Elméleti számítástudomány

- lényege
  - számítási folyamat hardverfüggetlen vizsgálata
  - megoldhatóság vizsgálata
  - számítási bonyolultság vizsgálata

Fogalmak

- halmaz:rendezetlen gyűjtemény
- függvény, reláció: leképezés egy halmazból egy másikba
- ábécé: Szimbólumok (betűk) véges halmaza.
- szó: Egy ábécé betűiből álló véges sorozat.
- nyelv: Adott ábécé feletti szavak véges vagy végtelen halmaza.

Véges automata: véges állapotszám, matematikai alapok

- részei: Q: állapotok, Σ (szigma) :bemeneti ábécé, $q_0 \in Q$: kezdőállapot, A: végállapotok, δ: állapotátmenet függvény
- [fogalom] kiterjesztett állapotátmenet függvény: Nem egy betű, hanem egy szó által okozott állapotátmenetet ír le. Jele: δ* („delta csillag”).
- Az $M=(Q, \Sigma, q_0, A, \delta)$ véges automata elfogadja az x szót, ha $\delta^*(q_0, x) \in A$. Az M által elfogadott nyelv: $L(M) = \{ x \in \Sigma^* \mid \delta^*(q_0, x) \in A \}$.
- véges automata állapotszáma
  - [definíció] Megkülönböztethetőség egy L nyelvre nézve: Az x és y szavak L-megkülönböztethetőek, ha létezik olyan z szó, hogy ($xz \in L$ **és** $yx \notin L$) **vagy** ($xz \notin L$ **és** $yx \in L$).
  - [tétel] Automata minimális állapotszáma: Ha L=L(M) valamilyen M automata esetén, és L-ben van n páronként egymástól L-megkülönböztethető szó, akkor M-nek legalább n állapota van.
  - Egy ábécé feletti L nyelvet ekvivalencia osztályokra oszthatunk L-megkülönböztethetőség alapján, az L nyelvet elfogadó véges automatának annyi állapotra van szüksége, mint az L szerinti megkülönböztethetőség ekvivalencia osztályainak száma.
- [tétel] Véges automatával nem elfogadható nyelvek: Ha egy L nyelvben végtelen sok páronként L-megkülönböztethető szó van, akkor $L \neq L(M)$ semmilyen M véges automata esetén.
- Pumpálási lemma: ha egy véges automata által elfogadott szó elég hosszú, akkor az automata kénytelen legalább egy állapotot többször felvenni. Formálisan: Ha az $L \subseteq \Sigma^*$ nyelvet elfogadja az $M=(Q, \Sigma, q_0, A, \delta)$ véges automata és $n = |Q|$, akkor minden olyan $x \in L$ szó, amely $|x| \geq n$, felírható $x = uvw$ alakban, ahol
  1. $|uv| \leq n$
  2. $|v| > 0$
  3. minden $i \geq 0$-ra $uv^iw \in L$

  Ha egy nyelv nem teljesíti a pumpálási lemma feltételeit, akkor nem lehet véges automatával elfogadni. A pumpálási lemma feltételeinek teljesülése **nem elégséges** feltétele a nyelvet elfogadó véges automata létezésének.

Reguláris nyelvek, reguláris kifejezések

- [definíció] Reguláris nyelv olyan nyelv, amely leírható reguláris kifejezésekkel.
- jelölésbeli különbségek (nyelv → kifejezés)
  - {} → ()
  - ⋃ → +
  - , → +

Nemdeterminisztikus véges automaták: olyan véges automaták, melyek egyszerre több állapotot is felvehetnek, vagy van bennük üresszó átmenet

- [definíció] Legyen $M=(Q, \Sigma, q_0, A, \delta)$ nemdeterminisztikus véges automata. Legyen $S \subseteq Q$ és jelölje $E(S)$ a következőt: $S \subseteq E(S)$ és minden $q \in E(S)$-re $\delta(q, \lambda) \subseteq E(S)$. $E(S) \subseteq Q$ az az állapothalmaz, amikbe S állapotaiból λ átmenetek mentén el lehet jutni.
- A nemdeterminisztikus automata által elfogadott $L(M)$ nyelv a következő: $L(M)=\{ w \in \Sigma^* \mid \delta^*(q_0, w) \cap A \neq \varnothing \}$
- nemdeterminisztikusság kiküszöbölésére algoritmus: az egyszerre felvett állapotok különböző halmazait egy állapotként vesszük

Generatív grammatika

- felépítés: $G(N, \Sigma, S, P)$, ahol N a nemterminális ábécé, Σ a terminális ábécé $S \in N$ kezdőszimbólum, és P a helyettesítési szabályok
- $L(G)$ a G által generált nyelv/szavak halmaza
- [definíció] levezetés fogalma formális alakban: Legyen $G(N, \Sigma, S, P)$. Ekkor egy u szó közvetlenül levezethető egy v szóból, ha
  - $v = v_1\alpha v_2$
  - $u = v_1\beta v_2$
  - létezik $\alpha \to \beta \in P$ szabály

  Ennek jelölése: $v \Rightarrow u$

  Szavak egy w_1, w_2, w_n sorozatára azt mondjuk, hogy w_n levezethető w_1-ből, ha $w_1 \Rightarrow w_2 \Rightarrow \ldots \Rightarrow w_n$. Ennek jelölése: $w_1 \Rightarrow^* w_n$.
- Egy $G(N, \Sigma, S, P)$ grammatika generálja a w szót, ha $S \Rightarrow^* w$
- Egy $G(N, \Sigma, S, P)$ grammatika által generált nyelv az $L(G) = \{ w \in \Sigma^* \mid S \Rightarrow^* w \}$
- Egy $G(N, \Sigma, S, P)$ akkor környezetfüggetlen, ha P minden szabálya $A \to B$ alakú, ahol $A \in N$ és $B \in \{ N \cup \Sigma \}^*$
- reguláris nyelveket generáló grammatikák
  - állapotátmenet <=> átírási szabály: $T \to x \to U$ <=> $T \to xU$
  - generálás abbahagyása törlő szabállyal (pl.: $B \to bA \mid aS \mid \lambda$) vagy terminálissal (pl.: $A \to bA \mid aB \mid a$)
  - [definíció] reguláris grammatika: Egy $G(N, \Sigma, S, P)$ környezetfüggetlen grammatika reguláris, ha a szabályok alakja $A \to aB$, $A \to \lambda$, vagy $A \to a$, ahol $A, B \in N$ és $a \in \Sigma$.
  - reguláris grammatika <=> automata konverzió: Egy $G(N, \Sigma, S, P)$ átalakítható az $M=(Q, T, q_0, \delta, F)$ véges automatává a következőképpen:
    - legyen $N=Q$
    - legyen q_0 a kezdőszimbólum
    - $q_i \to aq_j$ <=> $\delta(q_i, a) = q_j$
    - $q_i \to a$ <=> $\delta(q_i, a) = q$, ahol $q \in F$
    - $q_i \to \lambda$ <=> $q_i \in F$
  - [tétel] $L \subset \Sigma^*$ reguláris akkor és csak akkor, ha $L=L(G)$, ahol G reguláris grammatika.

Környezetfüggetlen grammatika

- levezetések
  - két levezetés lényegesen különbözik, ha a hozzájuk tartozó levezetési fák különböznek
  - [definíció] egyértelműség: Egy grammatika nem egyértelmű, ha van olyan szó, aminek több lényegesen különböző levezetése is van.
- Grammatika egyszerűsítése
  - törlőszabályok kiszedése
  - láncszabályok kiszedése
- pumpálási lemma
  - [definíció] Ha L környezetfüggetlen, akkor létezik p (S levezetési fájában van olyan út, amely a gyökértől a levélig hosszabb, mint a nemterminálisok száma a nyelvtanban), hogy ha $S \in L$ és $|S| > p$, akkor S felírható S=uvxyz alakban, hogy
    1. $uv^ixy^iz \in L$ minden $i \geq 0$-ra
    2. $|vy| > 0$
    3. $|vxy| \leq p$

Veremautomata

- véges automata veremmemóriával
- környezetfüggetlen grammatikákat képes leírni
- $M=(Q, T, \Gamma, q_0, Z_0, \delta, F)$: Q állapothalmaz, T bemeneti ábécé, Γ veremábécé, q_0 kezdőállapot, Z_0 kezdeti veremtartalom, δ állapotátmenet reláció, $F \subseteq Q$ végállapot halmaz
- készítés környezetfüggetlen grammatikából: $G(N, T, S, P) \rightarrow M(\{q_0, q_1, q_2\}, T, N \cup T \cup \{ Z_0 \}, q_0, \delta, \{ q_2 \})$
  $\delta(q_0, \lambda, Z_0) = \{(q_1, SZ_0)\}$
  $\delta(q_1, \lambda,$ [nemterminális]$) = \{(q_1,$ [nemterminális 1. szabálya]$), (q_1,$ [nemterminális 2. szabálya]$), \ldots\}$
  $\delta(q_1,$ [terminális], [terminális]$) = \{(q_1, \lambda)\}$
  $\delta(q_1, \lambda, Z_0) = \{(q_2, \lambda)\}$
- ha L környezetfüggetlen, akkor van olyan M veremautomata, hogy L=L(M)
- determinisztikus akkor, ha minden δ átmenet legfeljebb egy elemű

Szintaktikai elemzés és parsing

- felülről lefelé
  - LL(1) nyelvek
  - LL(k) nyelvek
- lentről felfelé
  - Adott egy G(N, T, S, P) grammatika. Legyen $M=(Q, T, T \cup N \cup \{ Z_0 \}, q_0, Z_0, \delta, q_2)$ ahol $Q= \{ q_0, q_2, q_2 \}$ és
    $\delta(q_0, x, X) = (q_0, xX)$
    minden $B \rightarrow \alpha$ szabályhoz legyen egy lépéssorozat: $(q_0, w, \alpha^R) \Rightarrow \ldots \Rightarrow (q_0, w, B)$
    végül: $\delta(q_0, \lambda, S) = (q_0, \lambda)$, $\delta(q_1, \lambda, Z_0) = (q_2, Z_0)$
    $(q_0, \lambda, B) \in \delta(q_0, \lambda, \alpha^R)$
  - LR(k) grammatikák
  - megengedett veremtartalmat felismerő véges automata konstruálása
    - LR elemek

Turing gépek

- $T(Q, \Sigma, \Gamma, q_0, \delta)$, ahol Q állapothalmaz, Σ bemeneti ábécé, Γ szalagábécé, $\Sigma \subseteq \Gamma$, $q_0 \in \Gamma$ kezdőállapot, δ állapotátmenet halmaz $(Q \times (\Gamma \cup \{ \Delta \}) \rightarrow (Q \times (\Gamma \cup \{ \Delta \}) \times \{R, L, S\})$
- konfiguráció: állapot, szalag tartalma, fej helye
- a T turing gép által elfogadott nyelv L(T)
- nem fogad el egy $w \in \Sigma^*$ szót, ha nem áll meg, vagy nem elfogadó állapotban áll meg
- egy Turing gép lehet k szalagos: Γ^k van Γ helyett az állapot átmenetben
- univerzális Turing gép
    - egy olyan Turing gép, amely minden más Turing gép működését szimulálni tudja
    - T_u
    - minden m állapotátmenetet ($\delta(p, a) = (q, b, D)$) az e(m) = s(p)1s(a)1s(q)1s(b)1s(D)1 string kódolja
    - minden s(x)-et egy {0} feletti szó kódol
    - minden z=z1z2z3z4z5… stringet, ahol $Zi \in S$, az e(z)=1s(z1)1s(z2)1...s(zk)1 string kódol
- programozható, általános célú számítási modell, matematikailag precíz → alkalmas az algoritmusok formális leírására → Church-Turing tézis: ha egy probléma megoldható mechanikus eljárássorozattal, akkor van Turing-gép, ami képes megoldani. Ha egy problémát nem lehet Turing-géppel megoldani, akkor nincs rá algoritmus.
    - A Turing-gép eldöntendő problémák megoldására alkalmas
    - pl. van Turing-gép annak eldöntésére, hogy egy szám prím-e, így az algoritmikusan megoldható probléma
    - a problémát akkor döntötte el a Turing-gép, ha az minden bemenetre megáll IGEN vagy NEM válasszal → egy probléma algoritmikusan megoldható, ha a hozzá tartozó nyelv rekurzív
    - egy rekurzívan felsorolható nyelv nem megoldható algoritmukusan

Rekurzív és rekurzívan felsorolható nyelvek

- [definíció] rekurzívan felsorolható nyelv: Egy L nyelv rekurzívan felsorolható, ha van olyan T turing gép, hogy L = L(T), vagyis a turing gép elfogadja a nyelvet.
- [definíció] rekurzív nyelv: Egy L nyelv rekurzív, ha van olyan T turing gép, hogy L = L(T), vagyis a turing gép elfogadja a nyelvet, **és a turing gép minden bemenetre megáll**.
- Ha van ilyen T turing gép, akkor azt mondjuk, T eldönti L-et.

Általános alakú grammatikák:

- nem csak egy nemterminális szerepelhet a szabályok bal oldalán, és lehet ott terminális is, de nem lehet csak terminális
- Turing géppel elfogadható nyelvekkel egyezik meg

Fogalmak

- Nyelv lezártja (iteráltja): Legyen L egy nyelv. $L * := \cup(i=0 \to \infty) L^i$
- Reguláris nyelvek az elemi nyelvek (∅, { ε }, { a } (a ∈ U [U az univerzális ábécé])) és az ezekből véges számú unióval, konkatenációval, és lezárással előálló nyelvek.
- Az X ábécé feletti reguláris kifejezések az elemi reguláris kifejezések ∅, ε, a (a ∈ X), és ha R_1 és R_2 reguláris kifejezések, akkor $(R_1 \cup R_2)$, (R_1R_2), és R_1^* is. Reguláris kifejezések az X feletti reguláris kifejezések X=U esetén.
- Rekurzívan felsorolható nyelv: Az L nyelv rekurzívan felsorolható ⟺ ha létezik A algoritmus, mely az elemeit (és csak azokat) felsorolja az outputjára.
- Rekurzív nyelv: Az L nyelv rekurzív, ha ⟺ létezik olyan A eldöntő algoritmus, melynek inputjára egy tetszőleges u szót helyezve eldönti, benne van-e az L nyelvben (mindig terminál, igen a válasz, ha u eleme az L nyelvnek, és nem a válasz ellenkező esetben).
- Generatív nyelvtan (grammatika): A G(N, T, S, P) négyest nyelvtannak nevezzük, ahol T a terminális, N a nyelvtani (nemterminális) jelek egymástól diszjunkt ábécéje, P (produkciós) szabályoknak egy véges halmaza, ahol minden p ∈ P szabály $a \to b$ alakú, $a, b \in (T \cup N)*$ és »a« tartalmaz legalább egy nyelvtani jelet, továbbá S ∈ N , melyet kezdőszimbólumnak nevezünk.
- Nyelvtan által generált nyelv: A G(N, T, S, P) nyelvtan által generált nyelv L(G) minden u szava u ∈ T* és minden u szó levezethető S-ből.
- Veremautomata: Egy $V=(Q, T, \Sigma, q_0, Z_0, \delta, F)$ 7-es, ahol Q az állapothalmaz, T a bemenő ábécé, Σ a veremábécé, $q_0 \in Q$ a kezdőállapot, $Z_0 \in \Sigma$ a kezdő veremtartalom, δ állapotátmenet függvény, F ∈ Q pedig a végállapotok halmaza. Egy véges automata veremmemóriával.
- Konfiguráció: Konfigurációnak nevezzük azoknak az adatoknak az összességét, melyektől a gép elkövetkezendő működése függ.
- Levezetés: $\alpha \to_G \to \beta$, vagyis α-ból közvetlenül levezethető β, ha létezik olyan γ_1, γ_2, és $(p \to q) \in (T \cup N)*$, hogy $\alpha = \gamma_1 p \gamma_2$ és $\beta = \gamma_1 q \gamma_2$. $\alpha \to_G * \to \beta$, vagyis α-ból levezethető β, ha létezik olyan $\gamma_1, \gamma_2, \ldots, \gamma_k$ sorozat, hogy $\alpha = \gamma_1$, $\beta = \gamma_k$, és minden γ_i, γ_{i+1} párosra $\gamma_i \to_G \to \gamma_{i+1}$.
- Turing-gép: egy olyan (elméleti) gép, mely egy vagy több végtelen hosszú szalagon dolgozik, a szalagról olvas, állapotot vált, ír a szalagra, és mozgatja az író-olvasó fejét a szalagon jobbra-balra. Alakja: $T=(Q, \Sigma, \Gamma, \delta, q_0, B, F)$, ahol Q az állapothalmaz, $\Sigma \in \Gamma$ a bemeneti ábécé, Γ a szalag szimbólumai, δ állapotátmenet fv., mely $(Q \times \Gamma) \to (Q \times \Gamma \times \{L, R, S\})$ alakú, $q_0 \in Q$ kezdőállapot, B az „üres” szimbólum a szalagon, $F \subseteq Q$ pedig az elfogadó állapotok halmaza.