Formális Nyelvek - 1.

Csuhaj Varjú Erzsébet

Algoritmusok és Alkalmazásaik Tanszék
Informatikai Kar
Eötvös Loránd Tudományegyetem
H-1117 Budapest
Pázmány Péter sétány 1/c

E-mail: csuhaj@inf.elte.hu

A kurzus célja, hogy megismerkedjünk a **formális nyelvek és automaták elméletének**, a **számítástudomány egyik tradícionális ágának** alapjaival.

Irodalom:

1. Csima Judit és Friedl Katalin: Nyelvek és automaták, BMGE jegyzet, 2013. (weben elérhető)

2. Révész György, Bevezetés a formális nyelvek elméletébe, Tankönyvkiadó, 1977.

3. Fülöp Zoltán, Formális nyelvek és szintaktikus elemzésük, Polygon, Szeged, 2004.

4. Bach Iván: Formális nyelvek, Typotex, 2001.

Irodalom:

1. György E. Révész, Introduction to Formal Languages, McGraw-Hill Book Company, 1983.

További irodalom:

2. A. Salomaa, Formal Languages, Acedemic Press, 1973.

3. K. Krithivasan, Rama, R., Introduction to Formal Languages, Automata Theory and Computation, Pearson, 2009.

4. J. E. Hopcroft, Rajeev Motwani, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation. Second Edition. Addison-Wesley, 2001.

5. M. Sipser: Introduction to the Theory of Computation, 2nd edition, Thomson Course Technology, 2006.

Tudnivalók

Az előadások alapjául az irodalomjegyzék magyar- és angol nyelvű elemei szolgálnak. Az táblaképek (a diák) segítik a felkészülést és kizárólag tanulási célokra használhatók.

A diák szerzői jogvédelem alatt állnak, mind részben, mind egészben kizárólag az előadó (Csuhaj Varjú Erzsébet) honlapján tehetők publikussá!!!

A kurzus vizsgával zárul, **a vizsga során az előadásokon elhangzott anyagot kérem számon**. Vizsgázni csak legalább elégséges gyakorlati jegy birtokában lehet. A vizsga előtt három héttel részletes információt adok a számonkérés alapjául szolgáló anyagról és a számonkérés módjáról.

Az előadások látogatása nem kötelező, de erősen ajánlott. A gyakorlatok látogatása kötelező, a lehetséges hiányzások számát a gyakorlatvezetők ismertetni fogják a gyakorlatokon. A gyakorlati jegy megszerzésének feltételeit is a gyakorlatvezetők fogják meghatározni és ismertetni.

Az előadásvázlatokat (a diákat) .pdf file formájában az előadások után felteszem a Neptun-meet-street-be.

Minden kedden 10-12 óra között fogadóórám van a Déli tömb 2.511-es hivatali szobámban, ahol az érdeklődőket szeretettel várom.

Ha bármilyen kérdésük van a tantárggyal kapcsolatban, keressenek meg emailben (csuhaj@inf.elte.hu) vagy a fogadóórán.

Jó tanulást kívánok!

Budapest, 2016. február

Csuhaj Varjú Erzsébet

tanszékvezető egyetemi tanár

A kurzus tartalmának rövid leírása

1. **Bevezetés, a formális nyelv fogalma:** alapvető fogalmak és jelölések,szavak, nyelvek, grammatikák, a grammatikák Chomsky-féle hierarchiája.

2. **Műveletek nyelveken:** definíciók, Chomsky-féle nyelvosztályok bizonyos zártsági tulajdonságai.

3. **Környezetfüggetlen grammatikák és nyelvek:** redukált grammatikák, normálformák, levezetési fa. Reguláris grammatikák, reguláris nyelvek, reguláris kifejezések. A generált nyelvek és nyelvosztályok bizonyos fontos tulajdonságai.

4. **Környezetfüggő- és mondatszerkezetű grammatikák:** hossz-nemcsökkentő grammatikák, normálformák. A generált nyelvek és nyelvosztályok tulajdonságai. Nyelvosztályok Chomsky-féle hierarchiája.

A kurzus tartalmának rövid leírása - folytatás

1. **Automaták és nyelvek:** véges automaták, veremautomaták. Az automaták tulajdonságai, a felismert nyelvosztályok, az automaták és a grammatikák kapcsolatai.

2. Lineárisan korlátozott automata, Turing gép.

3. **Szintaktikai elemzés:** kapcsolat szintaxis és szemantika között; $LL(k)$ és $LR(k)$ grammatikák.

A formális nyelvek és automaták elmélete - a gyökerek

A nyelv **grammatikájának** fogalma már kb. időszámításunk előtt az IV. században felmerült Indiában (Panini).

Fontosabb lépések:

- Axel Thue, Emil Post, matematika, a XX. század eleje.
- W. Mc Culloch, W. Pitts, 1943, az idegrendszer modellje - a véges állapotú gép;

  S.C. Kleene, 1956, neurális háló - a véges automata.
- Noam Chomsky, 1959, matematikai model, az angol nyelv grammatikájának matematikai modellje.
- Programnyelvek, ALGOL 60, 1960

Mivel foglalkozik a formális nyelvek és automaták elmélete?

A formális nyelvek elmélete szimbólumsorozatok halmazaival foglalkozik.

Célja - többek között - véges, tömör leírását adni az ilyen halmazoknak.

Az elmélet módszereket ad formális nyelvek definiálására, a formális elemek nyelvhez való tartozásának eldöntésére, a nyelvi elemek struktúrájának felismerésére és leírására.

A szimbólum fogalmát alapfogalomnak tekintjük, ezért nem definiáljuk.

Milyen tudományágakhoz kapcsolódnak a formális nyelvek és automaták?

- A természetes nyelvek gépi feldolgozása, matematikai modellezése, matematikai nyelvészet,

- programozási nyelvek, fordítóprogramok elmélete,

- kódelmélet,

- képfeldolgozás,

- mintafelismerés,

- fejlődő rendszerek modellezése (Lindenmayer rendszerek),

- molekuláris számítástudomány (DNS számítás),

- multi-ágens rendszerek formális leírásai,

- stb.

Alapfogalmak és jelölések - I

Szimbólumok véges nemüres halmazát **ábécének** nevezzük.

Példa: $V = \{a, b, c\}$

Egy V ábécé elemeiből képzett véges sorozatokat V feletti **szavaknak** vagy **sztringeknek**- más szóval füzéreknek- mondunk. A 0 hosszúságú sorozatot **üres szónak** nevezzük és ε-nal jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$ és akkor $aaabbbccc$ egy szó.

A V ábécé feletti szavak halmazát (beleértve az üres szót is) V^*-gal, a nemüres szavak halmazát V^+-**szal jelöljük.**

Alapfogalmak és jelölések - II

Legyen V egy ábécé és legyenek u, v V feletti szavak (azaz, legyen $u, v \in V^*$). Az uv szót az u és v szavak **konkatenáltjának** vagy más szóval összefűzésének nevezzük.

Példa: Legyen az ábécé $V = \{a, b, c\}$, legyenek $u = abb$ és $v = cbb$ szavak. Akkor $uv = abbcbb$ az u és v konkatenáltja.

Megjegyzés: A konkatenáció mint művelet asszociatív, de általában nem kommutatív.

Példa: Legyen $u = ab$, $v = ba$, akkor $uv = abba$ és $vu = baab$.

Alapfogalmak és jelölések - II - folytatás

Legyen V egy ábécé. Megállapíthatjuk, hogy V^* **zárt a konkatenáció műveletére nézve** (azaz, bármely $u, v \in V^*$ esetén $uv \in V^*$ teljesül), továbbá a **konkatenáció egységelemes művelet**, ahol az egységelem ε (azaz, bármely $u \in V^*$ esetén $u\varepsilon \in V^*$ és $\varepsilon u \in V^*$).

Alapfogalmak és jelölések - III

Legyen i nemnegatív egész szám és legyen w a V ábécé feletti szó ($w \in V^*$).

A w **szó i-edik hatványa** alatt a w szó i példányának konkatenálját értjük és w^i-vel jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$, és legyen $w = abc$. Akkor $w^3 = abcabcabc$.

Konvenció alapján minden $w \in V^*$ szóra $w^0 = \varepsilon$.

Alapfogalmak és jelölések - IV

Legyen V egy ábécé és legyen w egy V feletti szó (azaz, legyen $w \in V^*$).

A w **szó hosszán** a w szót alkotó szimbólumok számát értjük (azaz, w mint sorozat hosszát) és $|w|$-vel jelöljük.

Példa: Legyen az ábécé $V = \{a, b, c\}$ és legyen $w = abcccc$. Akkor w hossza 6.

Az üres szó hossza - nyilvánvalóan - 0, azaz $|\varepsilon| = 0$.

Alapfogalmak és jelölések - V

Egy V ábécé feletti két u és v szót azonosnak nevezünk, ha mint szimbólumsorozatok egyenlőek (azaz, mint sorozatok elemről-elemre megegyeznek.)

Legyen V ábécé 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 $u \neq v$ és $u \neq \varepsilon$.

Példa: Legyen $V = \{a, b, c\}$ ábécé és legyen $v = aabbbcc$ szó. Az $u = abbbc$ szó valódi részszava v-nek.

Alapfogalmak és jelölések - V - folytatás

Legyen V ábécé és legyenek u, v, x, y szavak V felett, továbbá legyen $v = xuy$. Ha $x = \varepsilon$, akkor u-t a v szó **prefixének** vagy kezdőszeletének, ha $y = \varepsilon$, akkor u-t a v szó **szufixének** vagy utótagjának hívjuk.

Legyen $v = aabbbcc$ szó. Az $u = aabbb$ szó prefixe, a $bbbcc$ szó szufixe v-nek.

Alapfogalmak és jelölések - VI

Legyen u egy V ábécé 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.

Legyen $u = a_1 \dots a_n$, $a_i \in V$, $1 \le i \le n$. Ekkor $u^{-1} = a_n \dots a_1$.

Alapfogalmak és jelölések - VII

Legyen V ábécé és legyen L tetszőleges részhalmaza V^*-nak. Akkor L-et egy V feletti **nyelvnek** nevezzük.

Az **üres nyelv** - amely egyetlen szót sem tartalmaz - jelölése $\emptyset$.

Egy V ábécé feletti nyelvet **véges nyelvnek** mondunk, ha véges számú szót tartalmaz, ellenkező esetben **végtelen nyelvről** beszélünk.

Példák nyelvekre

Legyen $V = \{a, b\}$ ábécé.

Akkor $L_1 = \{a, b, \varepsilon\}$ véges nyelv, $L_2 = \{a^i b^i \mid 0 \leq i\}$ végtelen nyelv.

Példa L_2-beli szavakra: $ab, aabb, aaabbb, \ldots$

Legyen $L_3 = \{uu^{-1} | u \in V^*\}$.

Példa L_3-beli szavakra: $u = ababb$, $u^{-1} = bbaba$ és $uu^{-1} = ababbbbaba$.

Nyelvekre vonatkozó műveletek - I

Legyen V egy ábécé és legyenek L_1, L_2 nyelvek V felett ($L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

- $L_1 \cup L_2 = \{u \mid u \in L_1$ vagy $u \in L_2\}$ - az L_1 és az L_2 nyelv **uniója**;
- $L_1 \cap L_2 = \{u \mid u \in L_1$ és $u \in L_2\}$ - az L_1 és az L_2 nyelv **metszete**;
- $L_1 - L_2 = \{u \mid u \in L_1$ és $u \notin L_2\}$ - az L_1 és az L_2 nyelv **különbsége**.
- Az $L \subseteq V^*$ nyelv **komplementere** a V ábécére vonatkozóan $\bar{L} = V^* - L$.

Nyelvekre vonatkozó műveletek - II

Legyen V ábécé és legyenek L_1, L_2 nyelvek V felett ($L_1 \subseteq V^*$ és $L_2 \subseteq V^*$).

- $L_1L_2 = \{u_1u_2 \mid u_1 \in L_1, u_2 \in L_2\}$ - az L_1 és az L_2 nyelv **konkatenációja**;
- Minden L nyelvre fennállnak a következő egyenlőségek:
  $\emptyset L = L\emptyset = \emptyset$ és $\{\varepsilon\}L = L\{\varepsilon\} = L$.
- L^i jelöli az L nyelv i**-edik hatványát**:
  $L^0 = \{\varepsilon\}$, $L^i = L^{i-1}L$, $i \geq 1$.
- Az L nyelv **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 $L^+ = L^*$.

Nyelvekre vonatkozó műveletek - III

Legyen V egy ábécé és $L \subseteq V^*$.

$L^{-1} = \{u^{-1} | u \in L\}$ a **tükörképe** (megfordítása) az L nyelvnek.

Tulajdonság:

$(L^{-1})^{-1} = L$ és $(L^{-1})^i = (L^i)^{-1}$, $i \geq 0$.

Nyelvekre vonatkozó leképezések:

Legyen V_1, V_2 ábécé.

A $h : V_1^* \to V_2^*$ leképezést **homomorfizmusnak** nevezzük, ha $h(uv) = h(u)h(v)$ minden $u, v \in V_1^*$ esetén.

A fenti tulajdonság alapján $h(\varepsilon) = \varepsilon$. (Minden $u \in V_1^*$-ra $h(u) = h(\varepsilon u) = h(u\varepsilon)$.)

Nyilvánvaló, hogy minden $u = a_1 a_2 \ldots a_n$ szóra, ahol $a_i \in V_1, 1 \leq i \leq n$, fennáll, hogy $h(u) = h(a_1)h(a_2)\ldots h(a_n)$.

Nyelvekre vonatkozó leképezések:

Legyen $h : V_1^* \to V_2^*$ homomorfizmus.

A h **homomorfizmus** ε**-mentes**, ha $h(u) \neq \varepsilon$ bármely $u \in V_1^*$ szóra, ahol $u \neq \varepsilon$.

Legyen $h : V_1^* \to V_2^*$ homomorfizmus. Az $L \subseteq V_1^*$ nyelv h**-homomorf képén** a

$$h(L) = \{w \in V_2^* \mid w = h(u), u \in L\}$$

nyelvet értjük.

Nyelvekre vonatkozó leképezések:

A h homomorfizmust **izomorfizmusnak** nevezzük, ha bármely u és v V_1^*-beli szóra teljesül, hogyha $h(u) = h(v)$, akkor $u = v$.

Egy példa az izomorfizmusra a decimális számok bináris reprezentációja:

$$V_1 = \{0, 1, 2, \ldots, 9\}, V_2 = \{0, 1\},$$
$$h(0) = 0000, \;\; h(1) = 0001, \;\; \ldots, h(9) = 1001$$