Szókombinatorikai alkalmazások

Kása Zoltán

Babeş-Bolyai Tudományegyetem, Kolozsvár

E-mail: `kasa@cs.ubbcluj.ro`

`http://www.cs.ubbcluj.ro/~kasa`

---

véges szó az $\mathcal{A}$ ábécé fölött:

$$w = w_1 w_2 \ldots w_N, \qquad w_i \in \mathcal{A}, 1 \leq i \leq N.$$

u *részszava* w-nek, ha : $\exists x, y \in \mathcal{A}^*$: $w = xuy$

$F(w)$ a w nem üres részszavainak a halmaza
$F_n(w)$ a w szó n hosszúságú részszavainak a halmaza

w *részszó-bonyolultsága*:

$$f_w(n) = \#F_n(w), \qquad 1 \leq n \leq |w|$$

$w = abaaba$
$F_1(w) = \{a, b\}$, $F_2(w) = \{ab, aa, ba\}$,
$F_3(w) = \{aba, baa, aab\}$

végtelen szó:

$u = u_0u_1u_2\dots u_n\dots,\quad u_i \in \mathcal{A}$

$f_u(n) = \#F_n(u)$ *részszó-bonyolultság*

Példák:

1) Fibonacci-szó: $\sigma(0) = 01$, $\sigma(1) = 0$

0

01

010

01001

01001010 $\quad u_F = \underbrace{01001010}\underbrace{01001}\dots$

$f_{u_F}(n) = n + 1$

2) Hatványszó:

$u_p = 010011000111\dots\underbrace{0\dots0}_{n}\underbrace{1\dots1}_{n}\dots$

$$f_{u_p}(n) = \frac{n(n+1)}{2} + 1$$

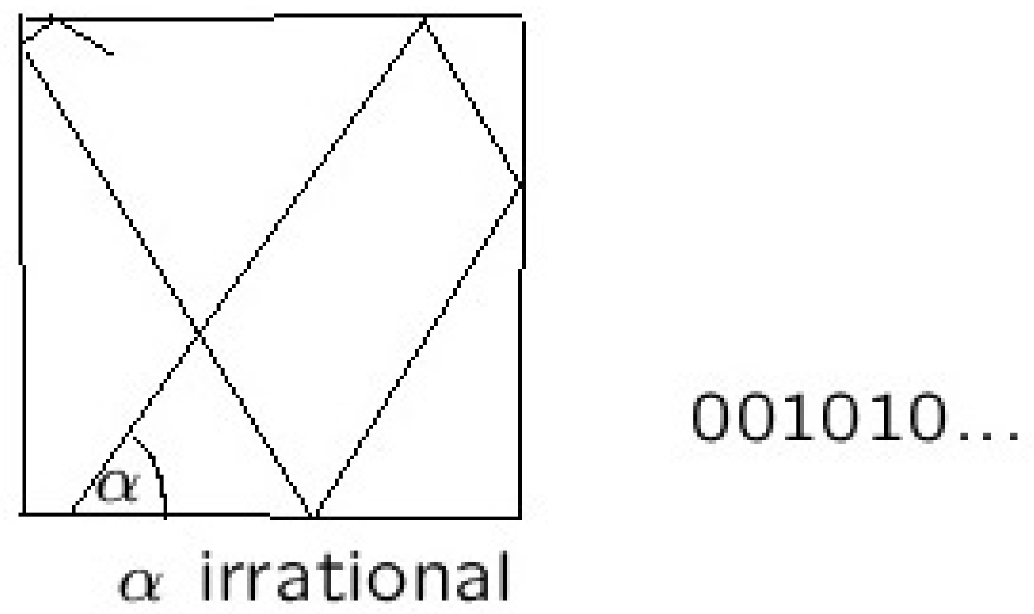
3) Champarnowne-szó:

$u_C = 0\ 1\ 10\ 11\ 100\ 101\ 110\ 111\ \dots$

$f_{u_C}(n) = 2^n$

$\underbrace{abc}\underbrace{abc}\underbrace{abc}\ldots\underbrace{abc}\ldots$ *periodikus*
$aaaaaba\,\underbrace{abc}\underbrace{abc}\underbrace{abc}\ldots\underbrace{abc}\ldots$ *végperiodikus*

Ha $f_u(n) \leq n$ minden $n \geq n_0$-ra, akkor u (vég)periodikus.
Sturm típusú szavak amelyekre $f_u(n) = n + 1$.



3-dimenzióban: $n^2 + n + 1$

S. Ferenczi, C. Mauduit: $u = u_0 u_1 u_2 \ldots u_n \ldots$, $u_i \in \{0, 1\}$, hozzárendeljük $\theta = 0.u_0 u_1 \ldots u_n \ldots$ 2-es alapú számrendszerben. *Ha u Sturm-szó*, akkor *θ transzcendens szám.*

Tribonacci-szó:

$\sigma(0) = 01, \;\; \sigma(1) = 02, \;\; \sigma(2) = 0$

0

01

0102

0102010

$\underbrace{0102010}\,\underbrace{0102}\,\underbrace{01}$

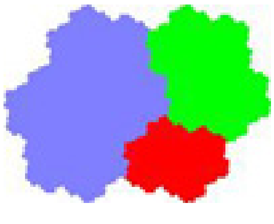
...

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

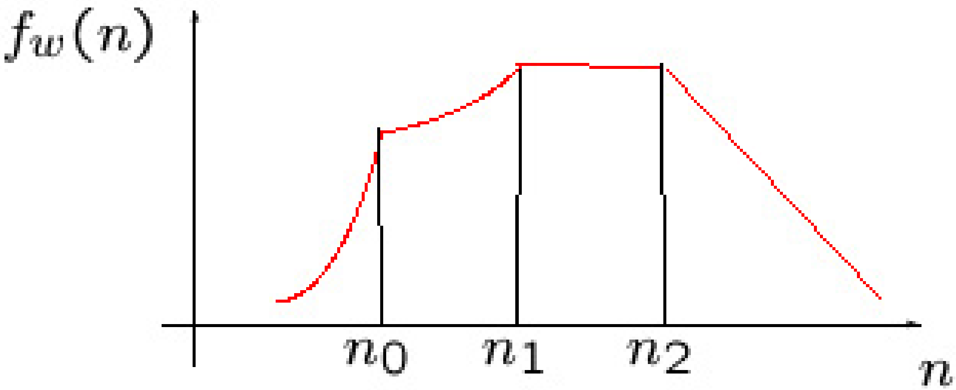
$$\begin{vmatrix} 1-x & 1 & 0 \\ 1 & -x & 1 \\ 1 & 0 & -x \end{vmatrix} = 0$$

karakterisztikus polinom, gyökei: $\beta > 1$, $\alpha, \overline{\alpha}$ komplexek

$$\mathcal{R} = \Big\{ \sum_{i \geq 0} \varepsilon_i \, \alpha^i ; \; \varepsilon_i = 0, 1; \; \varepsilon_i \varepsilon_{i+1} \varepsilon_{i+2} = 0 \Big\} \subset \mathbf{C}.$$



n	1	2	3	4	5
11111	1	1	1	1	1
11112	2	2	2	2	1
21122	2	4	3	2	1
21211	2	3	3	2	1
22112	2	4	3	2	1
22211	2	3	3	2	1



maximális bonyolultság:
$C(w) = \max\{f_w(n) \mid n \geq 1\}$
globális maximális bonyolultság $\mathcal{A}^N$-ben:
$K(N) = \max\{C(w) \mid w \in \mathcal{A}^N\}$

$R(N) = \{i \in \overline{1, N} \mid \exists w \in \mathcal{A}^N : f_w(i) = K(N)\}$

$M(N)$: azon $\mathcal{A}^N$-beli szavak száma, amelyek maximális bonyolultsága egyenlő a globális maximális bonyolultsággal

N	$K(N)$	$R(N)$	$M(N)$
1	1	1	2
2	2	1	2
3	2	1, 2	6
4	3	2	8
5	4	2	4
6	4	2, 3	36
7	5	3	42
8	6	3	48
9	7	3	40
10	8	3	16
11	8	3, 4	558
12	9	4	718
13	10	4	854
14	11	4	920
15	12	4	956
16	13	4	960
17	14	4	912
18	15	4	704
19	16	4	256
20	16	4, 5	79006

1. tétel
Ha $\#\mathcal{A} = q$ és $q^k + k \le N \le q^{k+1} + k$, akkor $K(N) = N - k$.

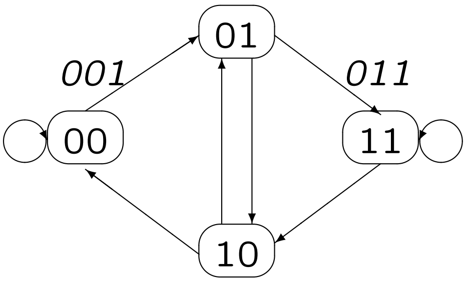
2. tétel
Ha $\#\mathcal{A} = q$ és $q^k + k < N < q^{k+1} + k + 1$, akkor $R(N) = \{k+1\}$;
ha $N = q^k + k$, akkor $R(N) = \{k, k+1\}$.

De Bruijn-gráfok

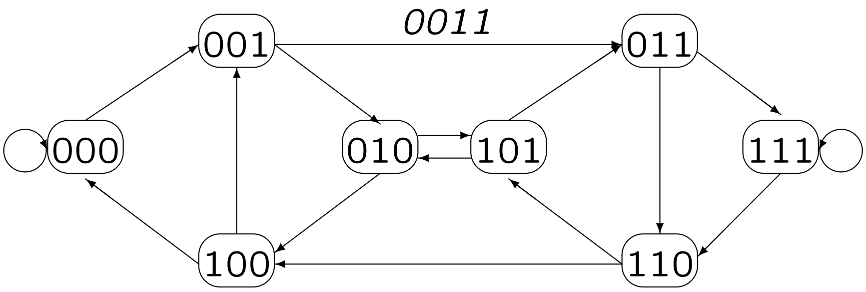
$\mathcal{A}$ q-betűs ábécé, a de Bruijn gráf::

$$B(q,k) = (V(q,k), E(q,k)),$$

ahol $V(q,k) = \mathcal{A}^k$ csúcsok halmaza,
$E(q,k) = \mathcal{A}^{k+1}$ élek halmaza.
Van él $x_1x_2\dots x_k$-ból $y_1y_2\dots y_k$-ba,
ha $x_2x_3\dots x_k = y_1y_2\dots y_{k-1}$,
és az él $x_1x_2\dots x_ky_k$.



$B(2,2)$

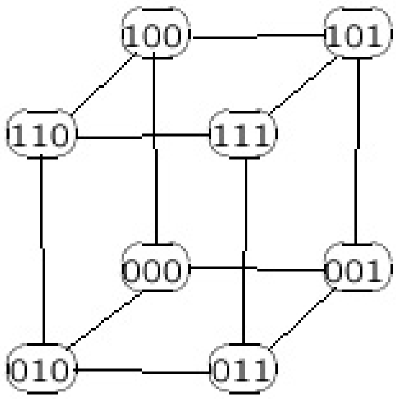


$B(2,3)$

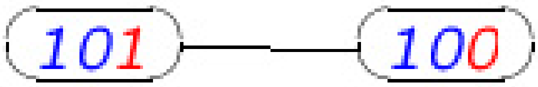
út: 001, 011, 111, 110 $\Longrightarrow$ szó: *001110*
Hamilton-út:
000, 001, 011, 111, 110, 101, 010, 100 $\Longrightarrow$
szó: *0001110100*

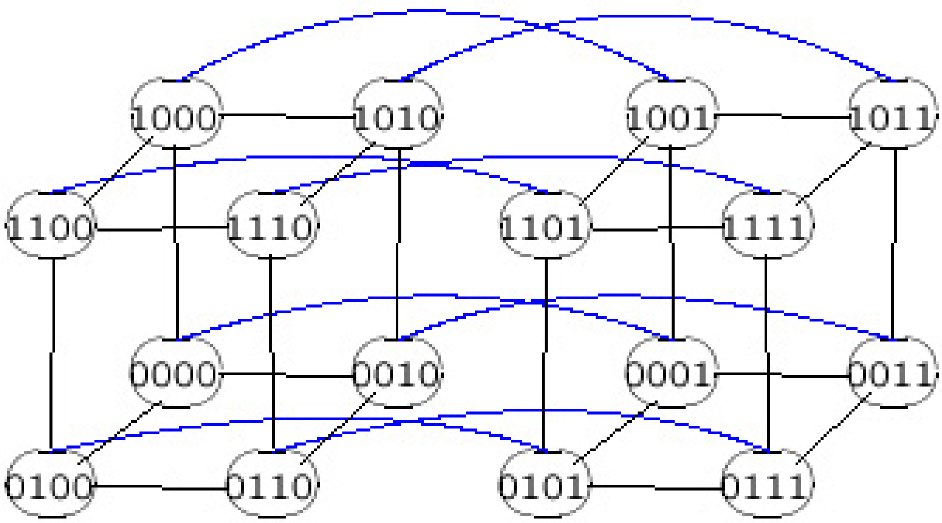
$B(q,k)$-ban leghosszabb út (amely Hamiltonian-út) egy de Bruijn-szót eredményez.

nem irányított de Bruijn-gráfok
hálózati modellek



(3-dimenziós hiper)kocka





4-dimenziós hiperkocka

3. tétel

Ha $\#\mathcal{A} = q$ és $q^k + k \leq N \leq q^{k+1} + k$, akkor $M(N)$ egyenlő a különböző $N - k - 1$ hosszúságú utakkal a $B(q, k+1)$ gráfban.

Pl.

$k = 2 : 2^2 + 2 \leq N \leq 2^3 + 2$

$N = 6$, $N - k - 1 = 3$, összesen 36 különböző 3 hosszúságú út van $B(2, 3)$-ban.

000, 111, 010, 101-ből kiindulva
4 különböző út van, máshonnan 5

4. tétel

Ha $N = 2^k + k - 1$, akkor $M(N) = 2^{2^{k-1}}$.

Általánosítás:

Ha $N = q^k + k - 1$, akkor $M(N) = (q!)^{q^{k-1}}$.

teljes bonyolultság: $\mathbf{K}(w) = \sum_{i=1}^{|w|} f_w(i)$

- $C \neq 1, 2, 4$, akkor $\exists$ egy nem triviális w úgy, hogy $\mathbf{K}(w) = C$
- $C \neq 1, 2, 4, 6, 10, 18, 22$, akkor $\exists$ egy nem triviális $w \in \mathcal{A}^*$, $\#\mathcal{A} = 2$, úgy, hogy $\mathbf{K}(w) = C$

$|w| = 5$

C	5	6	7	8	9	10	11	12	13	14	15
$f_5(C)$	5	0	0	0	60	0	200	400	1140	1200	120

$$f_k(C) \neq 0, \ \forall C\text{-re úgy, hogy } b_k \leq C \leq \frac{k(k+1)}{2}$$

$$k = \frac{\ell(\ell+1)}{2} + 2 + i, \ \ell \geq 2, \ 0 \leq i \leq \ell \Rightarrow$$

$$b_k = \frac{\ell(\ell^2-1)}{2} + 3\ell + 2 + i(\ell+1)$$

$k = 5 = \frac{2\cdot 3}{2} + 2 + 0$ so $\ell = 2, i = 0$, then $b_5 = \frac{2\cdot 5}{2} + 3\cdot 2 + 2 + 0 = 11$.

F. Levé, P. Séébold, 2000

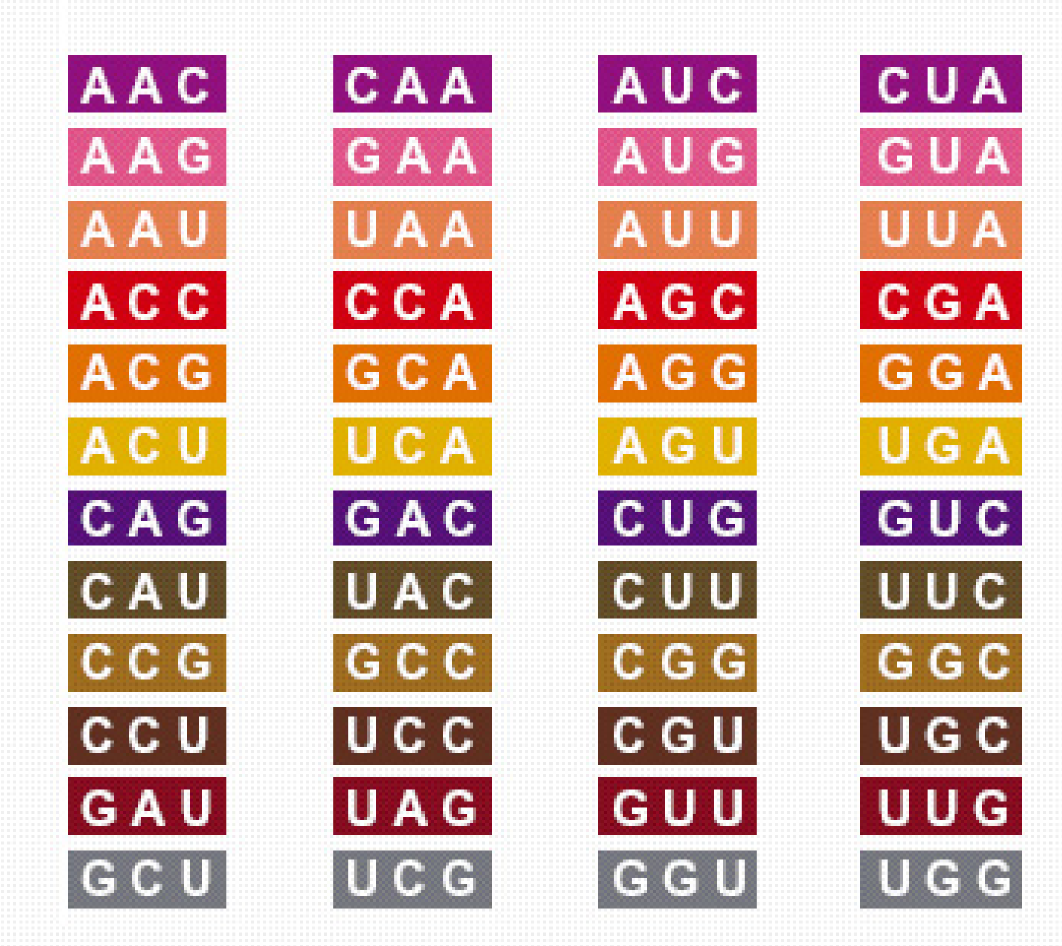
$$\mathbf{K}(w) = \sum_{k=1}^{|w|} f_w(k)$$

$$\mathbf{K}_u^+(n) = \max_i \ \mathbf{K}(u_i u_{i+1} \ldots u_{i+n-1})$$

$$\mathbf{C}(w) = \max_{k=1}^{|w|} f_w(k)$$

$$\mathbf{C}_u^+(n) = \max_i \ \mathbf{C}(u_i u_{i+1} \ldots u_{i+n-1})$$

u *nem végperiodikus:*	u *Sturm:*
$f_u(n) \geq n+1$	$f_u(n) = n+1$
$\mathbf{C}_u^+(n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1$	$\mathbf{C}_u^+(n) = \left\lfloor \frac{n}{2} \right\rfloor + 1$
$\mathbf{K}_u^+(n) \geq \left\lfloor \frac{n^2}{4} + n \right\rfloor$	$\mathbf{K}_u^+(n) = \left\lfloor \frac{n^2}{4} + n \right\rfloor$



genetikai kód

4 alapelem

DNS

A (adenine), *T* (thymine), *G* (guanine), *C* (cytosine)

RNS

A (adenine), *U* (uracil), *G* (guanine), *C* (cytosine)

. . . UGUCGUAAG. . . UGU, GUC, UCG, . . .

(B. Hayes: *The invention of the genetic code*, American Scientist, 1998, vol. 86 no. 1

jobbspeciális részszó többféleképpen lehet folytatni jobbról

0100101001001010010 10. . .
010 jobbspeciális részszó: *0100*, *0101*
100 nem jobbspeciális: *1001*

(bal-, jobb-, bi-) speciális részszavak szerepe DNS és RNS molekulákban

010 bispeciális
0100, *0101*
0010, *1010*

---

mintaillesztés

E-mail: kasa@cs.ubbcluj.ro
http://www.cs.ubbcluj.ro/~kasa