquier lenguaje L? Justificad la respuesta.

5. Demostrad que $(\Sigma^+)^c = \{\lambda\}$.

6. Demostrad que no existe ningún lenguaje L tal que $L = L^cL^c$. Al resolver este ejercicio, tened presente lo que pasa con λ.

7. Sean $L_1 = \{a, b, ab\}$ y $L_2 = \{aa, a, bb\}$. Contestad cada una de las siguientes cuestiones justificando siempre la respuesta:
a) ¿Es cierto que $L_1^* = \Sigma^*$?
b) ¿Es cierto que $L_2^* = \Sigma^*$?
c) ¿Es cierto que $(L_1^cL_1)^* = \Sigma^*$?
d) ¿Es cierto que $L_2 \subset L_2^cL_2$?

8. Todas las afirmaciones siguientes son falsas. Hallad contraejemplos que lo demuestren. (Nota: considerad lenguajes "simples" como Σ^*, Σ^+, $\{\lambda\}$, $\varnothing$, {a}, {a, b}, etc.)
a) $\forall L\ [(L^*)^+ = L^+]$.
b) $\forall L\ [LL^R \cup L^RL = \Sigma^*]$.
c) $\forall L\ [LL^c \cup L^cL = \Sigma^*]$.
d) $\forall L\ [(LL^c)^* = L^*]$.
e) $\forall L\ [L^+L^+ = L^+]$.

9. Dad expresiones formales para los siguientes lenguajes:
a) Palíndromos de longitud par con más símbolos *a* que símbolos *b*.
b) Palabras que tienen un factor que no es un palíndromo.
c) Palabras en las que ningún sufijo propio es un palíndromo.
d) Palabras que no contienen al mismo tiempo el factor *ab* y el factor *bba*.
e) Palabras que, si empiezan con el símbolo *a*, no acaban con el símbolo *b*.

Solucionario

Ejercicios de autoevaluación

1. Ninguna palabra de Σ^* puede verificar la igualdad $aw = wb$, porque el símbolo a aparece en aw una vez más que en wb. Sea n el número de ocurrencias del símbolo a en una palabra $w \in \Sigma^*$ cualquiera ($|w|_a = n$). Entonces $|aw|_a = n + 1$, pero $|wb|_a = n$. Por consiguiente, aw y wb no pueden ser la misma palabra.

2.
a) $a \in L^cL$ ya que $a = \lambda a$ ($\lambda \notin L$ y, por tanto, $\lambda \in L^c$, y $a \in L$).
b) $L^R = \{a, aa, ab\}$. Entonces $LL^R = \{aa, aaa, aab, aaaa, aaab, baa, baaa, baab\}$. Cualquiera de estas palabras es un ejemplo válido.
c) $aa \in (L^RL) \cap L$, ya que $aa \in L^RL$ (consultad la respuesta del apartado **b** de este ejercicio) y, también, $aa \in L$.
d) $\lambda \in (L^R)^c$, ya que $\lambda \notin L^R$.
e) $b \in (L^*)^c$. Para demostrar esta pertenencia, es suficiente demostrar que $b \notin L^*$. Esta demostración es una sencilla reducción al absurdo: supongamos que $b \in L^*$, entonces, dado que $L^* = L^0 \cup L \cup L^2 \cup ... \cup L^n$, debe existir un número p tal que $b \in L^p$. Sin embargo, dado que $\lambda \notin L$, p necesariamente tiene que ser 1. Pero esto es una contradicción, ya que $b \notin L_1 = L$. La contradicción origina que se pueda afirmar la negación del supuesto inicial: $b \notin L^*$.
Notad que la palabra más corta de L es a, la más corta de L^2 es aa, la más corta de L^3 es aaa... Esto nos permite deducir que una palabra de longitud 1 sólo puede pertenecer a L ($p = 1$).
f) $\lambda \in (L^c)^*$, porque $\lambda \notin L$. Y, por consiguiente, $\lambda \in L^c$ y $\lambda \in (L^c)^*$, ya que $L^c \subset (L^c)^*$ (cualquier lenguaje está incluido en su estrella de Kleene).
En este caso, el razonamiento puede ser aún más simple: λ pertenece a la estrella de Kleene de cualquier lenguaje.

3. Algunos ejemplos triviales son los siguientes:
- L_1 es cualquier lenguaje y $L_2 = \varnothing$, ya que $L_1\varnothing = \varnothing L_1 = \varnothing$.
- L_1 es cualquier lenguaje y $L_2 = \{\lambda\}$, ya que $L_1\{\lambda\} = \{\lambda\}L_1 = L_1$.
- $L_1 = \Sigma^*$ y $L_2 = \Sigma^+$, ya que $\Sigma^*\Sigma^+ = \Sigma^+\Sigma^* = \Sigma^+$.
- $L_1 = \{ab\}$ y $L_2 = \{abab\}$, ya que $L_1L_2 = L_2L_1 = \{ababab\}$.

Algunos ejemplos menos simples son los siguientes:

- $L_1 = \{w \in \Sigma^* \mid |w|_a = |w|_b\}$ (palabras con tantos símbolos a como símbolos b) y $L_2 = \{w \in \Sigma^* \mid |w|_a \neq |w|_b\}$ (palabras con un número diferente de símbolos a y símbolos b).
  Entonces $L_1L_2 = L_2L_1 = L_2$.
- $L_1 = \{w \in \Sigma^* \mid |w|_a = 2n, n \geq 0\}$ (palabras que tienen un número par de símbolos a) y $L_2 = \{w \in \Sigma^* \mid |w|_a = 2n + 1, n \geq 0\}$ (palabras con un número impar de símbolos a).
  Entonces $L_1L_2 = L_2L_1 = L_2$.

4. No es cierto que $\Sigma^*L = L\Sigma^*$ para cualquier lenguaje L. Para demostrarlo, es suficiente con encontrar un contraejemplo: un lenguaje tal que $\Sigma^*L \neq L\Sigma^*$.
Un posible contraejemplo es el lenguaje $\{b\}$, ya que $\Sigma^*\{b\}$ es el lenguaje de las palabras que acaban con el símbolo b, mientras que $\{b\}\Sigma^*$ es el lenguaje de las palabras que empiezan con el símbolo b. Evidentemente, se trata de lenguajes distintos.

5. Sabemos que $\Sigma^+ = \Sigma^* - \{\lambda\}$. La diferencia $\Sigma^* - \{\lambda\}$ se puede reescribir $\Sigma^* \cap \{\lambda\}^c$, de manera que $\Sigma^+ = \Sigma^* \cap \{\lambda\}^c$. Complementando los dos miembros de la igualdad, la expresión queda de la manera siguiente: $(\Sigma^+)^c = (\Sigma^* \cap \{\lambda\}^c)^c$. Operando: $(\Sigma^+)^c = (\Sigma^*)^c \cup \{\lambda\}^{cc}$. Ahora, teniendo en cuenta que $(\Sigma^*)^c = \varnothing$ y que $\{\lambda\}^{cc} = \{\lambda\}$, nos queda que $(\Sigma^+)^c = \varnothing \cup \{\lambda\} = \{\lambda\}$.
Una demostración alternativa a la anterior podría ser la siguiente:$w \in \Sigma^+ \Leftrightarrow |w| \geq 1$, entonces $w \notin \Sigma^+ \Leftrightarrow |w| < 1$. Pero si $|w| < 1$, esto significa que $|w| = 0$, y esto es lo mismo que decir que $w = \lambda$. De este modo, tenemos que $w \notin \Sigma^+ \Leftrightarrow w = \lambda$. O lo que es lo mismo: $w \in (\Sigma^+)^c \Leftrightarrow w = \lambda$.

6.
- Si $\lambda \in L$, entonces $\lambda \notin L^c$ y, por consiguiente, $\lambda \notin L^cL^c$. Así, L y L^cL^c discrepan, como mínimo, en la palabra λ.
- Si $\lambda \notin L$, entonces $\lambda \in L^c$ y, por tanto, $\lambda \in L^cL^c$. Otra vez L y L^cL^c discrepan, como mínimo, en la palabra λ.
- En resumen, L y L^cL^c no son nunca iguales, ya que cuando el primero contiene λ, el segundo no lo hace, y viceversa.

7.
a) Es cierto. Notad que L_1 contiene las palabras (símbolos) a y b. Así, $\Sigma \subset L_1$ y, por tanto, $\Sigma^* \subset L_1^*$. Sin embargo, Σ^* es el lenguaje universal sobre Σ (cualquier lenguaje sobre Σ es un subconjunto de Σ^*) y esto significa que $L_1^* \subset \Sigma^*$. Ahora bien, si $\Sigma^* \subset L_1^*$ y también $L_1^* \subset \Sigma^*$, entonces debemos concluir que $L_1^* = \Sigma^*$.
De hecho, cualquier lenguaje que contenga su alfabeto ($\Sigma \subset L$) tendrá por estrella de Kleene el lenguaje universal (Σ^*).

b) No es cierto. Para probarlo, es suficiente encontrar una palabra que pertenezca a uno de los lenguajes, pero que no pertenezca al otro. Dado que uno de los lenguajes es Σ^* (todas las palabras sobre Σ pertenecerán a éste), entonces será necesario encontrar una palabra que no pertenezca a L_2^*.
La palabra b no pertenece a L_2^* ($b \notin L_2^*$). L_2^* sólo contiene palabras con un número par de símbolos b.
c) La afirmación es cierta. Se puede razonar de forma análoga a como se ha hecho en la primer cuestión de este ejercicio. Será suficiente ver que $L_1^cL_1$ contiene Σ, porque, como ya se ha observado, cuando un lenguaje contiene Σ, su estrella de Kleene es Σ^*. Efectivamente, $\Sigma \subset L_1^cL_1$, ya que $\lambda \in L_1$ y entonces $\lambda a = a \in L_1^cL_1$ y $\lambda b = b \in L_1^cL_1$.
d) Esta afirmación también es cierta. Dado que $\lambda \in L_2^c$, tenemos que:
- $\lambda aa = aa \in L_2^cL_2$,
- $\lambda a = a \in L_2^cL_2$,
- $\lambda bb \in L_2^cL_2$.

Así pues, $L_2 = \{aa, a, bb\} \subset L_2^cL_2$.

8.
a) Se puede considerar el lenguaje $L = \{a\}$. Entonces: $L^* = \{\lambda, a, aa, aaa, aaaa,...\}$, $(L^*)^+ = \{\lambda, a, aa, aaa, aaaa,...\}$ y $L^+ = \{a, aa, aaa, aaaa,...\}$. Es inmediato darse cuenta de que $(L^*)^+ \neq L^+$, ya que $\lambda \in (L^*)^+$, pero $\lambda \notin L^+$.
De hecho, se puede considerar cualquier lenguaje L tal que $\lambda \notin L$. Veámoslo: $\lambda \in L^*$ (siempre) y, por tanto, $\lambda \in (L^*)^+$. Pero si $\lambda \notin L$, entonces $\lambda \notin L^+$.
Dado que $\lambda \in (L^*)^+$, y que $\lambda \notin L^+$, ya se puede concluir que $(L^*)^+ \neq L^+$.
b) En este caso se puede considerar $L = \emptyset$.
Dado que $\emptyset^R = \emptyset$, entonces $\emptyset\emptyset^R \cup \emptyset^R\emptyset = \emptyset\emptyset \cup \emptyset\emptyset = \emptyset \cup \emptyset = \emptyset \neq \Sigma^*$. Lenguajes como $\{\lambda\}$ o $\{a\}$ también serían contraejemplos.
c) Consideremos $L = \Sigma^*$ y recordemos primero que $(\Sigma^*)^c = \emptyset$ y, después, que $\emptyset L = L\emptyset = \emptyset$.
Así: $\Sigma^*(\Sigma^*)^c \cup (\Sigma^*)^c\Sigma^* = \Sigma^*\emptyset \cup \emptyset\Sigma^* = \emptyset \cup \emptyset = \emptyset \neq \Sigma^*$.
d) En este caso consideraremos el lenguaje $L = \{a\}$ y razonaremos que la palabra $ab \in (LL^c)^*$, pero que $ab \notin L^*$. Si $L = \{a\}$, entonces $b \notin L$.
Dado que $b \notin L$, entonces $b \in L^c$. Así, la palabra $ab \in LL^c$, y también $ab \in (LL^c)^*$. Sin embargo, $L^* = \{a\}^* = \{\lambda, a, aa, aaa,...\}$ y es evidente que $ab \notin L^*$.
e) En este caso, se puede volver a utilizar el lenguaje $L = \{a\}$ y ver que $a \in L^+$. Pero que $a \notin L^+L^+$. (La palabra más corta de L^+L^+ es aa).

9.
a) $L = \{w \mid (w = w^R) \wedge (|w| = 2n, n \geq 0) \wedge (|w|_a > |w|_b)\}$.
b) $L = \{w \mid \exists x, y, z \in \Sigma^* (w = xyz \wedge y \neq y^R)\}$.
c) $L^c = \{w \mid \exists x, y \in \Sigma^* (w = xy \wedge y = y^R \wedge y \neq \lambda \wedge y \neq w)$. Negando la condición de pertenencia: $L = \{w \mid \forall x, y \in \Sigma^* (w \neq xy \vee y \neq y^R \vee y = \lambda \vee y = w)\}$.
d) $L^c = \{w \mid \exists x, y, z \in \Sigma^* (w = xabybbaz \vee w = xbbayabz)\}$. Negando la condición de pertenencia: $L = \{w \mid \forall x, y, z \in \Sigma^* (w \neq xabybbaz \wedge w \neq xbbayabz)\}$.
e) Observad que la condición de pertenencia tiene forma de implicación (comenzar en $a \Rightarrow$ no acabar en b) y que, como cualquier implicación, se puede transformar en una disyunción (no comenzar en a o no acabar en b). En este punto, lo más sencillo es considerar el complementario (comenzar en a y acabar en b). $L^c = \{w \mid \exists x \in \Sigma^* (w = axb)\}$. Así: $L = \{w \mid \forall x \in \Sigma^* (w \neq axb)\}$.

Glosario

alfabeto
Conjunto finito y no vacío de símbolos.

cierre
Conjunto de todas las palabras que se pueden obtener concatenando palabras de este lenguaje un número arbitrario de veces.

concatenación
Operación que consiste en añadir los símbolos de una palabra detrás de los de otra.

estrella (de Kleene)
Unión de todas las potencias de un lenguaje ($L^0 \cup L^1 \cup L^2$ ...).

inversión
Operación que consiste en obtener una palabra como resultado de invertir el orden de los símbolos de otra.

lenguaje
Conjunto de palabras.

palabra
Secuencia de símbolos de un alfabeto.