



Notas de Curso

PALABRAS Y OPERACIONES PARA PALABRAS

TEORÍA y EJEMPLOS

Alberto A. Herrera Becerra

Octubre 2021

Notas de Curso

ICAT – UNAM, Facultad de Ingeniería - UNAM

Financiamiento interno

PALABRAS Y OPERACIONES PARA PALABRAS: Teoría y ejemplos

Alberto A. Herrera Becerra

Grupo Académico de Modelado y Simulación de Procesos, Departamento
de Instrumentación Científica e Industrial

Resumen

En este documento se presenta una introducción a los principales aspectos teóricos del tema de Secuencias Finitas y Palabras. Las nociones tratadas en este documento son fundamentales para estructurar un curso a nivel universitario acerca de la Teoría de los Lenguajes Formales. Este tema, a su vez, se ha vuelto básico en la preparación de los modernos ingenieros en computación al introducir los elementos matemáticos fundamentales para abordar cuestiones básicas acerca de la caracterización de las funciones y operaciones que pueden realizar, y no realizar, las computadoras digitales electrónicas modernas. Desafortunadamente, el tema no siempre es estudiado con la profundidad apropiada en los cursos impartidos en la Facultad de Ingeniería de nuestra Universidad. Adicionalmente, el tema es tratado de manera muy superficial y poco sistemática en muchos de los libros introductorios sobre la Teoría de los Lenguajes Formales. Este documento se presenta como una ayuda al alumno para subsanar tales deficiencias. El tema se aborda desde el punto de vista de las matemáticas discretas, empleando nociones básicas de teoría de conjuntos y álgebra abstracta, temas que ya son conocidos para los estudiantes que ya han cursado las materias de ciencias básicas. Los conceptos teóricos introducidos son ilustrados con un conjunto de ejemplos representativos y los conocimientos abordados se refuerzan con algunos ejercicios sugeridos para el alumno.

Índice

Introducción.....	4
-------------------	---

Parte 1

Teoría

1. Conceptos Iniciales.....	5
1.1. Indexación.....	5
1.2. Secuencias Finitas.....	8
1.3. Otras Entidades Matemáticas.....	10
2. Palabras.....	13
2.1. Las palabras como cadenas de contenedores.....	16
2.2. Las palabras como cadenas de símbolos.....	18
2.3. Construcción de palabras.....	21
3. Operaciones Para Palabras.....	23
3.1. Concatenación.....	24
3.2. Concatenación Repetida.....	26
3.3. Reverso.....	28
3.4. Factorización.....	30
3.5. Otras Operaciones Para Palabras.....	33
4. Algunos Conjuntos De Palabras.....	35
5. Semigrupos Y Monoides.....	36

Parte 2

Ejemplos

1. Palabras.....	38
2. Operaciones Para Palabras.....	43
3. Factorización.....	46
4. Palíndromos.....	47
Referencias Bibliográficas.....	54

Introducción

Todos los objetos lingüísticos que se presentan a continuación tienen al menos dos características en común:

Programas escritos en algún lenguaje de programación como C++.

Palabras inglesas.

Secuencias de símbolos que se usan para representar un valor entero positivo.

Frases escritas en algún lenguaje natural como el español o el inglés.

Primero, cada uno está compuesto por secuencias de símbolos tomados de alguna colección finita. En el caso de las palabras inglesas, la colección finita es el conjunto de letras del alfabeto inglés, junto con los símbolos que se usan para construir palabras en inglés, tales como el guion, el apóstrofe y otros por el estilo. De forma similar, la representación de enteros positivos son secuencias de caracteres del conjunto de los dígitos $\{0, 1, 2, \dots, 9\}$. Los programas de computadora escritos en C++ y las frases en español o inglés también están tomados de una colección finita. En estos dos casos, sin embargo, los conjuntos de símbolos son distintos. En el caso de los programas en C++, el conjunto de símbolos es una colección de identificadores legales de C++ (con una longitud menor o igual que una constante), palabras clave y palabras reservadas, símbolos especiales de C++ y espacios en blanco tales como el retorno de carro, el carácter de salto de línea y el espacio manual.

Segundo, en todos los casos las secuencias de símbolos que constituyen los elementos en cuestión tienen una longitud finita, aunque no existen limitaciones en cuanto a su tamaño.

La noción de secuencia finita de símbolos es el elemento principal a ser tratado en el presente documento. A continuación se introducen la notación y los nombres a usar para dichas secuencias.

Parte 1

Teoría

1 Conceptos Iniciales

Las funciones pueden conferir estructura a los conjuntos no ordenados al ajustar sus elementos a un molde que proporciona nuevas posibilidades para su procesamiento.

1.1 Indexación

Sea I un conjunto. Una *indexación* del conjunto D por I es una función $s: I \rightarrow D$. Tal función s también se denomina *familia de elementos indexados* por I .

Sea s una familia de elementos indexados por I . Así, s es un mapeo de la forma $i \mapsto s(i) = d$, donde $i \in I$ y $d \in D$. Se dice entonces que el elemento d ha sido señalado o seleccionado por medio del elemento i ; debido a esto, el elemento i se denomina un *índice* para el elemento $s(i) = d$ y, en consecuencia, el conjunto I se denomina el *conjunto de índices*. Una indexación s es una regla para señalar o seleccionar elementos del conjunto D .

Un elemento $s(i)$ de una indexación s usualmente se denota s_i . Con esto, la familia de elementos indexados s es el conjunto:

$$s = \{(i, s_i) \mid i \in I\}.$$

El conjunto anterior puede ser denotado de manera compacta $(s_i)_{i \in I}$, o simplemente (s_i) si el conjunto de índices es conocido. A manera de ejemplo, considérese el conjunto de índices $I = \{a, b, c, d, e, f\}$ y sea la familia de conjuntos $M = \{A, B, C, D, E, F, G, H\}$. Con estos conjuntos se puede construir la familia indexada s siguiente:

$$\begin{aligned} a &\mapsto s(a) = s_a = C; \\ b &\mapsto s(b) = s_b = H; \\ c &\mapsto s(c) = s_c = B; \\ d &\mapsto s(d) = s_d = E; \\ e &\mapsto s(e) = s_e = D; \\ f &\mapsto s(f) = s_f = A. \end{aligned}$$

La familia indexada es:

$$s = \{(a, s_a), (b, s_b), (c, s_c), (d, s_d), (e, s_e), (f, s_f)\}$$

$$= \{(a, C), (b, H), (c, B), (d, E), (e, D), (f, A)\}.$$

La familia anterior se representa de manera compacta como sigue:

$$s = (s_a, s_b, s_c, s_d, s_e, s_f) = (C, H, B, E, D, A).$$

Se debe destacar que el orden en el que se listan los elementos de s no es relevante.

Debido a esto, se tiene lo siguiente:

$$s = (s_f, s_d, s_e, s_b, s_a, s_c) = (A, E, D, H, C, B).$$

El rango de s se denota $\{s_i \mid i \in I\}$; esto permite hablar de una familia indexada en términos de conjuntos, precisamente el conjunto de imágenes definidas por la función s . De esto se puede construir una imagen mental muy sugerente sobre las familias indexadas. Una familia puede ser vista como un conjunto de “contenedores” tales que cada “contenedor” está identificado con uno y sólo un índice, y cada “contenedor” contiene uno y sólo un objeto del conjunto blanco. Esta imagen va a ser usada más adelante en las notas.

Adicionalmente, si el valor de s_i está dado por alguna expresión ξ que involucre a i , entonces el rango de s usualmente se denota $\{\xi \mid i \in I\}$, evitando con esto la necesidad de nombrar a s . Por ejemplo, $\{2n + 1 \mid n \in \mathbb{N}\}$ es, de hecho, $\{s_n \mid n \in \mathbb{N}\}$ donde $s: \mathbb{N} \rightarrow P$ está definida por $s(n) = 2n + 1$. (Aquí, $\mathbb{N} = \{0, 1, 2, \dots\}$ es el conjunto de los números naturales.)

Es muy importante destacar que un elemento $d \in D$ puede tener varios índices bajo s debido a que esta función no tiene porque ser inyectiva. Esto es, puede ocurrir que $s_i = s_j = d$ para $i, j \in I$ con $i \neq j$. Con base en esto, una familia contiene cualquier elemento exactamente una sola vez si, y sólo si, la función de indexación es inyectiva. Ahora bien, las funciones sobreyectivas y las familias son formalmente equivalentes, ya que cualquier función f con dominio I induce una familia $(f(i))_{i \in I}$: ser un elemento de una familia es equivalente a estar en el rango de la función. En la práctica, sin embargo, una familia usualmente es vista como una colección, no como una función.

A partir de las últimas aseveraciones se puede deducir que cualquier conjunto X da origen a una familia $(X_x)_{x \in X}$. Esto es, todo conjunto X puede ser indexado de manera trivial; en efecto, siempre se puede usar el mapeo identidad $1_X: X \rightarrow X$ para indizar a X . En consecuencia, todo conjunto puede ser considerado como una familia.

En resumen, cuando se trata con conjuntos se puede asumir, sin restringir la generalidad, de que los conjuntos están indexados. Pero si es necesario tener una precaución. Para cualquier familia $(A_i)_{i \in I}$ existe el conjunto de todos los elementos $\{A_i \mid i \in I\}$, pero esto no conlleva ninguna información acerca de la multiplicidad de apariciones de un cierto elemento o de la estructura que impone el conjunto I . Al usar un conjunto en lugar de la familia se puede perder algo de información.

A continuación, se describe una clase muy especial de familias indexadas, la clase fundamental para el estudio de los lenguajes formales.

1.2 Secuencias Finitas

Un caso especial de indexación surge cuando el conjunto de índices es un conjunto de números naturales de la forma general $I = \{k \in \mathbb{N} \mid k < n\}$ para algún $n \in \mathbb{N}$ dado. Para estos conjuntos es más conveniente adoptar la notación $\{0, 1, \dots, n-1\}$, obviamente más sugerente. Nótese que cuando $n = 0$, la especificación anterior denota al conjunto vacío $\emptyset$, y si n y m son dos números naturales diferentes, entonces $I = \{0, 1, \dots, n-1\}$ y $J = \{0, 1, \dots, m-1\}$ son dos conjuntos de índices diferentes, aunque no son conjuntos inconexos. Puede suceder que $n > m$, con lo que m pertenece al primer conjunto, pero no al segundo; se puede establecer un razonamiento análogo para el caso en que $m > n$.

Se define a una *secuencia finita* s como una indexación cuyo conjunto de índices es precisamente $I = \{0, \dots, n-1\}$ para algún número natural n dado y cuyo rango es un conjunto D . Esto es, se tiene que $s: \{0, \dots, n-1\} \rightarrow D$. El número natural n referido en la definición se denomina la *longitud* de s y, en consecuencia, se dice que s es una secuencia (finita) de longitud n ; la longitud de la secuencia se denota $|s|$, o $\text{long}(s)$. Los valores de s son denominados las *componentes* de s . Una secuencia finita s , de

longitud n , es denotada $s = (s_0, s_1, \dots, s_{n-1})$, con $s_i = s(i)$ para $0 \leq i \leq n-1$. Aquí es necesario establecer una regla de escritura: las componentes s_i se deben listar siguiendo el orden natural de sus índices.

En la definición de una secuencia finita también se pudo haber elegido el conjunto de índices $I = \{m, m+1, m+2, \dots, m+n-1\}$, para $m, n \in \mathbb{N}$ dados, con $m < n$. Nótese que también se tienen n números naturales consecutivos como índices. Sin embargo, el uso general de este conjunto de índices añade complicaciones innecesarias para los fines que se buscan. También se puede introducir y formalizar la noción de una secuencia infinita usando como conjunto de índices a todo el conjunto $\mathbb{N}$. Sin embargo, este concepto no va a ser usado en el presente documento y, en consecuencia, no se discutirá más.

Se define a la secuencia (finita) de longitud cero como la función de indexación cuyo dominio es el conjunto vacío y cuyo rango es D . Ya se sabe que existe exactamente una función como tal, a saber, la función vacía, por lo que sólo hay una secuencia de longitud cero. Esta secuencia será denominada la *secuencia nula* o la *secuencia vacía*, y será denotada ε . Así, $\varepsilon: \emptyset \rightarrow D$ y $|\varepsilon| = 0$. (Cabe señalar que otros autores utilizan notaciones alternativas, como e o como λ , para denotar a la secuencia vacía.)

Una propiedad fundamental de las secuencias finitas es la siguiente. Se dice que dos secuencias finitas s y t sobre el mismo conjunto D son *iguales*, lo cual se denota $s = t$, si y sólo si todas sus componentes coinciden. Esto es, se requiere que $(s_0, \dots, s_{n-1}) = (t_0, \dots, t_{m-1})$, pero esta igualdad sólo puede suceder si $n = m$, las secuencias tienen las mismas longitudes, y $s_i = t_i$ para $0 \leq i \leq n-1$, sus valores son iguales. Así pues, resulta que dos secuencias de longitudes diferentes no pueden ser iguales, y dos secuencias con la misma longitud todavía pueden ser diferentes si por lo menos una componente no coincide.

Nótese que se usa la notación (d, d') para denotar a una secuencia de longitud dos. Sin embargo, en la teoría de conjuntos se suele usar la expresión (d, d') como el

par ordenado $\{\{d\}, \{d, d'\}\}$, el cual es un objeto matemático diferente; su característica distintiva es que $(d, d') = (d_1, d'_1)$ si y sólo si $d = d_1$ y $d' = d'_1$. En el contexto de las secuencias finitas, resulta que la última igualdad también es válida, por lo que la ambigüedad no es muy dañina si se le maneja con precaución.

Ahora bien, una secuencia t de longitud uno, una secuencia unitaria, está completamente determinada por su valor $t(0) = t_0$. Muchas veces, sin embargo, se encontrará que es conveniente considerar que un objeto d puede ser usado como la secuencia unitaria (d) , cuya componente es precisamente d . Este uso permite simplificar muchas descripciones que, de manera más formal, se vuelven un tanto engorrosas. Nuevamente, el uso señalado no es muy dañino si se hace con precaución.

Si D es un conjunto, entonces se usa tanto $Seq_n(D)$ como D^n para denotar al conjunto de todas las secuencias de longitud n sobre D . También se usa $Seq(D)$ para denotar al conjunto de todas las secuencias finitas sobre D , y se usa $Seq^+(D)$ para denotar al conjunto de todas las secuencias finitas no vacías sobre D .

Obsérvese que $D^0 = \{\varepsilon\}$ para cualquier conjunto D . Con esto, se debe aceptar que $\emptyset^0 = \{\varepsilon\}$ y que $\emptyset^n = \emptyset$ para todo $n > 0$; en consecuencia, $Seq(\emptyset) = \{\varepsilon\}$. Estos resultados pueden parecer contra intuitivos, pero son fundamentales para establecer una teoría consistente.

Se debe señalar también que D^1 se puede identificar con D , mientras que D^2 puede ser identificado con $D \times D$ si se procede con las precauciones debidas. De hecho, estas identificaciones pueden ser sustentadas señalando la existencia de biyecciones entre D^1 y D y entre D^2 y $D \times D$; por ejemplo, se tiene la biyección $\Psi: D^1 \rightarrow D$ dada por $\Psi((d)) = d$ para toda $(d) \in D^1$.

En la sección que sigue se van a presentar, de manera muy breve, otras entidades matemáticas que tienen propiedades similares a las de las secuencias finitas.

1.3 Otras Entidades Matemáticas

Se recuerda que un par ordenado (a, b) , donde a, b son dos objetos cualesquiera, es una entidad matemática caracterizada por la propiedad fundamental $(a, b) = (c, d)$ si y

sólo si $a = c$ y $b = d$. Con base en esto, se dice que los objetos componentes del par ordenado (a, b) están ordenados de forma que a se distingue como el primer componente del par, mientras que b es el segundo componente. De esta manera, se tiene que $(a, b) \neq (b, a)$ si a y b son dos objetos distinguibles. Si sucede que $a = b$, todavía se puede hablar del par ordenado (a, a) en el que el objeto a aparece dos veces, una vez como el primer objeto, y una vez más como el segundo objeto.

A partir de la noción de par ordenado, se pueden concebir otras entidades matemáticas compuestas por n objetos ordenados. Para los fines de estas notas, las dos entidades de este tipo más interesantes se describen a continuación. En lo que sigue se asume que los objetos referidos se toman de un conjunto D dado de inicio.

Pares ordenados embebidos. Se dice que la entidad matemática $(a, \emptyset)$ es una entidad compuesta por un único componente, el objeto a . Estas entidades se caracterizan por la propiedad siguiente. Las entidades $(a, \emptyset)$ y $(b, \emptyset)$ se dice que son diferentes, lo cual se escribe $(a, \emptyset) \neq (b, \emptyset)$, si y sólo si a y b son objetos distinguibles; en caso contrario, se dice que las entidades son iguales.

A continuación, se dice que $(a_1, (a_2, \emptyset))$ es una entidad de dos componentes ordenados, el objeto a_1 , el cual se reconoce como el primer objeto, y por la entidad de un solo objeto $(a_2, \emptyset)$, cuyo único componente a_2 se reconoce como el segundo objeto. Esto es, la entidad compuesta $(a_1, (a_2, \emptyset))$ está formada por los objetos a_1 y a_2 de forma tal que a_1 es el primero, mientras que a_2 es el segundo. Ciertamente puede suceder que $a_1 = a_2 = a$, situación en la que se dice que el objeto a aparece dos veces, una vez como el primer objeto, una vez más como el segundo objeto. Se dice ahora que $(a_1, (a_2, \emptyset))$ y $(b_1, (b_2, \emptyset))$ son iguales, lo cual se escribe $(a_1, (a_2, \emptyset)) = (b_1, (b_2, \emptyset))$, si y sólo si se cumple que $a_1 = b_1$ y $(a_2, \emptyset) = (b_2, \emptyset)$, pero esta igualdad implica que $a_2 = b_2$. De esto se deduce que $(a_1, (a_2, \emptyset)) \neq (a_2, (a_1, \emptyset))$ si a_1 y a_2 son objetos distinguibles. Claramente, el orden de los objetos componentes es relevante.

Continuando la idea, ahora se considera la entidad compuesta $(a_1, (a_2, (a_3, \emptyset)))$ formada por el objeto a_1 , el cual se reconoce como el primer objeto, y por la entidad de dos componentes $(a_2, (a_3, \emptyset))$, cuyos componentes a_2 y a_3 se reconocen como el segundo y tercer objetos, respectivamente. Nuevamente, los objetos a_i , con $i = 1, 2, 3$, no tienen por qué ser diferentes entre sí. Con base en las nociones ya presentadas, no es difícil establecer que $(a_1, (a_2, (a_3, \emptyset))) = (b_1, (b_2, (b_3, \emptyset)))$ si y sólo si $a_i = b_i$ para $i = 1, 2, 3$. A partir de esta propiedad se deduce que el orden de los objetos componentes es relevante.

Generalizando las ideas recién descritas, se considera la siguiente entidad de n componentes:

$$(a_1, (a_2, (\dots, (a_n, \emptyset))))$$

donde los objetos $a_1, a_2, \dots, a_n$ se reconocen, respectivamente, como el primer objeto, el segundo objeto, y así sucesivamente hasta el n -ésimo objeto. Se acepta que los objetos componentes a_i , con $i = 1, \dots, n$, no necesariamente son diferentes entre sí. Claramente, la entidad está formada por un objeto, a_1 el primero, y una entidad de $(n - 1)$ objetos $a_2, \dots, a_n$, los restantes. La entidad se conoce como un *par ordenado embebido*, debido a la similitud que existe entre las propiedades de igualdad del par ordenado embebido y el par ordenado original.

Las n -adas. Una n -ada (del inglés n -tuple) es una entidad compuesta por n objetos ordenados $a_1, \dots, a_n$, no necesariamente diferentes entre sí, que se denota $(a_1, \dots, a_n)$, y que satisface la propiedad definitoria siguiente:

$$(a_1, \dots, a_n) = (b_1, \dots, b_n) \text{ si y sólo si } a_i = b_i \text{ para } i = 1, \dots, n.$$

En particular, la 0-ada (cero-ada) es una entidad compuesta por cero componentes; esta entidad se denota $\emptyset$, aunque no se debe confundir con el conjunto vacío. Por su parte, una 1-ada (uno-ada) es la entidad (a) formada únicamente por el objeto a . Aquí es importante distinguir la entidad (a) del objeto a . Claramente, sucede que $(a) = (b)$ si

y sólo si $a = b$. Por lo que respecta a las 2-adas, se tiene que $(a_1, a_2) = (b_1, b_2)$ si y sólo si $a_i = b_i$ para $i = 1, 2$. Adicionalmente, se deduce que $(a_1, a_2) \neq (a_2, a_1)$; claramente, el orden de aparición de los objetos componentes es relevante. Debe ser claro que las propiedades fundamentales de las 2-adas coinciden con las propiedades definitorias de los pares ordenados.

La última aseveración se puede usar de base para establecer que una n -ada es una generalización a n componentes de la noción de par ordenado.

Para finalizar esta sección se puede señalar lo siguiente. Se concibe una entidad matemática compuesta por n objetos tomados un conjunto D dado de antemano, algunos de los cuales se pueden repetir dos o más veces, y que aparecen ordenados de manera secuencial, esto es, uno de ellos es el primero, otro es el segundo, y así sucesivamente hasta alcanzar un último objeto. Esta entidad se puede construir de tres formas alternativas, a saber:

- Como una secuencia finita, o familia indexada por un conjunto de índices numéricos.
- Como un par ordenado embebido.
- Como una n -ada.

Las entidades construidas son estructuralmente diferentes, pero resultan ser equivalentes entre sí, esto es, tienen las mismas propiedades fundamentales.

En lo que sigue de las notas sólo se va a hacer uso de la noción de secuencia finita, pues es una descripción más conveniente para los fines que se buscan. Se va a continuar con la introducción de una clase especial de secuencias finitas.

2 Palabras

En esta sección se va a presentar y discutir una clase especial de secuencias finitas definidas sobre símbolos. Un *símbolo* es una marca, signo o palabra que indica, significa o se entiende que representa a una idea, objeto o relación. Los símbolos son un medio para la comunicación compleja que a menudo puede tener múltiples niveles de significación. Los símbolos son la base de toda la comprensión humana y sirven como vehículos de concepción para todo el conocimiento humano. Los símbolos facilitan la comprensión del mundo en el que vivimos, sirviendo en consecuencia como

un cimiento sobre el que realizamos nuestros juicios. En este sentido, las personas no sólo usan los símbolos para hacer sentido del mundo a su alrededor, sino también para identificar y cooperar en sociedad a través de una retórica constitutiva (creación de una identidad colectiva).

Se va a comenzar la exposición señalando que un *alfabeto* es un conjunto finito no vacío de símbolos. Por ejemplo, los conjuntos siguientes son considerados alfabetos:

$D = \{0, 1, \dots, 9\}$, el conjunto de los diez dígitos decimales.

$L = \{a, b, \dots, y, z\}$, el conjunto de las letras minúsculas del alfabeto español.

Algunos autores consideran que si un conjunto A es un alfabeto, entonces en A no se deben incluir símbolos que puedan formarse por yuxtaposición de otros símbolos de A ; esta restricción impide confusiones con la definición del concepto de palabra, que se dará más adelante.

Puesto que un alfabeto es simplemente un conjunto finito no vacío, dados los alfabetos A_1 y A_2 , se tiene que $A_1 \cup A_2$ también es un alfabeto. Es más, si los conjuntos $A_1 \cap A_2$, $A_1 - A_2$ y $A_2 - A_1$ son no vacíos, entonces también son alfabetos.

Una *palabra* de longitud n sobre un alfabeto A es una secuencia finita de longitud n de símbolos de A . Una palabra w de longitud n se denota $w = (w_0, w_1, \dots, w_{n-1})$ donde $w_i \in A$ para $0 \leq i \leq n-1$. Se debe recordar que los valores w_i no necesariamente son todos diferentes entre sí. Debido a esto, la *longitud* de la palabra w , denotada por $|w|$ o $long(w)$, es simplemente el número total de símbolos que la forman, incluyendo todas las posibles repeticiones de símbolos.

Ahora bien, debido a que se está considerando que cualquier símbolo que forme parte de un alfabeto no resulta de la combinación de otros símbolos, es más conveniente denotar a la palabra $w = (w_0, w_1, \dots, w_{n-1})$ simplemente como $w = w_0 w_1 \dots w_{n-1}$; esto es, se escriben los símbolos que forman a la palabra uno detrás de otro, sin marcadores que los separen. Debido a esto, en lo que sigue a menudo se usa el término *cadena* como sinónimo de palabra.

A manera de ejemplo, las cadenas **0102**, **656670**, **0** son algunas palabras sobre el alfabeto **D** definido antes, mientras que las cadenas **aaba**, **cdxjol**, **streetcar** son algunas palabras sobre el alfabeto **L**. Nótese el uso técnico que se le ha dado a los nombres alfabeto y palabra. Con esta terminología se habla del alfabeto **L** de las letras españolas más que de un vocabulario, y una palabra sobre **A** significa simplemente una secuencia finita de letras españolas. Cuando sea necesario usar el término palabra en su sentido no técnico, se hará explícito el lenguaje natural al que se está refiriendo, por ejemplo, palabra inglesa.

Es conveniente señalar que las palabras de longitud unitaria se escriben ahora simplemente a_0 , en lugar de (a_0) . Nótese que esto se presta a una mayor confusión con el símbolo a_0 del alfabeto en cuestión. Deberá ser el contexto el que ayude a resolver la posible ambigüedad; lo mejor es señalar siempre de manera explícita la diferencia.

También es conveniente recordar la existencia de una única secuencia finita de longitud cero, la cual se representa por ε . Con base en esto, para un alfabeto **A** dado se define una *palabra vacía* o *palabra nula* y se le representa también por ε . Ahora bien, dados dos alfabetos diferentes, digamos **A**₁ y **A**₂, para cada uno de ellos se puede definir una palabra vacía, pero las palabras vacías así definidas no son iguales, ya que son dos secuencias diferentes: una está definida sobre **A**₁, mientras que la otra está definida sobre **A**₂. Adicionalmente, el símbolo ε no debe identificarse erróneamente con el espacio en blanco que se encuentra en muchos vocabularios. De hecho, se debe cuidar que ε no sea un símbolo de ningún alfabeto.

Sea **w** una palabra sobre un alfabeto **A** y sea a un símbolo de **A**. Se define la *longitud a* de **w**, denotada $|w|_a$ o $long_a(w)$, como el número de veces que aparece a en **w**. Es claro que $|\varepsilon|_a = 0$ para todo $a \in A$ y que para toda palabra **w** de cualquier alfabeto **A** se satisface la igualdad:

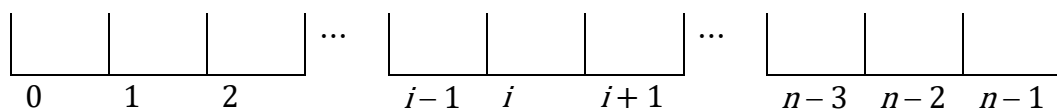
$$|w| = \sum_{a \in A} |w|_a.$$

Como una consecuencia más de que las palabras sean casos particulares de las secuencias finitas, se tiene que dos palabras u, v sobre un alfabeto A común son iguales, lo cual se denota $u = v$, si ambas palabras tienen la misma longitud y los mismos valores. Nótese que esto significa que las palabras tienen los mismos símbolos para los mismos índices o, dicho de manera coloquial, tienen los mismos símbolos en las mismas posiciones. Recuérdese que si dos palabras no tienen la misma longitud no pueden ser iguales, y aunque dos palabras tengan la misma longitud puede que no sean iguales porque por lo menos uno de sus valores no coincide.

Ahora bien, dado que cada componente w_i de una palabra $w = w_0 w_1 \dots w_{n-1}$ sobre un alfabeto A conlleva un significado individual intrínseco, se puede esperar entonces que la composición ordenada w conlleve un significado compuesto. Sin embargo, en la presentación que se va a realizar no se tomarán en cuenta ni los significados individuales de los símbolos disponibles, ni el posible significado compuesto de las palabras formadas con esos símbolos. Desde un punto de vista lingüístico, se va a realizar un estudio sintáctico de las palabras, esto es, sólo va a ser relevante el orden secuencial de los símbolos que forman a la palabra. En la sección que sigue se presentan las particularidades más relevantes de los estudios sintácticos de las palabras.

2.1 Las palabras como cadenas de contenedores

Para guiar el estudio sintáctico de las palabras se va a hacer uso de la idea de “contenedores” presentada en la sección de Indexación. Así, una palabra w de longitud n se puede visualizar como una colección finita y ordenada de n “contenedores”, cada uno de ellos etiquetado con un número natural. El ordenamiento de los “contenedores” se basa en el orden natural de los índices numéricos empleados.



Así pues, las palabras se visualizan como cadenas de “contenedores” consecutivos, uno de detrás de otro. En cada “contenedor” de la cadena se debe colocar uno y sólo

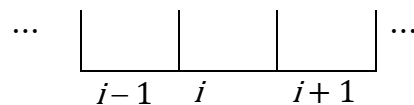
un símbolo de un alfabeto A previamente establecido. No pueden existir “contenedores” vacíos y no pueden existir “contenedores” con dos o más símbolos.

Con esta imagen, la palabra vacía se puede entender como una cadena vacía de “contenedores”. Dicho de otra manera, en la palabra vacía no hay “contenedor” alguno en el cual colocar símbolos.

Una característica distintiva fundamental de toda palabra no vacía de longitud n es que siempre deben existir un *primer* “contenedor” -el cual tiene la etiqueta 0 -, y un *último* “contenedor” -con la etiqueta $n - 1$. Un caso especial es la palabra de longitud unitaria, ya que su único “contenedor” es tanto el primero como el último.

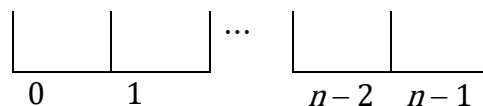


Otra característica distintiva es que todo “contenedor” debe poseer un “contenedor” antecesor inmediato y un “contenedor” sucesor inmediato, como se ilustra a continuación.



En este ejemplo, el “contenedor” con etiqueta $i - 1$ es el antecesor inmediato del “contenedor” i , mientras que el “contenedor” con etiqueta $i + 1$ es su sucesor inmediato.

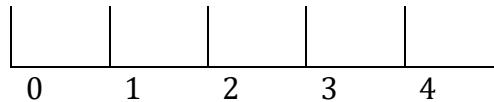
Existen dos excepciones a lo anterior. La primera excepción corresponde al primer “contenedor”, el cual no tiene un antecesor inmediato. La segunda excepción corresponde al último “contenedor”, el cual no tiene un sucesor inmediato.



Esto es, el primer “contenedor”, con etiqueta 0 , tiene como sucesor inmediato al “contenedor” con etiqueta 1 (el segundo “contenedor”), pero no tiene un “contenedor” que sea un antecesor inmediato. Por su parte, el último “contenedor”, con etiqueta $n - 1$, tiene al “contenedor” con etiqueta $n - 2$ (el penúltimo) como su antecesor

inmediato, pero no tiene un “contenedor” que sea su sucesor inmediato. Nuevamente, la palabra de longitud unitaria representa un caso especial, ya que es una palabra no vacía cuyo único “contenedor” no posee ni antecesor ni sucesor inmediatos.

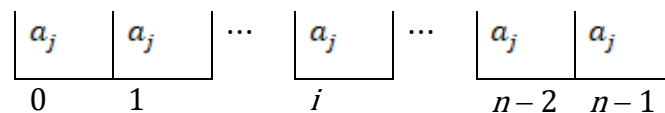
En toda palabra de longitud mayor o igual a 3 se pueden considerar “contenedores” antecesores o sucesores a un “contenedor” dado. Por ejemplo, considérese la palabra de longitud 5 ilustrada a continuación.



Si se toma como referencia al “contenedor” 1, se observa que tiene al “contenedor” 0 como su antecesor inmediato y único antecesor; que tiene al “contenedor” 2 como su sucesor inmediato; y que tiene a los “contenedores” 2, 3 y 4 como sus sucesores. Por su parte, el “contenedor” 2 tiene a los contenedores 0 y 1 como sus antecesores, siendo 1 su antecesor inmediato, y tiene a los contenedores 3 y 4 como sus sucesores, siendo 3 su sucesor inmediato.

2.2 Las palabras como cadenas de símbolos

Considérese un alfabeto A y considérese que a_j es un símbolo arbitrario de A . Si se considera ahora una cadena de n “contenedores”, se puede formar la palabra w de longitud n siguiente:



Se debe tener cuidado al interpretar la figura anterior. No se debe entender que en cada “contenedor” de la cadena se coloca un mismo símbolo, a_j en la figura. Lo que se quiere decir es que en cada “contenedor” se puede colocar un símbolo arbitrario a_j del alfabeto A . La elección es totalmente arbitraria, se puede escoger cualquier símbolo de A .

Con base en la imagen anterior, la palabra w se entiende que es una función de la forma $w: I \rightarrow A$, con $I = \{0, 1, \dots, n-1\}$, y que está dada por:

$$\begin{aligned} w(0) &= w_0 = a_{j_0} \\ w(1) &= w_1 = a_{j_1} \\ &\vdots \\ w(i) &= w_i = a_{j_i} \\ &\vdots \\ w(n-1) &= w_{n-1} = a_{j_{n-1}} \end{aligned}$$

Se debe entender que dos símbolos a_{j_k} y a_{j_l} , con $k \neq l$, pueden ser, o no ser, iguales. Dicho de otra manera, puede suceder que $w(k) = w(l)$, o puede suceder que $w(k) \neq w(l)$, para $k \neq l$.

La representación funcional anterior también se puede entender como el conjunto de pares ordenados siguiente:

$$w = \{(0, a_{j_0}), (1, a_{j_1}), \dots, (i, a_{j_i}), \dots, (n-2, a_{j_{n-2}}), (n-1, a_{j_{n-1}})\},$$

En este conjunto, los valores a_{j_i} se deben interpretar como el símbolo arbitrario a_j del alfabeto A con índice i , donde $0 \leq i \leq n-1$. En estas descripciones, se ha hecho uso de la noción de secuencia finita como un conjunto indizado por el conjunto de índices numéricos I .

En las aplicaciones, suele ser útil comenzar con la representación de conjunto de pares ordenados y luego construir la representación funcional.

A manera de ejemplo, considérese el alfabeto $A = \{a, b, c, d, e\}$. En este caso, los caracteres a, b, c, d y e no necesariamente se deben interpretar como letras de algún vocabulario; simplemente indican un alfabeto formado por cinco símbolos diferentes y distinguibles. Considérese ahora la cadena de símbolos de longitud 5 siguiente:

b	c	a	b	d
0	1	2	3	4

Esta figura se debe interpretar como sigue. En el “contenedor” 0 se colocó el símbolo b ; en el “contenedor” 1 se colocó el símbolo c ; en el “contenedor” 3 se colocó el símbolo a ;

y así sucesivamente. La cadena anterior también se puede representar de manera más compacta como sigue:

b_0	c_1	a_2	b_3	d_4
-------	-------	-------	-------	-------

Esta última figura también se puede interpretar de manera funcional. Así, la figura se puede interpretar como el conjunto de pares ordenados siguiente:

$$w = \{(0, b_0), (1, c_1), (2, a_2), (3, b_3), (4, d_4)\}.$$

En este caso, el par $(0, b_0)$ se interpreta diciendo que el símbolo b del alfabeto A está indizado con el número 0 . Los restantes pares ordenados del conjunto se interpretan de manera similar. Claramente, esta notación es un tanto redundante, por lo que se puede emplear la notación simplificada siguiente:

$$w = \{(0, b), (1, c), (2, a), (3, b), (4, d)\}.$$

A partir de esta representación, ya es fácil establecer la representación funcional siguiente:

$$\begin{aligned} w(0) &= w_0 = b \\ w(1) &= w_1 = c \\ w(2) &= w_2 = a \\ w(3) &= w_3 = b \\ w(4) &= w_4 = d \end{aligned}$$

En particular, nótese que $w(0) = w(3) = b$. Esto es, el símbolo b tiene dos índices diferentes, 0 y 3 . Dicho de otra manera, en la cadena w hay dos “contenedores” diferentes, 0 y 3 precisamente, que contienen copias del mismo símbolo, b en este caso.

De todo lo anterior, debería resultar claro que una palabra definida sobre un alfabeto A dado, es una entidad compuesta (de símbolos), cuyos componentes están ordenados (por una enumeración), y donde algunos componentes pueden aparecer repetidas veces. Se debería añadir, que no es necesario usar todos los símbolos del alfabeto para formar palabras.

Notaciones alternativas

Considérese que la palabra w se representa inicialmente en la forma de conjunto de pares ordenados:

$$w = \{(0, a_{j_0}), (1, a_{j_1}), \dots, (i, a_{j_i}), \dots, (n-2, a_{j_{n-2}}), (n-1, a_{j_{n-1}})\},$$

Esta palabra w también se puede expresar de manera simplificada como sigue:

$$w = (a_{j_0}, a_{j_1}, \dots, a_{j_{n-2}}, a_{j_{n-1}}).$$

Si además se satisfacen las recomendaciones establecidas para los símbolos de un alfabeto, se puede emplear la notación simplificada:

$$w = a_{j_0} a_{j_1} \dots a_{j_{n-2}} a_{j_{n-1}}.$$

Retomando el ejemplo de la sección anterior, la palabra particular w que se analizaba puede ser representada de manera muy compacta como sigue:

$$w = b_0 c_1 a_2 b_3 d_4.$$

Se pueden omitir los índices, llegando a la expresión siguiente:

$$w = bcabd.$$

En esta última forma de representación, el orden en el que se escriben los símbolos es fundamental. El primer símbolo de la palabra es el primero que se escribe; el segundo es el segundo que se escribe; y así sucesivamente hasta llegar al último símbolo.

2.3 Construcción de palabras

La construcción de palabras específicas se puede realizar de la manera siguiente.

Primero, es indispensable fijar el alfabeto A de referencia. Los símbolos de este alfabeto serán los únicos que se podrán emplear en la construcción.

Segundo, se deberán establecer las longitudes de las palabras que se desean construir. De esta manera se deja establecida la cantidad total de símbolos que se van a utilizar.

Tercero, se deberán establecer posibles restricciones al uso de los símbolos. Por decir, se debe establecer cuáles símbolos del alfabeto realmente van a ser utilizados, y cuáles no van a ser considerados. Se debe establecer cuáles de los símbolos realmente utilizados van a aparecer múltiples veces y señalar la cantidad de repeticiones. Se debe indicar si aparecen algunas combinaciones de símbolos (o patrones) y se debe fijar la cantidad de veces que aparece el patrón. Finalmente, se

deben señalar las posiciones en las que deben aparecer los símbolos o patrones distinguidos.

De lo anterior, debe resultar claro que el proceso de construcción de palabras particulares no es un proceso simple. Se debe tomar en cuenta que, en lo general, va a existir una cantidad muy grande de palabras diferentes que satisfacen las condiciones establecidas en el proceso de construcción. De entre todas estas palabras se deben seleccionar las palabras específicas que se desean. Esto quiere decir que todo problema de construcción de palabras involucra un cálculo combinatorio, y es bien conocido que los problemas combinatorios no son triviales y, en consecuencia, generalmente son muy engorrosos.

A manera de ejemplo, considérense las construcciones siguientes. Sea el alfabeto $A = \{a, b, c\}$. Si bien se han usado caracteres alfabéticos para denotar a los símbolos, no se debe entender que se van a realizar construcciones con alguna clase de significado lingüístico, simplemente se debe entender que se van a realizar construcciones con tres símbolos diferentes y distinguibles entre sí.

¿Cuántas palabras diferentes de longitud 5 se pueden construir con el alfabeto A ? Para responder a esta pregunta es conveniente traer a colación la imagen de cadena de contenedores. Así, se debe considerar una cadena de cinco contenedores de la forma $w = w(0)w(1)w(2)w(3)w(4)$; resta por determinar el contenido de cada uno de los contenedores $w(i)$. Para el primer contenedor $w(0)$ se puede usar uno, y sólo uno, de tres símbolos posibles; así, se tienen tres elecciones diferentes. Hecha la elección para el primer contenedor, se deduce que nuevamente hay tres elecciones diferentes para escoger el símbolo del segundo contenedor $w(1)$. La elección hecha para el primer contenedor no afecta en nada a la elección que se va a realizar para el segundo contenedor. Este razonamiento se puede extender a los tres contenedores restantes. De esta manera, se concluye que se tienen $3^5 = 243$ formas diferentes de llenar los cinco contenedores, considerando tres símbolos diferentes. Esto es, existen 243 palabras diferentes de longitud 5 sobre el alfabeto A .

Considérese ahora la situación siguiente. ¿Cuántas palabras diferentes de longitud 5 sobre el alfabeto A , que tienen una a en la posición 2, se pueden construir? En este nuevo caso se deben considerar palabras de la forma $w = w(0)w(1)aw(3)w(4)$, por lo que se debe considerar los posibles contenidos de cuatro contenedores. Siguiendo el razonamiento usado en el primer ejemplo, se concluye que existen $3^4 = 81$ formas diferentes de llenar los cuatro contenedores en cuestión. Así pues, hay 81 palabras diferentes sobre A que satisfacen las condiciones señaladas.

Considérese un último ejemplo. ¿Cuántas palabras diferentes de longitud 5 sobre el alfabeto A que tienen exactamente dos a s se pueden construir? Este problema de conteo es más elaborado. Primero, es necesario escoger dos de los cinco contenedores, pues en ellos se podrán las dos a s requeridas. Segundo, en los tres contenedores restantes sólo se pueden usar los símbolos b o c , ya no se puede usar ni una a más. Así, se tienen $\binom{5}{2} = 10$ formas diferentes de seleccionar dos contenedores de los cinco que se tienen; y se tienen $2^3 = 8$ formas diferentes de colocar dos símbolos diferentes en tres contenedores. Se concluye pues que se tienen $10 \times 8 = 80$ palabras diferentes sobre A que satisfacen las condiciones señaladas.

Se puede concluir de lo recién discutido, junto con las ilustraciones presentadas en las secciones anteriores, que las palabras son entidades *combinatorias*. En la Parte 2 de estas notas se presentan varios ejemplos más elaborados de problemas de construcción de palabras.

3 Operaciones Para Palabras

Si para construir palabras particulares siempre se debe seguir el método general que se describió en la sección anterior, rápidamente se enfrentarían problemas pragmáticos muy difíciles. Por lo tanto, es conveniente contar con estrategias alternativas que permitan obtener palabras particulares a partir del conocimiento de la existencia otras palabras.

Lo anterior no es un problema exclusivo del estudio de las palabras, en realidad se presenta en todas las áreas de las matemáticas. Tal problema general se puede plantear como sigue. En el estudio de los objetos matemáticos siempre es valioso contar con mecanismos que permitan construir nuevos objetos a partir de casos conocidos de los objetos. Las operaciones, o transformaciones, cumplen este cometido. A continuación, se describen las dos principales operaciones que se han definido para las palabras.

3.1 Concatenación

Sea el alfabeto A y sean las palabras x , y sobre A , de longitudes n y m respectivamente. En particular, considérese que dichas palabras tienen la forma siguiente:

$$x = \begin{array}{|c|c|c|c|c|} \hline a_{j_0} & a_{j_1} & \cdots & a_{j_{n-2}} & a_{j_{n-1}} \\ \hline \end{array}$$

Esto es, $x = a_{j_0} a_{j_1} \cdots a_{j_{n-2}} a_{j_{n-1}}$, donde a_j es un símbolo arbitrario de A .

$$y = \begin{array}{|c|c|c|c|c|} \hline b_{j_0} & b_{j_1} & \cdots & b_{j_{m-2}} & b_{j_{m-1}} \\ \hline \end{array}$$

Esto es, $y = b_{j_0} b_{j_1} \cdots b_{j_{m-2}} b_{j_{m-1}}$, donde b_j es un símbolo arbitrario de A . Se define ahora a la palabra $x \cdot y$, o simplemente xy , sobre A , la cual se denomina la *concatenación* de x con y , de la manera siguiente:

$$x \cdot y = \begin{array}{|c|c|c|c|c|} \hline a_{j_0} & \cdots & a_{j_{n-1}} & b_{j_0} & \cdots & b_{j_{m-1}} \\ \hline 0 & & n-1 & n & & n+m-1 \end{array}$$

Se debe observar que la concatenación $x \cdot y$ es una palabra de longitud $n + m$; es decir, se tiene que $|xy| = |x| + |y| = n + m$. Con base en lo anterior, no es difícil establecer que $|xy|_a = |x|_a + |y|_a$ para todo símbolo a .

Así pues, la concatenación entre dos palabras es una operación ordenada que adiciona los símbolos de la segunda palabra, en el orden en el que aparecen en dicha palabra, al final de la primera palabra. En términos de la imagen de “contenedores”, se puede entender a la concatenación como la acción de añadir tantos “contenedores”

como tiene y -en el orden en el que aparecen en y , junto con los símbolos contenidos- después del último “contenedor” de x .

En términos de la notación funcional, la concatenación $x \cdot y$ se expresa como sigue:

$$\begin{aligned} x \cdot y(0) &= x(0) &= a_{j_0} \\ &\vdots \\ x \cdot y(n-1) &= x(n-1) &= a_{j_{n-1}} \\ x \cdot y(n) &= y(0) &= b_{j_0} \\ &\vdots \\ x \cdot y(n+m-1) &= y(m-1) &= b_{j_{m-1}} \end{aligned}$$

La relación funcional anterior se puede expresar de manera compacta como sigue:

$$\begin{aligned} x \cdot y(i) &= xy(i) = x(i) \text{ si } 0 \leq i \leq n-1; \\ x \cdot y(i) &= xy(i) = y(i-n) \text{ si } n \leq i \leq n+m-1. \end{aligned}$$

De manera explícita, la concatenación $x \cdot y$ es la palabra:

$$\begin{aligned} x \cdot y &= xy = x_{0_0} \dots x_{n-1_{n-1}} y_{0_n} \dots y_{m-1_{n+m-1}}. \\ x \cdot y &= xy = a_{j_0} \dots a_{j_{n-1}} b_{j_0} \dots b_{j_{m-1}}. \end{aligned}$$

En este caso, la notación x_{ij} indica que el símbolo x_i de la secuencia x se encuentra en la posición j de la nueva secuencia; algo similar ocurre para la notación y_{ij} .

A manera de ejemplo ilustrativo, considérese el alfabeto $A = \{a, b, c\}$ y las palabras sobre A siguientes: $x = bccaba$, $y = aabbccabc$. En términos de la notación funcional se tiene lo siguiente.

$$\begin{aligned} x &= x(0)x(1)x(2)x(3)x(4)x(5) = bccaba. \\ y &= y(0)y(1)y(2)y(3)y(4)y(5)y(6)y(7)y(8) = aabbccabc. \end{aligned}$$

Con esto, la concatenación $x \cdot y$ es la palabra:

$$\begin{aligned} x \cdot y &= xy(0)xy(1)xy(2)xy(3)xy(4)xy(5)xy(6)xy(7)xy(8)xy(9)xy(10)xy(11)xy(12)xy(13)xy(14) \\ &= x(0)x(1)x(2)x(3)x(4)x(5)y(0)y(1)y(2)y(3)y(4)y(5)y(6)y(7)y(8) \\ &= bccabaaabbccabc \end{aligned}$$

Por su parte, la concatenación $y \cdot x$ es la palabra:

$$\begin{aligned}
y \cdot x &= yx(0)yx(1)yx(2)yx(3)yx(4)yx(5)yx(6)yx(7)yx(8)yx(9)yx(10)yx(11)yx(12)yx(13)yx(14) \\
&= y(0)y(1)y(2)y(3)y(4)y(5)y(6)y(7)y(8)x(0)x(1)x(2)x(3)x(4)x(5) \\
&= aabbccabcbccaba
\end{aligned}$$

Claramente se tiene que $x \cdot y \neq y \cdot x$, a pesar de que $|x \cdot y| = |y \cdot x|$ y de que $|x \cdot y|_i = |y \cdot x|_i$ para todo símbolo i de A .

A partir de la definición de la concatenación de palabras, no es difícil establecer las propiedades que se listan a continuación.

Sean las palabras x, y, z sobre un alfabeto común A .

1. En general, la concatenación no es una operación conmutativa. Esto es, en general ocurre que $x \cdot y \neq y \cdot x$.
2. Existe un elemento identidad. Siempre se cumple que $x \cdot \varepsilon = \varepsilon \cdot x = x$.
3. La concatenación es una operación asociativa. Siempre se cumple que $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.

Por la propiedad asociativa, las palabras $x \cdot (y \cdot z)$ o $(x \cdot y) \cdot z$ se pueden denotar simplemente $x \cdot y \cdot z$ o xyz . Sólo es necesario remarcar que el orden es fundamental.

Cabe señalar que la notación $x \cdot y$ típicamente se utiliza cuando se desea resaltar el acto de la operación, mientras que la notación xy típicamente se usa para denotar el resultado de la operación.

3.2 Concatenación repetida

La definición de concatenación de palabras no impide concatenar una palabra consigo misma, por lo que ahora se explora un poco más este hecho.

Así pues, dada una palabra x sobre un alfabeto A se define:

$$x^0 = \varepsilon; \text{ y}$$

$$x^i = x \cdot x^{i-1}, \text{ para todo número natural } i \geq 1.$$

Se tiene que $x^1 = x \cdot x^0 = x \cdot \varepsilon = x$; $x^2 = x \cdot x^1 = x \cdot x$; $x^3 = x \cdot x^2 = x \cdot x \cdot x$, etc. La palabra x^i se denomina la *potencia i* de x y se debe interpretar como la palabra que resulta al concatenar i veces la palabra x consigo misma.

Es conveniente señalar que también se pudo haber empleado la definición $x^i = x^{i-1} \cdot x$, obteniéndose las mismas conclusiones. Por lo tanto, se puede emplear

cualquiera de las dos alternativas, siempre y cuando se sea consistente con la elección realizada.

A manera de ejemplo, considérese la palabra $x = abaca$ sobre $A = \{a, b, c\}$. Con tal elección se tiene lo siguiente:

$$(abaca)^0 = \varepsilon; (abaca)^1 = abaca; (abaca)^2 = abacaabaca, \text{ etc.}$$

En esta forma de notación, los paréntesis se usan para remarcar que la concatenación repetida -la potencia- se aplica a la palabra encerrada entre los paréntesis, y sólo a esa palabra.

Con base en la notación recién empleada es muy conveniente hacer una observación adicional. Considérese la palabra xy construida a partir de las palabras x, y sobre el alfabeto A . Se establece entonces lo siguiente.

$$(xy)^0 = \varepsilon.$$

$$(xy)^1 = (xy) \cdot (xy)^0 = (xy) \cdot \varepsilon = xy.$$

$$(xy)^2 = (xy) \cdot (xy)^1 = (xy) \cdot (xy) = xyxy.$$

$$(xy)^3 = (xy) \cdot (xy)^2 = (xy) \cdot (xyxy) = xyxyxy.$$

Etcétera.

Nuevamente se debe resaltar la importancia del uso de los paréntesis, ya que, de no hacerlo, podrían resultar confusiones importantes. Por decir, lo anterior se debe distinguir de lo siguiente:

$$xy^0 = x\varepsilon = x.$$

$$xy^1 = xy.$$

$$xy^2 = xyy.$$

Etcétera.

Resultan claras las coincidencias y las diferencias.

La definición de la potencia de una palabra permite establecer el resultado siguiente. Para cualquier palabra x y para todos los números naturales $i, j \geq 0$ se cumple la igualdad $x^{i+j} = x^i \cdot x^j = x^j \cdot x^i$. Esto es, se cumple una especie de ley de potencias, similar a la que se encuentra para los números.

3.3 Reverso

Haciendo uso de la imagen de cadena de “contenedores”, considérese la siguiente palabra x de longitud n sobre el alfabeto A :

$$x = \begin{array}{|c|c|} \hline a_{j_0} & a_{j_1} \\ \hline 0 & 1 \\ \hline \end{array} \cdots \begin{array}{|c|c|} \hline a_{j_{n-2}} & a_{j_{n-1}} \\ \hline n-2 & n-1 \\ \hline \end{array}$$

Se debe recordar que a_j es un símbolo arbitrario de A . Se define ahora la palabra *reversa* de x (o palabra *inversa* o *imagen especular* o, a veces, palabra *transpuesta*), denotada x^R , como la palabra siguiente:

$$x^R = \begin{array}{|c|c|} \hline a_{j_{n-1}} & a_{j_{n-2}} \\ \hline 0 & 1 \\ \hline \end{array} \cdots \begin{array}{|c|c|} \hline a_{j_1} & a_{j_0} \\ \hline n-2 & n-1 \\ \hline \end{array}$$

En términos de la notación funcional, se han establecido las igualdades siguientes:

$$\begin{aligned} x^R(0) &= x(n-1) = a_{j_{n-1}} \\ x^R(1) &= x(n-2) = a_{j_{n-2}} \\ &\vdots \\ x^R(n-2) &= x(1) = a_{j_1} \\ x^R(n-1) &= x(0) = a_{j_0} \end{aligned}$$

La descripción funcional anterior se expresa de manera compacta como sigue:

$$x^R(i) = x(n-1-i), \text{ para } 0 \leq i \leq n-1.$$

Esto es, la reversa de x es la palabra que resulta al escribir todos los símbolos de x en orden inverso.

A partir de la definición anterior, no resulta difícil verificar que $|x^R| = |x|$ y que $|x^R|_a = |x|_a$ para todo $a \in A$. En la operación reversa no se adicionan nuevos símbolos, simplemente los símbolos que ya había se acomodan en otro orden.

A manera de ejemplo, considérese el alfabeto $A = \{a, b\}$ y considérense las palabras $x = babbaa$ y $y = aabaabb$. Se tiene lo siguiente:

$$x = x(0)x(1)x(2)x(3)x(4)x(5) = babbaa.$$

$$\begin{aligned}
x^R &= x^R(0)x^R(1)x^R(2)x^R(3)x^R(4)x^R(5) \\
&= x(5)x(4)x(3)x(2)x(1)x(0) \\
&= aabbab.
\end{aligned}$$

$$y = y(0)y(1)y(2)y(3)y(4)y(5)y(6) = aabaabb.$$

$$\begin{aligned}
y^R &= y^R(0)y^R(1)y^R(2)y^R(3)y^R(4)y^R(5)y^R(6) \\
&= y(6)y(5)y(4)y(3)y(2)y(1)y(0) \\
&= bbaabaa.
\end{aligned}$$

Así $x^R = aabbab$ y $y^R = bbaabaa$. Claramente se cumple que $|x^R| = |x|$ y $|x^R|_j = |x|_j$ para $j = a, b$, e igualdades similares para y .

De la definición anterior se deduce de inmediato que $\varepsilon^R = \varepsilon$ y que $a^R = a$ para cualquier palabra unitaria a sobre A . Pero, en lo general, sucede que $x^R \neq x$; las excepciones a esto último resultan ser palabras con propiedades muy especiales, son las llamadas palabras *palíndromas*. En la Parte 2 de estas notas se harán algunas acotaciones acerca de las palabras palíndromas.

A partir de la definición de palabra reversa, no resulta difícil probar el resultado siguiente: $(x^R)^R = x$ para toda palabra x sobre algún alfabeto A . En efecto, si x es una palabra de longitud n , se tiene que:

$$\begin{aligned}
(x^R)^R(i) &= x^R(n-1-i) = x(i), \\
&\text{para } 0 \leq i \leq n-1.
\end{aligned}$$

Si una palabra es el resultado de la concatenación de otras palabras, su reverso tiene características especiales como ahora se muestra. Considérense tres palabras x, y, z sobre un alfabeto A tales que $x = yz$. Entonces la reversa de x está dada por:

$$x^R = (yz)^R = z^R y^R.$$

No es difícil probar este resultado. Supóngase que $|x| = n$ y que $|y| = n_1$ y $|z| = n_2$, con lo que $n = n_1 + n_2$. Se tiene pues que $x^R(i) = (yz)(n-1-i)$ para $0 \leq i \leq n-1$; nótese que $n-1 = n_1 + n_2 - 1$. Ahora bien, se tiene que $(yz)(i) = y(i)$ si $0 \leq i \leq n_1 - 1$, mientras que $(yz)(i) = z(i - n_1)$ si $n_1 \leq i \leq n_1 + n_2 - 1$. De esta manera, se tiene que:

$$x^R(i) = (yz)(n_1 + n_2 - 1 - i) = z(n_2 - 1 - i) \text{ si } 0 \leq i \leq n_2 - 1.$$

$$x^R(i) = (yz)(n_1 + n_2 - 1 - i) = y(n_1 + n_2 - 1 - i) \text{ si } n_2 \leq i \leq n_2 + n_1 - 1.$$

Si ahora se toma en cuenta lo siguiente:

$$y^R(i) = y(n_1 - 1 - i) \text{ para } 0 \leq i \leq n_1 - 1.$$

$$z^R(i) = z(n_2 - 1 - i) \text{ para } 0 \leq i \leq n_2 - 1.$$

El resultado buscado se obtiene al combinar los cuatro últimos resultados.

La operación reversa se puede expresar en términos de la operación de concatenación si se procede como se indica a continuación.

$$x^R = \begin{cases} \varepsilon & \text{si } x = \varepsilon; \\ y^R a & \text{si } x = ay. \end{cases}$$

En la igualdad anterior se debe entender que a es una palabra unitaria sobre el alfabeto empleado.

Por ejemplo, si se vuelve a considerar la palabra $x = babbbaa$ que se había utilizado previamente, la definición anterior conduce al cálculo siguiente:

$$\begin{aligned} x^R &= (babbbaa)^R \\ &= (abbbaa)^R b \\ &= (bbbaa)^R ab \\ &= (baa)^R bab \\ &= (aa)^R bbab \\ &= (a)^R abbab \\ &= (\varepsilon)^R aabbab \\ &= \varepsilon aabbab \\ &= aabbab. \end{aligned}$$

Este resultado coincide con lo que se había encontrado previamente.

Debido a esta forma de conceptualizar a la operación reversa, algunos autores consideran que sólo es necesario definir una operación para palabras: la concatenación.

3.4 Factorización

Primer enfoque

Sea x una palabra sobre un alfabeto A , la cual se puede expresar como $x = y_1 y_2 \dots y_r$, donde y_i es alguna palabra sobre el mismo alfabeto A para $1 \leq i \leq r$; podría suceder que algunas de palabras y_i sean vacías. Si sucede lo anterior, entonces se dice que x está *factorizada* en r factores y las r palabras y_i se denominan los *factores* de x .

Se puede establecer una ilustración gráfica del concepto anterior en términos muy similares a la idea de cadena de “contenedores”. En el caso presente, sin embargo, no se hablará de “contenedores” de símbolos sino de “bloques” de palabras completas. Así, una palabra factorizada se entiende como una cadena de “bloques”, tal y como se ilustra a continuación.

$$x = \begin{array}{|c|c|c|c|} \hline y_1 & y_2 & \cdots & y_r \\ \hline \end{array}$$

La palabra x está factorizada en r bloques, aunque algunos de ellos contengan a la palabra vacía. Con base en la ilustración anterior, se puede deducir el resultado siguiente:

$$|x| = \sum_{i=1}^r |y_i|.$$

Similarmente, también se puede deducir el resultado siguiente:

$$|x|_a = \sum_{i=1}^r |y_i|_a,$$

para todo $a \in A$.

De particular importancia en muchas aplicaciones de las palabras son las factorizaciones en términos de tres factores. En este caso, si $x = uvw$, entonces u y w se denominan *prefijo* y *sufijo*, respectivamente, de x , mientras que v suele denominarse *subpalabra* de x . En términos de la imagen de cadena de “bloques”, se tiene la ilustración siguiente:

$$x = \begin{array}{|c|c|c|} \hline u & v & w \\ \hline \end{array}$$

En la descripción anterior, algunos autores señalan que v es realmente una subpalabra sólo si u y w son palabras no vacías, aunque ciertamente no todos los autores se ajustan a esta forma de denominación.

Los autores que aceptan la factorización en tres factores señalan que los prefijos, subpalabras y sufijos son *propios* si sucede que $x \neq u$, $x \neq v$ y $x \neq w$, respectivamente. Esto es, se dice que u es un prefijo propio de x si sucede que $x = uy$, donde $y \neq \varepsilon$; esto

quiere decir que la subpalabra o el sufijo, pero no ambos, puede ser vacío. Una descripción similar aplica en los otros dos casos.

Segundo enfoque

Desafortunadamente, los conceptos de prefijo, subpalabra y sufijo no tienen un uso estandarizado y, por ello, las definiciones dadas en los párrafos anteriores no son usadas tal cual por todos los autores. Una variante de las definiciones anteriores, empleado principalmente en textos de los años 1980, es como sigue. Dadas dos palabras x, y sobre un alfabeto A , se dice que y es un *prefijo* de x si existe una palabra z , también sobre A , tal que $x = yz$. Este concepto se puede ilustrar con la notación funcional como sigue.

$$\begin{array}{rcl}
 x(0) & = & yz(0) & = & y(0) \\
 x(1) & = & yz(1) & = & y(1) \\
 & \vdots & & & \\
 x(m-1) & = & yz(m-1) & = & y(m-1) \\
 x(m) & = & yz(m) & = & z(0) \\
 x(m+1) & = & yz(m+1) & = & z(1) \\
 & \vdots & & & \\
 x(m+n-1) & = & yz(m+n-1) & = & z(n-1)
 \end{array}$$

El concepto anterior de prefijo puede ser interpretado como una relación de igualdad parcial por la izquierda entre las palabras x, y . En efecto, si $x = yz$, entonces cada uno de los primeros m símbolos de x necesariamente coinciden con cada uno de los m símbolos de y .

Se dice además que y es un *prefijo propio* de x si $y \neq \varepsilon$, $y \neq x$, donde y es un prefijo de x . Lo que diferencia esta definición de la definición de tres factores es que en este caso se acepta que ε siempre es un prefijo de cualquier palabra x (incluyendo a la misma palabra vacía) y una palabra x es siempre un prefijo de sí misma; son prefijos, pero no prefijos propios.

De manera análoga se definen los conceptos de sufijo y sufijo propio. Esto es, dadas dos palabras x, z sobre el alfabeto A , se dice que z es un *sufijo* de x si existe una palabra y sobre A tal que $x = yz$, y se dice que z es un *sufijo propio* de x si $z \neq \varepsilon$, $z \neq x$, donde z es un sufijo de x . Similarmente, z es un sufijo de x si se presenta una igualdad

parcial por la derecha entre cada uno de los últimos n símbolos de x y cada uno de los n símbolos de z .

A manera de ejemplo, considérese la palabra $w = cbcabaacb$ sobre el alfabeto $A = \{a, b, c\}$. Los prefijos de w son las palabras siguientes:

$\varepsilon, c, cb, cbc, cbca, cbcab, cbcaba, cbcabaa, cbcabaac, cbcabaacb.$

Salvo la primera y la última palabra del listado anterior, todas las demás palabras son prefijos propios de w . Por su parte, los sufijos de w son las palabras siguientes:

$\varepsilon, b, cb, acb, aacb, baacb, abaacb, cabaacb, bcabaacb, cbcabaacb.$

Nuevamente, salvo la primera y la última palabra, las demás palabras son sufijos propios de w .

Si y es un prefijo de x , entonces se puede escribir $y = xz^{-1}$ y se dice que y es el *cociente por la derecha* de x por z . Dicho de manera coloquial, la palabra y es lo que queda si a la palabra x se le quita su parte final z . Similar a esto, si z es un sufijo de x , entonces se escribe $z = y^{-1}x$ y se dice que z es el *cociente por la izquierda* de x por y . Así, coloquialmente se dice que la palabra z es lo que queda de la palabra x si se le quita su parte inicial y .

Algunos autores establecen además el concepto siguiente. Si una palabra w sobre un alfabeto A puede expresarse como $w = u_1v_1u_2v_2 \dots u_nv_n$, para algún número natural n , y palabras u_i, v_i , para $1 \leq i \leq n$, también sobre A , algunas de las cuales posiblemente sean vacías, pero no todas, entonces las palabras $u = u_1u_2 \dots u_n$ y $v = v_1v_2 \dots v_n$ son denominadas *subpalabras dispersas* de w . Esto es, una subpalabra dispersa no necesariamente es una cadena de símbolos consecutivos de la palabra w , sino que consiste en pequeñas cadenas tomadas de aquí y de allá de w , pero respetando el orden en el que aparecen los símbolos.

3.5 Otras Operaciones Para Palabras

En las secciones anteriores se han definidos dos operaciones básicas para palabras, la concatenación, que es una operación que permite añadir símbolos, y la reversa, que es una operación que altera el orden de los símbolos. Con estas operaciones se pueden

construir operaciones compuestas para palabras que permitan realizar otras manipulaciones de símbolos. A manera de ejemplo, se va a presentar una estrategia para construir una operación con la que se pueden quitar algunos símbolos de una palabra.

Considérese que $pref_n(u)$ es el prefijo de longitud n de u ; similarmente, $suf_n(u)$ es el sufijo de longitud n de u . Considérese ahora que u es una palabra de longitud n sobre el alfabeto A , y considérense los números n_1 y n_2 tales que $n_1 + n_2 = n$. Se puede entonces establecer la igualdad siguiente:

$$u = pref_{n_1}(u) \cdot suf_{n_2}(u).$$

Esto es, si $u = u(0) \cdots u(n_1 - 1)u(n_1) \cdots u(n_1 + n_2 - 1)$, entonces se tiene que:

$$pref_{n_1}(u) = u(0) \cdots u(n_1 - 1);$$

$$suf_{n_2}(u) = u(n_1) \cdots u(n_1 + n_2 - 1).$$

Coloquialmente, se puede decir que en $pref_{n_1}(u)$ se están considerando los primeros n_1 símbolos de u , mientras que en $suf_{n_2}(u)$ se están considerando los últimos n_2 símbolos de u .

Por ejemplo, sea la palabra $u = aabbccaabbcc$ sobre el alfabeto $A = \{a, b, c\}$; claramente se tiene que $|u| = 12$. Considérense ahora los números $n_1 = 4$ y $n_2 = 8$, con $n_1 + n_2 = 12$. Con esto, se tiene que $pref_4(u) = aabb$ y $suf_8(u) = ccaabbcc$. Claramente se tiene que $pref_4(u) \cdot suf_8(u) = aabb \cdot ccaabbcc = aabbccaabbcc = u$.

Aprovechando las palabras $pref$ y suf , se puede realizar la construcción:

$$u' = pref_{n_1-1}(u) \cdot suf_{n_2}(u).$$

De hecho, se tiene lo siguiente:

$$u = u(0) \cdots u(n_1 - 2)u(n_1 - 1)u(n_1) \cdots u(n_1 + n_2 - 1);$$

$$pref_{n_1-1}(u) = u(0) \cdots u(n_1 - 2);$$

$$suf_{n_2}(u) = u(n_1) \cdots u(n_1 + n_2 - 1);$$

$$u' = u(0) \cdots u(n_1 - 2)u(n_1) \cdots u(n_1 + n_2 - 1).$$

Nótese que $|u'| = n_1 + n_2 - 1 = n - 1$, esto es u' tiene un símbolo menos que u . Se puede decir que u' es la palabra que resulta de haber quitado un símbolo a la palabra u , precisamente el símbolo $u(n_1 - 1)$.

Así pues, se puede construir una función para palabras con la que se pueden eliminar algunos símbolos consecutivos de una palabra dada. Tal función, llamada *poda* (en inglés prune) de k símbolos y representada $Prune_{n_1, n_2}^k(u)$, se puede establecer como sigue:

Sea la palabra u sobre el alfabeto A , tal que $|u| = n$, con $n > 0$.

Sean los números n_1 y n_2 tales que $n_1 + n_2 = n$.

Sea el número $k > 0$ tal que $k < n_1$.

Se define $Prune_{n_1, n_2}^k(u) = pref_{n_1-k}(u) \cdot suf_{n_2}(u)$.

Esto es, $Prune_{n_1, n_2}^k(u)$ es la palabra que resulta de eliminar los últimos k símbolos del factor $pref_{n_1}(u)$ y concatenar la palabra obtenida con $suf_{n_2}(u)$. El resultado es una palabra que tiene k símbolos menos que la palabra original u .

A manera de ejemplo, considérese de nuevo la palabra $u = aabbccaabbcc$ sobre el alfabeto $A = \{a, b, c\}$. Con ella, se tiene lo siguiente:

$$\begin{aligned} Prune_{7,5}^3(aabbccaabbcc) &= pref_4(aabbccaabbcc) \cdot suf_5(aabbccaabbcc) \\ &= aabb \cdot abbcc \\ &= aabbabbcc. \end{aligned}$$

Procediendo de manera parecida a cómo se acaba de hacer, se pueden definir muchas otras funciones para palabras.

4 Algunos Conjuntos De Palabras

Se definen ahora a los conjuntos A^* (el cual se lee “a estrella”) como el conjunto de todas las palabras sobre A formadas por cero o más símbolos de A , y A^+ (el cual se lee “a más”) como el conjunto de todas las palabras sobre A formadas con por lo menos un símbolo A . Los conjuntos A^* y A^+ también se les denominan las *cerraduras* de A por la razón que se verá en otra parte de las notas; en particular, el conjunto A^* también se denomina el *lenguaje universal* sobre A . De las definiciones anteriores se deduce que

$A^* = A^+ \cup \{\varepsilon\}$. Nótese que A^* y A^+ son conjuntos infinitos, pero sus elementos son palabras finitas de símbolos.

Ahora bien, nótese que la concatenación de palabras es una operación binaria cerrada tanto en A^* como en A^+ . Esto es, si $u, v \in A^* (A^+)$, entonces $uv \in A^* (A^+)$. Además, para todas las palabras $u, v, w \in A^* (A^+)$ se cumple que $u(vw) = (uv)w = uvw$. Esto es, la concatenación es una operación binaria asociativa. También se cumple que $\varepsilon \in A^*$ (pero $\varepsilon \notin A^+$) y se cumple que $\varepsilon v = v\varepsilon = v$ para todo $v \in A^*$; esto es, ε es el elemento identidad de la operación de concatenación. De esta manera, el par $(A^*, \cdot)$ es un monoide y, de hecho, es el monoide libre generado por A . Por otro lado, el par $(A^+, \cdot)$ es un semigrupo. El enfoque algebraico de las palabras será discutido muy brevemente a continuación.

5 Semigrupos Y Monoides

Un sistema algebraico abstracto está definido en términos de un conjunto arbitrario y un cierto número de operaciones sobre el conjunto. Se asume que estas operaciones tienen propiedades que son tomadas como axiomas del sistema. Ninguna otra propiedad es asumida de inicio. Cualquier conclusión válida que se deduzca a partir de los axiomas es un teorema del sistema. Tales teoremas son válidos para cualquier sistema algebraico en el que se satisfagan los axiomas.

En este caso la muy breve discusión siguiente se va a restringir a un par de sistemas algebraicos con una operación binaria. Si no se imponen restricciones sobre la operación binaria se obtiene un sistema algebraico que es demasiado general como para que resulte de ayuda. Luego, se van a considerar sistemas algebraicos que consisten de un conjunto y una operación binaria asociativa sobre el conjunto. Después de describir tales sistemas, se consideran aquellos sistemas que poseen una operación binaria asociativa y un elemento identidad. Estos sistemas algebraicos son denominados semigrupos y monoides respectivamente y tienen aplicaciones fundamentales en las áreas de la aritmética computacional y en las teorías de los lenguajes formales y de las máquinas secuenciales.

Sea S un conjunto no vacío y sea $\circ$ una operación binaria sobre S . El sistema algebraico $(S, \circ)$ se denomina un *semigrupo* si la operación $\circ$ es asociativa. En otras palabras, $(S, \circ)$ es un semigrupo si para cualesquiera $x, y, z \in S$ sucede que $(x \circ y) \circ z = x \circ (y \circ z)$.

Un semigrupo puede ser descrito proporcionando la tabla de composición de la operación $\circ$ cuando S es finito y cuando su orden no es demasiado grande; en el caso contrario, puede ser descrito por medio de alguna regla para la operación $\circ$ sobre S . Nótese que un semigrupo puede tener o no un elemento identidad con respecto a la operación $\circ$. Si existe un elemento identidad, entonces se tiene la definición siguiente.

Un semigrupo $(M, \circ)$ con un elemento identidad con respecto a la operación $\circ$ se denomina un *monoide*. En otras palabras, un sistema algebraico $(M, \circ)$ se denomina un monoide si para cualesquiera $x, y, z \in M$ sucede que $(x \circ y) \circ z = x \circ (y \circ z)$, y existe un elemento $e \in M$ tal que para cualquier $x \in M$ se cumple que $e \circ x = x \circ e = x$.

Se puede mostrar que si existe un elemento identidad para cualquier operación binaria, entonces debe ser único. De esta manera, un monoide tiene un elemento único, o distinguible, denominado la *identidad* del monoide. A veces es conveniente representar a un monoide en la forma $(M, \circ, e)$ para enfatizar el hecho de que e es un elemento distinguido para tal monoide.

Para un monoide $(M, *)$, la existencia del elemento identidad garantiza que en la tabla de composición no van a existir dos columnas idénticas, ya que si e es la identidad, entonces para cualesquiera $a_i, a_j \in M$, con $a_i \neq a_j$, se tiene que $e * a_i = a_i \neq e * a_j = a_j$. De manera similar, debido a que $a_i * e = a_i$, no existen dos renglones idénticos. Nótese que esta propiedad se cumple para todo monoide independientemente de si puede ser descrito por una tabla de composición o no.

De estas definiciones debe resultar claro que el sistema $(V^+, \cdot)$, donde V es un alfabeto y $\cdot$ es la operación (binaria) de concatenación de palabras, es un semigrupo; la operación de concatenación es asociativa. También es claro que este sistema no contiene un elemento identidad, esto es, no hay ninguna palabra γ en V^+ tal que para

cualquier palabra ω de V^+ se cumpla que $\gamma \cdot \omega = \omega \cdot \gamma = \omega$. A diferencia de esto, el sistema $(V^*, \cdot)$ es un monoide, ya que el elemento $\varepsilon \in V^*$, la palabra vacía, es un elemento identidad: para todo $\omega \in V^*$ se cumple que $\varepsilon \cdot \omega = \omega \cdot \varepsilon = \omega$.

De la definición de la operación (binaria) de concatenación, resulta claro que esta operación no es conmutativa, excepto para el caso en el que el alfabeto V esté formado por un único símbolo. También resulta que ningún elemento de V^* es invertible, a excepción de la palabra vacía ε . (Si $*$ es una operación binaria sobre el conjunto X , entonces un elemento $a \in X$ se denomina invertible si existe un elemento $a^{-1} \in X$ tal que $a^{-1} * a = a * a^{-1} = e$, donde e es el elemento identidad.) Similarmente, ningún elemento de V^* es idempotente, nuevamente a excepción de la palabra vacía. (Si $*$ es una operación binaria sobre el conjunto X , entonces un elemento $a \in X$ se denomina idempotente si $a * a = a$.) Sin embargo, en el sistema $(V^*, \cdot)$ todo elemento satisface la propiedad de cancelación. Para ver esto, considérense dos palabras α y β de V^* tales que $\alpha = \alpha_1 x$ y $\beta = \beta_1 x$, donde x es una palabra unitaria (un símbolo). Si $\alpha = \beta$, entonces $\alpha_1 x = \beta_1 x$, lo cual implica que $\alpha_1 = \beta_1$.

El hecho de que el monoide $(V^*, \cdot)$ sea, en lo general, no conmutativo, permite establecer muchas propiedades un tanto extrañas que no se observan en los sistemas conmutativos. Esto hace que el álgebra de palabras sea un área muy interesante de estudiar en el contexto de la teoría de semigrupos y monoides. Estos temas, sin embargo, ya pertenecen a un curso intermedio o avanzado de lenguajes formales.

PARTE 2

EJEMPLOS

1. PALABRAS

Problemas de estructura y combinatoria

1. Sea el alfabeto $A = \{a, b, c, d, e\}$. ¿Cuántas palabras de A^* tienen una longitud no mayor a diez?

Se está preguntando por el número de palabras que forman al conjunto siguiente:

$$A^{\leq 10} = \bigcup_{i=0}^{10} A^i.$$

Cada conjunto A^i está formado por todas las palabras de longitud i sobre el alfabeto dado. Se sabe que el conjunto A^0 está formado por una única palabra, la palabra vacía. Para los restantes conjuntos, cada una de sus palabras tiene la forma siguiente:

$$x_{j_0} x_{j_1} \dots x_{j_{i-1}}, \text{ donde cada } x_j \text{ es alguno de los símbolos de } A.$$

En este caso, la notación x_{i_j} indica que el símbolo x_i del alfabeto se encuentra en la posición j de la secuencia. Ahora bien, en la primera posición de la palabra se puede colocar cualquiera de los símbolos del alfabeto; hay cinco elecciones posibles. Colocado el primer símbolo, en la segunda posición se puede colocar nuevamente a cualquiera de los símbolos del alfabeto; y nuevamente hay cinco elecciones posibles. Resulta entonces que la elección de un símbolo para una posición de la cadena no depende de las elecciones realizadas previamente, y no influye sobre las elecciones que se van a realizar después. Todas las elecciones son independientes entre sí. Así, el número de palabras que contiene un conjunto A^i es 5^i . Y, con esto, el número de palabras que contiene $A^{\leq 10}$ es:

$$1 + \sum_{i=1}^{10} 5^i = 1 + 5^1 + 5^2 + \dots + 5^{10}.$$

El conjunto es finito, pero ciertamente muy grande.

Ejercicio. ¿Cómo cambia el número de palabras si disminuye el número de símbolos del alfabeto? ¿Y si aumenta el número de símbolos?

En los siguientes ejemplos se considera el alfabeto $A = \{0, 1, 2, 3\}$.

2. Considere las secuencias de A^* de longitud nueve que no tienen ceros. A estas secuencias se les añade una posición más exactamente al final y en esa posición se coloca un cero. ¿Cuál es la cantidad de secuencias nuevas que se obtienen?

En primer término, se están considerando secuencias de la forma siguiente:

$$x_{i_0} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5} x_{i_6} x_{i_7} x_{i_8}.$$

Aquí cada símbolo x_i pertenece al subconjunto $\{1, 2, 3\}$. Dado que la elección de los símbolos es independiente, se tendrán 3^9 secuencias diferentes.

En segundo término, se consideran secuencias de la forma siguiente:

$$x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4}x_{i_5}x_{i_6}x_{i_7}x_{i_8}0.$$

Aquí cada uno de los primeros nueve símbolos puede ser elegido de $\{1, 2, 3\}$. Nótese que nuevamente se tienen 3^9 secuencias diferentes.

La cantidad de secuencias de ambas clases es la misma, pero se trata de secuencias diferentes: unas son de longitud nueve, las otras son de longitud diez; no pueden ser secuencias iguales.

Ejercicio, ¿Cambiarían las cantidades si la nueva posición se coloca en cualquier otro lugar de la secuencia original?

3. ¿Cuántas secuencias de A^* de longitud diez tienen a lo más un solo cero?

Se está preguntando por secuencias de la forma general siguiente:

$$x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4}x_{i_5}x_{i_6}x_{i_7}x_{i_8}x_{i_9}.$$

Aquí, cada $x_i \in A$. En particular, se pregunta por secuencias de dos tipos: unas en las que no se encuentra ningún cero, otras en las que se encuentra exactamente un cero. En el primer caso, cada símbolo x_i debe pertenecer al subconjunto $\{1, 2, 3\}$ y, en consecuencia, en cada posición de la secuencia se puede elegir uno de tres símbolos posibles. Así, se tienen 3^{10} diferentes secuencias sin un cero. En el segundo caso, una, y solo una de las posiciones, debe contener un cero; se tienen 10 diferentes opciones. Las nueve restantes posiciones pueden contener uno de los símbolos $\{1, 2, 3\}$; estas nueve posiciones se pueden construir de 3^9 diferentes formas. En total, se pueden construir $10 \cdot 3^9$ secuencias posibles con un solo cero. De esta manera, se tienen $3^{10} + 10 \cdot 3^9 = 13 \cdot 3^9$ secuencias diferentes con a lo más un cero.

Ejercicio. ¿Cómo cambia el número de secuencias si la longitud de las secuencias disminuye en una unidad? ¿Y si aumenta en una unidad?

4. Determínese el número de secuencias de A^* de longitud diez con exactamente dos ceros.

Para empezar, de una secuencia de longitud diez se deben escoger exactamente dos posiciones. Tales posiciones se pueden elegir de:

$$\binom{10}{2} = \frac{10!}{8! \cdot 2!} = \frac{10 \cdot 9}{2} = 45 \text{ formas diferentes.}$$

Las restantes ocho posiciones deben contener un símbolo de $\{1, 2, 3\}$. Esto nos da 3^8 combinaciones diferentes. Así, se tienen $45 \cdot 3^8$ secuencias diferentes de longitud diez con exactamente dos ceros.

Ejercicio. ¿En cuánto cambia la cantidad de secuencias si la longitud disminuye en una unidad? ¿Y si la longitud aumenta en una unidad?

5. Determínese el número de secuencias de A^* de longitud diez con un número par de ceros.

Se debe considerar el número de secuencias de cada una de las clases siguientes. Secuencias sin ceros, secuencias con dos, cuatro, seis, ocho y diez ceros. Para el primer caso se tienen 3^{10} diferentes secuencias. Para las clases siguientes primero se deben considerar las diferentes posiciones en las que se pueden colocar los ceros. Así pues, se tienen las combinaciones siguientes:

$$\binom{10}{2} = \frac{10!}{8! \cdot 2!} = \frac{10 \cdot 9}{2} = 45 \text{ combinaciones diferentes.}$$

$$\binom{10}{4} = \frac{10!}{6! \cdot 4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2} = 210 \text{ combinaciones diferentes.}$$

$$\binom{10}{6} = \frac{10!}{4! \cdot 6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2} = 210 \text{ combinaciones diferentes.}$$

$$\binom{10}{8} = \frac{10!}{2! \cdot 8!} = \frac{10 \cdot 9}{2} = 45 \text{ combinaciones diferentes.}$$

Y sólo hay una manera de elegir diez posiciones de diez posibles. Con estos resultados, se tienen las cantidades de secuencias siguientes:

$45 \cdot 3^8$ secuencias diferentes con dos ceros.

$210 \cdot 3^6$ secuencias diferentes con cuatro ceros.

$210 \cdot 3^4$ secuencias diferentes con seis ceros.

$45 \cdot 3^2$ secuencias diferentes con ocho ceros.

1 secuencias con diez ceros.

De esta manera, el número total de secuencias con un número par de ceros es:

$$1 + 45 \cdot (3^8 + 3^2) + 210 \cdot (3^6 + 3^4) + 3^{10}.$$

Ejercicio. Comparar la cantidad recién obtenida con la cantidad total de secuencias de longitud diez.

6. Determine el número de secuencias de longitud diez con exactamente un cero y un uno.

Primero se deben seleccionar dos posiciones de las diez posiciones de la secuencia, las dos posiciones donde se colocarán el cero y el uno. Como ya se ha calculado previamente, hay 45 combinaciones posibles. Nótese que el orden en el que se coloquen el cero y el uno no es relevante, se puede colocar primero el cero y después el uno, o viceversa. Así, hay $2 \cdot 45 \cdot 3^8$ secuencias diferentes con un cero y un uno.

Ejercicio. ¿Cómo cambian las cantidades de secuencias si aumenta o disminuye el número de símbolos del alfabeto?

7. Determine el número de secuencias de longitud diez con exactamente un cero y un uno, donde el uno siempre aparece antes que el cero.

Para empezar, se debe establecer la posición donde se coloca el único uno de la secuencia. Dada esta elección, se selecciona ahora la posición donde se coloca el cero. Hechas las dos elecciones, en las ocho restantes posiciones se colocan los dos restantes símbolos en todas las combinaciones posibles. Procediendo de esta manera se obtienen los resultados siguientes:

El uno se coloca en la primera posición; el cero puede ser colocado en cualquiera de las nueve posiciones restantes. Hay 9 combinaciones posibles.

El uno se coloca en la segunda posición; el cero no puede ser colocado en la primera posición, sólo puede ser colocado en alguna de las restantes ocho posiciones. Hay 8 combinaciones posibles.

El uno se coloca en la tercera posición; el cero no puede ser colocado en ninguna de las dos primeras posiciones, sólo puede ser colocado en alguna de las siete posiciones restantes. Hay 7 combinaciones posibles.

Repitiendo el razonamiento anterior se puede concluir que hay $9 + 8 + 7 + 6 + 5 + 4 + 3 + 2 + 1 = 45$ combinaciones posibles para acomodar los dos símbolos cero y uno. De esta manera, hay $45 \cdot 3^8$ secuencias diferentes de longitud con las características señaladas.

Ejercicio. ¿Qué relación existe entre este ejemplo y el anterior?

2. OPERACIONES PARA PALABRAS

Problemas de operaciones para palabras

En los siguientes ejemplos se considera un alfabeto arbitrario A .

8. ¿Cuál es el resultado de concatenar palabras de longitud seis caracterizadas por presentar un solo cero con palabras de longitud cuatro también caracterizadas por presentar un solo cero?

Por un lado, se están considerando secuencias de la forma:

$$w_1 = x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4}x_{i_5}.$$

(Recuerde que la notación x_{i_j} indica que el símbolo x_i del alfabeto se encuentra en la posición j de la secuencia.) Estas secuencias se caracterizan por que en una de sus posiciones se encuentra un cero, mientras que en las restantes cinco posiciones se encuentran los otros tres símbolos, en todos los órdenes posibles. Por el otro lado, las otras secuencias consideradas son de la forma:

$$w_2 = y_{i_0}y_{i_1}y_{i_2}y_{i_3}.$$

Estas secuencias también contienen un solo cero, una posición está ocupada por ese símbolo, las restantes tres posiciones contienen a los símbolos restantes en todos los órdenes posibles.

Las concatenaciones señaladas se pueden realizar de dos formas diferentes. En la primera se obtiene la secuencia siguiente:

$$w_1w_2 = x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4}x_{i_5}y_{i_0}y_{i_1}y_{i_2}y_{i_3}.$$

(La notación x_{i_3} , por ejemplo, indica que el símbolo x_i del alfabeto, que se encontraba en la posición 3 de la secuencia original, ahora se encuentra en la posición 3 de la

nueva secuencia; por su parte, $y_{i_{28}}$ indica que el símbolo y_i del alfabeto, que se encontraba en la posición 2 de la secuencia original, ahora se encuentra en la posición 8 de la nueva secuencia.) Estas secuencias son de longitud diez y se caracterizan por contener exactamente dos ceros, uno de los cuales se encuentra en alguna de las primeras seis posiciones, mientras que el segundo se encuentra en alguna de las últimas cuatro posiciones. Nótese que los dos ceros no pueden estar ambos en las primeras seis posiciones, ni en las últimas cuatro posiciones. (Ejercicio. Determine la validez de esta afirmación.)

La segunda forma de realizar la concatenación produce la secuencia siguiente:

$$w_2 w_1 = y_{i_{00}} y_{i_{11}} y_{i_{22}} y_{i_{33}} x_{i_{04}} x_{i_{15}} x_{i_{26}} x_{i_{37}} x_{i_{48}} x_{i_{59}}.$$

Nuevamente se obtienen secuencias de longitud diez que contienen exactamente dos ceros. El primer cero se encuentra en alguna de las primeras cuatro posiciones, mientras que el segundo cero se encuentra en alguna de las últimas seis posiciones. Nótese que ahora los dos ceros no pueden estar ambos en las primeras cuatro posiciones, ni en las últimas seis posiciones. (Ejercicio. Verifique también esta afirmación.)

9. ¿En qué condiciones son iguales las dos clases de secuencias concatenadas obtenidas en el ejemplo anterior?

Se tiene la secuencia:

$$w_1 w_2 = x_{i_{00}} x_{i_{11}} x_{i_{22}} x_{i_{33}} x_{i_{44}} x_{i_{55}} y_{i_{06}} y_{i_{17}} y_{i_{28}} y_{i_{39}}.$$

Y se tiene la secuencia:

$$w_2 w_1 = y_{i_{00}} y_{i_{11}} y_{i_{22}} y_{i_{33}} x_{i_{04}} x_{i_{15}} x_{i_{26}} x_{i_{37}} x_{i_{48}} x_{i_{59}}.$$

Las dos secuencias son de la misma longitud, por lo que son iguales si, y sólo si, tienen los mismos símbolos en las mismas posiciones. En particular, se sabe que la segunda secuencia $w_2 w_1$ tiene un cero en alguna de sus primeras cuatro posiciones. Así, en la misma posición de la primera secuencia $w_1 w_2$ también se debe encontrar un cero, si es que las secuencias van a ser iguales. Recordemos que esa primera secuencia se obtuvo a partir de una secuencia w_1 que tiene un cero en alguna de sus seis posiciones. Si se requiere que dicho cero se encuentre en alguna de las primeras cuatro

posiciones, entonces las dos posiciones restantes no pueden tener un cero. Adicionalmente, se sabe que en la primera secuencia w_1w_2 se encuentra otro cero en alguna de sus últimas cuatro posiciones. Pero esas cuatro posiciones corresponden a las cuatro últimas posiciones de la secuencia original w_1 . Esto es, la secuencia w_1 debe tener un cero en alguna de sus primeras cuatro posiciones y en algunas de sus últimas cuatro posiciones. Esto implica, que la secuencia original w_1 debe tener un cero en la tercera o en la cuarta posiciones. Así pues, se deben considerar las dos clases de secuencias siguientes:

$$(w_1w_2)' = x_{i_{00}}x_{i_{11}}0_{22}x_{i_{33}}x_{i_{44}}x_{i_{55}}y_{i_{06}}y_{i_{17}}y_{i_{28}}y_{i_{39}}.$$

$$(w_1w_2)'' = x_{i_{00}}x_{i_{11}}x_{i_{22}}0_{33}x_{i_{44}}x_{i_{55}}y_{i_{06}}y_{i_{17}}y_{i_{28}}y_{i_{39}}.$$

(Recuerde que, por la forma de numeración adoptada, las posiciones 2 y 3 en la secuencia original corresponden a las entradas tercera y cuarta de la secuencia.) Con esto, la segunda secuencia w_2w_1 debe ser de alguna de las dos formas siguientes:

$$(w_2w_1)' = y_{i_{00}}y_{i_{11}}0_{22}y_{i_{33}}x_{i_{04}}x_{i_{15}}0_{26}x_{i_{37}}x_{i_{48}}x_{i_{59}}.$$

$$(w_2w_1)'' = y_{i_{00}}y_{i_{11}}y_{i_{22}}0_{33}x_{i_{04}}x_{i_{15}}x_{i_{26}}0_{37}x_{i_{48}}x_{i_{59}}.$$

Y con todo lo anterior, la primera secuencia w_1w_2 debe ser de alguna de las dos formas siguientes:

$$(w_1w_2)' = x_{i_{00}}x_{i_{11}}0_{22}x_{i_{33}}x_{i_{44}}x_{i_{55}}y_{i_{06}}y_{i_{17}}0_{28}y_{i_{39}}.$$

$$(w_1w_2)'' = x_{i_{00}}x_{i_{11}}x_{i_{22}}0_{33}x_{i_{44}}x_{i_{55}}y_{i_{06}}y_{i_{17}}y_{i_{28}}0_{39}.$$

Nótese pues que las dos clases de secuencias concatenadas no pueden ser iguales.

Ejercicio. Repita los Ejemplos 8 y 9 considerando ahora dos palabras de longitud cinco, cada una de ellas caracterizada por tener un único cero.

10. Calcule los reversos de las dos palabras concatenadas obtenidas en los dos ejemplos anteriores.

Recuerde que el reverso de una palabra concatenada se calcula de la forma siguiente:

$$(wv)^R = v^Rw^R.$$

De esta manera, para la palabra concatenada w_1w_2 se obtiene lo siguiente:

$$\begin{aligned}
(w_1 w_2)^R &= w_2^R w_1^R \\
&= (y_{i_0} y_{i_1} y_{i_2} y_{i_3})^R (x_{i_0} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5})^R \\
&= y_{i_3} y_{i_2} y_{i_1} y_{i_0} x_{i_5} x_{i_4} x_{i_3} x_{i_2} x_{i_1} x_{i_0}.
\end{aligned}$$

Nótese que, en lo general, $(w_1 w_2)^R \neq w_2 w_1$, aunque ambas palabras tienen la misma longitud y contienen a los mismos símbolos, pero dichos símbolos están ordenados de forma diferente en cada palabra. Las dos palabras aludidas contienen exactamente dos ceros, donde uno de los ceros se encuentra en alguna de las primeras cuatro posiciones, mientras que el segundo cero se encuentra en alguna de las seis últimas posiciones.

Por su parte, para la palabra concatenada $w_2 w_1$ se obtiene lo siguiente:

$$\begin{aligned}
(w_2 w_1)^R &= w_1^R w_2^R \\
&= (x_{i_0} x_{i_1} x_{i_2} x_{i_3} x_{i_4} x_{i_5})^R (y_{i_0} y_{i_1} y_{i_2} y_{i_3})^R \\
&= x_{i_5} x_{i_4} x_{i_3} x_{i_2} x_{i_1} x_{i_0} y_{i_3} y_{i_2} y_{i_1} y_{i_0}.
\end{aligned}$$

Nuevamente nótese que, en lo general, $(w_2 w_1)^R \neq w_1 w_2$, aunque nuevamente ambas palabras tienen la misma longitud y contienen a los mismos símbolos, pero dichos símbolos están ordenados de forma diferente en cada palabra. Las dos palabras aludidas también contienen exactamente dos ceros, donde uno de los ceros se encuentra en alguna de las primeras seis posiciones, mientras que el segundo cero se encuentra en alguna de las cuatro últimas posiciones.

Ejercicio. Repita este ejemplo, pero ahora considere las palabras de longitud cinco que uso en el ejercicio del Ejemplo 9.

3. FACTORIZACIÓN

Problemas de factorización

Considere el alfabeto $A = \{a, b, \dots, z\}$ (las letras minúsculas del español) y considere la secuencia $\omega = \text{aardvark}$ sobre este conjunto.

11. Liste los prefijos y sufijos de la palabra ω en orden de longitud.

Los prefijos de ω son:

$\varepsilon, a, aa, aar, aard, aardv, aardva, aardvar, aardvark.$

Los sufijos de ω son:

$\varepsilon, k, rk, ark, vark, dvark, rdvark, ardvark, aardvark.$

Ejercicio. Distinga los prefijos y sufijos propios de los no propios.

12. Liste los prefijos y sufijos de la palabra ω^R en orden de longitud.

La palabra ω^R es *kravdraa*. Sus prefijos son:

$\varepsilon, k, kr, kra, krav, kravd, kravdr, kravdra, kravdraa.$

Sus sufijos son:

$\varepsilon, a, aa, raa, draa, vdraa, avdraa, ravdraa, kravdraa.$

Ejercicio. ¿Qué relación existe entre los resultados de los dos últimos ejemplos?

Sobre el mismo alfabeto A anterior, considere ahora a la palabra $\omega = \text{sometemos}$.

13. Liste los prefijos y sufijos de ω en orden de longitud.

Los prefijos de ω son:

$\varepsilon, s, so, som, some, somet, somete, sometem, sometemo, sometemos.$

Los sufijos de ω son:

$\varepsilon, s, os, mos, emos, temos, etemos, metemos, ometemos, sometemos.$

Ejercicio. ¿Qué relación existe entre los prefijos y los sufijos?

4. PALÍNDROMOS

Problemas con palabras palíndromas

En el ejemplo siguiente se está considerando un alfabeto arbitrario A con por lo menos dos símbolos.

14. Construya la forma general de las palabras palíndromas sobre A de longitud no mayor a cinco.

Recuerde que una palabra w es palíndroma si satisface la igualdad $w = w^R$. A partir de esto, se tienen que considerar los casos siguientes:

a. Para la palabra vacía se tiene que $\varepsilon = \varepsilon^R = \varepsilon$. La palabra vacía es palíndroma.

b. Una palabra unitaria tiene la forma general x_{i_0} , donde x_i es un símbolo del alfabeto dado. El reverso de la palabra anterior es $(x_{i_0})^R = x_{i_0_0}$. Recuerde, la notación x_{i_0} indica que el símbolo x_i del alfabeto se encuentra en la posición 0 de la secuencia; por su parte, la notación $x_{i_0_0}$ indica que el símbolo x_i que ocupa la posición 0 en la secuencia original, ahora se encuentra en la posición 0 de la nueva secuencia; así pues, en las posiciones 0 de las dos palabras se encuentra el mismo símbolo x_i . De esta manera, todas las palabras unitarias x_{i_0} son palíndromas.

c. Las palabras de longitud dos tienen la forma general $x_{i_0}x_{i_1}$, donde x_i es un símbolo del alfabeto dado. El reverso de estas palabras tiene la forma general $(x_{i_0}x_{i_1})^R = x_{i_1_0}x_{i_0_1}$. Las dos palabras son iguales si, y sólo si, se cumple que $x_{i_0} = x_{i_1_0}$ (el símbolo en la posición 0 de la palabra original coincide con el símbolo de la posición 1 de la misma palabra) y $x_{i_1} = x_{i_0_1}$ (el símbolo en la posición 1 de la palabra original coincide con el símbolo en la posición 0 de la misma palabra). Esto es, se requiere que $x_{i_0} = x_{i_1}$. De esta manera, las palabras palíndromas de longitud dos tienen la forma general x_0x_1 ; son las palabras con dos símbolos repetidos.

d. Las palabras de longitud tres tienen la forma general $x_{i_0}x_{i_1}x_{i_2}$, donde x_i es un símbolo del alfabeto dado. El reverso de estas palabras tiene la forma general:

$$(x_{i_0}x_{i_1}x_{i_2})^R = x_{i_2_0}x_{i_1_1}x_{i_0_2}.$$

Las dos palabras son iguales si, y sólo si, se cumple que $x_{i_0} = x_{i_2_0}$, $x_{i_1} = x_{i_1_1}$ y $x_{i_2} = x_{i_0_2}$. Esto es, se requiere únicamente que $x_{i_0} = x_{i_2}$. De esta manera, las palabras palíndromas de longitud tres tienen la forma general $x_0x_{i_1}x_2$, donde x_i es cualquier símbolo del alfabeto; son las palabras en las que el primer y el tercer símbolo coinciden.

e. Las palabras de longitud cuatro tienen la forma general $x_{i_0}x_{i_1}x_{i_2}x_{i_3}$, donde x_i es un símbolo del alfabeto dado. El reverso de estas palabras tiene la forma general:

$$(x_{i_0}x_{i_1}x_{i_2}x_{i_3})^R = x_{i_3_0}x_{i_2_1}x_{i_1_2}x_{i_0_3}.$$

Las dos palabras son iguales si, y sólo si, $x_{i_0} = x_{i_{30}}, x_{i_1} = x_{i_{21}}, x_{i_2} = x_{i_{12}}$ y $x_{i_3} = x_{i_{03}}$.

Esto es, se requiere que $x_{i_0} = x_{i_3}$ y que $x_{i_1} = x_{i_2}$. De esta manera, las palabras palíndromas de longitud cuatro tienen la forma general $x_{i_0}y_{i_1}y_{i_2}x_{i_3}$, donde x_i, y_i son símbolos, no necesariamente iguales, del alfabeto dado. Las palabras palíndromas de longitud cuatro son las palabras en las que el primer y el cuarto símbolo coinciden, y el segundo y tercer símbolo también coinciden.

f. Las palabras de longitud cinco tienen la forma general $x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4}$, donde x_i es un símbolo del alfabeto dado. El reverso de estas palabras tiene la forma general:

$$(x_{i_0}x_{i_1}x_{i_2}x_{i_3}x_{i_4})^R = x_{i_{40}}x_{i_{31}}x_{i_{22}}x_{i_{13}}x_{i_{04}}.$$

Las dos palabras son iguales si, y sólo si, se cumplen las igualdades:

$$x_{i_0} = x_{i_{40}}, x_{i_1} = x_{i_{31}}, x_{i_2} = x_{i_{22}}, x_{i_3} = x_{i_{13}}, x_{i_4} = x_{i_{04}}.$$

Esto es, se requiere que $x_{i_0} = x_{i_4}$ y que $x_{i_1} = x_{i_3}$. De esta manera, las palabras palíndromas de longitud cinco tienen la forma general $x_{i_0}y_{i_1}z_{i_2}y_{i_3}x_{i_4}$, donde x_i, y_i, z_i son símbolos, no necesariamente iguales, del alfabeto dado. Las palabras palíndromas de longitud cinco son las palabras en las que el primer y el quinto símbolo coinciden, y el segundo y cuarto símbolo también coinciden; el tercer símbolo es arbitrario.

Resumiendo. Las palabras palíndromas cuyas longitudes van de cero hasta cinco tienen las formas generales siguientes:

Sólo hay una palabra palíndroma de longitud cero, la palabra vacía ε .

Todas las palabras unitarias son palíndromas; todas son de la forma general a , donde a es un símbolo arbitrario del alfabeto.

Todas las palabras palíndromas de longitud dos son de la forma general aa , donde a es un símbolo arbitrario del alfabeto.

Todas las palabras palíndromas de longitud tres son de la forma general aba , donde a y b son símbolos del alfabeto, no necesariamente diferentes entre sí.

Todas las palabras palíndromas de longitud cuatro son de la forma general $abba$, donde a y b son símbolos del alfabeto, no necesariamente diferentes entre sí.

Todas las palabras palíndromas de longitud cinco son de la forma general $abcba$, donde a , b y c son símbolos del alfabeto, no necesariamente diferentes entre sí.

A manera de verificación: $abcba = (abcba)^R = abcba$.

Ejercicio. Continúe estos resultados hasta considerar palabras palíndromas de longitud doce.

15. Se considera ahora el alfabeto $A = \{0, 1, 2, 3\}$. Con base en los resultados del ejemplo anterior, obtenga las palabras palíndromas sobre A de longitudes no superiores a cinco.

a. Solo hay una palabra palíndroma de longitud cero, la palabra vacía ε .

b. Las palabras palíndromas de longitud unitaria son 0 , 1 , 2 y 3 .

c. Las palabras palíndromas de longitud dos son:

$00, 11, 22, 33$.

d. Las palabras palíndromas de longitud tres son:

$000, 010, 020, 030,$

$101, 111, 121, 131,$

$202, 212, 222, 232,$

$303, 313, 323, 333$.

e. Las palabras palíndromas de longitud cuatro son:

$0000, 0110, 0220, 0330,$

$1001, 1111, 1221, 1331,$

$2002, 2112, 2222, 2332,$

$3003, 3113, 3223, 3333$.

f. Las palabras palíndromas de longitud cinco son:

$00000, 00100, 00200, 00300,$

$01010, 01110, 01210, 01310,$

$02020, 02120, 02220, 02320,$

$03030, 03130, 03230, 03330,$

10001,10101,10201,10301,
11011,11111,11211,11311,
12021,12121,12221,12321,
13031,13131,13231,13331,
20002,20102,20202,20302,
21012,21112,21212,21312,
22022,22122,22222,22322,
23032,23132,23232,23332,
30003,30103,30203,30303,
31013,31113,31213,31313,
32023,32123,32223,32323,
33033,33133,33233,33333.

Ejercicio. Complete la lista anterior hasta considerar palabras palíndromas de longitud siete.

16. Con base en los resultados del ejemplo anterior, describa las características principales de las palabras palíndromas de longitud cinco.

En el ejemplo anterior se estableció que las palabras palíndromas de longitud cinco tienen la forma general *abcba*, donde *a*, *b* y *c* son símbolos arbitrarios del alfabeto. Se pueden establecer las tres características siguientes.

Los símbolos *a*, *b* y *c* son diferentes entre sí. Nótese que los símbolos *a* y *b* se repiten exactamente dos veces.

Dos de los símbolos *a*, *b* y *c* son iguales. De esta manera, se pueden tener palabras palíndromas de la forma *aacaa* o de la forma *ababa*. Nótese que en el primer caso, el símbolo *a* se repite cuatro veces, mientras que en el segundo caso el símbolo *a* se repite tres veces, y el símbolo *b* se repite dos veces.

Los tres símbolos *a*, *b* y *c* son iguales entre sí. De esta manera se obtiene la palabra palíndroma trivial *aaaaa*. Aquí, el símbolo *a* se repite cinco veces.

De los tres símbolos que se requieren para formar las palabras palíndromas de longitud cinco se pueden tener los casos siguientes:

Un solo símbolo se repite cinco veces.

Un símbolo se repite cuatro veces y se usa un símbolo más.

Dos de los símbolos se repiten dos veces cada uno de ellos y se usa un símbolo más.

Recuerde que los símbolos empleados no se pueden acomodar de manera arbitraria, deben acomodarse en posiciones bien definidas. A manera de ejemplos particulares de los casos anteriores se tienen las palabras siguientes: **00000**, **11311**, **21012**. Nótese que para formar a estas palabras palíndromas se requieren a lo más tres símbolos diferentes; y el alfabeto que se está usando tiene cuatro símbolos diferentes.

Ejercicio. Repita este ejemplo considerando ahora las palabras palíndromas de longitudes siete y once. (Recuerde que el alfabeto referido tiene cuatro símbolos diferentes.)

Considérese al conjunto **PALÍNDROMO**, el conjunto de todas las palabras palíndromas, sobre el alfabeto $A = \{a, b\}$.

17. Proporcione evidencia de que **PALÍNDROMO** contiene tantas palabras de longitud $2n$ como palabras de longitud $2n - 1$.

Considerando los resultados obtenidos en los Ejemplos 14, 15 y 16, se puede establecer de inmediato que hay dos palabras palíndromas de longitud unitaria (**a** y **b**), y también se tienen dos palabras palíndromas de longitud dos (**aa** y **bb**). Continuando, se tienen cuatro palabras palíndromas de longitud tres (**aaa**, **aba**, **bab**, **bbb**), y también se tienen cuatro palabras palíndromas de longitud cuatro (**aaaa**, **abba**, **baab**, **bbbb**). Como un último ejemplo, se tienen seis palabras palíndromas de longitud cinco (**aaaaa**, **aabaa**, **ababa**, **abbbba**, **baaab**, **babab**, **bbbbbb**), y se tienen seis palabras palíndromas de longitud seis (**aaaaaa**, **aabbaa**, **abaaba**, **abbbba**, **baaaab**, **babbab**, **bbbbbb**).

Con base en los resultados anteriores se puede establecer una generalización de la manera siguiente. Si **w** es una palabra de longitud n , entonces **ww^R** es una palabra

palíndroma de longitud $2n$; en efecto, se tiene que $(ww^R)^R = (w^R)^R w^R = ww^R$. Se deduce que, dada la palabra w , la palabra reversa w^R ya está fija y, por lo tanto, la palabra palíndroma ww^R sólo depende de w . Ahora bien, dado que w es de longitud n , se tienen 2^n palabras diferentes (recordando que el alfabeto que se está usando tiene dos símbolos diferentes) y, en consecuencia, también se tienen 2^n palabras palíndromas diferentes de longitud $2n$.

Considérese ahora una palabra w de longitud $n - 1$. Con esta palabra se puede construir una palabra palíndroma de la forma wcw^R de longitud $2n - 1$, donde c es una palabra unitaria formada con un símbolo arbitrario del alfabeto. Se verifica rápidamente que $(wcw^R)^R = (w^R)^R c^R w^R = wcw^R$. Claramente, la palabra palíndroma depende únicamente de la palabra w y de la palabra unitaria c , y dado que w tiene longitud $n - 1$, hay 2^n palabras diferentes de la forma wc . Así pues, se tienen 2^n palabras palíndromas diferentes de longitud $2n - 1$.

Ejercicio. Parafrasee el enunciado del resultado que se acaba de obtener considerando ahora palabras palíndromas de longitudes $2n$ y $2n + 1$.

18. Muestre que si la secuencia x se encuentra en *PALÍNDROMO*, entonces la secuencia x^n , para cualquier n , también se encuentra ahí.

Primero vamos a considerar un poco de evidencia sobre la validez del resultado aludido. Si x es una palabra palíndroma, entonces $x = x^R$. Ahora bien, para la palabra x^2 se puede establecer que $(x^2)^R = (xx)^R = x^R x^R = xx = x^2$. Esto es, la palabra x^2 también es una palabra palíndroma. Considérese ahora la palabra $x^3 = xx^2$. Para esta palabra se puede establecer que $(x^3)^R = (xx^2)^R = (x^2)^R x^R = x^2 x = xx^2 = x^3$. La palabra x^3 vuelve a ser una palabra palíndroma.

Con base en estos resultados se puede establecer el razonamiento inductivo siguiente. Si x es una palabra palíndroma, entonces $x^0 = \varepsilon$ también es una palabra palíndroma. Supóngase ahora que x^n es una palabra palíndroma para algún $n \geq 0$. Considérese la palabra $xx^n = x^{n+1}$. Para esta palabra se cumple lo siguiente:

$$(xx^n)^R = (x^n)^R x^R = x^n x = xx^n = x^{n+1}.$$

Así, x^{n+1} también es una palabra palíndroma. Y con esto se valida la hipótesis.

Ejercicio. Haga un resumen con las propiedades de las palabras palíndromas que se han obtenido en los ejemplos previos.

Referencias Bibliográficas

Cohen, D. I. A. (1986) **Introduction to Computer Theory**. New York, John Wiley & Sons, Inc.

Epp, S. S. (2012) **Matemáticas Discretas con Aplicaciones**. Cuarta edición, México, Cengage Learning.

Fejer, P. A. and Simovici, D. A. (1990) **Mathematical Foundations of Computer Science. Volume 1: Sets, Relations, and Induction**. Texts and Monographs in Computer Science, Volume X, New York, Springer-Verlag.

Kelley, D. (1995) **Teoría de Autómatas y Lenguajes Formales**. Madrid, Prentice Hall.

Mateescu, A. and Salomaa, A. (1997) Formal Languages: an Introduction and a Synopsis. En: Rozenberg, G. and Salomaa, A., eds. **Handbook of Formal Languages, Volume 1: Word, Language, Grammar**. Berlin-Heidelberg, Springer-Verlag, p. 1-39.

Tremblay, J. P. and Manohar R. (1987) **Discrete Mathematical Structures with Applications to Computer Science**. New York, McGraw-Hill Book Company.