Formele talen 12

uitgebreid

verkeerslicht?

신호등을지킵시다

(Automatische) Vertaling van het Koreaans

You should observe the traffic lights

is Koreaans een formele taal? nee natuurlijk niet! alleen, voor iemand die de taal niet kan lezen, doet het ‘formeel’ aan, ik zie vorm, geen betekenis.

is een formele taal zonder betekenis? ook al niet! rekenkundige expressies bv hebben een semantiek.

taal …

letter = symbool

alfabet Σ : eindige (niet-lege) verzameling letters

string = woord : (geordend) rijtje letters (uit alfabet)
alle strings Σ^*
lege string λ lengte nul

taal : verzameling woorden
[dus deelverzameling Σ^*]
alle talen $P(\Sigma^*)$

op verzamelingen kunnen we de bekende boolese operaties toepassen, maar op talen ook concatenatie, macht en ster (dwz achter elkaar zetten +herhaling)

Een *alfabet* is een eindige, niet-lege, verzameling *letters*.

Σ = { a,b,c }
B = { 0,1 }
C = { а,б,в,г,д,е,ж,з,и,й,к,л, … э,ю,я }
P = { if, else, while, do, … }
V = { … , appel, koek, ei , … }

strings

Σ alfabet.
Een *string/woord* (over Σ) is een eindig geordend rijtje letters uit Σ.

ab, abca, abcbabcba
0, 1, 00, 01, 10, 11, 000, 001, …
Приключения, Астерикса
"de appel valt niet ver"

Σ^*, lege string λ, lengte $|x|$

$|abcbac| = 6$ $|\lambda| = 0$ $a^6b^3 = aaaaaabbb$
$B^* = \{ \lambda, 0, 1, 00, 01, 10, 11, 100, 101, \dots \}$

Een *taal* (over Σ) is een verzameling strings over Σ

Σ^* alle strings

PAL = { λ, aa, bb, abba, baab, abaaba, …}

BIN = { 0, 1, 10, 11, 100, 101, 110, 111, 1000, … }

K = { a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … }

= { $x \in \{a,b\}^*$ | x eindigt op een a }

L = { λ, aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … }

= { $x \in \{a,b\}^*$ | x heeft even lengte }

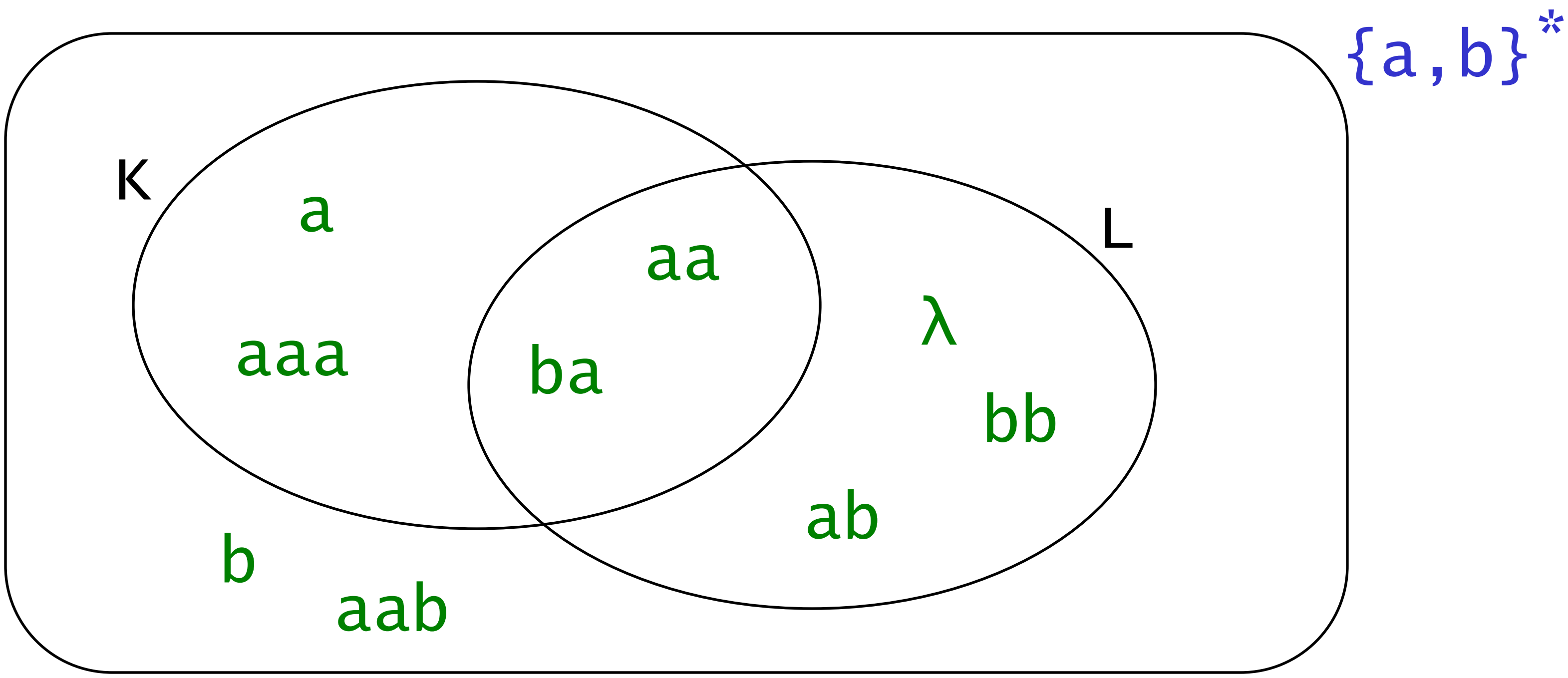
$\varnothing$

$\mathcal{P}(\Sigma^*)$ talen over Σ

Een *taal* (over Σ) is een verzameling strings over Σ

K = { a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … }
= { x∈{a,b}* | x eindigt op een a }

L = { λ, aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … }
= { x∈{a,b}* | x heeft even lengte }



operaties

standaardoperaties op woorden en talen

concatenatie = achter elkaar plakken

macht = idem, vast aantal herhalingen

ster = idem, willekeurig herhaald

concatenatie

$\cdot : \Sigma^* \times \Sigma^* \to \Sigma^*$

$a_1 \ldots a_m \cdot b_1 \ldots b_n = a_1 \ldots a_m b_1 \ldots b_n$

λ één $\lambda \cdot x = x = x \cdot \lambda$

associatief $(x \cdot y) \cdot z = x \cdot (y \cdot z)$

niet commutatief bei·aard ≠ aard·bei

$|x \cdot y| = |x| + |y|$

$\cdot$ wordt vaak weggelaten

prefix

$x, y \in \Sigma^*$
x is een *deelwoord* van y als $y = u \cdot x \cdot v$ voor $u, v \in \Sigma^*$
… *prefix* … als $y = x \cdot v$ voor $v \in \Sigma^*$
… *suffix* … als $y = u \cdot x$ voor $u \in \Sigma^*$

óók: *sub*woord

lekkerkerker deelwoord kerk (twee *voorkomens*)

aarzelaar aar prefix én suffix (als λ)

$x, y \in \Sigma^*$

x is een *deelwoord* van y als $y = u \cdot x \cdot v$ voor $u, v \in \Sigma^*$

prefix … als $y = x \cdot v$ voor $v \in \Sigma^*$

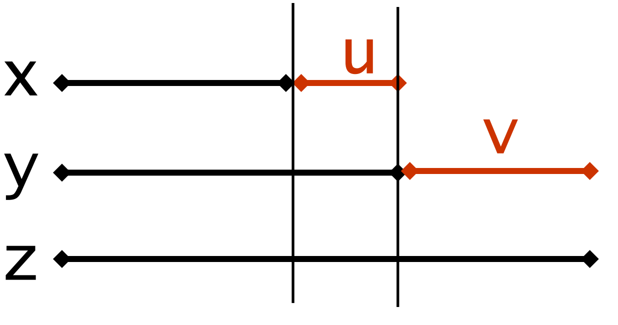
suffix … als $y = u \cdot x$ voor $u \in \Sigma^*$

reflexief $x \preceq x$ voor alle x

anti-symmetrisch als $x \preceq y$ en $y \preceq x$ dan $x = y$

transitief als $x \preceq y$ en $y \preceq z$ dan $x \preceq z$

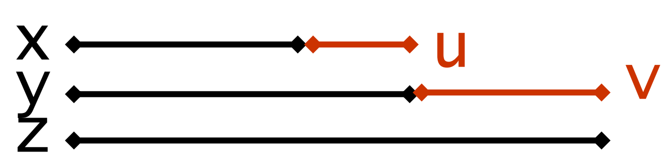
$\Rightarrow$ partiële ordening



eigenschappen van prefix

$x, y \in \Sigma^*$
x is een *deelwoord* van y als $y = u \cdot x \cdot v$ voor $u, v \in \Sigma^*$
prefix … als $y = x \cdot v$ voor $v \in \Sigma^*$
suffix … als $y = u \cdot x$ voor $u \in \Sigma^*$

reflexief $x \preceq x$ voor alle x
anti-symmetrisch als $x \preceq y$ en $y \preceq x$ dan $x = y$
transitief als $x \preceq y$ en $y \preceq z$ dan $x \preceq z$
$\Rightarrow$ partiële ordening



prefix

$x = x \cdot \lambda$
$y = x \cdot u$ en $x = y \cdot v$ dan $|x| \leq |y|$ en $|y| \leq |x|$
daarom $|x| = |y|$ en dus $u = v = \lambda$, $x = y$
$y = x \cdot u$ en $z = y \cdot v$ dan $z = (x \cdot u) \cdot v = x \cdot (uv)$

macht

$x^n = x \cdot \ldots \cdot x$ (n maal)

inductief:

$x^0 = \lambda$

$x^{n+1} = x^n \cdot x$

$x^{m+n} = x^m \cdot x^n$

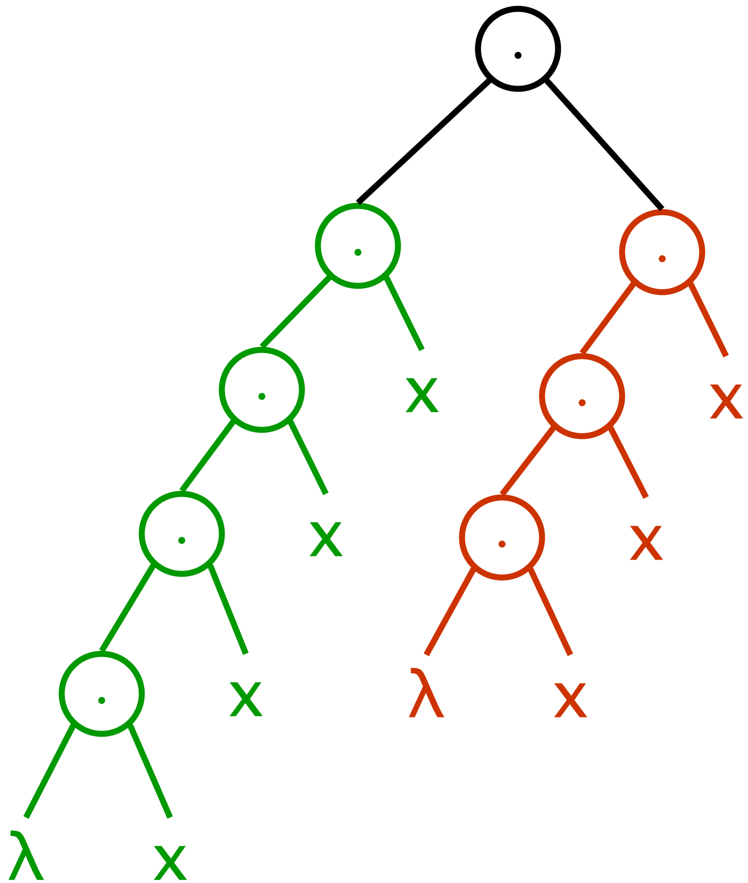
$|x^n| = n \cdot |x|$

$(\text{ker})^5 = \text{kerker} \cdot \text{kerk} \cdot \text{erker}$

intuïtie formeel gemaakt

inductief:

$$x^0 = \lambda$$
$$x^{n+1} = x^n \cdot x$$



$$x^4 \cdot x^3 \stackrel{?}{=} x^7$$

$((((\lambda x)x)x)x) \cdot$

$(((\lambda x)x)x)$

associatief !

en λ is één

inductief:

$\mathrm{mir}(\lambda) = \lambda$

$\mathrm{mir}(xa) = a \cdot \mathrm{mir}(x)$

ook x^R

palindroom $x = \mathrm{mir}(x)$

snorfrons netebeten

$\mathrm{mir}(\mathrm{mir}(x)) = x$

$\mathrm{mir}(xy) = \mathrm{mir}(y) \cdot \mathrm{mir}(x)$

$\mathrm{mir}(x^n) = \mathrm{mir}(x)^n$

taal: Boolese operaties

Een *taal* (over Σ) is een verzameling strings over Σ

$K = \{$ a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … $\}$
$= \{ x \in \{a,b\}^* \mid x$ eindigt op een a $\}$
$L = \{ \lambda,$ aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … $\}$
$= \{ x \in \{a,b\}^* \mid x$ heeft even lengte $\}$

$\mathcal{P}(\Sigma^*)$ talen over Σ
vereniging, doorsnede, complement (tov. Σ^*)

$K \cap L = \{$ aa, ba, aaaa, aaba, abaa, abba, baaa, … $\}$
$K - L = \{$ a, aaa, aba, baa, bba, aaaaa, aaaba, … $\}$
$L - K = \{ \lambda,$ ab, bb, aaab, aabb, abab, abbb, baab, … $\}$
$\{a,b\}^* - (K \cup L) = \{$ b, aab, abb, bab, bbb, aaaab, … $\}$

verzamelingsrekenregels

commutatief	$A \cap B = B \cap A$	$A \cup B = B \cup A$
associatief	$(A\cap B)\cap C = A\cap(B\cap C)$	$(A\cup B)\cup C = A\cup(B\cup C)$
distributief	$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$	$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$
wetten van De Morgan	$(A\cap B)^C = A^C\cup B^C$	$(A\cup B)^C = A^C\cap B^C$
absorptie	$A\cap(A\cup B) = A$	$A\cup(A\cap B) = A$
idempotentie	$A\cap A = A$	$A\cup A = A$
nulelement	$A\cap\varnothing = \varnothing$	$A\cup\varnothing = A$
éénelement	$A\cap U = A$	$A\cup U = U$
dubbel complement	$(A^c)^c = A$	
complementregels	$A\cap A^c = \varnothing$	$A\cup A^c = U$

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{ x \cdot y \mid x \in K \text{ en } y \in L \}$

$\{ a, ab \} \cdot \{ a, ba \} = \{ a \cdot a, a \cdot ba, ab \cdot a, ab \cdot ba \}$
$= \{ aa, aba, abba \}$

$\{ \lambda, a, b, aa, ab, ba, bb, aaa, aab, \ldots \} \cdot \{ a \} =$
$\{ a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, \ldots \}$

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{ x \cdot y \mid x \in K \text{ en } y \in L \}$

{ a,ab,abb,abbb, … }·{ a,ba,aba,baba,ababa, … }
= { aa,aba,aaba,abba,ababa,abbba,aababa,abbaba, … }

	a,	ba,	aba,	baba,	ababa,	…
a	aa,	aba,	aaba,	ababa,	aababa,	…
ab	aba,	abba,	ababa,	abbaba,	abababa,	…
abb	abba,	abbba,	abbaba,	abbbaba,	abbababa,	…
abbb	abbba,	abbbba,	abbbaba,	abbbbaba,	…	
…						

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{ x \cdot y \mid x \in K \text{ en } y \in L \}$

$\{\lambda\}$ één $\qquad \{\lambda\} \cdot L = L \cdot \{\lambda\} = L$

$\varnothing$ nul $\qquad \varnothing \cdot L = L \cdot \varnothing = \varnothing$

associatief $\quad (K \cdot L) \cdot M = K \cdot (L \cdot M)$

niet commutatief

$K^n = K \cdot \ldots \cdot K$ (n maal)
inductief:
$K^0 = \{ \lambda \}$
$K^{n+1} = K^n \cdot K$

$\{\lambda, a, ab\}^1 = \{\lambda, a, ab\}$

$\{\lambda, a, ab\}^2 =$
$\{\lambda, a, aa, ab, a \cdot ab, aba, abab\}$

$\{\lambda, a, ab\}^3 =$
$\{\lambda, a, aa, ab, aaa, aab, aba, aaab, a \cdot ab \cdot a, abaa,$
$abab, aabab, abaab, ababa, ababab\}$

$$L = L^2$$

- voorbeelden
- wat weten we van L

ster & plus

$K^n = K \cdot \ldots \cdot K$ (n maal)

inductief:

$K^0 = \{ \lambda \}$

$K^{n+1} = K^n \cdot K$

$K^* = \bigcup_{n \geq 0} K^n$, $K^+ = \bigcup_{n > 0} K^n$

ab* = {a}·{b}* = { a,ab,abb,abbb,abbbb … }

(ab*)* = { a,ab,abb,abbb,abbbb … }* =

= { λ,a,aa,ab,aaa,aab,aba,abb,aaaa,aaab,… }

= {λ} ∪ {a}·{a,b}*

$\{\lambda,a,ab\}^* = \{a,ab\}^* =$

{ x | voor elke b staat een a }

ster & plus

$K^n = K \cdot \ldots \cdot K$ (n maal)

inductief:

$K^0 = \{ \lambda \}$

$K^{n+1} = K^n \cdot K$

$K^* = \bigcup_{n\geq 0} K^n$, $K^+ = \bigcup_{n>0} K^n$

alternatief:

K^* is de kleinste verzameling die λ bevat en gesloten is onder concatenatie met K

oftewel (inductief):

- $\lambda \in K^*$
- als $x \in K^*$ en $w \in K$ dan $xw \in K^*$

reguliere expressies in PERL

```
$greeting = "World";
print "It matches\n" if "Hello World" =~ /$greeting/;

"cats and dogs" =~ /dog|cat|bird/; # matches "cat"

/item[0-9]/; # matches 'item0' or ... or 'item9' \d
/(a|b)b/;    # matches 'ab' or 'bb'
   a* = match 'a' 0 or more times
   a? = match 'a' 1 or 0 times
   period '.' matches any character but "\n" (eol)

$time =~ /(\d\d):(\d\d):(\d\d)/; # match hh:mm:ss
$hours = $1; $minutes = $2; $seconds = $3;

# match a number, get $1 = whole number, $2 = exponent
/([+-]?\ *(?:\d+(?:\.\d*)?|\.\d+)(?:[eE]([+-]?\d+))?)/;
```

rekenkundige expressies

$D = \{ 0,1,2, \ldots , 9 \}$

i. ieder element van $D^* - \{\lambda\}$ behoort tot R

ii. als $x \in R$, dan $(-x) \in R$

als $x \in R$ en $y \in R$,

dan $(x+y) \in R$, $(x-y) \in R$, $(x \cdot y) \in R$, $(x/y) \in R$

alfabet $\{ 0,1,2, \ldots ,9 \} \cup \{ +,-,\cdot,/ \} \cup \{ (,) \}$

syntax: geheel getal

```
⟨geheel⟩ ::= ⟨teken⟩⟨natuurlijk⟩ | ⟨natuurlijk⟩
⟨natuurlijk⟩ ::= ⟨cijfer⟩ | ⟨cijfer⟩⟨natuurlijk⟩
⟨cijfer⟩ ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
⟨teken⟩  ::= + | -
```

$$\{ +,- \} \cdot \{ 0,1, \ldots ,9 \}^+ \cup \{ 0,1, \ldots ,9 \}^+ =$$
$$\{ +,-,\lambda \} \cdot \{ 0,1, \ldots ,9 \}^+$$

alfabet $\{a,b\}^*$

voor en na elke a staat een b

rekenregels 'algebra'

$\varnothing \cdot L = L \cdot \varnothing = \varnothing$

$\{\lambda\} \cdot L = L \cdot \{\lambda\} = L$

$(K \cdot L) \cdot M = K \cdot (L \cdot M)$

$K \cdot (L \cup M) = K \cdot L \cup K \cdot M$

$(K \cup L) \cdot M = K \cdot M \cup L \cdot M$

$(K^*)^n = K^*$ voor alle $n \geq 1$

$(K^*)^* = K^*$

$(K^* L^*)^* = (K \cup L)^* = (K^* \cup L^*)^*$

$K \subseteq L$

dan $K \cup M \subseteq L \cup M$

$K \cdot M \subseteq L \cdot M$, $M \cdot K \subseteq M \cdot L$

$K^n \subseteq L^n$ voor alle $n \geq 1$

$K^* \subseteq L^*$

oneindige vereniging … ?

eindig:associatief!

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = ((\cdots(A_1 \cup A_2) \cup \cdots) \cup A_n)$$

$$\boxed{\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c}$$

$$\neg(\varphi \vee \psi) \equiv \neg\varphi \wedge \neg\psi$$

$$\left(\bigcup_{i=1}^{n+1} A_i\right)^c = \left(\bigcup_{i=1}^{n} A_i \cup A_{n+1}\right)^c =$$

$$\left(\bigcup_{i=1}^{n} A_i\right)^c \cap A_{n+1}^c = \bigcap_{i=1}^{n} A_i^c \cap A_{n+1}^c = \bigcap_{i=1}^{n+1} A_i^c$$

$$\bigcup_{i=1}^{\infty} A_i = \{x \in U \mid (\exists i)(x \in A_i)\}$$

oneindig ?

$$\boxed{\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c}$$

$$\neg(\exists i)\psi(i) = (\forall i)\neg\psi(i)$$

oneindig

bij talen en hun operaties gebruiken we soms oneindige verenigingen. Als we twee verzamelingen verenigen geldt de regel van de Morgan.
Die geldt ook voor meerdere (eindig veel) verzamelingen, dat kunnen we laten zien door de regels herhaald toe te passen (en wat associativiteit).

Ook oneindige vereniging kent De Morgan, maar daar moeten we een beetje logica bij gebruiken, met de begrippen ‘er is’ een ‘voor alle’.
Bv. een woord x zit in de vereniging van A_i als er een index i is zodat x in A_i.

❖ beschrijven, genereren

K: w_0, w_1, w_2, w_3, …

- recursieve definitie :
- syntax-diagram
- **grammatica**
- operaties $*$ $\cup$ $\cdot$
- logische formule

❖ herkennen, algoritme

$w \in K$? ja/nee

- **automaat**, Turing machine
- parser

grammatica

- MI ∈ L
- als xI ∈ L dan xIU ∈ L ♣
  als Mx ∈ L dan Mxx ∈ L ♦
  als xIIIy ∈ L dan xUy ∈ L ♥
  als xUUy ∈ L dan xy ∈ L ♠
- L bevat geen andere elementen

axioma MI

'herschrijf'regels

I	→	IU	mits uiteinde
Mx	→	Mxx	mits hele woord (?!)
III	→	U	in elke context
UU	→	λ	

MI ⇒ MII ⇒ MIIII ⇒ MUI ⇒ MUIUI ⇒ MUIUIU ⇒…

Chomsky: type regels *families*
rechts-lineair, context-vrij, monotoon, type-0

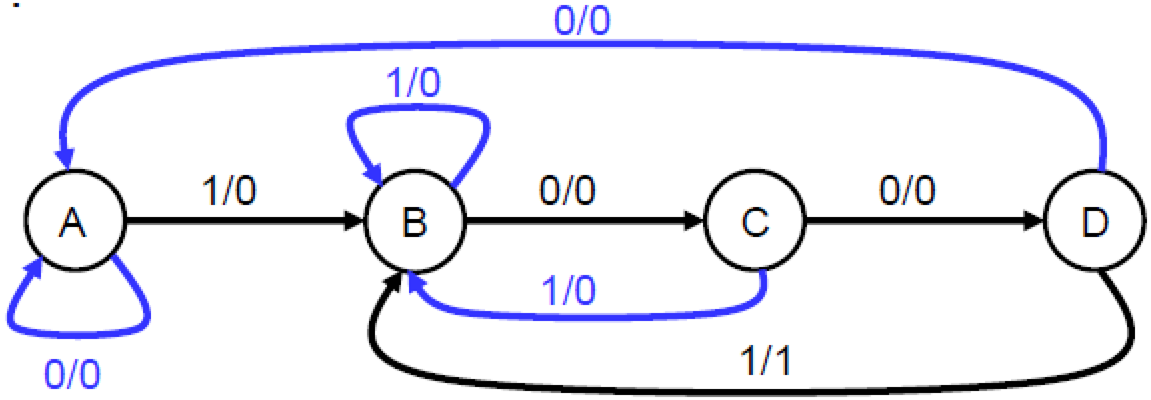
reguliere talen

kleinste familie van talen die
- $\varnothing$ $\{\lambda\}$ $\{a\}$ bevat
- gesloten is onder $*$ $\cup$ $\cdot$

- $\varnothing$ $\{\lambda\}$ $\{a\}$ $\in$ REG
- als K, L $\in$ REG dan
  - $K \cup L \in$ REG
  - $K \cdot L \in$ REG
  - $K^* \in$ REG
- verder geen talen in REG

Step 1: Deriving the State Table (cont.)

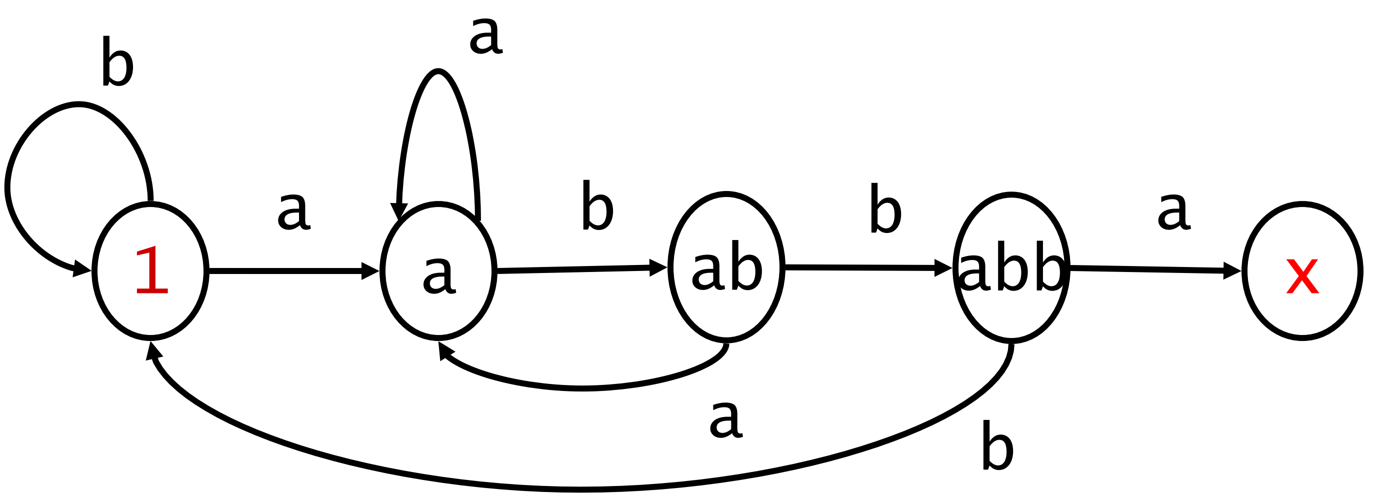
- Remember that we need *two* outgoing arrows for each node, to account for the possibilities of X = 0 and X = 1.
- The remaining arrows we need are shown in blue. They also allow for the correct detection of overlapping occurrences of 1001.



State	Meaning
A	None of the desired pattern (1001) has been input yet.
B	We've already seen the first bit (1) of the desired pattern.
C	We've already seen the first two bits (10) of the desired pattern.
D	We've already seen the first three bits (100) of the desired pattern.

patroonherkenning

{ x∈{a,b}* | x heeft geen subwoord abba }



van 1 naar 1

({b}∪{a}{a,ba}*{bbb})*

· ({λ}∪{a}{a,ba}*{λ,b,bb})

volgt. over twee weken staan eindige automaten op het programma, een methode om met grafen talen te specificeren.

spiegelbeeld

$mir(K) = \{ mir(x) \mid x \in K \}$ ook K^R

$mir(mir(K)) = K$

$mir(K \cup L) = mir(K) \cup mir(L)$

$mir(K \cdot L) = mir(L) \cdot mir(K)$

$mir(K^n) = mir(K)^n$

$mir(K^*) = mir(K)^*$

- $\varnothing$ $\{\lambda\}$ $\{a\}$ $\in$ REG
- als K, L $\in$ REG dan
  - $K \cup L \in$ REG
  - $K \cdot L \in$ REG
  - $K^* \in$ REG

REG is gesloten onder mir (volgt uit 1,3,5)

dwz: K regulier dan ook mir(K) regulier

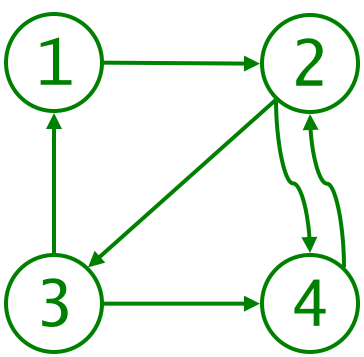
probleem = taal

HPP: Hamilton Path Problem
gegeven: gerichte graaf G, begin- & eindpunt
gevraagd: is er een Hamilton pad (elke knoop éénmaal)

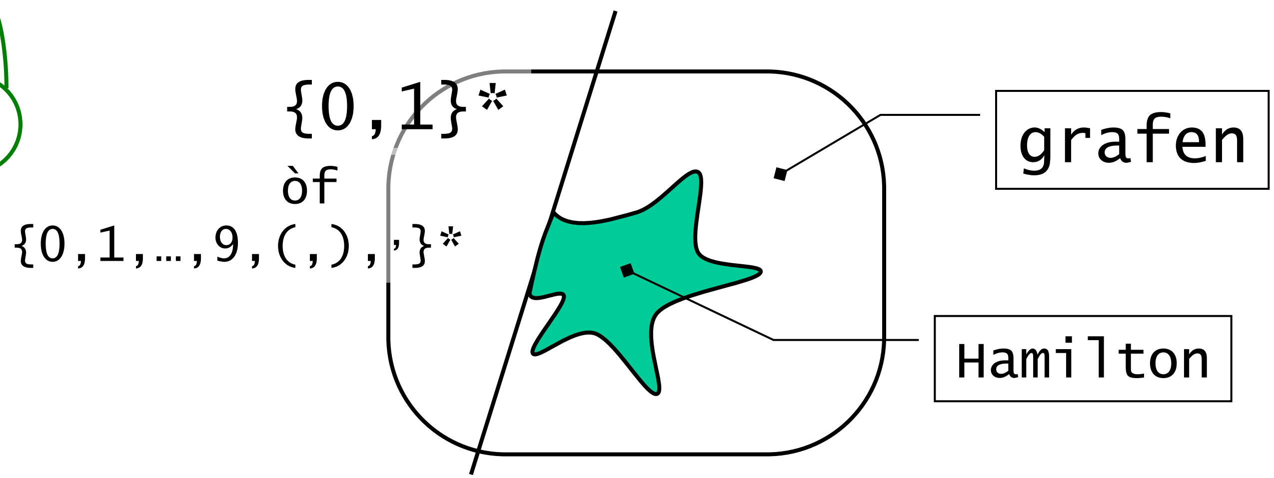
aantal punten pijlen begin & eind

4 (1,2)(2,3)(2,4)(3,1)(3,4)(4,2) 1,2

000 100 001010 010011 010100 011001 011100 100010 000 001010



G graaf -> woord w(G)
Hamilton grafen -> taal Ham



probleem = taal

bij het vak Complexiteit komt de formele definitie van het begrip NP-compleet aan de orde. Hiervoor moeten we het begrip ‘probleem’ omzetten naar een ‘taal’.

Hoeveel tijd kost het om instantie X van probleem P op te lossen wordt dan: hoeveel tijd kost het om te beslissen of X tot de taal P behoort.

end...

Regular Expressions

http://xkcd.com/208/

$\boxed{L = L^2}$

als $\lambda \in L$ dan $L \subseteq L^2$
dus nog testen of $L^2 \subseteq L$?

$\{\lambda\}$
$\varnothing$
Σ^*

- heeft subwoord/prefix/suffix x
- heeft even lengte/ veelvoud n
- subwoord x én/of subwoord y

- ? heeft géén subwoord x

- is een palindroom
- L^* van *willekeurige* L (!)
  $\{ ab^2, ab^3, ab^5, ab^7, \dots \}$
  $\{ ab, aabb, aaabbb, \dots \}$
  $= \{ a^n b^n \mid n \geq 1 \}$

$\boxed{L = L^*}$