Limbaje Formale și Automate
Tutoriat 1

Gabriel Majeri

1 Introducere

Cursul este destul de teoretic, iar la examen veți avea atât o parte de teorie cât și o parte de exerciții. Pentru teorie ar fi bine să știți **algebră** și câteva noțiuni legate de **grafuri** pentru că demonstrațiile conțin noțiuni din aceste materii.

Laboratorul explorează aplicațiile **practice** ale noțiunilor discutate la curs, cum ar fi utilizarea expresiilor regulate pentru căutarea în text, sau importanța gramaticilor în compilatoare.

Cuprins

2 Notare

Acestea sunt câteva sfaturi practice legate de notarea la această materie (relevante în special pentru seria 13).

Examen

Timpul alocat pentru rezolvarea subiectelor este **2 ore**. Modele de examen din anii trecuți se găsesc pe palcu/fmi.

Teorie

Pe partea de teorie, întrebările sunt de forma:

- Să demonstrezi o teoremă sau o propoziție din curs;

  Exemple:

  - Demonstrați lema de pompare pentru limbaje independente de context;
  - Demonstrați prima implicație a teoremei Myhill-Nerode;
  - Demonstrați trei proprietăți ale limbajelor regulate.

- Să justifici dacă o anumită afirmație este sau nu adevărată;

  Exemple:

  - Limbajele regulate sunt închise la complementare;
  - Fie limbajele $\mathcal{L}_1, \mathcal{L}_2$ cu $\mathcal{L}_1 \subset \mathcal{L}_2$ și $\mathcal{L}_2 \in \mathrm{REG}$. Atunci $\mathcal{L}_1 \in \mathrm{REG}$.

- Să justifici dacă o anumită problemă este sau nu decidabilă.

  Exemple:

  - Argumentați dacă este decidabilă egalitatea între două limbaje regulate;
  - Justificați dacă determinarea intersecției a două limbaje independente de context este decidabilă.

Exerciții

Exemple de exerciții pe care le puteți primii la examen:

- Se dă un limbaj, dacă acesta este regulat/independent de context scrieți un automat sau o expresie regulată/o gramatică care să îl accepte. Dacă nu, demonstrați că nu este regulat/nu este independent de context (folosind lema de pompare respectivă).
- Se dau două DFA/NFA/λ-NFA: să se minimizeze, să se verifice dacă acceptă același limbaj, sau să se calculeze intersecția/reuniunea lor.
- Să se dea exemplu de o gramatică independentă de context care respectă anumite condiții: să aibă un anumit număr de terminali/neterminali, un anumit număr de producții, să genereze un limbaj finit, etc.
- Se dă o gramatică independentă de context, să se aducă la o formă normală Chomsky, sau să se elimine producțiile unitate, etc.
- Să se scrie un automat push-down determinist/nedeterminist care să accepte un limbaj prin stare vidă/prin stare finală/prin stare finală și stivă vidă.

Laborator

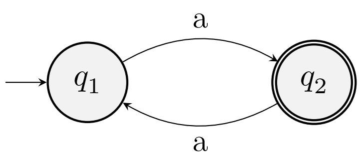
Îți alegi teme (exerciții) pe care le rezolvi individual și apoi le prezinți la laborator. În general, temele mai grele (care necesită mai mult efort) au note maxime mai mari (dar și penalitate mai mare dacă nu le faci).

3 Limbaje formale

Orice problemă de decizie (adică cu răspuns „da” sau „nu”) din informatică poate fi redusă la determinarea apartenenței la un limbaj. De exemplu, verificarea că un număr k este *prim*, se poate face verificând dacă cuvântul $\underbrace{aaa\ldots aaa}_{k \text{ ori}}$ aparține limbajului $\{\, a^p \mid p \text{ prim} \,\}$.

3.1 Limbaje regulate

Cum le recunoaștem: Principala limitare a DFA/NFA/λ-NFA este că nu au memorie. În formula pentru limbajele regulate pot apărea doar condiții liniare, de forma a^{nk+m}, și dacă apar mai mulți indici la putere, aceștia nu sunt corelați.



Un DFA care acceptă limbajul $\mathcal{L} = \{\, a^{2k+1} \mid k \in \mathbb{N} \,\}$

Observație. Orice limbaj finit este regulat. Putem construi un DFA care să aibă câte o stare finală pentru fiecare cuvânt. Astfel se obține un trie.

Proprietăți: închise la intersecție, reuniune, complement, diferență de mulțime. Sunt închise și la concatenare și la stelare, și la morfisme și morfisme inverse.

3.2 Limbaje independente de context

Cum le recunoaștem: Putem avea indici corelați (de exemplu $a^{2k}b^{3k}$).

Dacă apar mai mulți indici corelați, aceștia ar trebui să se grupeze asemenea parantezelor corect închise. De exemplu, se poate arăta că $a^n b^m c^m d^n$ este independent de context dar $a^n b^m c^n d^m$ **nu** este.

De asemenea, nu pot fi mai mult de două variabile corelate. De exemplu $a^n b^n c^n$ este exemplul clasic de limbaj care **nu** este independent de context.

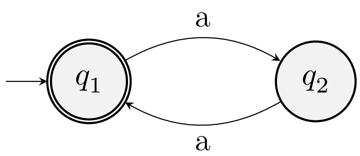
Proprietăți: închise la toate operațiile menționate mai sus **cu excepția** intersecție, complement, sau diferență. Sunt închise totuși la intersecția *cu un limbaj regulat* (acest lucru ne ajută în exerciții).

4 Exerciții

Exercițiu 1. Scrieți un automat finit determinist care să accepte limbajul

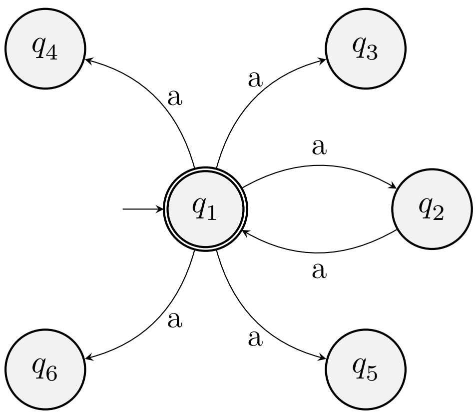
$$\mathcal{L} = \{ a^{2k} \mid k \geq 0 \}$$

Rezolvare. Cel mai simplu automat care acceptă acest limbaj este: □



Exercițiu 2 (exercițiul 7 din examen iunie 2011)**.** Dați exemplu de un NFA-care nu este nici λ-NFA, nici DFA. Automatul trebuie să aibe minim 6 stări accesibile din starea inițială. Transformați automatul într-un DFA.

Rezolvare. Ni se cere ca stările să fie accesibile, nu neapărat utile. Putem să modificăm automatul de mai sus:



Folosim metoda tabelului pentru a obține DFA-ul echivalent:

	a
Notăm cu $A = \{ q_1 \}$	$\{ q_2, \dots, q_6 \}$
Notăm cu $B = \{ q_2, \dots, q_6 \}$	$\{ q_1 \}$

Obținem DFA-ul inițial, unde q_1 este A și q_2 este B. □

Exercițiu 3 (bazat pe exercițiul 9 din examen iunie 2011)**.** Construiți o gramatică independentă de context care să genereze limbajul

$$\mathcal{L} = \{ a^{2k} b^l a^k \mid k \geq 0, l \geq 1 \}$$

Rezolvare. Vom genera partea exterioară a cuvântului (a-urile) prin A și oricâte b-uri prin B.

$$\begin{aligned} S &\to A \\ A &\to aaAa \\ A &\to B \\ B &\to bB \\ B &\to b \end{aligned}$$

□

Exercițiu 4 (exercițiul 8 din examen iunie 2013)**.** Construiți o gramatică independentă de context care să genereze limbajul

$$\mathcal{L} = \{\ 0^{4k}1^l0^k \mid k \geq 0, l \geq 1\ \} \cdot \{\ 0^i1^{j+3} \mid i \neq j\ \}$$

Rezolvare. Pentru a genera prima mulțime, putem folosi producțiile:

$$\begin{aligned} S_1 &\to 0^4S0 \mid A \\ A &\to 1A \mid 1 \end{aligned}$$

Pentru a genera un număr diferit de 0 și de 1, trebuie să generăm unul din două cazuri: $i > j$ sau $i < j$.

$$\begin{aligned} & S_2 \to U \mid V \\ & U \to 0U \mid 0T && (i > j) \\ & V \to 1V \mid 1T && (i < j) \\ & T \to 0T1 \mid \lambda \end{aligned}$$

Le putem unii prin producția:

$$S \to S_1S_21^3$$

□

Exercițiu 5 (exercițiul 10 din examen iunie 2018)**.** Construiți o gramatică independentă de context care să genereze limbajul

$$\mathcal{L} = \{\ a^{m+n}b^ka^{m+k+i}b^n \mid i, k, m, n \geq 1\ \}$$

Rezolvare. Putem rescrie limbajul ca:

$$\mathcal{L} = \{\ a^na^mb^ka^ka^ma^ib^n \mid i, k, m, n \geq 1\ \}$$

Acum se vede mult mai ușor că se poate genera limbajul:

$$
\begin{aligned}
&S \to aSb \mid aAb && \text{(generează } a^n \dots b^n) \\
&A \to BC && \\
&B \to aBa \mid aDa && \text{(generează } a^m \dots a^m) \\
&C \to aC \mid a && \text{(generează } a^i) \\
&D \to bDa \mid ba && \text{(generează } b^k a^k)
\end{aligned}
$$

□

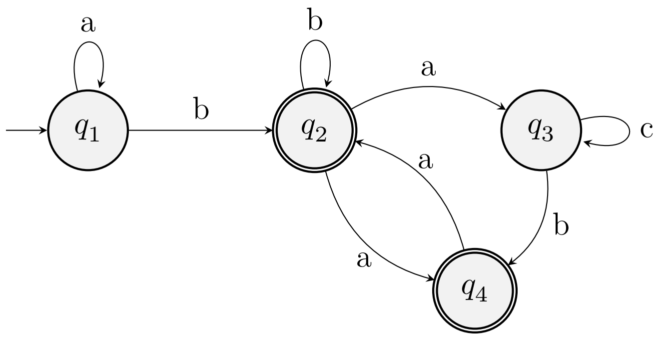
Exercițiu 6 (exercițiul 7, subpunctul *a* din examen iunie 2016)**.** Dați o gramatică independentă de context cu 7 producții, 2 din ele să fie producții unitare (unit production), și care are cel puțin 2 simboluri neterminale (nonterminating) și un simbol inaccesbil (unreachable).

Rezolvare.

$$
\begin{aligned}
S &\to AB \mid C \mid \lambda && \\
A &\to a, B \to b && (A, B \text{ simboluri neterminale}) \\
C &\to A, C \to B && \text{(producții unitare)} \\
X &\to Y && \text{(două simboluri inaccesibile)}
\end{aligned}
$$

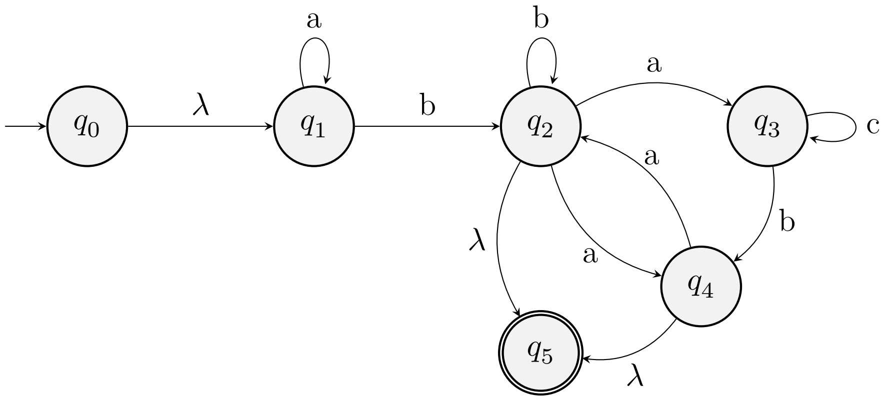
□

Exercițiu 7 (exercițiul 7 din examen iunie 2017)**.** Pentru următorul automat dați fiecare pas din algoritmul de construire al expresiei regulate echivalente:

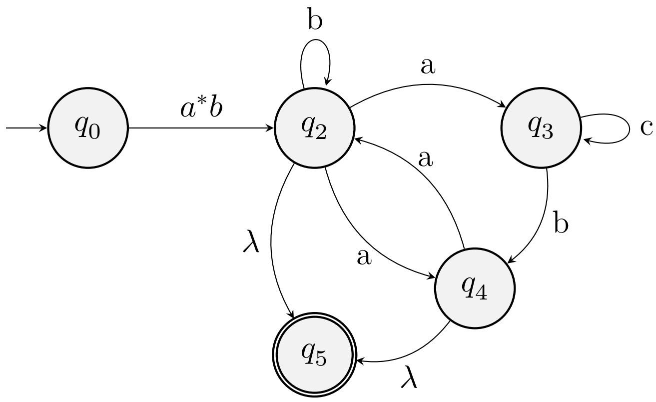


Rezolvare.

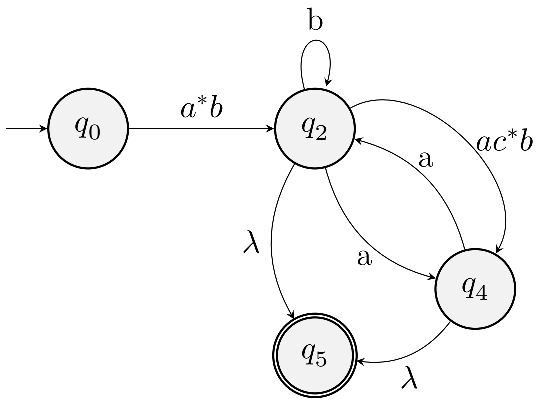
Începem prin a adăuga o nouă stare inițială și modificăm automatul să aibă o singură stare finală:



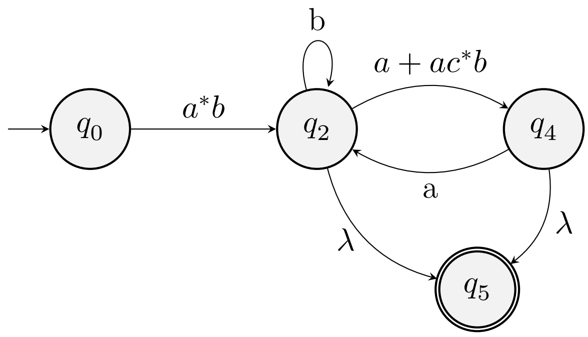
Începem să eliminăm din stări. Începem cu q_1:



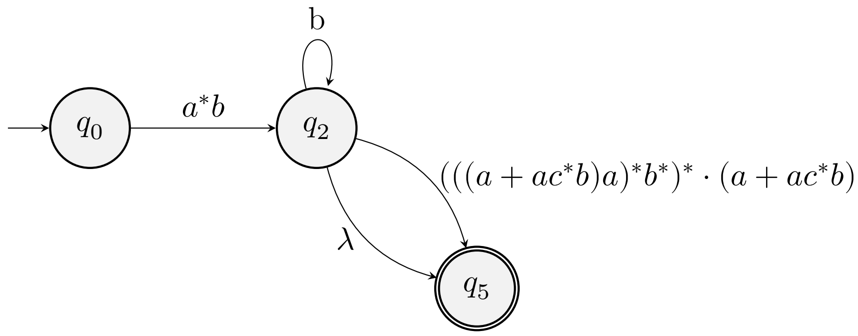
Eliminăm q_3:



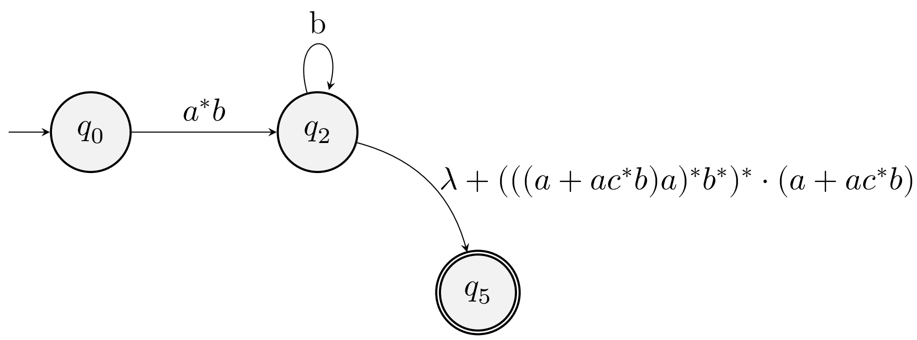
Reunim muchiile de la q_2 la q_4:



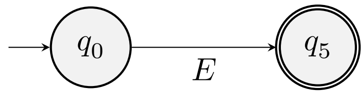
Eliminăm starea q_4:



Reunim muchiile de la q_2 la q_5:



Eliminăm starea q_2:



Unde E este expresia regulată cerută:

$$a^*bb^* \cdot (\lambda + (((a + ac^*b)a)^*b^*)^* \cdot (a + ac^*b))$$

$\square$