Formális nyelvek és automaták vizsga jegyzet

Szerkesztette: Szabó Dániel

2014. június 5.

1. feladathoz való fogalmak:

- **ε :** A 0 hosszúságú sorozatot üres szónak nevezzük és ε-al jelöljük.
- **V*:** A V abc feletti szavak halmazát beleértve az üres szót is V*-al jelöljük.
- **V^+:** A V abc feletti nem üres szavak halmazát V^+ -al jelöljük
- **Részszó:** Legyen V abc és legyenek u és v szavak V felett. Az u szót a v szó részszavának nevezzük, ha v = xuy teljesül valamely x és y V feletti szavakra. Az u szót a v szó valódi részszavának mondjuk, ha u != v és u != ε.
- **Prefix, Szufix:** Legyen V abc és legyenek u,v,x,y szavak V felett, továbbá legyen v = xuy. Ha x = ε, akkor u-t a v szó prefixének vagy kezdőszeletének, ha y = ε, akkor u-t a v szó szufixének vagy utótagjának hívjuk.
- **Szó hossza:** A w szó hosszán a w szót alkotó szimbólumok számát értjük (azaz,
- w mint sorozat hosszát) és |w|-vel jelöljük.
- **Szó i-edik hatványa (0. első i-edik):** Legyen i nem negatív egész szám és legyen w a V abc feletti szó (w ∈ V*). A w szó i-edik hatványa alatt a w szó i példányának konkatenáltját értjük és w^i-vel jelöljük. Konvenció alapján minden w ∈ V* szóra $w^0 = ε$.
- **Tükörkép:** Legyen u egy V abc feletti szó. Az u szó tükörképe vagy fordítottja alatt azt a szót értjük, amelyet úgy kapunk, hogy u szimbólumait megfordított sorrendben írjuk. Az u szó tükörképét u^{-1} -gyel jelöljük.
- **Levezetés:** Legyen G = (N,T,P,S) egy generatív grammatika és legyen u,v ∈ (N ∪ T)*. Azt mondjuk, hogy a v szó k lépésben levezethető az u szóból G-ben, k ≥ 1, ha létezik olyan $u_1,\dots,u_{k+1}$ ∈ (N ∪ T)* szavakból álló sorozat, amelyre $u = u_1$, $v = u_{k+1}$, valamint $u_i \Longrightarrow_G u_{i+1}$, $1 \le i \le k$ teljesül. A v szó levezethető az u szóból G-ben, ha vagy u = v, vagy létezik olyan k ≥ 1 szám, hogy a v szó az u szóból k lépésben levezethető.
- **Közvetlen Levezetés**: Legyen G = (N,T,P,S) egy generatív grammatika és legyen u,v ∈ (N ∪ T)*. Azt mondjuk, hogy a v szó közvetlenül vagy egy lépésben levezethető az u szóból G-ben és ezt $u \Rightarrow_G v$ módon jelöljük, ha u = u1xu2, v = u1yu2, u1,u2 ∈ (N ∪ T)* és x → y ∈ P.
- **Konkatenáció:** Legyen V egy abc és legyenek u,v V feletti szavak (azaz, legyen u,v ∈ V*). Az uv szót az u és v szavak konkatenáltjának nevezzük. PL: Legyen az abc V = {a,b,c}, legyenek u = abb és v = cbb szavak. Akkor uv = abbcbb az u és v konkatenáltja. A konkatenáció mint művelet asszociatív, de általában nem kommutatív.
- **Reguláris műveletek:** Az unió, a konkatenáció, valamint a lezárás (a ∗) műveleteket együttesen reguláris műveleteknek nevezzük.
- **Nyelv:** Legyen V abc és legyen L tetszőleges részhalmaza V *-nak. Akkor L-et egy V feletti nyelvnek nevezzük. Az üres nyelv - amely egyetlen szót sem tartalmaz – jelölése ∅. Egy nyelvet véges nyelvnek mondunk, ha véges számú szót tartalmaz, ellenkező esetben végtelen nyelvről beszélünk.
- **Üres Nyelv *:** Nem üres, tartalmazza az epszilont.
- **De-Morgan:**
- **Asszociatív:**
- **Disztributív:**

2. feladathoz való fogalmak: (Generatív grammatikák)

- **Mit nevezünk generatív grammatikának:**
  - Egy G generatív grammatikán (grammatikán vagy (generatív) nyelvtanon) egy (N,T,P,S) négyest értünk, ahol:
    - N és T diszjunkt abc-k, a nemterminális és a terminális szimbólumok abc-i
    - $S \in N$ a kezdőszimbólum (axióma)
    - P véges halmaza (x,y) rendezett pároknak, ahol $x,y \in (N \cup T)^*$ és x legalább egy nemterminális szimbólumot tartalmaz. A P halmaz elemeit átírási szabályoknak (röviden szabályoknak) vagy produkcióknak nevezzük. Az (x,y) jelölés helyett használhatjuk az $x \rightarrow y$ jelölést is, ahol a $\rightarrow$ szimbólum nem eleme az $(N \cup T)$ halmaznak.
- **Levezethetőséggel kapcsolatos fogalmak:** LSD 1. feladat szabályok.
- **Generált nyelv:** Legyen G = (N,T,P,S) egy tetszőleges generatív grammatika. A G grammatika által generált L(G) nyelv alatt az $L(G) = \{w | S \Longrightarrow^*_G w, w \in T^*\}$ szavakból álló halmazt értjük. Azaz, a G grammatika által generált nyelv a T^* halmaz azon elemei, amelyek levezethetők a G grammatika S kezdőszimbólumából.
- **Ekvivalencia:** Két generatív grammatikát ekvivalensnek nevezünk, ha ugyanazt a nyelvet generálják, gyengén ekvivalensnek, ha legfeljebb az üres szóban különböznek.
- **1,2,3 as típusu grammatika:**
  - A G = (N,T,P,S) generatív grammatikát i-típusúnak mondjuk, (i = 0,1,2,3) ha P szabályhalmazára teljesülnek a következők:
    - i = 0: Nincs korlátozás.
    - i = 1: P minden szabálya $u1Au2 \rightarrow u1vu2$ alakú, ahol $u1,u2,v \in (N \cup T)^*$ és $A \in N$, és $v \mathrel{!}= \varepsilon$, kivéve az $S \rightarrow \varepsilon$ alakú szabályt, feltéve, hogy P-ben ilyen szabály létezik. Ha P tartalmazza az $S \rightarrow \varepsilon$ szabályt, akkor S nem fordul elő P egyetlen szabályának jobboldalán sem.
    - i = 2: P minden szabálya $A \rightarrow v$ alakú, ahol $A \in N$ és $v \in (N \cup T)^*$
    - i = 3: P minden szabálya vagy $A \rightarrow uB$ vagy $A \rightarrow u$, alakú, ahol $A,B \in N$ és $u \in T^*$.
- **Mondat szerkezetű (rekurzívan felsorolható) grammatika:** A 0-s típusú grammatikát rekurzívan felsorolható, mondat szerkezetű grammatikának is mondjuk.
- **Környezet függő Grammatika:** Az 1-típusú grammatikát környezet függő grammatikának is mondjuk.
- **Környezet független Grammatika:** A 2-típusú grammatikát környezet független grammatikának is mondjuk.
- **Reguláris grammatika:** A 3-típusú grammatikát reguláris vagy véges állapotú grammatikának is mondjuk.
- **Mikor aktív egy nem terminális szimbólum:** A környezetfüggetlen grammatika egy nemterminálisát inaktívnak vagy nem aktívnak nevezzük, ha nem vezethető le belőle terminális szó; egyébként aktívnak mondjuk.
- **Mikor elérhető egy nem terminális szimbólum:** A környezetfüggetlen grammatika egy nemterminálisát nem elérhetőnek nevezzük, ha nem fordul elő egyetlen olyan mondatformában sem, amely a kezdőszimbólumból levezethető; egyébként elérhetőnek mondjuk.
- **Mikor hasznos egy nem terminális szimbólum:** Egy nemterminálist nem hasznosnak mondunk, ha vagy inaktív, vagy nem elérhető, vagy mindkét tulajdonság teljesül esetében. Egy nemterminálist hasznosnak nevezünk, ha aktív és elérhető.
- **Redukált grammatika jelentése:** Egy környezetfüggetlen grammatika redukált, ha minden nemterminálisa aktív és elérhető.

3. feladathoz való fogalmak: (Reguláris nyelvek)

- **Chomsky normálforma fogalma:** A G = (N,T,P,S) környezetfüggetlen grammatikát
- Chomsky-normálformájúnak mondjuk, ha minden egyes szabálya vagy:
  - o $X \to a$, ahol $X \in N$, $a \in T$, vagy
  - o $X \to Y Z$, ahol $X,Y,Z \in N$ alakú
- **Reguláris kifejezés fogalma:** Legyenek V e´s V ′ = {ε,·,+,∗,(,)} diszjunkt abc-k. A $V \cup V'$ abc feletti reguláris kifejezéseket rekurzív módon a következőképpen definiáljuk:
  - o ε reguláris kifejezés V felett.
  - o Minden $a \in V$ reguláris kifejezés V felett.
  - o Ha R reguláris kifejezés V felett, akkor $(R)*$ is reguláris kifejezés V felett.
  - o Ha Q és R reguláris kifejezések V felett, akkor $(Q)\cdot(R)$ e´s $(Q)+(R)$ is reguláris kifejezések V felett.

  A reguláris kifejezések az L3 nyelvosztályt a grammatikáktól eltérően határozzák meg; megadunk formulák egy osztályát, ahol minden formula egy nyelvet jelöl, és megmutatjuk, hogy a formulák által jelölt nyelvek családja L3.

- **Lineáris grammatika fogalma:** Egy G = (N, T, P, S) környezetfüggetlen grammatikát lineárisnak nevezünk, ha minden szabálya vagy
  - o $A \to u$, $A \in N$, $u \in T^*$ vagy
  - o $A \to u1Bu2$ alakú, ahol $A, B \in N$ és $u1, u2 \in T^*$
- **Lezártak fogalma:** Az L nyelv az lezártja (Kleene-lezártja) alatt az $L^* = \bigcup_{i\geq 0} L^i$ nyelvet értjük.
  A megfelelő műveletet lezárásnak vagy ∗ műveletnek mondjuk.
  Az L^+ nyelv alatt az $L^+ = \bigcup_{i\geq 1} L^i$ nyelvet értjük. Nyilvánvalóan, ha $\varepsilon \in L$, akkor $L^+ = L^*$.

Legyen V egy abc és legyenek L1,L2 nyelvek V felett ($L1 \subseteq V^*$ és $L2 \subseteq V^*$).

- **Reguláris nyelvek metszete:** $L1 \cap L2 = \{u \mid u \in L1 \text{ és } u \in L2\}$ - az L1 és az L2 nyelv metszete.
- **Reguláris nyelvek uniója:** $L1 \cup L2 = \{u \mid u \in L1 \text{ vagy } u \in L2\}$ - az L1 és az L2 nyelv uniója
- **Reguláris nyelvek konkatenáltja:** $L1L2 = \{u1u2 \mid u1 \in L1, u2 \in L2\}$ - az L1 és az L2 nyelv konkatenációja.
- **Reguláris nyelvek különbsége:** $L1 - L2 = \{u \mid u \in L1 \text{ és } u \,!\!\in L2\}$ - az L1 és az L2 nyelv különbsége.
- **Reguláris nyelvek tükörképe:** $L^{-1} = \{u^{-1} | u \in L\}$ a tükörképe (megfordítása) az L nyelvnek. $(L^{-1})^{-1} = L$ és $(L^{-1})^i = (L^i)^{-1}$, $i \geq 0$.
- **Reguláris nyelvek komplemense:** Az $L \subseteq V^*$ nyelv komplemense a V abc-re vonatkozóan L felül vonás = $V^* - L$.
- **i-edik hatvány:** L^i jelöli az L nyelv i-edik hatványát, ahol $i \geq 0$.($L0 = \{\varepsilon\}$, $L^n = L^{n-1}L$, $n \geq 1$.)
- Minden L nyelvre fennállnak a következő egyenlőségek: $\emptyset L = L\emptyset = \emptyset$ és $\{\varepsilon\}L = L\{\varepsilon\} = L$.
- **Homomorfizmus:** A $h : V1^* \to V2^*$ leképezést homomorfizmusnak nevezzük, ha $h(uv) = h(u)h(v)$ minden $u,v \in V1^*$ esetén. A fenti tulajdonság alapján $h(\varepsilon) = \varepsilon$. (Minden $u \in V1^*$ -ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.)
- **Izomorfizmus:** A h homomorfizmust izomorfizmusnak nevezzük, ha bármely u és v V1*-beli szóra teljesül, hogyha h(u) = h(v), akkor u = v.

4. feladathoz való fogalmak: (Véges automaták)

- **Determinisztikus automata fogalma:** A (determinisztikus) véges automata egy rendezett ötös, $A = (Q,T,\delta,q0,F)$
  - ahol Q állapotok véges nem üres halmaza,
  - T input szimbólumok abc-je
  - δ egy leképezés, az ún. állapot-átmeneti (röviden átmeneti) függvény a $Q \times T$ halmazból a Q halmazba,
  - $q0 \in Q$ az ún. kezdeti állapot (kezdő állapot)
  - $F \subseteq Q$ az ún. elfogadó állapotok halmaza.

  Szokás az elfogadó állapot helyett végállapotot is mondani.

  A δ függvény egyértékű, ezért minden egyes (q,a) párra, ahol $(q,a) \in Q \times T$ egyetlen olyan p állapot létezik, amelyre $\delta(q,a) = p$ teljesül. Ezért ezt a véges automata´t determinisztikusnak nevezzük.

- **Nem determinisztikus automata fogalma:** Ha többértékű állapot-átmeneti függvényt is megengedünk, azaz δ a $Q \times T$ halmazból a 2Q halmazba való leképezés, akkor nemdeterminisztikus véges automatáról beszélünk. Ebben az esetben aktuális állapotnak egy állapothalmaz valamely elemét, mintsem egyetlen állapotot tekinthetünk. Ez azt jelenti, hogy a kezdeti állapot helyettesíthető egy $Q0 \subseteq Q$ kezdő állapothalmazzal (kezdeti állapothalmazzal). Az is előfordulhat, hogy egy a input szimbólum esetén $\delta(q,a)$ üres az aktuális állapotok mindegyikére.
- **Közvetlen lépés:** Legyen $A = (Q,T,\delta,Q0,F)$ egy véges automata és legyenek $u,v \in QT^*$ szavak (ahol QT^* a Q és T^* konkatenációját jelöli). Azt mondjuk, hogy az A automata az u szót a v szóra redukálja egy lépésben vagy közvetlenül ($u \Longrightarrow A\ v$), ha van olyan $qa \rightarrow p$ szabály $M\delta$-ban (azaz, $\delta(q,a) = p$) és van olyan $w \in T^*$ szó, amelyre $u = qaw$ és $v = pw$ teljesül.
- **Redukció:** Az $A = (Q,T,\delta,Q0,F)$ véges automata az $u \in QT^*$ szót a $v \in QT^*$ szóra redukálja ($u \Longrightarrow^*_A v$), ha vagy $u = v$, vagy van olyan $z \in QT^*$, $u \Longrightarrow^*_A z$ és $z \Longrightarrow^*_A v$ teljesül. A $\Longrightarrow_A$ reláció, valamint tranzitív és reflexív lezártja, $\Longrightarrow^*_A$ lényegében a grammatikák elméletéből ismert levezetésnek felel meg.
- **Összefüggőség:** Nem létezik olyan reguláris nyelv, amely nemdeterminisztikus véges automatával felismerhető, de nem ismerhető fel determinisztikus véges automatával

5. feladathoz való fogalmak: (Veremautomata)

- **Veremautomata fogalma, felépítése:**
  - A veremautomata egy $A = (Z, Q, T, \delta, z0, q0, F)$ rendezett hetes, ahol
    - Z : a veremszimbólumok véges halmaza,
    - Q: az állapotok véges halmaza
    - T : az inputszimbólumok véges halmaza
    - δ: leképezése a $Z \times Q \times (T \cup \{\varepsilon\})$ halmaznak $Z^* \times Q$ véges részhalmazaiba, az ún. átmeneti függvény,
    - $z0 \in Z$ a kezdeti (kezdő) veremszimbólum
    - $q0 \in Q$ a kezdeti (kezdő) állapot
    - $F \subseteq Q$ az elfogadó állapotok halmaza.
- **Veremautomata által elfogadott nyelv:**

6. feladathoz való fogalmak: (Bizonyítás)