3. Expresiones Regulares

3.1. Expresiones regulares (ER)

Fernando Rosa Velardo

Traducción y adaptación de transparencias de Ananth Kalyanaraman
(http://www.eecs.wsu.edu/~ananth/)

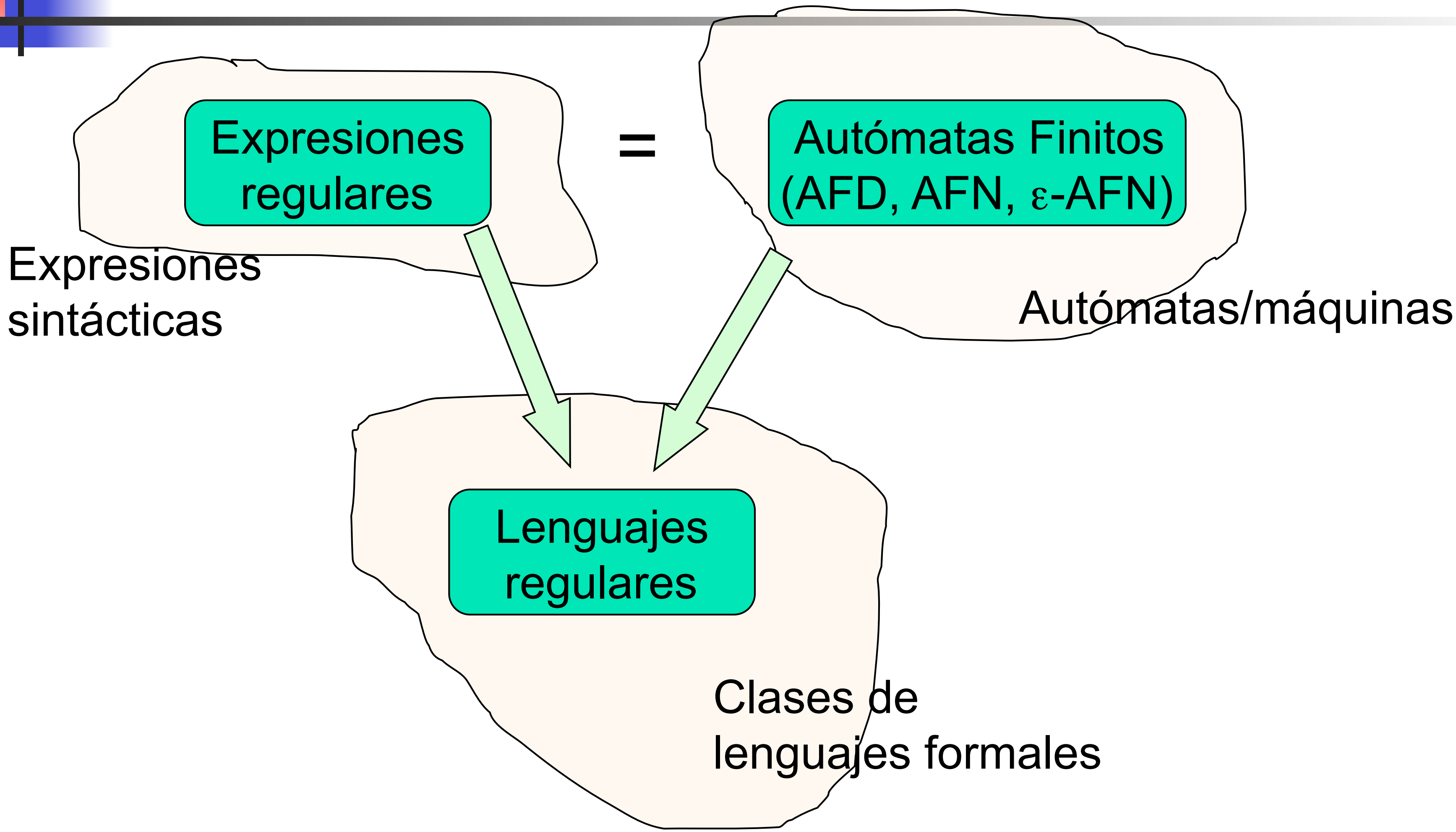
Expresiones Regulares vs. Autómatas Finitos

- Las expresiones regulares son una forma concisa de identificar cadenas que siguen cierto patrón
  - Por ejemplo, 01*+ 10*

- Autómatas => estilo “máquinas abstractas”
- Expresiones regulares => estilo “lenguajes de programación”

- Entornos Unix usan expresiones regulares
  - P. ej, bash, grep, vi & otros editores, sed
- Scripts Perl – para el procesamiento de cadenas
- Analizadores léxicos, como Lex o Flex

Expresiones regulares



Operaciones entre lenguajes (recordatorio)

- Unión de dos lenguajes:
  - **L U M** = cadenas que están en L o M

- Concatenación de dos lenguajes:
  - **L . M** = cadenas de la forma *xy* con $x \in L$ e $y \in M$
  - Normalmente el *punto* se omite
    - es decir, **LM** es lo mismo que L.M

Estrella de Kleene (operación *)

- Estrella de Kleene de un lenguaje L:
  - $L^0 = \{\varepsilon\}$
  - $L^1 = \{w \mid \text{para } w \in L\}$
  - $L^2 = \{ w_1 w_2 \mid w_1 \in L, w_2 \in L$ (pueden ser iguales)$\}$
  - $L^i = \{ w_1 w_2 \ldots w_i \mid w\text{'s en} \in L$ (puede haber repeticiones)$\}$
  - $L^* = \bigcup_{i \geq 0} L^i$ (cualquier número de concatenaciones)

Ejemplo:

- Sea $L = \{ 1, 00\}$
  - $L^0 = \{\varepsilon\}$
  - $L^1 = \{1, 00\}$
  - $L^2 = \{11, 100, 001, 0000\}$
  - $L^3 = \{111, 1100, 1001, 10000, 000000, 00001, 00100, 0011\}$
  - $L^* = L^0 \cup L^1 \cup L^2 \cup \ldots$

Clausura de Kleene (observaciones)

- L^* es infinito si y sólo si $L \neq \Phi$ y $L \neq \{\varepsilon\}$
- Si $L=\{\varepsilon\}$ entonces $L^* = \{\varepsilon\}$
- Si $L = \Phi$ entonces $L^* = \{\varepsilon\}$
- Σ^* es el conjunto de todas las palabras sobre el alfabeto Σ

Expresiones Regulares (definición)

- Definimos simultáneamente E y L(E), el lenguaje generado por E

- Casos base:
  - Φ es una ER y $L(\Phi) = \Phi$
  - ε es una ER y $L(\varepsilon) = \{\varepsilon\}$
  - $a \in \Sigma$ es una ER y $L(a)=\{a\}$

Expresiones Regulares (definición)

- Casos inductivos:
- Si E y F son ER entonces:
  - (E) es una ER y L((E))= L(E)
  - E+F es una ER y L(E + F) = L(E) U L(F)
  - EF es una ER y L(E F) = L(E) L(F)
  - E* es una ER y L(E*) = (L(E))*

Ejemplo

- L = { w | w es una cadena de 0s y 1s alternos}
  - Por ejemplo, w = 01010101 está en L, pero w = 10010 no
- <u>Objetivo:</u> Construir una expresión regular para L
- Cuatro casos para w:
  - Caso A: w comienza por 0 y |w| es par
  - Caso B: w comienza por 1 y |w| es par
  - Caso C: w comienza por 0 y |w| es impar
  - Caso D: w comienza por 1 y |w| es impar
- Expresión regular para cada uno de los casos:
  - Caso A: (01)*
  - Caso B: (10)*
  - Caso C: 0(10)*
  - Caso D: 1(01)*
- Como L es la unión de los cuatro casos:
  - Expresión regular para L: (01)* + (10)* + 0(10)* + 1(01)*
- Si usamos ε podemos simplificarla:
  - Expresión regular para L: (ε +1)(01)*(ε +0)

Precedencia de operadores

- De mayor a menor precedencia
  - * (estrella de Kleene)
  - . (concatenación)
  - +

- Ejemplo:
  - 01* + 1 = (0 ((1)*)) + 1

Propiedades de las Expresiones Regulares

- Conmutativa:
  - E+F = F+E
- Asociativa:
  - (E+F)+G = E+(F+G)
  - (EF)G = E(FG)
- Elemento neutro:
  - E+Φ = E
  - ε E = E ε = E
- Elemento nulo:
  - ΦE = EΦ = Φ

Propiedades…

- Distributiva:
  - $E(F+G) = EF + EG$
  - $(F+G)E = FE+GE$
- Idempotente: $E + E = E$
- Propiedades relativas a *:
  - $(E^*)^* = E^*$
  - $\Phi^* = \varepsilon$
  - $\varepsilon^* = \varepsilon$
  - $E^+ = EE^*$

¿Cierto o falso?

Sean R y S dos expresiones regulares. Entonces:

1. $((R^*)^*)^* = R^*$? ✓
2. $(R+S)^* = R^* + S^*$? × $R=a$ $S=b$

   $\in$ aba $\notin$ aba
3. $(RS + R)^* RS = (RR^*S)^*$? × $R=a$ $S=b$

   $\notin \varepsilon$ $\in \varepsilon$

3. Expresiones Regulares

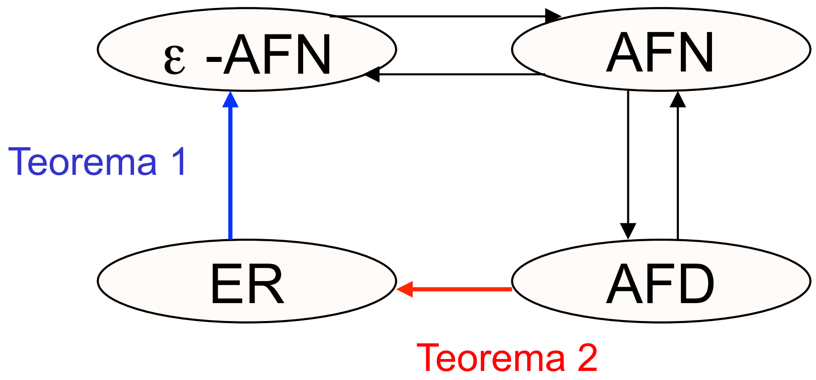
3.2. De las Expresiones Regulares a los ε -AFN

Fernando Rosa Velardo

Traducción y adaptación de transparencias de Ananth Kalyanaraman
(http://www.eecs.wsu.edu/~ananth/)

Autómatas Finitos y Expresiones Regular (ER)

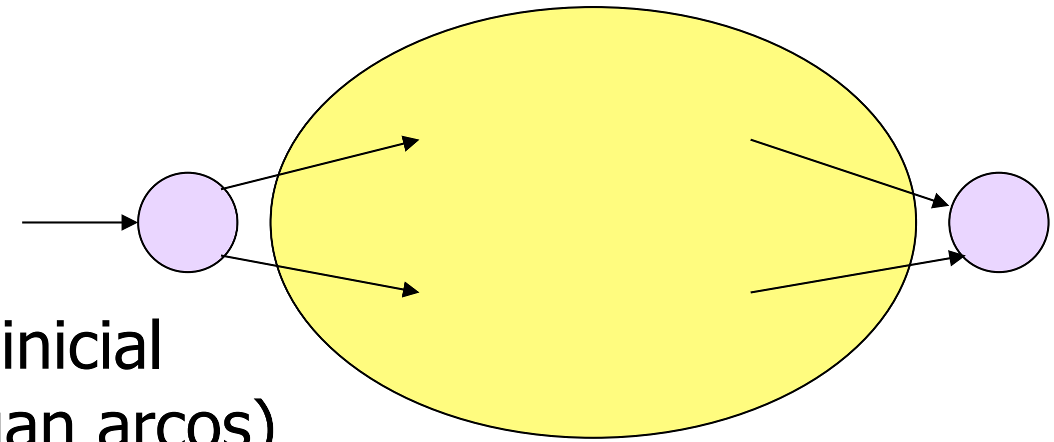
- Para probar que son equivalentes, considera los siguientes resultados
  - Teorema 1: Si R es una expresión regular existe un ε -AFN E tal que L(E)=L(R)
  - Teorema 2: Si A es un AFD existe una expresión regular R tal que L(R)=L(A)



Conversión de ER a ϵ-AFN

- Construcción por inducción sobre el número de operadores (+, concatenación, *) en la ER.
- Siempre construimos un autómata con una forma especial (próxima transparencia).

Forma de los є-AFN construidos

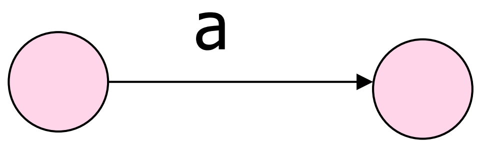


Estado inicial
(no llegan arcos)

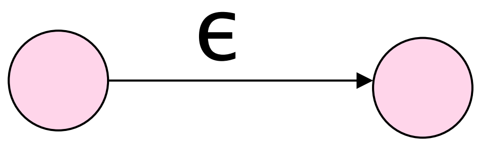
Estado final
(no salen arcos)

Casos base

- Símbolo **a**:

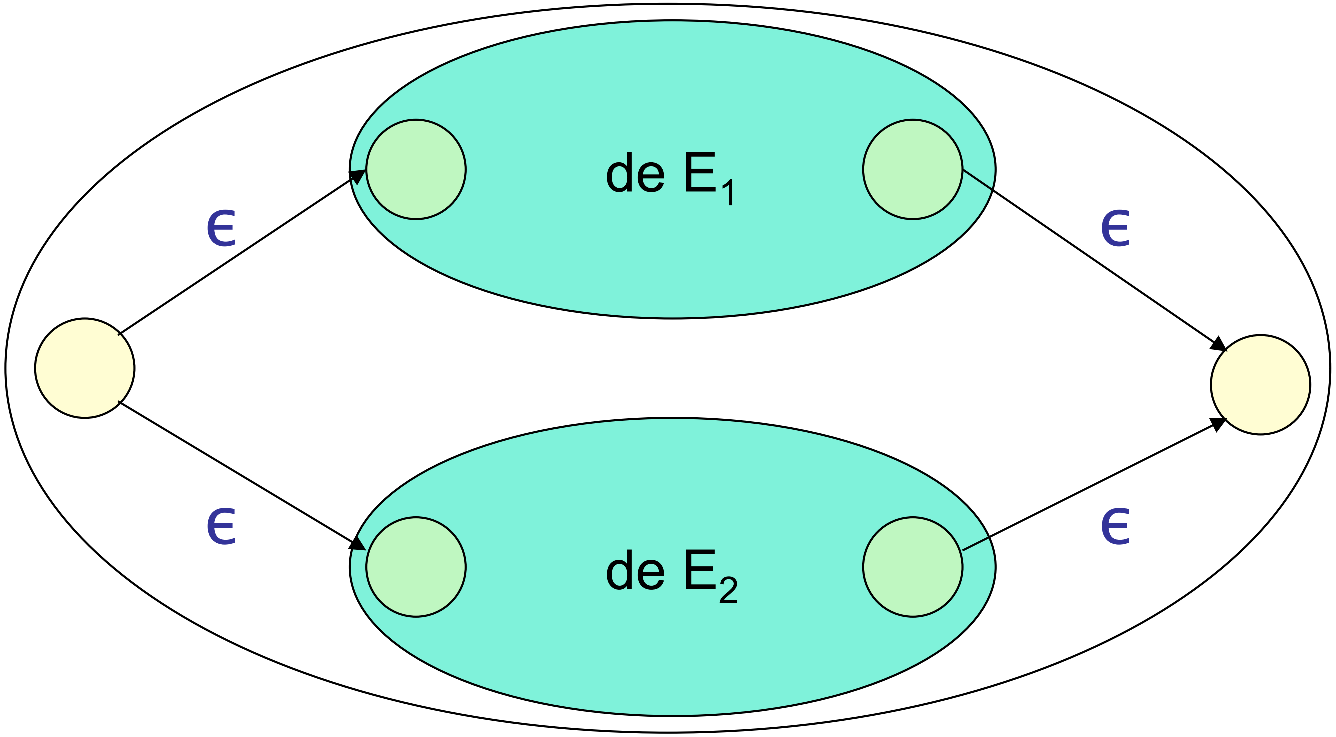


- ϵ:



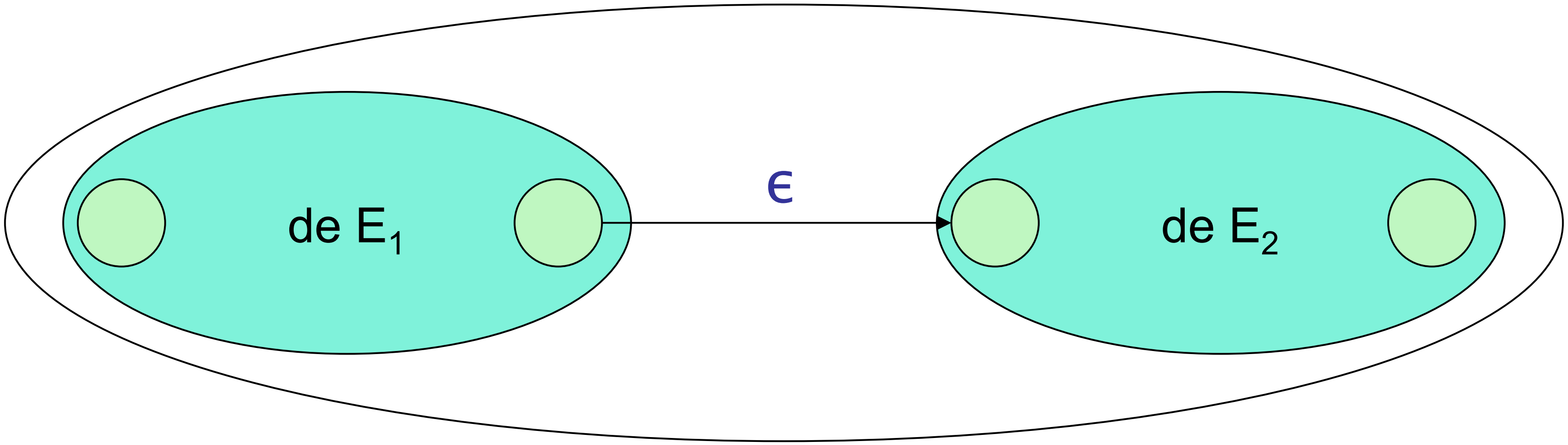
- ∅:

Inducción 1 – Suma



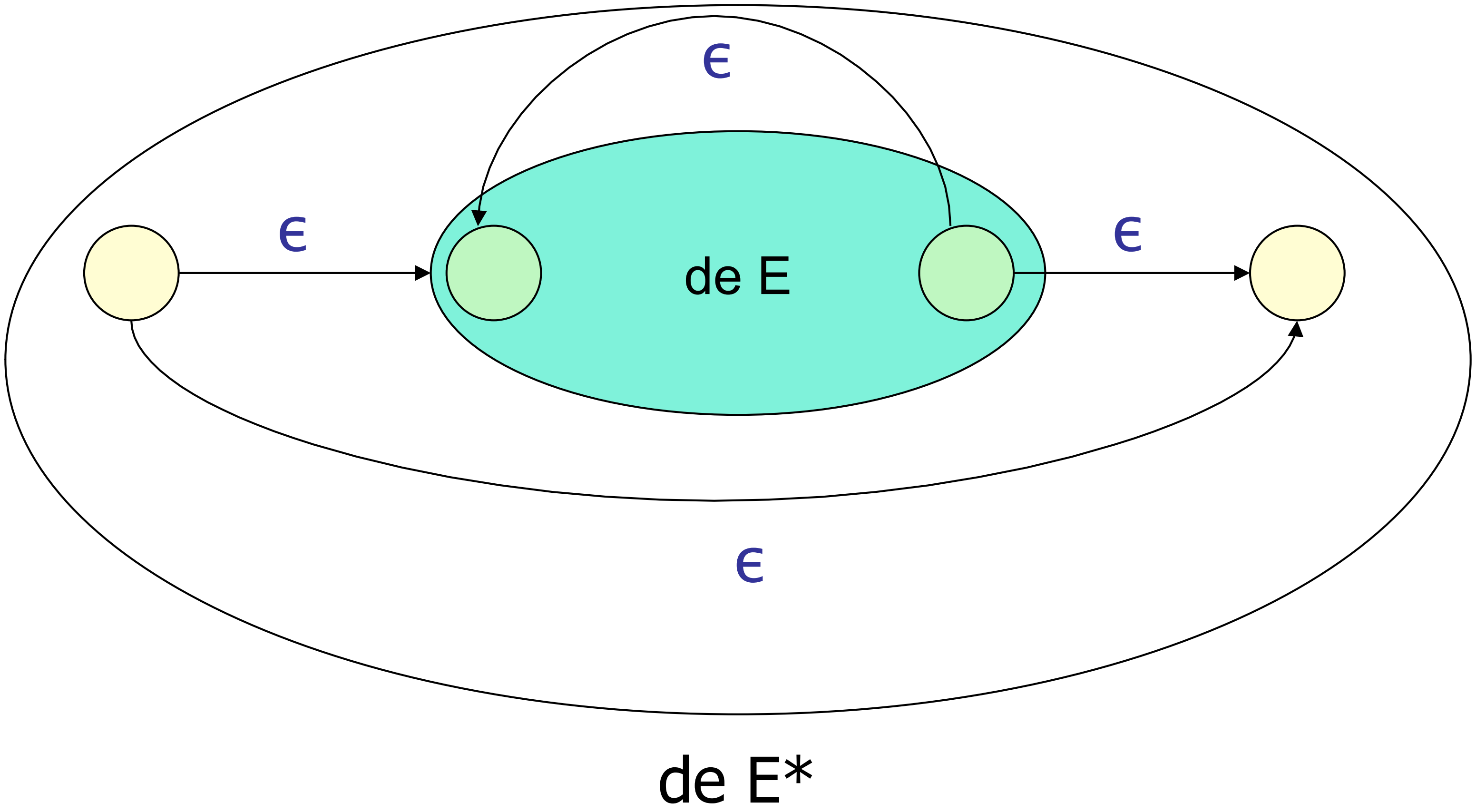
de E_1+E_2

Inducción 2 – Concatenación



de E_1E_2

Inducción 3 – *



3. Expresiones Regulares

3.3. De los AFD a las expresiones regulares

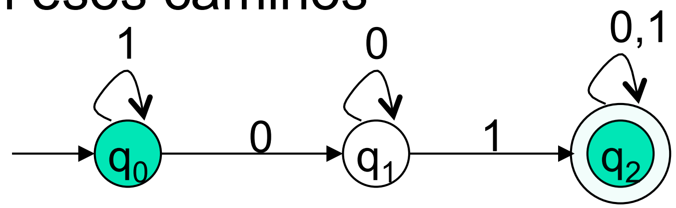
Fernando Rosa Velardo

Traducción y adaptación de transparencias de Ananth Kalyanaraman (http://www.eecs.wsu.edu/~ananth/)

De los AFD a las ER

- Intuitivamente, seguir todos los caminos desde el estado inicial a todos los estados finales, identificando las expresiones que generan esos caminos

Ejemplo:



$$\underbrace{(\varepsilon + 1)^*}_{1^*}\ \underbrace{0\ (\varepsilon + 0)^*}_{00^*}\ \underbrace{1}_{1}\ \underbrace{(\varepsilon + 0 + 1)^*}_{(0+1)^*}$$

$$\Downarrow$$

$$1^*00^*1(0+1)^*$$

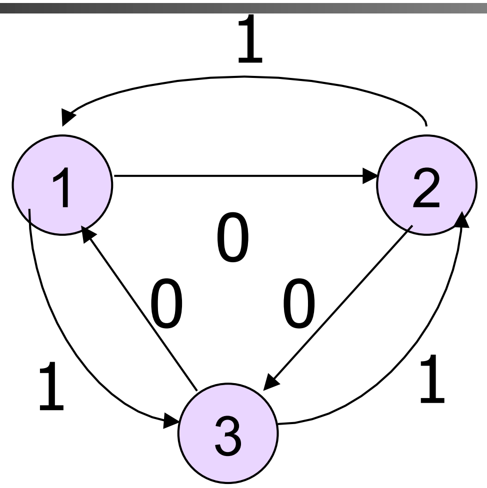
De AFD a expresión regular

- Suponemos que los estados del AFD son 1,2,…,n.
- Construimos ER que generan “ciertos” caminos
  - Base: un sólo arco, o ningún arco.
  - Inducción: caminos que no pueden pasar por algunos estados.

k-camino

- Un k-camino es un camino en el grafo de un AFD que no pasa por ningún estado mayor que k.
  - Es decir, los estados intermedios están en {1,…,k}
- Los extremos no están limitados, pueden ser cualquier estado.

Ejemplo: k-caminos



0-camino de 2 a 3: **0**.

1-caminos de 2 a 3:
0+**11**.

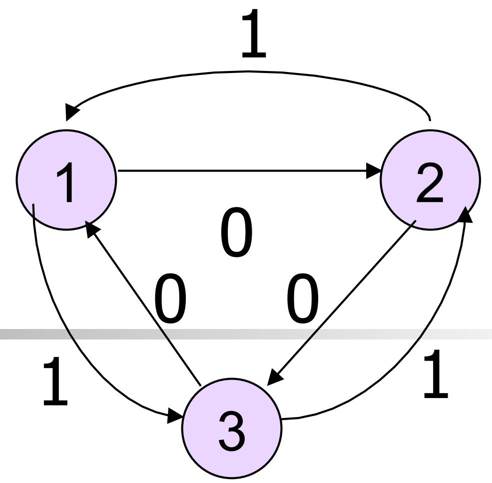
2-caminos de 2 a 3:
(**10**)*(**0**+**11**)

3-caminos de 2 a 3: ??

k-caminos (Inducción)

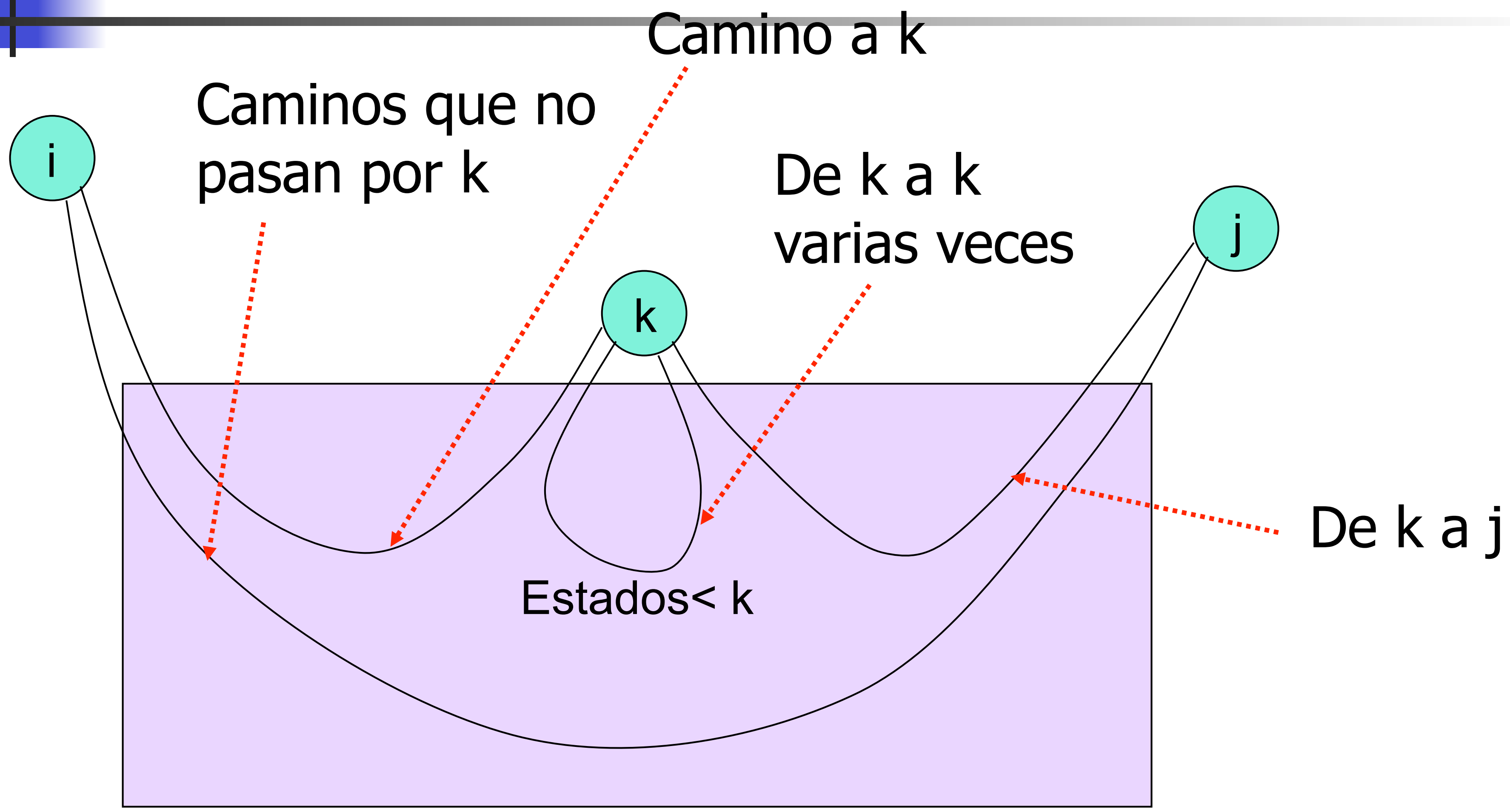
- Sea R_{ij}^{k} la expresión regular que caracteriza los k-caminos de i a j.
- Base: k=0. R_{ij}^{0} = suma de las etiquetas de los arcos que van de i a j.
  - $\emptyset$ si no existe tal arco.
  - Se añade ϵ si i=j.

Ejemplo: Base



- $R_{12}^{0} = \mathbf{0}$.
- $R_{11}^{0} = \varnothing + \epsilon = \epsilon$.

Ilustración de la Inducción



k-camino (Caso inductivo)

- Un k-camino de i a j:
  1. Nunca pasa por k, o bien
  2. Pasa por k una o más veces.

$$R_{ij}^{k} = R_{ij}^{k-1} + R_{ik}^{k-1}(R_{kk}^{k-1})^{*} R_{kj}^{k-1}.$$

No pasa por k

Visita k por primera vez

Cero o más veces de k a k

De k a j, sin pasar por k

Simplificaciones de la fórmula general

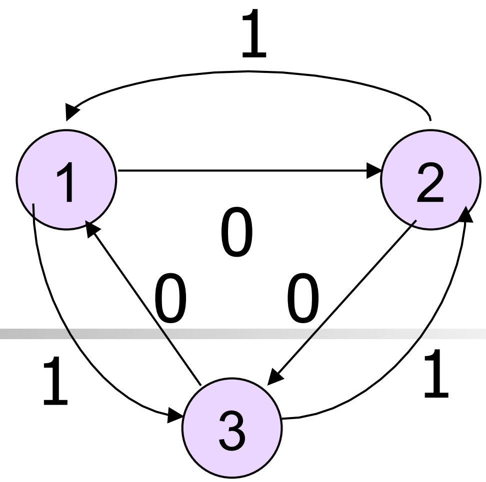
si i=k: $R_{kj}^{k} = (R_{kk}^{k-1})^* \, R_{kj}^{k-1}$

si j=k: $R_{ik}^{k} = R_{ik}^{k-1} (R_{kk}^{k-1})^*$

Último paso

- La expresión regular es la suma de todas las R_{ij}^{n}, donde:
  1. n es el número de estados; es decir, los caminos no están restringidos.
  2. i es el estado inicial.
  3. j es un estado final.

Ejemplo



- $R_{23}^{3} = R_{23}^{2}(R_{33}^{2})^{*}$
- $R_{23}^{2} = (R_{22}^{1})^{*}R_{23}^{1} = (\mathbf{10})^{*}(\mathbf{0+11})$
- $R_{33}^{2} = R_{33}^{1} + R_{32}^{1}(R_{22}^{1})^{*}\, R_{23}^{1}$

  $= (\mathbf{01})^{*} + (\mathbf{1+00})(\mathbf{10})^{*}(\mathbf{0+11})$

- $R_{23}^{3} = [(\mathbf{10})^{*}(\mathbf{0+11})][(\mathbf{01})^{*} + (\mathbf{1+00})(\mathbf{10})^{*}(\mathbf{0+11})]^{*}$