Separating Words Problem

Önálló Projekt 3

Csányi Dávid

Témavezető: Pálvölgyi Dömötör

2023. május

1. Bevezetés

Előző félévben a Separating Words Problem-mel foglalkoztam a [3] összefoglaló alapján [2]. Ezen munka folytatásaként ebben a félévben a felső korlátok megismerésén és általánosításán dolgoztam.

1.1. Definíció. Jelölje $\text{sep}_l(a_1, \dots a_l)$ azt a legkisebb k számot, amire létezik egy k állapotú véges automata, ami az $a_1, a_2, \dots a_l$ szavak esetén különböző végállapotokba kerül.

1.2. Definíció. $S_l(n) = \max\{\text{sep}_l(a_1, \dots a_l) : a_1, \dots a_l$ a Σ abc feletti maximum n hosszú páronként különböző szavak.$\}$

A separting words problem célja az $S_2(n)$ függvényre alsó és felső korlátokat találni. Az eddigi legjobb ismert felső korlát $S_2(n) = \tilde{\mathcal{O}}(n^{\frac{1}{3}})$, amelyet Chase bizonyított be 2020-ban[1]. A legjobb ismert alsó korlát $S_2(n) = \Omega(\log(n))$.

A korábban belátott állítások alapján [2] a feladatot elég a $\Sigma = \{0, 1\}$ abc felett vizsgálni, ezért a továbbiakban feltesszük, hogy $\Sigma = \{0, 1\}$. A beszámoló során bemutatok az $S_2(n)$-re eddig bebizonyított felső becslések közül néhányat, köztük a legerősebbet is. Továbbá a gyökös becslés általánosításában elért eredményemet fogom leírni.

1.1. Segédállítások

Az alábbi, prímszámokkal kapcsolatos tételt a későbbiekben sokszor fogjuk használni:

1.3. Lemma. [7] Minden $n \geq 2$ természetes számra létezik egy $p \leq 4,4 \log(n)$ prím, amelyre igaz, hogy $p \nmid n$.

1.4. Következmény. Ha $0 \leq i, j \leq n, n \geq 2$ és $i \neq j$, akkor létezik egy p prím, amelyre $p \leq 4,4 \cdot \log(n)$ és $i \not\equiv j \pmod{p}$.

A bizonyítás vázlata [2]-ben, a teljes bizonyítás a megjelölt [7] forrásban megtalálható.

2. Felső korlátok

2.1. Gyökös felső becslés

Először egy egyszerűbb $\tilde{\mathcal{O}}(\sqrt{n})$-es felső korlátot ismerhetünk meg, melynek bizonyításához a körosztási polinomok egyes tulajdonságait fogjuk használni.

2.1. Definíció. Legyen $a = a_0 \dots a_{n-1} \in \{0,1\}^n$ és $w = w_0 w_1 \dots w_{l-1} \in \{0,1\}^l$. Ekkor $\mathrm{pos}_w(a)$-val jelöljük a w részszó kezdőpozícióinak halmazát az a szóban. Formálisan $\mathrm{pos}_w(a) = \{j : 0 \leq j \leq n-l, a_j = w_0, \dots a_{j+l-1} = w_{l-1}\}$.

2.2. Definíció. Legyen $m \in \mathbb{Z}^+$ és i egy maradékosztály modulo m. Ekkor $\mathrm{pos}_w(a)_{m,i}$-vel jelöljük a w részszó azon j kezdőpozícióinak halmazát, amelyekre $j \equiv i \mod m$.

Ha az a és b bináris szavak különbözőek, akkor $\mathrm{pos}_1(a) \neq \mathrm{pos}_1(b)$. Következő lépésként azt fogjuk belátni, hogy ha egy p prímre és egy i maradékosztályra igaz, hogy $|\mathrm{pos}_1(a)_{p,i}| \neq |\mathrm{pos}_1(b)_{p,i}|$, akkor az a és b szavakat meg tudjuk különböztetni egy automatával. Ezután meggondoljuk, hogy létezik nem túl nagy ilyen p és hozzá megfelelő i.

2.3. Tétel (Chase [1])**.** $S_2(n) = \mathcal{O}(\sqrt{n} \cdot \log(n)^{1.5})$

Bizonyítás. Tegyük fel, hogy találtunk egy p prímet, i egy maradékosztályt, melyekre $|\mathrm{pos}_1(a)_{p,i}| \neq |\mathrm{pos}_1(b)_{p,i}|$. (A következő állításban fogjuk megmutatni, hogy tudunk találni ilyen p-t és i-t.) Tudjuk, hogy $|\mathrm{pos}_1(a)_{p,i}|, |\mathrm{pos}_1(b)_{p,i}| \leq n$ és a 1.4 következmény miatt létezik egy $q = \mathcal{O}(\log(n))$ prím, amire a két érték osztási maradéka különböző. Készítsünk egy olyan automatát, amely egy a szó esetén a $|\mathrm{pos}_1(a)_{p,i}| \mod q$ értéket számolja. Ezt megtehetjük, ha q darab p hosszú kört használunk és ha egy kör i-edik állapotában 1-est olvasunk, akkor a következő kör $i+1$-edik állapotába lépünk, egyébként ugyanazon a körön a következő állapotba. Ez egy pq állapotú automata, amely megkülönbözteti a két szót. A következő számelméleti állítás alapján létezik egy p prím, amely $\mathcal{O}(\sqrt{n \log(n)})$ méretű és hozzá egy i maradékosztály, amelyek megfelelőek. Ebből következik, hogy bármelyik kettő n hosszú bináris szóhoz létezik egy $pq = \mathcal{O}(\sqrt{n}\log(n)^{1.5})$ állapotú automata, amely megkülönbözteti őket. □

2.4. Definíció. Legyen $A \subseteq \{0,1,\dots n-1\}$. Ekkor definiáljuk az $A_{p,i} = \{j \in A : j \equiv i \mod p\}$ halmazt.

2.5. Állítás (Chase [1])**.** Ha $A \neq B \subseteq \{0,1\dots n-1\}$, akkor létezik egy $p = \mathcal{O}(\sqrt{n\log(n)})$ prím és i maradékosztály, amelyekre $|A_{p,i}| \neq |B_{p,i}|$.

2.6. Definíció. [4] Az n-edik körosztási polinom az a normált polinom, amelynek gyökei pontosan az n-edik primitív komplex egységgyökök, mindegyik egyszeresen. $\Phi_n(x) = (x - \xi_1)\dots(x - \xi_{\varphi(n)})$, ahol $\xi_1, \dots \xi_{\varphi(n)}$ a primitív n-edik egységgyökök.

2.7. Definíció. Ha $a \in \{0,1\}^n$ egy bináris szó, akkor jelölje $A(x) = \sum_{i=0}^{n-1} a_i x^i$ az egyesek generáló függvényét. Ha $A \subseteq \{0,\dots n-1\}$, akkor $A(x) = \sum_{i=0}^{n-1} \mathbb{1}_A(i) x^i$.

2.8. Lemma ([8])**.** Tekintsük az $A \neq B \subseteq \{0,\dots n-1\}$ halmazokat és az $m \in \mathbb{Z}^+$ számot. Ha minden $i \in \{0,\dots m-1\}$ maradékosztályra $|A_{m,i}| = |B_{m,i}|$, akkor $\Phi_m(x)$ osztja az $(A(x) - B(x))$ polinomot.

Bizonyítás. Elég megmutatni, hogy $x^m - 1$ osztja az $A(x) - B(x)$ polinomot, ugyanis $\Phi_m(x)$ osztja $x^m - 1$-et. Rögzített i-re tekintsük $A(x)$ azon monomjait, amelyek x^{i+km} alakúak valamely $k \in \mathbb{Z}^+$-ra, ezek száma $|A_{m,i}|$. Hasonlóan $B(x)$ azon monomjainak száma, amelyek x^{i+km} alakúak $|B_{m,i}|$. A lemma feltétele miatt $|A_{m,i}| = |B_{m,i}|$, ezért az $A(x)$ és $B(x)$ ilyen monomjai párbaállíthatók olyan módon, hogy egy $A(x)$-beli $x^{i+k_1 m}$ alakú tag párja $B(x)$-ben egy $x^{i+k_2 m}$ legyen. Könnyen ellenőrizhető, hogy $(x^m-1)|(x^{i+k_1 m} - x^{i+k_2 m})$. Ezeket felhasználva adódik, hogy $x^m - 1$ osztja az $A(x) - B(x)$ polinomot. □

Bizonyítás. (2.5 Állítás) Tegyük fel, hogy egy adott $k \in \mathbb{Z}^+$-ra minden $p \leq k$ prím és $i \in \{0, \dots p-1\}$ maradékosztály esetén $|A_{p,i}| = |B_{p,i}|$. Az előző lemmából az következik, hogy $\Phi_p(x)|(A(x) - B(x))$ minen $p \leq k$ prímre. Az $A(x) - B(x)$ polinom nem azonosan nulla és a körosztási polinomik relatív prímek, ezért az $A(x) - B(x)$ foka legalább a $\Phi_p(x)$ körosztási polinomok fokainak összege kell legyen.

$$n \geq \deg(A(x) - B(x)) \geq \sum_{p \leq k} \deg(\Phi_p(x)) = \sum_{p \leq k} (p-1) \sim \frac{k^2}{\log k}$$

Azt kaptuk, hogy $n \log(n) \geq n \log(k) \geq k^2$, azaz $k \leq \sqrt{n \log n}$, amiből következik az állítás. □

2.2. A gyökös becslés általánosítása

2.9. Tétel. $S_l(n) = \mathcal{O}(l^5 \sqrt{n} \log(n)^{1.5})$

2.10. Definíció. Legyen $l \in \mathbb{Z}^+$. Ha minden $h \neq j \in \{1, 2, \dots l\}$ párhoz tartozik egy $i_{h,j} \in \{0, 1, \dots l-1\}$ maradékosztály, akkor ezek $i_{1,2}, i_{1,3}, \dots i_{1,l}, i_{2,3}, \dots i_{l-1,l}$ gyűjteményét (összesen $\binom{l}{2}$ darab) egy maradékosztály $\binom{l}{2}$-esnek nevezzük modulo p.

2.11. Lemma. Tekintsük az $A_1, A_2, \dots A_l \subseteq \{0, \dots n-1\}$ páronként különböző halmazokat és az $m \in \mathbb{Z}^+$ számot. Ha minden $i_{1,2}, \dots i_{l-1,l} \in \{0, \dots m-1\}$ maradékosztály $\binom{l}{2}$-esre valamelyik h, j-re igaz, hogy $|(A_h)_{m,i_{h,j}}| = |(A_j)_{m,i_{h,j}}|$, akkor $\Phi_m(x)$ osztja az $(A_1)(x), \dots (A_l)(x)$ polinomok közül valamelyik kettő különbségét.

Bizonyítás. Ha létezik egy $h \neq j$, hogy minden i maradékosztályra $|(A_h)_{m,i}| = |(A_j)_{m,i}|$, akkor a 2.8 Lemma miatt $\Phi_m(x)$ osztja az $((A_j)(x) - (A_h)(x))$ polinomot.

Ha nem létezik ilyen, akkor minden lehetséges $h \neq j$-hez van egy $i_{h,j}$, amelyre $|(A_h)_{m,i_{h,j}}| \neq |(A_j)_{m,i_{h,j}}|$, ezeket összegyűjtve egy $i_{1,2}, \dots i_{l-1,j}$ maradékosztály $\binom{l}{2}$-esbe a lemma feltételének ellentmondó példát találunk, tehát ez az eset nem lehetséges. □

2.12. Állítás. Ha $A_1, A_2, \dots A_l \subseteq \{0, 1, \dots n-1\}$ és páronként eltérőek, akkor létezik egy $p = \mathcal{O}(l\sqrt{n \log(n)})$ prím és egy $i_{1,2}, \dots i_{l-1,l}$ maradékosztály $\binom{l}{2}$-es, hogy minden $0 \leq h \neq j \leq l-1$-re $|(A_h)_{p,i_{h,j}}| \neq |(A_j)_{p,i_{h,j}}|$.

Bizonyítás. Tegyük fel, hogy egy $k \in \mathbb{Z}^+$-ra minden $p \leq k$ prím és $i_{1,2}, \dots i_{l-1,l} \in \{0, \dots p-1\}$ maradékosztály $\binom{l}{2}$-es esetén valamely h, j-re $|(A_h)_{p,i_{h,j}}| = |(A_j)_{p,i_{h,j}}|$. Felhasználva az előző lemmát a 2.5 Állításhoz hasonlóan azt kapjuk, hogy:

$$\binom{l}{2} n \geq \deg(\prod_{h \neq j} ((A_h)(x) - (A_j)(x))) \geq \sum_{p \leq k} \deg(\Phi_p(x)) = \sum_{p \leq k} (p-1) \sim \frac{k^2}{\log k}$$

Tehát $k \leq \sqrt{\binom{l}{2} n \log(n)}$. □

2.9. Tétel. $S_l(n) = \mathcal{O}(l^5 \sqrt{n} \log(n)^{1.5})$

Bizonyítás. Az előző állítást alkalmazva a $\mathrm{pos}_1(a_1), \mathrm{pos}_1(a_2), \ldots \mathrm{pos}_1(a_l) \subseteq \{0, 1, \ldots n-1\}$ páronként különböző halmazokra azt kapjuk, hogy található egy $p = \mathcal{O}(l\sqrt{n \log(n)})$ prím és egy $i_{1,2}, \ldots i_{l-1,l}$ maradékosztály $\binom{l}{2}$-es, melyekre minden $h \neq j$-re $|\mathrm{pos}_1(a_h)_{p,i_{h,j}}| \neq |\mathrm{pos}_1(a_j)_{p,i_{h,j}}|$.

Definiáljuk egy a szóra az alábbi értéket:

$$V(a) = |\mathrm{pos}_1(a)_{p,i_{1,2}}| + n \cdot |\mathrm{pos}_1(a)_{p,i_{1,3}}| + n^2 |\mathrm{pos}_1(a)_{p,i_{1,4}}| + \cdots + n^{\binom{l}{2}-1} |\mathrm{pos}_1(a)_{p,i_{l-1,l}}|$$

Könnyen látható, hogy $V(a_h) \neq V(a_j)$, ha $0 \leq h \neq j \leq l-1$, ugyanis a $V(a_h)$ és $V(a_j)$ értékeket n alapú számrendszerben felírva az $i_{h,j}$-hez tartozó számjegy különböző lesz.

Tudjuk, hogy $V(a) \leq n^{\binom{l}{2}+1}$ és a 1.4 következmény miatt létezik egy q prím, melyre $q = \mathcal{O}\left(\left(\binom{l}{2}+1\right)\binom{l}{2}\log(n)\right) = \mathcal{O}(l^4 \log(n))$ és ami nem osztja a $\prod_{h \neq j}(V(a_h) - V(a_j))$ szorzatot. Tehát a $V(a_1), \ldots V(a_l)$ értékek különböző maradékot adnak modulo q.

Készítsünk egy olyan automatát, amely egy a szó esetén a $V(a) \mod q$ értéket számolja. Ezt megtehetjük, ha q darab p hosszú kört használunk és ha egy kör $i_{h,j}$-edik állapotában 1-est olvasunk, akkor a (modulo q értve) n^x-el későbbi kör $i+1$-edik állapotába lépünk, egyébként ugyanazon a körön a következő állapotba. (Ahol n^x a $V(a)$ képletében a $|\mathrm{pos}_1(a)_{p,i_{h,j}}|$ tag együtthatója.) Ez egy pq állapotú automata, amely megkülönbözteti a szavakat. Tehát bármelyik n hosszú bináris szó l-eshez létezik egy $pq = \mathcal{O}\left(l^5 \sqrt{n} \log(n)^{1.5}\right)$ állapotú automata, amely megkülönbözteti őket. □

2.3. Második gyökös becslés

2.13. Definíció. Egy $a \in \Sigma^n$ szó p-periodikus, ha $a_i = a_{i+p}$ igaz minden $0 \leq i \leq n-1-p$ indexre. A legkisebb ilyen p-t nevezzük a szó periódusának. Egy szót periodikusnak nevezünk, ha a periódusa nem nagyobb, mint a hossza fele.

2.14. Lemma (Robson [6])**.** Az $w0$ és $w1$ szavak közül legalább az egyik nem periodikus. (Ahol $w0$ azt jelöli, hogy az w mögé írunk egy 0-t.)

Bizonyítás. Jelölje l a w szó hosszát, továbbá legyen p_0 a $w0$ periódusa és p_1 a $w1$ periódusa. Indirekt tegyük fel, hogy $w0$ és $w1$ is periodikus, azaz $p_0 \leq \frac{l+1}{2}$ és $p_1 \leq \frac{l+1}{2}$. Ekkor $w_{l+ap_1 \mod p_0} = 0$ és $w_{l+bp_0 \mod p_1} = 1$ minden $a, b \in \mathbb{Z}$-re. Az $ap_1 + bp_0 = -x \cdot \gcd(p_0, p_1)$ választással, ahol $x \in \mathbb{Z}$ és $0 \leq l - x \cdot \gcd(p_0, p_1) < p_0, p_1$ azt kapjuk, hogy $0 = w_{l-x\cdot\gcd(p_0,p_1)} = 1$.

□

2.15. Lemma (Robson [6])**.** Legyen $w \in \{0,1\}^l$, $a \in \{0,1\}^n$ és $l < n$. Ha w periódusa p és $a_i \ldots a_{i+l-1} = w = a_j \ldots a_{j+l-1}$, akkor $|j - i| \geq p$.

Bizonyítás. Trivi. □

2.16. Lemma (Robson [6]). Minden $\alpha < 1$-re ha az $a_{i-l+1} \dots a_i$ részszó nem periodikus és $l \leq n^\alpha$, akkor létezik egy j, hogy $a_{k-l+1} \dots a_k \neq a_{i-l+1} \dots a_i$ igaz minden olyan k-ra, amelyre $k \equiv i \mod j$, de $k \neq i$.

Továbbá létezik egy c konstans, ami csak α-tól függ, hogy j választható $c\frac{n\log(n)}{l}$-nél nem nagyobbnak.

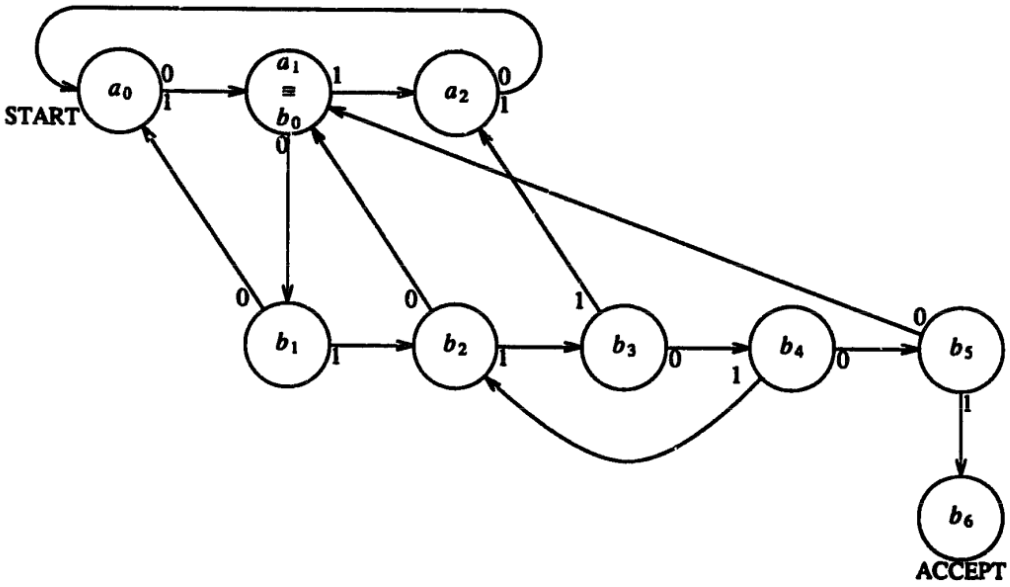
Bizonyítás. Az $a_i \dots a_{i-l+1}$ periódusa nagyobb, mint $\frac{l}{2}$. A 2.15 lemma miatt tehát ezen szó előfordulásainak száma a-ban maximum $\frac{2n}{l}$.

Minden ilyen előfordulás kevesebb, mint $(1-\alpha)^{-1}$ darab $j > \frac{n}{l}$ prímszámra ronthatja el a lemmában szereplő egyenlőtlenséget. Ez azért igaz, mert $|k-i|$-nek nem lehet $(1-\alpha)^{-1}$ darab $\frac{n}{l}$-nél nagyobb prím osztója. Ekkor ugyanis $\frac{n}{l} \geq \frac{n}{n^\alpha} = n^{1-\alpha}$ miatt az $(1-\alpha)^{-1}$ darab $n^{1-\alpha}$-nál nagyobb prím osztó szorzata már n-nél nagyobb lenne, de $|k-i| \leq n$.

Így lennie kell egy j-nek az $\frac{n}{l}$-nél nagyobb első $\frac{2n}{l(1-\alpha)}$ prím között amire igaz a lemma első fele. A nagy prímszámtételből következik a lemma második fele. □

2.17. Tétel (Robson [6]). $S_2(n) = \mathcal{O}(\sqrt{n\log(n)})$

Bizonyítás. Legyen $a \neq b \in \{0,1\}^n$. Jelölje i azt az indexet, ahol a két szó először eltér. Ha $i \leq \sqrt{n\log(n)}$, akkor a [2]-ben szereplő eltérés a szavak elején rész miatt a két szó megkülönböztethető egy $\mathcal{O}(\sqrt{n\log(n)})$ állapotú automatával. Ellenkező esetben az 2.14 lemma miatt az $a_{i-\sqrt{n\log(n)}} \dots a_i$ és $b_{i-\sqrt{n\log(n)}} \dots b_i$ szavak közül legalább az egyik nem periodikus. Tegyük fel, hogy az a-ban lévő nem periodikus. Válasszunk egy $j \leq c\frac{n\log(n)}{\sqrt{n\log(n)}+1} = \mathcal{O}(\sqrt{n\log(n)})$ számot az 2.16 lemma alapján. Ekkor minden $k \equiv i \mod j, k \neq i$ számra $a_{k-\sqrt{n\log(n)}} \dots a_k \neq a_{i-\sqrt{n\log(n)}} \dots a_i$.



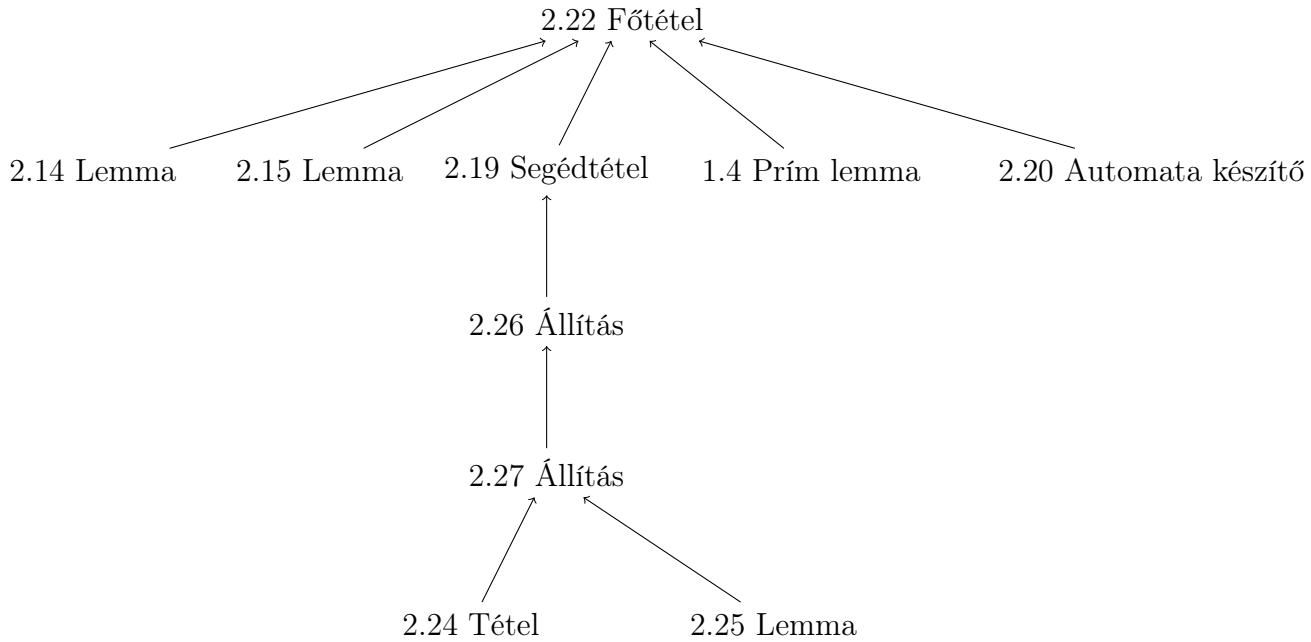
1. ábra. Példa a bizonyítás során készített automatára a 011001 részszó és $j = 3$ esetén, ha a részszó kezdőpozíciójára azt írjuk elő, hogy 1-el legyen kongruens modulo 3. Forrás: [6]

Készítsünk egy automatát, ami az $a_{i-\sqrt{n\log(n)}} \dots a_i$ részszó olyan előfordulását keresi, amelynek p kezdőpozíciója igaz, hogy $p \equiv i - \sqrt{n\log(n)} \mod j$. Az automata két részből fog állni. Az A rész az eddig beolvasott karakterek számát tartja számon modulo j. A $count \equiv i - \sqrt{n\log(n)} \mod j$ állapot megegyezik a B rész kezdőállapotával(nulladik állapot). A B automata x-edik állapotában akkor vagyunk, ha az előző x karakter a részszó x hosszú prefixe volt és a megfelelő pozícióban kezdődött modulo j. Ez A rész j állapotból és a B rész további $\sqrt{n\log(n)} + 1$ állapotból áll és az egész automata $\mathcal{O}(\sqrt{n\log(n)})$ állapottal megkülönbözteti a két szót. □

1989-ben Robson ezt a gondolatot javítva bizonyította be az $\tilde{\mathcal{O}}(n^{2/5})$ felső korlátot, amely a [6] cikkben olvasható.

2.4. A legerősebb ismert becslés

Zachary Chase 2020-ban bebizonyította az eddig ismert legjobb felső korlátot [1], amely szerint $S_2(n) = \tilde{\mathcal{O}}(n^{1/3})$. Ebben a fejezetben ezen eredmény bizonyításának lépéseit fogom bemutatni. Az alábbi ábra mutatja, hogy a felhasznált tételekből hogyan jutunk el a fő eredményhez. Úgy gondolom ez segítséget nyújthat az egyes lépések közötti összefüggések megértéséhez és a teljes kép könnyebb átlátásához.



2.18. Definíció. Az $A \subseteq \{1, \dots n\}$ halmazt d-szeparáltnak nevezzük, ha minden $i \neq j \in A$-ra $|j - i| \geq d$.

2.19. Tétel (Chase [1])**.** Ha $A \neq B \subseteq \{1, \dots n\}$ és $n^{1/3}$-szeparáltak, akkor létezik egy $p = \mathcal{O}(n^{1/3}\log^6(n))$ prím és i maradékosztály modulo p, melyekre $|A_{p,i}| \neq |B_{p,i}|$.

Először ezen tétel felhasználásával bebizonyítom a fő eredményt, majd a kövötkező részben bemutatom a segédtétel bizonyításának lépéseit.

2.20. Lemma (Chase [1]). Legyen $m \in \mathbb{Z}^+$, $i \in [m]_0$ egy maradékosztály modulo m, q prímszám és $a \in [q]_0$ maradékosztály modulo q. Ekkor létezik egy $2mq$ állapotú automata, amely pontosan azokat az $x \in \{0,1\}^n$ szavakat fogadja el, amelyekre $|\mathrm{pos}_w(x)_{m,i}| \equiv a \mod q$.

Bizonyítás. Készítsünk egy q darab m hosszú körből álló automatát, melyben minden állapotból kettő van, egy 0-ás és egy 1-es. Mindegyik kör azt ellenőrzi, hogy a w szót olvassuk-e a megfelelő pozícióban. Azaz ha a kör i-edik állapotában a w_1-et olvassuk a kör $i+1$-edik állapotának az 1-es változatába kerülünk. Ameddig a w következő betűjét olvassuk a következő állapot 1-es változatába kerülünk. Ha valamelyik lépésben nem a w megfelelő betűje következik, akkor a következő állapot 0-ás változatába lépünk. Ha egy kör $i + len(w) - 1$-edik állapotának az 1-es példányában a w utolsó betűjét olvassuk akkor átkerülönk a következő kör $i+1$-es állapotába, ellenkező esetben ugyanazon a körön maradunk. Az a-adik kör állapotai lesznek az elfogadó állapotok. □

Fel fogjuk használni a Robson-féle gyökös becslésnél bebizonyított 2.14 és 2.15 lemmákat:

2.14. Lemma (Robson [6]). Az $w0$ és $w1$ szavak közül legalább az egyik nem periodikus. (Ahol $w0$ azt jelöli, hogy az w mögé írunk egy 0-t.)

2.15. Lemma (Robson [6]). Legyen $w \in \{0,1\}^l$, $a \in \{0,1\}^n$ és $l < n$. Ha w periódusa p és $a_i \dots a_{i+l-1} = w = a_j \dots a_{j+l-1}$, akkor $|j - i| \geq p$.

2.21. Következmény. Ha w szó periódusa p, akkor a $\mathrm{pos}_w(a)$ halmaz p-szeparált.

2.22. Tétel (Chase [1]). $S_2(n) = \mathcal{O}(n^{1/3} \log^7(n))$.

Bizonyítás. Legyen $a, b \in \{0,1\}^n$ két különböző bináris szó. Ha az első $2n^{1/3}$ pozíció valamelyikén eltér a két szó, akkor a [2]-ben lévő eltérés a szavak elején rész miatt készen vagyunk. Különben legyen $k > 2n^{1/3}$ az a pozíció, ahol a két szó először eltér. Legyen $w' = a_{k-2n^{1/3}+1} \dots a_{k-1} = b_{k-2n^{1/3}+1} \dots b_{k-1}$ az ezen pozíció előtti közös $2n^{1/3} - 1$ hosszú részszavuk. Ez előző kettő lemma miatt választhatunk egy $w \in \{w'0, w'1\}$ szót, amelyre a $\mathrm{sep}_w(a), \mathrm{sep}_w(b)$ halmazok $n^{1/3}$-szeparáltak. Emellett $\mathrm{sep}_w(a) \neq \mathrm{sep}_w(b)$ is igaz lesz.

A 2.19 segédtételt felhasználva azt kapjuk, hogy létezik egy $p = \mathcal{O}(n^{1/3} \log^6(n))$ prím és $i \in [p]_0$ maradékosztály, melyekre $|\mathrm{pos}_w(a)_{p,i}| \neq |\mathrm{pos}_w(b)_{p,i}|$. A 1.4 következmény miatt létezik egy $q = \mathcal{O}(\log(n))$ prím, melyre $|\mathrm{pos}_w(a)_{p,i}| \not\equiv |\mathrm{pos}_w(b)_{p,i}| \mod q$.

Ezután felhasználva a 2.20 lemmát kapunk egy $2pq = \mathcal{O}(n^{1/3} \log^7(n))$ állapotú automatát, amely megkülönbözteti a két szót. □

A 2.19 segédtétel bizonyítása

2.23. Definíció. Definiáljuk a komplex polinomok egy részhalmazát az alábbi módon: $\mathcal{P}_n = \{p(x) = 1 - \sigma x^d + \sum_{j=n^{1/3}}^{n} a_j x^j \in \mathbb{C}[x] : 1 \leq d \leq n^{1/3}, \sigma \in \{0,1\}, |a_j| \leq 1 \forall j\}$

2.24. Tétel (Chase [1]). Létezik egy C_1 abszolút konstans, melyre igaz, hogy ha $n \geq 2$ és $p \in \mathcal{P}_n$, akkor $\max_{x \in [1-n^{-2/3},1]} |p(x)| \geq \exp(-C_1 n^{1/3} \log^5(n))$.

Itt $p(x)$ egy komplex polinom, aminek a maximumát egy olyan intervallumon nézzük, ami a valós számok részhalmaza.

2.25. Lemma ([5]). Tegyük fel, hogy $p(x) = \sum_{j=0}^{n} a_j x^j \in \mathbb{C}[n]$ komplex polinomra igaz, hogy $|a_j| \leq 1$ minden j-re. Ekkor ha $(x-1)^k | p(x)$, akkor $\max_{x \in [1-\frac{k}{9n}, 1]} |p(x)| \leq (n+1)\left(\frac{e}{9}\right)^k$.

A lemma azt a gondolatot használja fel, hogy ha egy polinomnak az 1 többszörös gyöke, akkor a polinom deriváltjai eltűnnek az 1 helyen, azaz itt sima. A $p(1) = \sum_{j=0}^{n} a_j \leq n+1$ és az 1 pont körüli simaság miatt az 1 közelében a polinom értékei nem lehetnek túl magasak.

A továbbiakban a 2.24 Tétel és a 2.25 Lemma felhasználásával bebizonyítjuk a segédállítást. Ezen felhasznált két tétel bizonyítására nem térek ki és a megjelölt forrásokban megtalálhatóak.

2.26. Állítás (Chase [1]). Létezik egy $C > 0$ abszolút konstans, hogy $\forall n \geq 2$ és $p \in \mathcal{P}_n$-re igaz $(x-1)^{\lfloor Cn^{1/3}\log^5(n) \rfloor}$ nem osztja $p(x)$-et.

Bizonyítás. Legyen C egy megfelelően nagy konstans. Indirekt tegyük fel, hogy valamely n-re és $p \in \mathcal{P}_n$-re $(x-1)^{\lfloor Cn^{1/3}\log^5(n) \rfloor}$ osztja $p(x)$-et. A 2.25 Lemma és a 2.24 Tétel felhasználásával az alábbi egyenlőtlenség lánc vezethető le:

$$\begin{aligned}(n+1)\left(\frac{e}{9}\right)^{\lfloor Cn^{1/3}\log^5(n) \rfloor} &\geq \max_{x \in [1-\frac{C}{9}n^{-2/3}\log^5(n), 1]} |p(x)| \\ &\geq \max_{x \in [1-n^{-2/3}, 1]} |p(x)| \\ &\geq \exp(-Cn^{1/3}\log^5(n)) \\ &= \left(\frac{1}{e}\right)^{Cn^{1/3}\log^5(n)}\end{aligned}$$

Ha C elég nagy, akkor ez ellentmondás, ugyinais $\frac{e}{9} < \frac{1}{e}$. □

2.27. Állítás (Chase [1]). Legyenek az $A \neq B \subseteq \{1, \dots n\}$ halmazok $n^{1/3}$-szeparáltak. Ekkor létezik egy $m = \mathcal{O}(n^{1/3}\log^5(n))$ egész szám, amelyre $\sum_{a \in A} a^m \neq \sum_{b \in B} b^m$.

Bizonyítás. Definiáljuk az f polinomot a következő módon: $f(x) = \sum_{j=0}^{n} \epsilon_j x^j$, ahol $\epsilon_j = \mathbb{1}_A(j) - \mathbb{1}_B(j)$. Legyen $r \in \mathbb{Z}^+$ az a legnagyobb szám, amelyre $\epsilon_0 = \dots = \epsilon_{r-1} = 0$. Ekkor az $\tilde{f}(x) = \frac{f(x)}{x^j} = \epsilon_r + \epsilon_{r+1}x + \dots \epsilon_n x^{n-r}$ polinom főegyütthatója 1 vagy -1. Feltehető, hogy $\epsilon_r = 1$ (ellenkező esetben az A és B halmazt megcseréljük). Az A és B halmazok $n^{1/3}$-szeparálhatóak, amiből az következik, hogy $\tilde{f}(x) \in \mathcal{P}_n$.

Az 2.26 Állítás miatt $(x-1)^{\lfloor Cn^{1/3}\log^5(n) \rfloor}$ nem osztja az $\tilde{f}(x)$ polinomot, emiatt $f(x)$-et sem. Ez azt jelenti, hogy létezik egy $0 \leq k < \lfloor Cn^{1/3}\log^5(n) \rfloor$ egész szám, amelyre $f(x) = (x-1)^k g(x)$ és $(x-1) \nmid g(x)$, azaz $g(1) \neq 0$. Ekkor az f függvény k-adik deriváltja $f^{(k)}(x) = (k!)g(x) + (x-1)(...)$, tehát $f^{(k)}(1) \neq 0$ és k a legkisebb ilyen szám. Ha $k = 0$, akkor $0 \neq f^{(0)}(1) = f(1) = |A| - |B|$, tehát $m = 0$-val igaz az állítás.

Ha $k > 0$, akkor indukcióval belátjuk, hogy minden $m < k$-ra $\sum_{a \in A} a^m = \sum_{b \in B} b^m$ és $\sum_{a \in A} a^k \neq \sum_{b \in B} b^k$. A kezdőlépéshez be kell látni, hogy $m = 0$-ra igaz az egyenlőség, azaz $|A| = |B|$. Ez igaz, mert $0 = f^{(0)}(1) = f(1) = |A| - |B|$.

Az indukciós lépéshez tegyük fel, hogy $0, \dots m-1$-re igaz az egyenlőség és be szeretnénk látni, hogy $\sum_{a \in A} a^m = \sum_{b \in B} b^m$, ha $m < k$ vagy $\sum_{a \in A} a^m \neq \sum_{b \in B} b^m$ ha $m = k$.

Tudjuk, hogy $c = f^{(m)}(1) = \sum_{j=0}^{n} j(j-1)\dots(j-m+1)\epsilon_j$ és $c = 0$, ha $m < k$, $c \neq 0$, ha $m = k$. Rendezzük a j^m-es tagokat a bal oldalra, ekkor azt kapjuk, hogy $\sum_{j=0}^{n} j^m \mathbb{1}_A(j) - \sum_{j=0}^{n} j^m \mathbb{1}_B(j) = \sum_{i=0}^{m-1} \alpha_i \left(\sum_{j=0}^{n} j^i \mathbb{1}_A(j) - \sum_{j=0}^{n} j^i \mathbb{1}_B(j) \right) - c = -c$. Itt az α_i-k a megfelelő együtthatók és felhasználjuk, hogy $\sum_{j=0}^{n} j^i \mathbb{1}_A(j) = \sum_{a \in A} a^m$. Ezzel beláttuk az állítást.

□

2.19. Tétel (Chase [1])**.** Ha $A \neq B \subseteq \{1, \dots n\}$ és $n^{1/3}$-szeparáltak, akkor létezik egy $p = \mathcal{O}(n^{1/3} \log^6(n))$ prím és i maradékosztály modulo p, melyekre $|A_{p,i}| \neq |B_{p,i}|$.

Bizonyítás. Az előző állítás alapján választható egy olyan $m = \mathcal{O}(n^{1/3} \log^5(n))$ egész szám, amelyre $\sum_{a \in A} a^m \neq \sum_{b \in B} b^m$. A 1.4 következmény alkalmazásához felhasználjuk, hogy $\sum_{a \in A} a^m \leq n \cdot n^m$ és $\sum_{b \in B} b^m \leq n \cdot n^m$. Így azt kapjuk, hogy létezik egy $p = \mathcal{O}(\log(n^{m+1})) = \mathcal{O}((m+1)\log(n)) = \mathcal{O}(n^{1/3} \log^6(n))$ prím, amelyre $\sum_{a \in A} a^m \not\equiv \sum_{b \in B} b^m \mod p$.

Megfigyelhetjük, hogy $\sum_{a \in A} a^m \equiv \sum_{i=1}^{p-1} |A_{p,i}| i^m \mod p$ és ugyanez igaz B-re is. Ezt felhasználva adódik, hogy $\sum_{i=1}^{p-1} |A_{p,i}| i^m \not\equiv \sum_{i=1}^{p-1} |B_{p,i}| i^m \mod p$. Ebből következik, hogy valamely $i \in \{0, \dots p-1\}$-re $|A_{p,i}| \neq |B_{p,i}|$. □

Hivatkozások

[1] Zachary Chase. „A new upper bound for separating words". (2020). URL: `https://arxiv.org/abs/2007.12097`.

[2] Csányi Dávid. „Separating Words Problem (Projekt2)". (2022). URL: `https://math-projects.elte.hu/media/works/266/report/csanyi_david_projekt2_beszamolo.pdf`.

[3] Erik D. Demaine, Sarah Eisenstat, Jeffrey Shallit és David A. Wilson. „Remarks on Separating Words". (2011). URL: `https://arxiv.org/abs/1103.4513v1`.

[4] Kiss Emil. *Bevezetés az algebrába.* 2007.

[5] P.Borwein, T. Erdélyi és G. Kós. „Littlewood type problems on [0,1]". *Proc. London Math. Soc.* 79.1 (1999), 22–46. old.

[6] J. M. Robson. „Separating strings with small automata". (1989).

[7] J. Shallit és Y. Breitbart. „Automaticity I: Properties of a measure of descriptional complexity." *J. Comput. System Sci.* 53 (1996), 10–25. old.

[8] M. N. Vyalyi és R. A. Gimadeev. „Separating Words by Occurrences of Subwords". (2014). URL: `https://link.springer.com/content/pdf/10.1134/S1990478914020161.pdf`.