Diszkrét matematika 2

10. előadás
Kódelmélet

Mérai László
merai@inf.elte.hu
https://sites.google.com/view/laszlomerai

Komputeralgebra Tanszék

2023 ősz

Forráskódolás

Emlékeztető

Kódolás: $\varphi : \mathcal{X} \to \mathcal{Y}^*$ injektív függvény. Szavak kódolása betűnként:
$\varphi(u_1 u_2 \ldots u_r) = \varphi(u_1)\varphi(u_2)\ldots\varphi(u_r)$

Felbontható kódolás: ha egyértelműen dekódolható: $\mathbf{u} \neq \mathbf{v}$, akkor

$$\varphi(u_1)\varphi(u_2)\ldots\varphi(u_r) \neq \varphi(v_1)\varphi(v_2)\ldots\varphi(v_s).$$

Prefix kódolás

Célunk elégséges feltételt adni a felbonthatóságra.

Definíció

Egy $\mathbf{u} = \mathbf{abc}$ szó

- prefixe: $\mathbf{a}$, $\mathbf{ab}$, $\mathbf{abc}$;
- infixe: $\mathbf{b}$, $\mathbf{ab}$, $\mathbf{bc}$: $\mathbf{abc}$;
- szuffixe: $\mathbf{c}$, $\mathbf{bc}$, $\mathbf{abc}$.

Definíció

Egy φ kódolás prefix kód (vagy prefixmentes kód), ha nem léteznek olyan $\mathbf{u}, \mathbf{v}$ különböző kódszavak, hogy $\mathbf{u}$ prefixe $\mathbf{v}$-nek.

Példa

- ASCII és UTF-8 prefix kódok.
- Morze-kód nem prefix kód:
  $\varphi(e) = \cdot$ és $\varphi(i) = \cdot\cdot$ prefixei a $\varphi(s) = \cdot\cdot\cdot$ kódszónak.

Prefix kódolás

Prefix kód: nincsenek olyan $\mathbf{u}, \mathbf{v}$ kódszavak, hogy $\mathbf{u} = \mathbf{vc}$ valamely $\mathbf{c}$ szóra.

Tétel

Minden prefix kód felbontható.

Bizonyítás.

- Legyen $\mathbf{v} = v_1 v_2 \dots v_s \in \mathcal{Y}^*$ egy üzenet kódolása. (Azaz létezik olyan $\mathbf{u}$, hogy $\varphi(\mathbf{u}) = \mathbf{v}$.)
- Vizsgáljuk meg a prefixeit:
  - v_1
  - $v_1 v_2$
  - $v_1 v_2 v_3$
  - $\dots$.
- Ha találunk egy $v_1 v_2 \dots v_i$ szót, ami egy betű kódszavak, azt dekódolhatjuk. Mivel a kód prefix, ez nem lehet más betű kódjának prefixe.
- Az eljárást folytathatjuk a $v_{i+1} v_{i+2} \dots v_s$ kóddal.

□

Kódfa

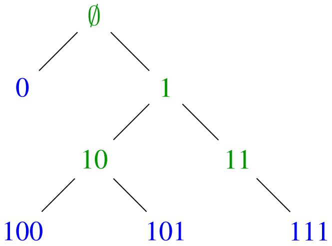
Definíció

Egy φ kód kódfája egy olyan fa, melynek csúcsai a kódszavak és azok prefixei és az $y_1y_2\dots y_s$ és $y_1y_2\dots y_sy_{s+1}$ csúcsok vannak összekötve.

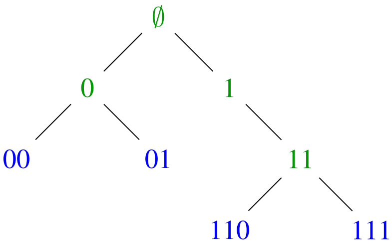
Állítás: Egy kód prefix, ha csak a levelek a kódszavak.

Példa

A $\{0, 100, 101, 111\}$ kódszóhalmaz kódfája:



A $\{00, 01, 110, 111\}$ kódszóhalmaz kódfája:



Prefix kódolás

Prefix kód: nincsenek olyan $\mathbf{u}, \mathbf{v}$ kódszavak, hogy $\mathbf{u} = \mathbf{vc}$ valamely $\mathbf{c}$ szóra.

Példa

- A $\{0, 100, 101, 111\}$ kód prefix.
- A $\{100, 10, 11\}$ kód nem prefix: 10, 100 szavak is kódszavak.

Elégséges feltételek a prefix tulajdonságra:

Definíció

Legyen $\mathcal{C} \subset \mathcal{Y}^*$ a kódszavak véges halmaza. Ekkor

- A $\mathcal{C}$ kód egyenletes (blokk kód), ha minden $\mathbf{c} \in \mathcal{C}$ kódszó azonos hosszú.
- A $\mathcal{C}$ kód vesszős kód, ha van olyan $\mathbf{v} \in \mathcal{Y}^*$ nemüres szó („vessző”), mely szuffixe minden $\mathbf{c} \in \mathcal{C}$ kódszónak, de nem prefixe, ill. infixe semelyik kódszónak.

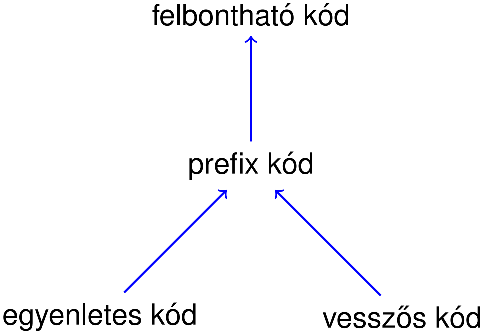
Példa

- A $\{000, 010, 111, 101\}$ kód egyenletes.
- A $\{0100, 100, 1100\}$ kód vesszős.

Felbontható kód mégegyszer

Tétel (Biz. HF.)

Minden egyenletes ill. vesszős kód prefix.



Példák

Példa

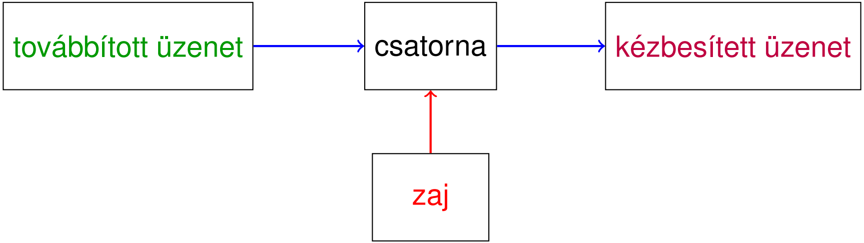
- ASCII: egyenletes kód: minden karakter 7 biten kódolt.
- UTF-8: prefix kód:

`0xxxxxxx`	ASCII karakterek
`110yyyyy 10xxxxxx`	nem ASCII karakterek
`1110zzzz 10xxxxxx 10yyyyyy`	
`11110www 10zzzzzz 10xxxxxx 10yyyyyy`	

- Ha az első karakter `0` $\rightarrow$ ASCII karakter
- Ha az első karakter `1` $\rightarrow$ nem ASCII karakter. Ekkor a kódszó több byte, első blokk `1`-ek száma a byte-ok száma, `1`-ek után `0`, minden további byte `10`-val kezdődik.

Csatornakódolás

Csatornakódolás



Hiba fajták

- karakter módosulás
- karakterek tölése
- karakterbeszúrás

Lehetséges módszerek

- hibajelzés
- hibájavítás

Példák kódokra

- Kódismétlés: $0 \mapsto 000$, $1 \mapsto 111$
  Képes egy hibát javítani, két hibát jelezni

- ISBN (könyvek és egyéb kiadványok egyedi azonosítója).

  Az első tipus (10 számjegyű, 2007-ig).
  Ha $I_1, I_2, \ldots, I_{10}$ az ISBN, akkor ez helyes, ha

  $$1 \cdot I_1 + 2 \cdot I_2 + \cdots + 10 \cdot I_{10} \equiv 0 \mod 11.$$

  Példa ISBN 0-246-024682
  $1\cdot 0+2\cdot 2+3\cdot 4+4\cdot 6+5\cdot 0+6\cdot 2+7\cdot 4+8\cdot 6+9\cdot 8+10\cdot 0 = 220 \equiv 0 \mod 11$
  Képes egy hibát ill. két szomszédos számjegy felcserélését jelezni.

- Paritásbit Legyen $\mathbf{u} \in \{0,1\}^k$. $\mathbf{u} \mapsto (u_1, \ldots, u_k, u_1 + u_2 + \cdots + u_k \mod 2)$.
  Képes egy hibát jelezni.

Kódszavak, Hamming-távolság

Mostantól karakter módosulás tipusú hibákra fókuszálunk!

Definíció

Legyen Σ egy véges halmaz (ábécé) és tekintsük az n hosszú szavak halmazát Σ^n. Ekkor a $\mathcal{C} \subset \Sigma^n$ részhalmaz egy kód, elemei a kódszavak.

Tipikusan $\Sigma = \mathbb{F}_2$ vagy általában $\mathbb{F}_{2^k}$.

Definíció

Legyen $\mathbf{u}, \mathbf{v} \in \Sigma^n$ két szó. A szavak Hamming-távolsága: $d(\mathbf{u}, \mathbf{v}) = \#\{i : u_i \neq v_i\}$.

Példa

- $d(000, 111) = 3$, $d(012, 210) = 2$
- $d(0000, 0001) = 1$, $d(0000, 0009) = 1$, $d(1234, 0123) = 4$
- általában: $0 \leq d(\mathbf{u}, \mathbf{v}) \leq n$

Hamming távolság

Hamming-távolság:
Legyen $\mathbf{u}, \mathbf{v} \in \Sigma^n$ két szó. A szavak Hamming-távolsága: $d(\mathbf{u}, \mathbf{v}) = \#\{i : u_i \neq v_i\}$.

A Hamming-távolság d valóban egy távolság-függvény:

Tétel

Legyen $d : \Sigma^n \times \Sigma^n \to \{0, 1, 2 \ldots\}$ a Hamming-távolság. Ekkor

- $d(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{u} = \mathbf{v}$.
- $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.
- $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{c}) + d(\mathbf{c}, \mathbf{v})$ (háromszög egyenlőtlenség).

Hamming távolság

Tétel:

- $d(\mathbf{u},\mathbf{v}) = 0 \iff \mathbf{u} = \mathbf{v}$.
- $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$.
- $d(\mathbf{u},\mathbf{v}) \leq d(\mathbf{u},\mathbf{c}) + d(\mathbf{c},\mathbf{v})$ (háromszög egyenlőtlenség).

Bizonyítás.
Az **első két tulajdonság** közvetlenül adódik a definícióból.

Háromszög egyenlőtlenség: a bizonyítás koordinátánként.

Terjesszük ki a d függvényt a koordinátákra: $u, v \in \Sigma$ esetén $d(u, v) = 0$ ha $u = v$ és $d(u, v) = 1$ ha $u \neq v$.

Ekkor $d(\mathbf{u},\mathbf{v}) = \sum_i d(u_i, v_i)$. Adott i-re,

- ha $u_i = v_i$, akkor $0 = d(u_i, v_i) \leq d(u_i, c_i) + d(c_i, v_i)$;
- ha $u_i \neq v_i$, akkor $c_i \neq u_i$ vagy $c_i \neq v_i$. Így $d(u_i, c_i) + d(c_i, v_i) \geq 1 = d(u_i, v_i)$.

$\square$

Kódtávolság

Definíció

Egy $\mathcal{C}$ kód **kódtávolsága** a kódszavak közti minimális távolság: $d(\mathcal{C}) = \min\{d(\mathbf{u}, \mathbf{v}) : \mathbf{u}, \mathbf{v} \in \mathcal{C}, \mathbf{u} \neq \mathbf{v}\}$.

Példa

- Az ismétlő kód ($0 \mapsto 000$, $1 \mapsto 111$) távolsága $d = 3$.
- Paritásbit ($\mathbf{u} \mapsto (u_1, \ldots, u_k, u_1 + u_2 + \cdots + u_k \bmod 2)$) távolsága $d = 2$.

Tétel (Biz.: HF)

Egy $\mathcal{C}$ kód $d = d(\mathcal{C})$ kódtávolsággal:

- $d - 1$ hibát tud **jelezni**;
- $t = \lfloor (d-1)/2 \rfloor$ hibát tud **javítani**.