Limbaje formale şi automate

Seriile 13 si 15

Săptămâna 5, 24 martie 2025

Andrei Păun

Cuprinsul cursului:
24 martie 2025

- Noțiuni despre limbajele DFA
- Lema de pompare pentru limbajele DFA
- Exerciții cu lema de pompare
- Exerciții din examenele anterioare

Noțiuni despre limbajele DFA

- Sunt toate limbajele modelate de către DFA-uri?
- Nu!
  – Majoritatea limbajelor NU sunt regulate/DFA
- De ce?
  – Automatul finit are memorie finită.
- Cum putem arăta că un limbaj este regulat?
- Cum putem arăta că un limbaj nu este regulat?

Limbajele finite

- ***Teoremă***: Orice limbaj finit L este modelat de un DFA

- ***Demonstrație***:
  - Pentru fiecare cuvânt din L se poate construi câte un automat (linie) cu stări disjuncte
  - Se mai definește o stare, cu λ-mişcări către stările inițiale ale automatelor
  - Se transformă λ-NFA în NFA şi apoi în DFA

Limbajele infinite

- Limbajele infinite pot fi regulate/DFA sau neregulate

- Exemplu de limbaj infinit regulat:

$$L=\{a^n \mid n\geq 0\}$$

- Exemplu de limbaj infinit ne-regulat :

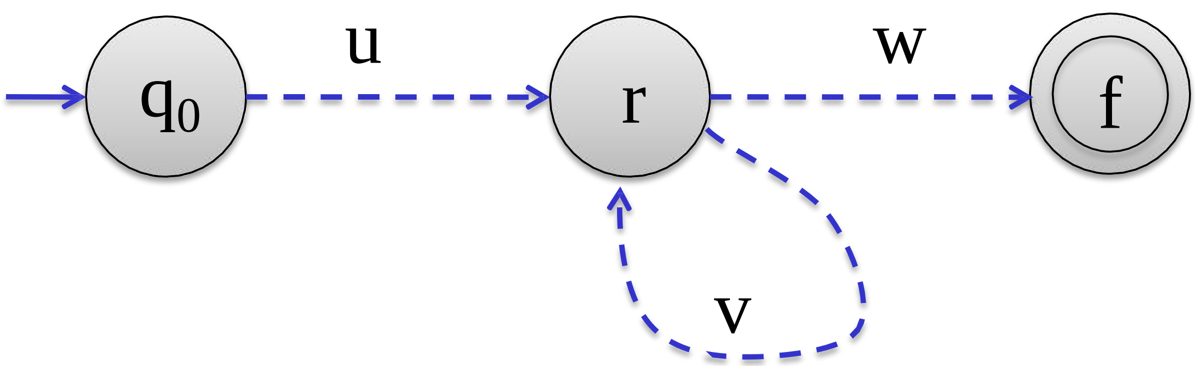
$$L = \{a^n b^n \mid n\geq 0\}$$

Lema de pompare pentru limbajele DFA

- Teoremă importantă în zona de automate finite
- Apare la examen în mod frecvent
- Orice limbaj regulat/DFA satisface această lema/proprietate
- Dacă un limbaj nu are această proprietate atunci acel limbaj NU este regulat

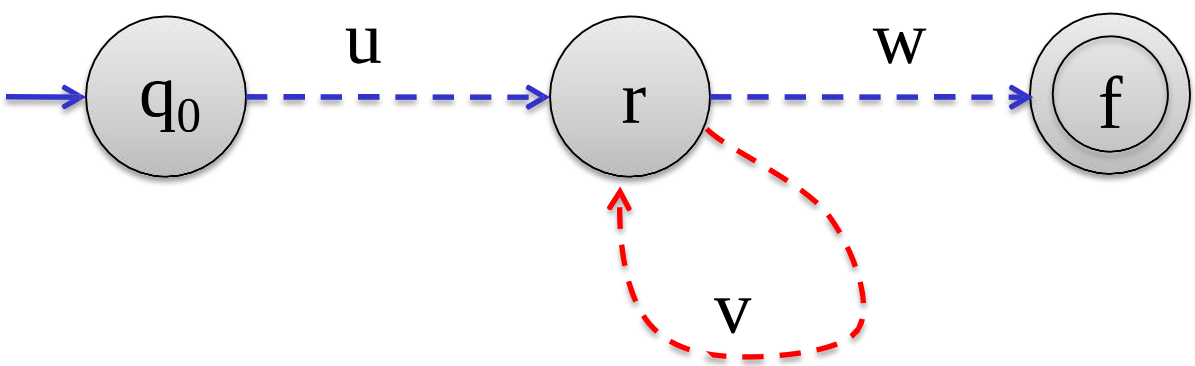
Ideea de bază

- Să presupunem că numărul de stări din automat este p
- Orice cuvânt z, $|z| \geq p$ trebuie să aibă un ciclu
- Notăm cuvântul procesat prin acest ciclu cu v, cuvântul până în ciclu cu u şi cuvântul după ciclu cu w



Lema de pompare (cont.)

- Ciclul poate fi repetat de oricâte ori şi apoi să se meargă cu w către o stare finală
- Deci partea v poate fi "pompată"
- Așadar mergem prin u până în starea care începe ciclul respectiv, apoi mergem de oricâte ori prin ciclu cu v, şi în final, mergem cu w către o stare finală şi acceptăm şi acest cuvânt.



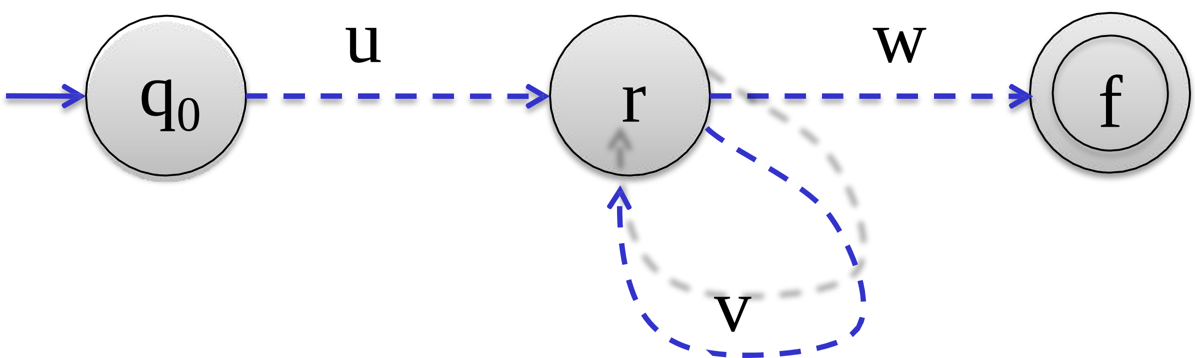
Lema de pompare

Lema: Fie L un limbaj regulat, există atunci un număr p astfel încât **orice** cuvânt $z \in L$, cu proprietatea că $|z| \geq p$, **poate** fi descompus/spart/împărțit în trei părți $z=uvw$ astfel încât:

1. $|uv| \leq p$

2. $|v|>0$

3. Pentru orice $i \geq 0$ avem că $uv^iw \in L$



Descriere informală

- Pentru fiecare limbaj regulat există un număr/o constantă p (care nu se modifică) pentru care lema de pompare spune că orice cuvânt de lungime mai mare decât această constantă se poate descompune în 3 componente (uvw) iar partea din mijloc v poate fi pompată
- În plus partea din mijloc (care e pompată) are cel puțin o literă
- Prefixul şi/sau sufixul (u sau w) pot fi λ
- Partea pompată face parte din primele p litere ale cuvântului.

Demonstrație formală

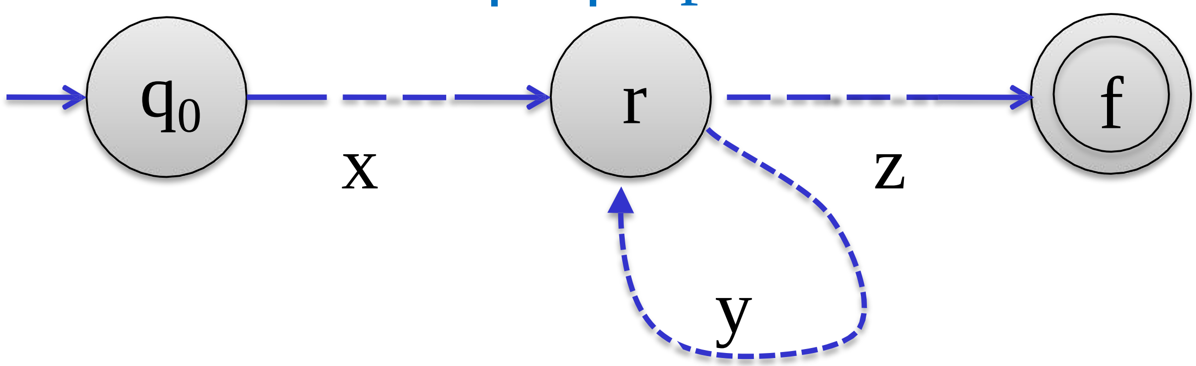
- Fie un automat DFA M automatul care modelează limbajul L. Fie p numărul de stări ale automatului M.
- Fie z un cuvânt din L de lungime mai mare sau egală cu p.
- Considerăm prefixul de lungime exact p litere al lui z şi procesarea prin automatul M a acestui prefix.
- Plecăm din starea inițială q_0 şi apoi fiecare stare va procesa câte o literă din prefixul de p litere considerat
- Pentru că considerăm cuvântul de lungime p, vom ajunge în starea q_p, deci trecem prin p+1 stări

Demonstrație formală (cont.)

- Din principiul cutiei şi pentru că avem p stări în automatul M, dar prefixul de lungime p al cuvântului va fi procesat de $n+1$ stări rezultă că cel puțin o stare apare de două ori
- Notăm o asemenea stare care apare de două ori cu r.
- Fie cuvântul u prefixul lui z cu care se ajunge în prima apariție a lui r
- Fie v subșirul din z (care apare imediat după u în z) prin care se ajunge din r în aceeași stare r.
- Fie w sufixul lui z pe care îl citim după v. Adică uv concatenat cu w ne dă z

Demonstrație formală (cont.)

- *w ne va duce din starea r intr-o stare finală*
- *u este un cuvânt care ciclează din r în aceeași stare r.*
- *de asemenea v are cel puțin o literă pentru că ne mutăm din stare în stare cu cel puțin o literă; uv fac parte din primele p litere ale cuvântului z, deci* $|uv| \leq p$



Demonstrație formală (cont.)

- De vreme ce cuvântul z e acceptat rezultă că există un drum de la starea r la o stare finală f, $q_0 \rightarrow r \rightarrow f$; asta înseamnă că cuvântul uw nu va merge prin ciclu şi va ajunge la starea finală f. ($uv^0w=uw$).
- Pe de altă parte, se poate merge un număr arbitrar de ori prin ciclu şi tot se ajunge în starea finală f: (uv^iw).

Demonstrație formală (cont.)

- așadar cuvântul z se poate descompune în cele trei părți u,v,w cu proprietățile

$|uv| \leq p$ deoarece am ales prefixul de exact p litere şi din acest prefix am definit v ca fiind bucata pe ciclul respectiv

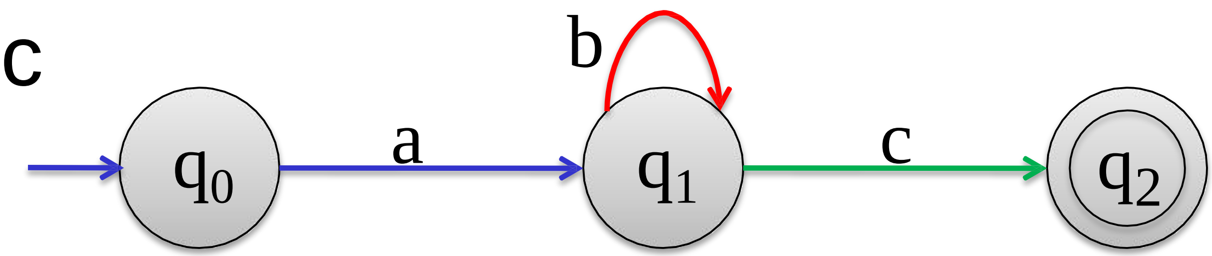
$|v|>0$ pentru că ciclul are cel puțin o literă

Pentru orice $i \geq 0$ avem că $uv^iw \in L$

Q.E.D.

Exemplu de limbaj

L= ab*c



pentru orice cuvânt de lungime mai mare de 3 din L putem să îl scriem ca uvw, luam ab^7c

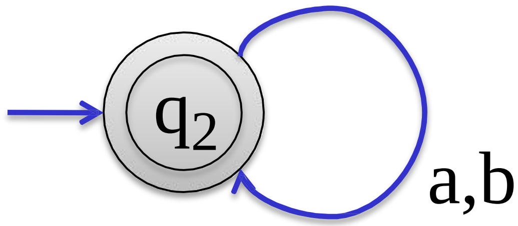
în cazul de faţă

$u=a$; $v=b$ (sau b^k cu $k>1$) si $w=c$ (sau $w= b^lc$)

Evident, orice cuvânt $uv^iw \in L$

Descompunerea e unică?

Un alt exemplu

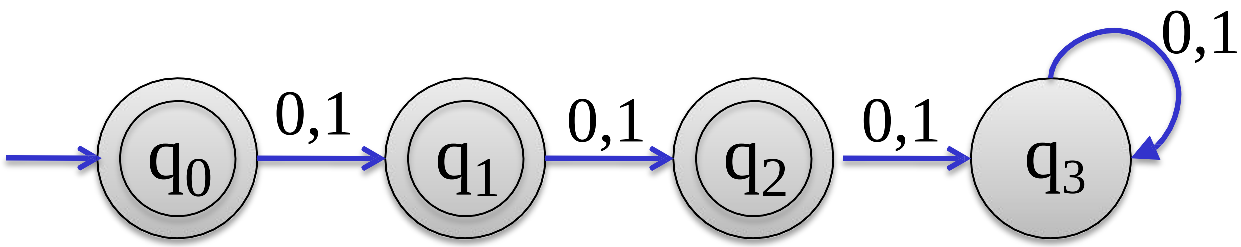


acceptă toate cuvintele peste alfabetul $\Sigma=\{a,b\}$, deci $L=\Sigma^*$

în acest caz putem lua u şi w cuvântul vid λ

Cazul limbajelor finite

O imagine grafică ar fi un automat de genul următor:



A are 4 stări. *L(A) conține toate cuvintele de lungime 2 sau mai puțin*. *L(A) este regulat doar că lema de pompare nu pare să funcționeze aici. De ce*???

Teoremă: Dacă un DFA A are limbajul acceptat finit şi numărul de stări din A este n, atunci lungimea celui mai lung cuvânt din L este mai mică decât n (|z|< n)

Dem: dacă ∃(w∈L şi |w|≥n) atunci ∃(w=xyz) astfel încât 1+2+3 există, proprietatea 3 îmi dă L infinit ceea ce ar fi o contradicție.

Utilizarea Lemei de Pompare

Utilizarea Lemei este de a demonstra că un limbaj NU ESTE regulat, dacă vrem să demonstrăm că un limbaj ESTE regulat, e suficient să dăm un automat pentru acel limbaj.

Tehnica de demonstrație: prin reducere la absurd

1.Presupunem prin absurd că limbajul ESTE regulat.

2.Atunci există un p fix pentru L şi căutam/găsim un cuvânt z de lungime mai mare decât p

3.Demonstrăm că nu există nici o descompunere a lui z în z=uvw astfel încât avem toate cele 3 proprietăți din L.P. De fapt proprietățile 1 şi 2 ne dau restricții pentru u dar mai ales pentru v iar din proprietatea 3 vom avea contradicția

Atentie!

- Lema: dacă L este regulat $\exists n$, a.î. $\forall (w \in L$, şi $|w| \geq n)$
  $\exists$ (w=xyz) a.î. 1+2+3
- regulat(L) $\Rightarrow$ lema (L).
- lema (L) $\not\Rightarrow$ regulat(L).
- ~lema(L) $\Rightarrow$ ~regulat(L).
- Ca să arătăm că L nu e regulat vom arăta că $\exists (w \in L$ cu proprietatea $|w| \geq n)$ a.î. $\forall$(w=xyz) , ~1 or ~2 or ~3. (deobicei 1 şi 2 şi ~3)

Greşeli comune

- $\exists z \in L$ a.î. lema(z) $\Rightarrow$ regulat(L)
- $\forall z \in L$, lema(z) $\Rightarrow$ regulat(L)
- $\exists z \in L$ şi $\exists (z=uvw)$ a.î. $uv^iw \notin L$ $\Rightarrow$ ~regulat(L)

(trebuie arătat pentru toate posibilele descompuneri)

- $\exists z \in L$ si $\forall (z=uvw)$ $uv^iw \notin L$ $\Rightarrow$ ~regulat(L)

(|w| e posibil să fie mai mică decât p)

$$L = \{0^m 1^m \mid m>0\}$$

- Arătăm cu Lema de pompare că limbajul următor $L = \{0^m 1^m \mid m>0\}$ nu e regulat.
- Presupunem prin absurd că L este regulat.
- Fie p numărul dat de lema de pompare pentru L (este numărul de stări din eventualul automat care îl acceptă).
- Consider $z= 0^p 1^p$.
- w este evident din L, $|z|>p$ așadar din LP z poate fi descompus in cel puțin o descompunere $z=uvw$.

$L = \{0^m 1^m \mid m>0\}$ (cont.)

- u şi v sunt ambele în prefixul de lungime p, deci amândouă au doar 0-uri în componenţă.
- Notăm lungimea lui u cu *s* şi lungimea lui v cu *t*.

$$\underbrace{000\ldots0}_{s}\underbrace{00}_{t}00111\ldots111$$

- Deci :
  – $u=0^s$ $s\geq0$
  – $v=0^t$ $t\geq1$, $s+t\leq n$
  – $w=0^{n-s-t}1^n$

•Din LP avem că uv^iw este în L pentru orice i.

$L = \{0^m 1^m \mid m>0\}$ (cont.)

- Luăm i=2.
- Cuvântul obținut $uv^2w = 0^s0^{2t}0^{n-s-t}1^n = 0^{n+t}1^n$.
- Acest cuvânt nu este în L pentru $t\geq1$: numărul de 0-uri e mai mare decât numărul de 1-uri.
- Am găsit o contradicție cu LP: nu există descompuneri ale lui $w= 0^p1^p$ cu proprietățile 1,2,3
- Adică: presupunerea că L e regulat e greșită.
- L nu e limbaj regulat. Q.E.D.

$L=\{a^{k!} \mid k\geq 0\}$

- Este $L=\{a^{k\,!} \mid k\geq 0\}$ regulat?
- Demonstrăm folosind Lema de pompare că NU este regulat.
- Presupunem că L este regulat (prin absurd).
- Fie *p* valoarea din lema de pompare
- Alegem $z= a^n$ unde $n=\max(p!,2)$
- z este în L, $|z|\geq p$ deci din LP w poate fi descompus în trei părţi z=uvw.

$L=\{a^{k!} \mid k \geq 0\}$

- u şi v au doar a-uri în componenţă.
- Fie lungimea lui u *s* şi lungimea lui v *t*.
- Atunci avem:

  - $u=a^s$, $s \geq 0$
  - $v=a^t$ $t>0$, $s+t \leq p$ (din 1 si 2)
  - $w=a^{p!-t-s}$

- Din proprietatea 3: uv^iw este în L pentru orice i.

$$L=\{a^{k!} \mid k\geq 0\}$$

- Fie i=2.
- Cuvântul obținut este $uv^2w = a^{p!+t}$.
- Știm din 2 şi 1 că $t>0$ şi $t\leq p$, deci avem

$$p!+t \leq p!+p < p!+(p+1) < p!*(p+1) = (p+1)!$$

- Deci pentru i=2 cuvântul nu este în limbaj (este între două factoriale)
- Așadar orice descompunere a cuvântului $a^{p!}$ care satisface 1 şi 2 nu îl va satisface pe 3, deci avem contradicție, așadar presupunerea că L este regulat a fost falsă

$L=\{0^m1^n \mid m>n\}$

- Presupunem că L este regulat.
- Fie *p* numărul promis de Lema de pompare.
- Fie $z=0^{p+1}1^p$; evident $|z|>p$
- *z* se poate descompune in $z=uvw$.
- din 1: $|uv|\leq p$, rezultă că *v* este compus doar de 0-uri, iar din 2 $|v|>0$.
- $\Rightarrow \; |(uv^0w)|_0 \leq |(uv^0w)|_1$
- $\Rightarrow uv^0w \notin L$ ceea ce contrazice punctul 3 din LP $\Rightarrow$ L nu e regulat.

$$L = \{w \mid |w|_1 = |w|_0\}$$

- Este regulat L?

- **idee de demonstrație**:
  - Presupunem prin absurd că este regulat.
  - din LP avem *p* a.î. … 1,2,3
  - Alegem $z=0^n1^n \in L$
  - Pompăm *v* spre un $v' \notin L$
- **Problemă**: dacă alegem $v=0^n1^n$ $v' \in L$??

$$L = \{w \mid \#_1(w) = \#_0(w)\}$$

- Este L regulat? (nu)

- **Proof** :
  - REG e închisă la intersecție.
  - Pp prin absurd că e REG.
  - Atunci L'=($0^*1^* \cap L$) este şi el regulat.
  - Dar L'=$\{0^n1^n | n>0\}$ nu e REG

  - L nu e regulat

$L = \{ww \mid w \in \{0,1\}^* \}$

- Este regulat? (nu)

- **Dem**:
  - Pp prin absurd că e REG.
  - Avem lema de pompare…
  - alegem $w = 0^n 1 0^n 1 \in L$
  - Pompăm *w* spre un $w' \notin L$

Problemă din examen vechi

9. (10 puncte) Spuneţi dacă limbajul următor este sau nu regulat. Dacă limbajul este regulat construiţi un automat finit determinist care să îl accepte, dacă nu, demonstraţi folosind lema de pompare pentru REG că limbajul nu este regulat $L = \{a^k wcw \mid w \in \{a,b,c\}^*, k \geq 4\}$.
ALTERNATIV pentru max 5 puncte: $L = \{a^{k-1}b^{2l+3} \mid k, l \geq 5\}$.

- Deci $L_1=\{a^k wcw \mid k>3\}$
- $L_2=\{a^{k-1}b^{2l+3} \mid k,l>4\}$

$L_2=\{a^{k-1}b^{2l+3}|k,l>4\}$

- Se poate construi un DFA pentru
  - $L_3=\{a^k| k>3\}$: un ciclu de lungime 3 şi de asemenea pentru
  - $L_4=\{b^{2l+3}| l>4\}$: un ciclu de lungime 2 precedat de un drum (“o linie”) de lungime 3
- Se poate face concatenarea lui L3 cu L4 şi acesta este DFA-ul pentru limbajul respectiv

$L_1=\{a^k wcw \mid k>3\}$

- Limbajul nu e REG
- Demonstrație prin lema de pompare
- Alegem cuvântul $a^4b^pa^pcb^pa^p$