Poznámky z přednášek
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

# Datové struktury I

Peter Černo, 2010
petercerno@gmail.com

**Garant**: doc. RNDr. Václav Koubek, DrSc.
**E-mail**: Vaclav.Koubek@mff.cuni.cz

**Anotace**: Přednáška navazuje na přednášky Algoritmy a datové struktury I a II a Programování I a II bakalářského studia. Bude věnována dvěma základním datovým strukturám, hašování a $(a, b)$-stromům (tato struktura se také nazývá $B$-stromy). Popisují se zde základní vlastnosti těchto struktur a jejich složitost. Na závěr přednášky se provede stručné zhodnocení třídicích algoritmů.

**Sylabus**:

1. Předpoklady: znalost a schopnost pracovat s uspořádanými množinami, stromy, průběh a odhady funkcí a řad, znalost základních termínů z pravděpodobnosti.

2. Míry složitosti – časová složitost, paměťová složitost, složitost v nejhorším případě, složitost v průměrném případě, amortizovaná složitost.

3. Hašování – řešení konfliktů, očekávané délky řetězců v jednotlivých typech hašování. Univerzální hašování, jeho vlastnosti a způsob jeho použití. Návrh perfektního hašování. Hašování na vnějších mediích.

4. Vyhledávání v uspořádaném poli. Unární, binární, interpolační a kvadratické vyhledávání.

5. $(a, b)$-stromy – operace MEMBER, INSERT, DELETE, JOIN, SPLIT, hledání $k$-tého prvku. $A$-sort a jeho použití. Paralelní implementace algoritmů (vyvažování shora dolů). Propojené $(a, b)$-stromy s prstem. Počet a rozdělení aktualizačních operací.

6. Binární vyhledávací stromy. Jejich vyvážené verze – červeno-černé stromy, AVL-stromy, $BB(\alpha)$-stromy. Operace MEMBER, INSERT, DELETE, JOIN, SPLIT a hledání $k$-tého prvku.

7. Haldy. $d$-regulární haldy, leftist haldy, binomiální haldy a jejich líná verze, Fibonacciho haldy. Operace INSERT, MIN, DELETEMIN, DECREASEKEY, INCREASEKEY, DELETE, MERGE, MAKEHEAP. Jejich aplikace.

8. Třídicí algoritmy. MERGESORT, HEAPSORT, QUICKSORT a jejich analýza. Verze BUCKETSORTU a HYBRIDSORTU a jejich analýza. Rozhodovací stromy. Srovnání algoritmů.

**Cíl předmětu**: Naučit základní datové struktury a algoritmy včetně teoretické analýzy jejich chování.

**Literatura**:


1. K. Mehlhorn: Data Structures and Algorithms 1: Sorting and Searching, Springer 1984

   `http://www.mpi-inf.mpg.de/~mehlhorn/DatAlgbooks.html`

2. J. S. Vitter, W.-Ch. Chen: Design and Analysis of Coalesced Hashing, Oxford Univ. Press, 1987

3. M. A. Weiss: Data Strucutres and Algorithm Analysis, The Benjamin/Cummings Publishing Company, Inc., Redwood City, California, 1992

This page is intentionally left blank.

10.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## HAŠOVANIE SO SEPAROVANÝMI REŤAZCAMI

UNIVERZUM $U$, $S \subseteq U$, $h: U \to \{0, \dots, m-1\}$

$i$-TÁ POLOŽKA = SPOJOVÝ ZOZNAM $S_i$ : $s \in S \mid h(s) = i$

REŤAZCE SÚ PROSTÉ!

PREDPOKLADÁME, ŽE $h(x)$ MOŽNO SPOČÍTAŤ V ČASE $O(1)$

NAJHORŠIA ČAS. ZLOŽITOSŤ $O(|S|)$, PAM. ZL.: $O(m + |S|)$

ODHAD OČAKÁVANEJ DĽŽKY REŤAZCA:

(1) $h$ JE RÝCHLO SPOČÍTATEĽNÁ

(2) $h$ ROZDEĽUJE $U$ ROVNOMERNE $\left(-1 \le |h^{-1}(i)| - |h^{-1}(j)| \le 1\right)$

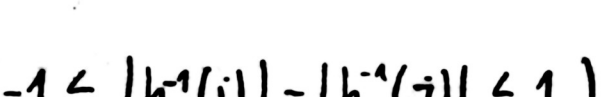
▽ (3) $S$ JE NÁHODNE VYBRANÁ Z $U$, PRESNEJŠIE Z $\binom{U}{n}$, $n = |S|$

○ (4) KAŽDÝ PRVOK Z $U$ MÁ ROVNAKÚ PRST. BYŤ ARGUMENTOM

OZNAČENIE: $n = |S|$, $m$ VEĽKOSŤ TAB., $\ell(i) = |S_i|$, FAKT. NAP. $\alpha = \frac{n}{m}$

PLATÍ $P(h(x) = i) = \frac{1}{m}$ PRE NÁHODNE ZVOLENÉ $x \in U$

$P(|S_i| = \ell) = \binom{n}{\ell} \left(\frac{1}{m}\right)^{\ell} \left(1 - \frac{1}{m}\right)^{n-\ell}$




ZA PREDPOKLADU NEKONEČNOSTI $U$ ...

AK $|U| = N$, POTOM PRE $\ell \le \min\left\{n, \frac{N}{m}\right\}$

$$P(\ell(i) = \ell) = \frac{\binom{N/m}{\ell}\binom{N - N/m}{n-\ell}}{\binom{N}{n}} =$$

$$= \frac{n!}{\ell!(n-\ell)!} \frac{\prod_{i=0}^{\ell-1}(N/m - i) \prod_{i=0}^{n-\ell-1}(N - N/m - i)}{\prod_{i=0}^{n-1}(N - i)} =$$

$$= \underbrace{\frac{n!}{\ell!(n-\ell)!} \frac{(N/m)^{\ell} \cdot (N - N/m)^{\ell}}{N^n}}_{= \binom{n}{\ell}\left(\frac{1}{m}\right)^{\ell}\left(1 - \frac{1}{m}\right)^{n-\ell}} \cdot \underbrace{\frac{\prod_{i=0}^{\ell-1}\left(1 - i\frac{m}{N}\right) \prod_{i=0}^{n-\ell-1}(\dots)}{\prod_{i=0}^{n-1}(1 - i/N)}}_{\to 1 \text{ PRE } N \gg n^2 m}$$

$= \binom{n}{\ell}\left(\frac{1}{m}\right)^{\ell}\left(1 - \frac{1}{m}\right)^{n-\ell}$ = DOBRÁ APROXIMÁCIA PRE KONEČNÉ $U$

$$E\ell = \sum_{\ell=0}^{n} \ell \cdot p_{n,\ell} = \sum_{\ell=0}^{n} \ell \cdot \underbrace{\binom{n}{\ell}\left(\frac{1}{m}\right)^{\ell}\left(1-\frac{1}{m}\right)^{n-\ell}}_{p_{n,\ell} = P(\ell(i)=\ell)} = \underline{\underline{\frac{n}{m}}} = \alpha$$

$$\operatorname{var} \ell = \underline{\underline{n \cdot \frac{1}{m}\left(1-\frac{1}{m}\right)}}$$

TOTIŽ $\underline{\underline{\ell \sim Bi(n, 1/m)}}$

ENP = OČAKÁVANÁ DĽŽKA <u>NAJHORŠIEHO REŤAZCA</u>

<u>TRIK</u>: $P(\max_i \ell(i) = j) = P(\max_i \ell(i) \geq j) - P(\max \ell(i) \geq j+1)$

$$ENP = \sum_j j P(\max_i \ell(i) = j) = \sum_{j=1}^{n} j P(\dots \geq j) - \sum_{j=1}^{n} j P(\dots \geq j+1)$$

$$\boxed{= \sum_j P(\max_i \ell(i) \geq j)}$$

$$P(\max_i \ell(i) \geq j) = P(\ell(1) \geq j \vee \ell(2) \geq j \vee \dots \vee \ell(m-1) \geq j)$$

$$\leq \sum_{i=0}^{m-1} P(\ell(i) \geq j) \leq m \cdot \binom{n}{j}\left(\frac{1}{m}\right)^{j} =$$

$$= \frac{\prod_{k=0}^{j-1}(n-k)}{j!} \cdot \left(\frac{1}{m}\right)^{j-1} \leq \boxed{n \cdot \left(\frac{n}{m}\right)^{j-1} \frac{1}{j!}}$$

PRE $\alpha \leq 1$ JE TO KLES. FCIA. OD $j$

$$\Rightarrow P(\max_i \ell(i) \geq j) \leq \min\left\{1, n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!}\right\}$$

ZA PREDPOKLADU $\alpha = \frac{n}{m} \leq 1$ ODHADNEME $j_0 = \min_j \left(n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!} \leq 1\right)$

DOKÁŽEME $j_0 = O\left(\frac{\log n}{\log \log n}\right)$.

$$\Rightarrow ENP = \sum_j P(\max_i \ell(i) \geq j) \leq \sum_j \min\left\{1, n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!}\right\} =$$

$$= \sum_{j=1}^{j_0} 1 + \sum_{j=j_0+1} \left(n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!}\right) \leq j_0 + \sum_{j=j_0+1}^{\infty} \frac{n}{j!} \leq$$

$$\leq j_0 + \frac{n}{j_0!}\sum_{j=j_0+1}^{\infty}\frac{j_0!}{j!} \rightsquigarrow \text{AK } \frac{n}{j_0!} \leq 1 \rightsquigarrow \leq j_0 + \sum_{j=j_0+1}^{\infty}\left(\frac{1}{j_0+1}\right)^{j-j_0} =$$

$$= j_0 + \frac{1}{j_0} = O(j_0)$$ .. AK $0.5 \leq \alpha \leq 1$ JE TO ZÁROVEŇ AJ DOLNÝ ODHAD

10.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

**ODHAD HODNOTY** $j_0 = \min_j \left( n \left(\frac{n}{m}\right)^{j-1} \frac{1}{j!} \le 1 \right)$

$(\alpha = \frac{n}{m} \le 1)$

(1) UKÁŽEME, ŽE PRE VEĽKÉ $n$ JE $j_0 \le \frac{8 \log n}{\log \log n}$

$\frac{n}{m} \le 1$, $\boxed{(j/2)^{j/2} \le j!}$ $\rightsquigarrow$ $n \left(\frac{n}{m}\right)^{j-1} \frac{1}{j!} \le \frac{n}{j!} \le \frac{n}{(j/2)^{j/2}}$ (2) (1)

$\Rightarrow j_0 \le \min_j \left( \frac{n}{(j/2)^{j/2}} \le 1 \right) = \boxed{\min_j \left( n \le (j/2)^{j/2} \right)}$ ← OZNAČME AKO $k+1$

$\boxed{\left(\frac{k}{2}\right)^{k/2} < n \le \left(\frac{k+1}{2}\right)^{\frac{k+1}{2}}} \Rightarrow \left(\frac{k}{2}\right) \log\left(\frac{k}{2}\right) < \log n \le \frac{k+1}{2} \log\left(\frac{k+1}{2}\right)$

$\log\left(\frac{k}{2}\right) + \log\log\left(\frac{k}{2}\right) < \log\log n \le \log\left(\frac{k+1}{2}\right) + \log\log\left(\frac{k+1}{2}\right)$

$\Rightarrow \log\log n < 2 \log\left(\frac{k+1}{2}\right)$

TAKŽE ZA PREDPOKLADU $k \ge 3$ JE:

$$\frac{\log n}{\log\log n} > \frac{k}{4} \frac{\log\left(\frac{k}{2}\right)}{\log\left(\frac{k+1}{2}\right)} > \frac{k}{8}$$

(2) **SOFISTIKOVANEJŠIA METÓDA**:

PRIAMO ODHADNEME $k_0 = \min_j \left( \frac{n}{j!} \le 1 \right) = \min_j \left( n \le j! \right)$

ZO STIRLINGOVEJ APROXIMÁCIE: $m! = \sqrt{2\pi m} \left(\frac{m}{e}\right)^m \left(1 + \frac{1}{12m} + O\left(\frac{1}{m^2}\right)\right)$

! PLYNIE $\log m! = (1 + o(1)) m \log m$

$\Rightarrow \boxed{\forall c > 1 \ \exists m_c : \forall m \ge m_c \quad \frac{1}{c} m \log m \le \log m! \le c m \log m}$

ZAFIXUJME $c > 1$.

Z DEFINÍCIE $k_0$: $(k_0-1)! < n \leq k_0!$

PRE DOSTATOČNE VEĽKÉ $n$ JE

$$\frac{1}{c}(k_0-1)\log(k_0-1) \leq \log(k_0-1)! < \log n \leq$$
$$\leq \log k_0! \leq c\,k_0 \log k_0$$

$$\Rightarrow \log(k_0-1) - \log c + \log\log(k_0-1) < \log\log n \leq$$
$$\leq \log k_0 + \log c + \log\log k_0$$

$c$ JE KONŠTANTA, $k_0$ RASTIE NAD VŠETKÝ MEDZE S RASTÚCIM $n$

$\Rightarrow \exists n_c \ \forall n \geq n_c$:

$$\boxed{\log(k_0-1) < \log\log n < \log k_0}$$

$$\frac{1}{c^2}(k_0-1) < \frac{1}{c}(k_0-1)\frac{\log(k_0-1)}{\log k_0} < \frac{\log n}{\log\log n} < c k_0 \frac{\log k_0}{\log(k_0-1)} < c^2 k_0$$

$$\Rightarrow k_0 = (1+o(1))\frac{\log n}{\log\log n} \Rightarrow \underline{\underline{j_0 = O\left(\frac{\log n}{\log\log n}\right)}}$$

<u>VETA</u>: ZA PREDPOKLADU $\alpha = \frac{n}{m} \leq 1$ JE PRI NAJMOVANÍ SO SEPAR. REŤAZCAMI HORNÝ ODHAD OČAKÁVANEJ DĽŽKY MAXIMÁLNEHO REŤAZCA $O\left(\frac{\log n}{\log\log n}\right)$. KEĎ $0{,}5 \leq \alpha \leq 1$, JE TO ZÁROVEŇ AJ DOLNÝ ODHAD.

10.02.2009 $T_4$
PETER ČERNO

# NTINOO6 DATOVÉ STRUKTÚRY I

## OČAKÁVANÝ POČET TESTOV

TEST = POROVNANIE ARGUMENTU OPERÁCIE S PRVKOM NA DANOM MIESTE REŤAZCA, POPR. ZISTENIE, ŽE VYŠETROVANÝ REŤAZEC JE PRÁZDNY

ÚSPEŠNÉ VYHĽADÁVANIE = HĽADANÝ PRVOK $\in S$
NEÚSPEŠNÉ VYHĽADÁVANIE = HĽADANÝ PRVOK $\notin S$

OČAKÁVANÝ POČET TESTOV PRI NEÚSP. VYHĽ.

$$ET = \underbrace{P(\ell(i)=0)}_{} + \sum_{\ell} \ell \cdot P(\ell(i)=\ell) = \left(1-\frac{1}{m}\right)^n + \frac{n}{m} \approx \underline{\underline{e^{-\alpha} + \alpha}}$$

ZISTENIE, ŽE DANÝ REŤAZEC JE PRÁZDNY = 1 TEST

## OČAKÁVANÝ POČET TESTOV PRI ÚSPEŠNOM VYHĽ.

MAJME REŤAZEC DĽŽKY $\ell$. POČET TESTOV PRI VYHĽ. VŠETKÝCH PRVKOV JE $1+2+\ldots+\ell = \binom{\ell+1}{2}$

OČAKÁVANÝ POČET TESTOV PRI VYHĽADANÍ VŠETKÝCH PRVKOV V DANOM REŤAZCI JE $\sum_{\ell} \binom{\ell+1}{2} P(\ell(i)=\ell)$

$\Rightarrow$ OČ. POČET TESTOV PRI VYHĽ. $\forall$ PRVKOV V TAB. JE $m\times\ldots$

$\Rightarrow$ OČAKÁVANÝ POČET TESTOV PRE 1 PRVOK JE

$$\frac{m}{n}\sum_{\ell=0}^{n}\binom{\ell+1}{2} P_{n,\ell} = \ldots = \frac{n-1}{2m} + 1 \approx \underline{\underline{1+\frac{\alpha}{2}}}$$

## ŠPECIÁLNA METÓDA

PREDPOKLAD: TESTY PRI ÚSPEŠNOM VYHĽADÁVANÍ $s \in S$ SÚ (AŽ NA TEN POSLEDNÝ) TOTOŽNÉ S TESTAMI, KTORÉ BOLI VYKONANÉ PRI NEÚSPEŠNOM VYHĽ. $s$ V OKAMŽIKU, KEĎ SA TENTO PRVOK UKLADAL DO TABUĽKY

OČAKÁVANÝ POČET TESTOV PRI ÚSPEŠNOM VYHĽADÁVANÍ x
= 1 + OČAKÁVANÁ DĽŽKA RETAZCA V OKAMŽIKU VKLADANIA x :

MÁME n PRVKOV; PRVOK x BOL DO TABUĽKY VLOŽENÝ V i-TOM KROKU ($i = 0, \dots, n-1$, UDÁVA POČET PRVKOV PRED VLOŽENÍM x) S PRAVDEPODOBNOSTOU $\frac{1}{n}$

V i-TOM KROKU (PRED VLOŽENÍM x) JE OČAKÁVANÝ POČET TESTOV PRI NEÚSPEŠNOM VYHĽADÁVANÍ x = $i/m$

$\Rightarrow$ OČAKÁVANÝ POČET TESTOV PRI ÚSPEŠNOM VYHĽ. x JE

$$\sum_{i=0}^{n-1} \frac{1}{n} \cdot \left(1 + \frac{i}{m}\right) = \underline{\underline{1 + \frac{n-1}{2m}}}$$

VŠIMNIME SI, ŽE OČAKÁVANÝ POČET TESTOV PRI NEÚSP. VYHĽ. JE MENŠÍ NEŽ OČAKÁVANÝ POČET TESTOV PRI ÚSP. VYHĽ. AK $\alpha \leq 1$. DÔVOD: POČET TESTOV PRI ÚSP. VYHĽ. POČÍTAME AKO PRIEMER CEZ n, KDEŽTO PRI NEÚSP. CEZ m.

## HAŠOVANIE S USPORIADANÝMI RETAZCAMI

PRVKY V RETAZCOCH SÚ USPORIADANÉ

POČET OČAKÁVANÝCH TESTOV PRI ÚSPEŠNOM VYHĽADÁVANÍ JE ROVNAKÝ AKO PRI NEUSPOR. RET. = $\underline{\underline{1 + \frac{\alpha}{2}}}$

PRI NEÚSP. VYHĽ. KONČÍME SKOREJ =

$$\underline{\underline{e^{-\alpha} + 1 + \frac{\alpha}{2} - \frac{1}{\alpha}\left(1 - e^{-\alpha}\right)}}$$

(BEZ DÔKAZU)

NEVÝHODA SEP. RETAZCOV: NEEFEKTÍVNE VYUŽITIE PAMÄTI

10.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## HAŠOVANIE S PRENIESTNOVANÍM

REŤAZCE SÚ IMPLEMENTOVANÉ AKO DVOJSMERNÉ ZOZNAMY A UKAZATELIA NA PREDCH. A NÁSL. PRVOK SÚ URČENÍ POLOŽKAMI PRE PRÁCU S TABUĽKOU (next, previous)

PRI KOLÍZII PRENIESTNIME KOLIDUJÚCI PRVOK Z INÉHO REŤAZCA NA INÝ VOĽNÝ RIADOK V TABUĽKE

```
MEMBER (x)
i := h(x)
if (i.previous ≠ NIL) ∨ i.key = NIL then
    return x ∉ S
while i.next ≠ NIL & i.key ≠ x do i ← i.next
if i.key = x then return x ∈ S else return x ∉ S
```

NA RIADKU i SA NACHÁDZA PRVOK Z INÉHO REŤAZCA

OČAKÁVANÝ POČET TESTOV JE ROVNAKÝ AKO PRI HAŠOVANÍ SO SEPAROVANÝMI REŤAZCAMI

## HAŠOVANIE S DVOMA UKAZATEĽMI

NEVÝHODA HAŠOVANIA S PRENIESTNOVANÍM JE, ŽE V OPERÁCIACH INSERT A DELETE – PRENIESTNENIE PRVKOV V TABUĽKE JE ČASOVO NÁROČNEJŠIE NEŽ VYKONANIE TESTU.

REŤAZEC JE IMPLEMENTOVANÝ AKO JEDNOSMERNÝ ZOZNAM, KTORÝ NEMUSÍ ZAČÍNAŤ NA SVOJOM MIESTE

TJ. $S_j = \{s \in S \mid h(s) = j\}$ NEMUSÍ ZAČÍNAŤ NA j-TOM RIADKU

NAMIESTO previous MÁME begin → KDE ZAČÍNA REŤAZEC PRÍSLUŠNÝ DANÉMU RIADKU ...

!

POLOŽKA begin V j-TOM RIADKU JE NEPLNENÁ
$\Leftrightarrow$ REPREZENTOVANÁ MNOŽINA S OBSAHUJE
$s \in S : h(s) = j$.

!

```
MEMBER (x)
i := h(x)
if i.begin = NIL then return x ∉ S else i ← i.begin
while i.next ≠ NIL & i.key ≠ x do i ← i.next
if i.key = x then return x ∈ S else return x ∉ S
```

JE RÝCHLEJŠÍ AKO HAŠOVANIE S PREDIESTNOVANÍM, AVŠAK ZAČIATOK REŤAZCA V INOM MIESTE TABUĽKY PRIDÁVA NAVIAC JEDEN TEST

E #TESTOV V ÚSP. VYHĽ.: $1 + \frac{(n-1)(n-2)}{6m^2} + \frac{n-1}{2m} \approx 1 + \frac{\alpha^2}{6} + \frac{\alpha}{2}$

E #TESTOV V NEÚSP. VYHĽ.: $1 + \frac{\alpha^2}{2} + \alpha + e^{-\alpha}(2+\alpha) - 2$

## ZRASTAJÚCE HAŠOVANIE

POUŽÍVA IBA UKAZATEĽ JEDNOSMERNÉHO SPOJOVÉHO ZOZNAMU, REŤAZCE NIE SÚ SEPAROVANÉ, ALE ZRASTAJÚ
KEĎ MÁME PRIDAŤ PRVOK s, TAK HO ZARADÍME DO REŤAZCA, KT. OBSAHUJE h(s)-TÝ RIADOK TABUĽKY

(1) ŠTANDARDNÉ ZRASTAJÚCE HAŠOVANIE BEZ POMOCNEJ PAMÄTI: LISCH, EISCH

(2) ZRASTAJÚCE HAŠOVANIE S POMOCNOU PAMÄŤOU:
LICH, VICH, EICH

10.02.2003 T4
PETER ČERNO

# NTIN 066 DATOVÉ STRUKTÚRY I

LISCH = LATE-INSERTION STANDARD COALESCED HASHING
PRIDÁVA ZA POSLEDNÝ PRVOK REŤAZCA

EISCH = EARLY-INSERTION STANDARD COALESCED HASHING
PRIDÁVA ZA PRVÝ PRVOK REŤAZCA

PRVKY $s \in S$ TAKÉ, ŽE $h(s) = i$ SÚ UMIESTNENÉ V REŤAZCI OD i-TEHO RIADKU (AJ KEĎ REŤAZEC MÔŽE ZAČÍNAŤ SKÔR)

PRIRODZENÁ EFEKTÍVNA METÓDA DELETE PRE STANDARDNÉ ZRASTAJÚCE HAŠOVANIE NIE JE ZNÁMA

HLAVNÝ PROBLÉM, KTORÝ OBMEDZUJE ALGORITMY PRE OPERÁCIU DELETE JE, AKO ZACHOVÁVALI NÁHODNÉ USPORIADANIE PRVKOV V REŤAZCOCH...

PRI ÚSPEŠNOM VYHĽADÁVANÍ JE METÓDA EISCH RÝCHLEJŠIA NEŽ METÓDA LISCH.
(PRÁCA S NOVÝM PRVKOM JE PRAVDEPODOBNEJŠIA)
OČAKÁVANÝ POČET TESTOV PRE NEÚSPEŠNÉ VYHĽADÁVANIE JE PRE OBE METÓDY ROVNAKÝ.

## ANALÝZA ZLOŽITOSTI

MÁME ULOŽENÚ $S = \{s_1, \dots, s_n\} \subseteq U$ DO TABUĽKY VEĽKOSTI $m$ A JE DANÝ PRVOK $s_{n+1}$.
NAŠOU ÚLOHOU JE ZISTIŤ, ČI $s_{n+1} \in S$
OZNAČME $a_i = h(s_i)$ $\forall i = 1, \dots, n+1$
(1) VŠETKY POSTUPNOSTI ADRIES $a_1, \dots, a_{n+1}$ SÚ ROVNAKO PRAV.
(2) DETERMINISTICKÝ SPÔSOB VYBERANIA PRÁZDNYCH RIADKOV

# NEÚSPEŠNÉ VYHĽADÁVANIE

$C(a_1, \dots, a_n ; a_{n+1})$ = POČET TESTOV PRI NEÚSP. V. V $S$

OČAKÁVANÝ POČET TESTOV JE: $\frac{\sum C(a_1, \dots, a_n ; a_{n+1})}{m^{n+1}}$

<u>REŤAZEC DĽŽKY $l$</u> V MNOŽINE $S$ JE MAXIMÁLNA POSTUPNOSŤ ADRIES $(b_1, b_2, \dots, b_l)$ TAKÁ, ŽE $b_i.next = b_{i+1}$ $\forall i = 1..l-1$

KEĎ ADRESA $a_{n+1}$ JE $i$-TÝ PRVOK V REŤAZCI, POTOM POČET VYKONANÝCH TESTOV JE $l-i+1$

UVAŽUJME ĽUBOVOĽNÚ, ALE PEVNE ZVOLENÚ REPREZENTÁCIU $(a_1, a_2, \dots, a_n) = \vec{a}$

TÁ NÁM JEDNOZNAČNE URČUJE MNOŽINU DISJUNKTNÝCH REŤAZCOV $R_1, \dots, R_k$ DĽŽOK $l_1, \dots, l_k$ V TABUĽKE + $c$ PRÁZDNYCH RIADKOV

SKÚSIME JEDNOTLIVÉ $a_{n+1} = 0, \dots, m-1$:

KAŽDÉ JEDNO TAKÉ $a_{n+1}$ SA BUĎ TRAFÍ DO PRÁZDNEHO RIADKU $\rightarrow$ =1 TEST, ALEBO NIEKAM DO NEJAKÉHO REŤAZCA (VŽDY NIEKAM INAM + VŠETKY PRVKY VŠETKÝCH REŤAZCOV BUDÚ ZASIAHNUTÉ PRÁVE 1x + KAŽDÉ PRÁZDNE POLÍČKO BUDE TIEŽ ZASIAHNUTÉ PRÁVE 1x)

$$\Rightarrow \sum_{a_{n+1}=0}^{m-1} C(a_1, \dots, a_n ; a_{n+1}) = c + \binom{l_1+1}{2} + \binom{l_2+1}{2} + \dots + \binom{l_k+1}{2}$$

PLATÍ
$l_1 + l_2 + \dots + l_k = n$
$n + c = m \Rightarrow c = m - n$

OZNAČME $c_{\vec{a}}(l)$ POČET REŤAZCOV DĽŽKY $l$ V REPR. $\vec{a}$

POTOM: $\sum_{a_{n+1}=0}^{m-1} C(a_1, \dots, a_n ; a_{n+1}) =$

$$\underbrace{c_{\vec{a}}(0)}_{=c} + \sum_{l=1}^{n} \binom{l+1}{2} c_{\vec{a}}(l)$$

PLATÍ
$\sum_{i=1}^{n} l c_{\vec{a}}(l) = l_1 + \dots + l_k = n$, $c_{\vec{a}}(0) = m - n$

10.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

OZNAČME $c_n(\ell) = \sum_{\vec{a}} c_{\vec{a}}(\ell)$ = POČET VŠETKÝCH REŤAZCOV DĹŽKY $\ell$ VO VŠETKÝCH REPREZENTÁCIACH $\vec{a} = (a_1, \dots, a_n)$ $n$-PRVKOVÝCH MNOŽÍN

$c_n(0)$ = POČET PRÁZDNYCH RIADKOV VO $\forall$ REPR. $\vec{a}$

POTOM $\sum C(a_1, \dots, a_n; a_{n+1}) = c_n(0) + \sum_{\ell=1}^{n} \binom{\ell+1}{2} c_n(\ell) =$

$$= c_n(0) + \underline{\underline{\sum_{\ell=1}^{n} \ell\, c_n(\ell)}} + \sum_{\ell=1}^{n} \binom{\ell}{2} c_n(\ell) \qquad (\underline{\text{TRIK}})$$

REPREZENTÁCIA S MÁ $m-n$ PRÁZDNYCH RIADKOV (VŽDY)
MÁME $m^n$ RÔZNYCH REPREZENTÁCIÍ

$$\Rightarrow \underline{\underline{c_n(0) = (m-n) \cdot m^n}}$$

KEĎŽE $\sum_{\ell=1}^{n} \ell \cdot c_{\vec{a}}(\ell) = n$, NUTNE $\underline{\underline{\sum_{\ell=1}^{n} \ell \cdot c_n(\ell) = n \cdot m^n}}$

CHCEME SPOČÍTAŤ $\underline{S_n = \sum_{\ell=1}^{n} \binom{\ell}{2} c_n(\ell)}$

KEĎ PRIDÁME V REPREZENTÁCII $\vec{a}$ NOVÝ PRVOK S ADRESOU $a_{n+1}$, POTOM REŤAZEC DĹŽKY $\ell$ SA NEZVÄČŠÍ, KEĎ $a_{n+1}$ NELEŽÍ V TOMTO REŤAZCI, V OPAČNOM PRÍPADE DĹŽKA NARASTIE NA $\ell+1$

UVAŽUJME REŤAZEC $R$ DĹŽKY $\ell$ V REPREZENTÁCII $\vec{a}$
KEĎ SKÚŠAME PRIDAŤ $a_{n+1} = 0, 1, \dots, m-1$

TAK V $m-\ell$ REPREZENTÁCIACH $\vec{a}, a_{n+1}$ ZOSTANE TENTO REŤAZEC ZACHOVANÝ

A V $\ell$ REPREZENTÁCIACH DOSTANEME VŽDY NOVÝ $R'$ DĹŽKY $\ell+1$

$$\Rightarrow \underline{\underline{\sum_{a_{n+1}=0}^{m-1} c_{\vec{a}, a_{n+1}}(\ell) = (m-\ell)\, c_{\vec{a}}(\ell) + (\ell-1)\, c_{\vec{a}}(\ell-1)}}$$

KEĎ TO NASČÍTAME CEZ VŠETKY $\vec{\alpha}$, POTOM:

$$c_{n+1}(\ell) = (m-\ell)\, c_n(\ell) + (\ell-1)\, c_n(\ell-1) \qquad \text{PRE } \ell \geq 1$$

POZOR, $c_{n+1}(0) = c_n(0) - 1$.

$$S_n = \sum_{\ell=1}^{n} \binom{\ell}{2} c_n(\ell) =$$

$$\left[\sum_{\ell=1}^{n} \binom{\ell}{2} (m-\ell)\, c_{n-1}(\ell)\right] + \left[\sum_{\ell=0}^{n-1} \binom{\ell+1}{2} \ell\, c_{n-1}(\ell)\right] =$$

$$= \left[\cancel{\binom{n}{2}(m-n)\, c_{n-1}(n)}\right] + \left[\sum_{\ell=1}^{n-1} \left(\binom{\ell}{2}(m-\ell) + \binom{\ell+1}{2}\ell\right) c_{n-1}(\ell)\right] +$$

$$\binom{\ell}{2} m - \cancel{\binom{\ell}{2}\ell} + \cancel{\binom{\ell}{2}\ell} + \ell \cdot \ell \to \frac{\ell\cdot(\ell-1)}{2} \cdot \frac{2\cdot(\ell-1)+2}{\ell-1}$$

$$= \qquad + \left[\cancel{\binom{1}{2} 0\, c_{n-1}(0)}\right] = \left[\sum_{\ell=1}^{n-1} \left(\binom{\ell}{2}(m+2) + \ell\right) c_{n-1}(\ell)\right] =$$

$$= (m+2)\, S_{n-1} + (n-1)\, m^{n-1}$$

$$S_n = (m+2)\, S_{n-1} + (n-1)\, m^{n-1} = (m+2)\left[(m+2)\, S_{n-2} + (n-2)\, m^{n-2}\right] + (n-1)\, m^{n-1} =$$

$$= (m+2)^2 S_{n-2} + (m+2)(n-2)\, m^{n-2} + (n-1)\, m^{n-1} =$$

$$= (m+2)^3 S_{n-2} + (m+2)^2 (n-3)\, m^{n-3} + (m+2)(n-2)\, m^{n-2} + (n-1)\, m^{n-1} =$$

$$= \cancel{(m+2)^{n-1} S_0} + \sum_{i=0}^{n-1} (m+2)^i (n-1-i)\, m^{n-1-i} =$$

$$= (m+2)^{n-1} \sum_{i=0}^{n-1} (n-1-i) \left(\frac{m}{m+2}\right)^{n-1-i} =$$

$$\boxed{= (m+2)^{n-1} \sum_{i=1}^{n-1} i \left(\frac{m}{m+2}\right)^{i}}$$

11. 02. 2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

OZNAČME: $T_c^n = \sum_{i=1}^{n} i c^i$, $c \neq 1$

$$(c-1)T_c^n = cT_c^n - T_c^n = \left[\sum_{i=2}^{n+1} (i-1) c^i\right] - \left[\sum_{i=1}^{n} i c^i\right] =$$ TRIK

$$= nc^{n+1} + \left[\sum_{i=2}^{n} [(i-1)-i] c^i\right] - c =$$

$$= nc^{n+1} - \left[\sum_{i=0}^{n} c^i\right] + 1 = nc^{n+1} - \frac{c^{n+1}-1}{c-1} + 1 =$$

$$= nc^{n+1} - \frac{c^{n+1}-c}{c-1} \Rightarrow \boxed{\frac{nc^{n+2} - (n+1)c^{n+1} + c}{(c-1)^2} = T_c^n}$$

$$\Rightarrow S_n = (m+2)^{n-1} \cdot \frac{(n-1)\left(\frac{m}{m+2}\right)^{n+1} - n\left(\frac{m}{m+2}\right)^n + \left(\frac{m}{m+2}\right)}{\left(\frac{m}{m+2} - 1\right)^2} =$$

$$= \frac{1}{4}(m+2)^{n+1}\left[(n-1)\left(\frac{m}{m+2}\right)^{n+1} - n\left(\frac{m}{m+2}\right)^n + \left(\frac{m}{m+2}\right)\right] =$$

$$= \frac{1}{4}\left[(n-1)m^{n+1} - n(m+2)m^n + m(m+2)^n\right] =$$

$$= \frac{1}{4}\left[m(m+2)^n - m^{n+1} - 2nm^n\right]$$

$\Rightarrow$ OČAKÁVANÝ POČET TESTOV PRI NEÚSPEŠNOM VYHĽ.:

$$\frac{(m-n)m^n + nm^n + \frac{1}{4}\left[m(m+2)^n - m^{n+1} - 2nm^n\right]}{m^{n+1}} =$$

$$= \frac{m^{n+1} + \frac{1}{4}\left[m(m+2)^n - m^{n+1} - 2nm^n\right]}{m^{n+1}} =$$

$$= 1 + \frac{1}{4}\left[\left(1+\frac{2}{m}\right)^n - 1 - \frac{2n}{m}\right] \sim \boxed{1 + \frac{1}{4}\left[e^{2\alpha} - 1 - 2\alpha\right]}$$

# OČAKÁVANÝ POČET TESTOV PRI ÚSPEŠNOM VYHĽADÁVANÍ

V MODELI LISCH MÔŽEME POUŽIŤ ŠPECIÁLNU METÓDU

PRI VYHĽADÁVANÍ PRVKU $s_{n+1} \in S$ JE POČET TESTOV ROVNÝ 1 + POČET POROVNANÍ KĽÚČOV PRI OP. INSERT($s_{n+1}$)

a) $s_{n+1}$ JE VLOŽENÝ NA MIESTO $h(s_{n+1})$ $\rightsquigarrow$ 1 TEST

b) $s_{n+1}$ JE VLOŽENÝ NA $i$-TE MIESTO REŤAZCA DĺŽKY $\ell$ $\rightsquigarrow$ $\ell - i + 1 + 1$ TESTOV

OČAKÁVANÝ POČET POROVNANÍ KĽÚČOV PRI NEÚSP. VYHĽ. JE

$$\frac{1}{m^{n+1}} \sum_{\ell=1}^{n} \binom{\ell+1}{2} c_n(\ell) = \frac{1}{m^{n+1}} \left[ n m^n + \frac{1}{4} \left( m(m+2)^n - m^{n+1} - 2n m^n \right) \right] =$$

$$= \frac{1}{4} \left[ \left(1 + \frac{2}{m}\right)^n - 1 + \frac{2n}{m} \right]$$

ZVOLÍME $x \in S$. PRVOK $x$ BOL DO TABUĽKY VLOŽENÝ V $i$-TOM KROKU ($i = 0, 1, \dots, n-1$ = UDÁVA POČET PRVKOV V TAB. TESNE PRED VLOŽENÍM $x$) S PRAVDEPODOBNOSŤOU $\frac{1}{n}$

OČAKÁVANÝ POČET POROVNANÍ KĽÚČOV PRI VKLADANÍ $x$ V $i$-TOM KROKU JE $\frac{1}{4}\left[\left(1+\frac{2}{m}\right)^i - 1 + \frac{2i}{m}\right] \Rightarrow$ POČET TESTOV $= +1$

STREDNÁ HODNOTA CEZ VŠETKY $x \in S$ JE POTOM:

$$\frac{1}{n} \sum_{i=0}^{n-1} \left[ 1 + \frac{1}{4} \left[ \left(1 + \frac{2}{m}\right)^i - 1 + \frac{2i}{m} \right] \right] = \dots$$

$$\dots = 1 + \frac{m}{8n} \left[ \left(1 + \frac{2}{m}\right)^n - 1 - \frac{2n}{m} \right] + \frac{n-1}{4m} \sim$$

$$\sim \boxed{1 + \frac{1}{8\alpha}\left(e^{2\alpha} - 1 - 2\alpha\right) + \frac{\alpha}{4}}$$

PRE METÓDU EICH JE TO: $\frac{m}{n}\left[\left(1+\frac{1}{m}\right)^n - 1\right] \sim \boxed{\frac{1}{\alpha}\left(e^{\alpha} - 1\right)}$

11.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## METÓDY LICH, EICH, VICH

PAMÄŤ ROZDELENÁ NA 2 ČASTI:
(a) PRIAMO ADRESOVATEĽNÁ (m RIADKOV, PRÍSTUP. CEZ HASH. F.)
(b) POMOCNÁ (NEMÁME PRÍSTUP CEZ HASH. FCIU.)

KEĎ PRI PRIDÁVANÍ NOVÉHO PRVKU VZNIKNE KOLÍZIA, TAK SA NAJSKÔR VYBERIE VOĽNÝ RIADOK Z POMOCNEJ ČASTI .. AŽ PO ZAPLNENÍ POMOCNEJ ČASTI SA POUŽIJE ADRESOV. PAM.
⇝ ODDÁLUJEME ZRASTANIE REŤAZCOV

LICH = LATE-INSERTION COALESCED HASHING
EICH = EARLY-INSERTION COALESCED HASHING
VICH = VARIED-INSERTION COALESCED HASHING

LICH VKLADÁ VŽDY NA KONIEC REŤAZCA
EICH BUĎ NA $h(x)$ (AK PRÁZDNY), ALEBO TESNE ZA $h(x)$
VICH ZA POSLEDNÝ PRVOK, KT. JE EŠTE V POMOCNEJ ČASTI. AK REŤAZEC NEOBSAHUJE ŽIADNY RIADOK Z POM. ČASTI, POUŽIJE EARLY-INSERTION

TIETO METÓDY NEPODPORUJÚ PRIRODZENÉ EFEKTÍVNE ALGORITMY PRE OPERÁCIU DELETE.

ZLOŽITOSŤ: $n = |S|$, $m = |$ADRES. ČASŤ$|$, $m' = |$CELÁ TABUĽKA$|$,
$\alpha = n/m'$ FAKTOR ZAPLNENIA, $\beta = m/m'$ ADRESOVACÍ FAKTOR
$\lambda$ = JEDINÉ NEZÁPORNÉ RIEŠENIE $e^{-\lambda} + \lambda = 1/\beta$

ROZLIŠUJEME VŽDY 2 PRÍPADY: $\alpha \leq \lambda\beta$, $\alpha \geq \lambda\beta$
(1) $\alpha \leq \lambda\beta$ .. NIE JE ZAPLNENÁ POMOCNÁ ČASŤ PAMÄTI
NEÚSPEŠNÝ PRÍPAD: $e^{-\alpha/\beta} + \frac{1}{\beta}$
ÚSPEŠNÝ PRÍPAD: $1 + \alpha/2\beta$
PODOBNE AKO PRI SEPAROVANÝCH REŤAZCOCH

RIADOK TAB. MÁ IBA key

# HAŠOVANIE S LINEÁRNYM PRIDÁVANÍM

PRI KOLÍZII NÁJDEME PRVÝ NASLEDUJÚCI VOĽNÝ RIADOK.
METÓDA NEPODPORUJE EFEKTÍVNU IMPL. OPERÁCIE DELETE.

VETA: OČAKÁVANÝ POČET TESTOV JE
V NEÚSPEŠNOM PRÍPADE CCA: $\frac{1}{2}\left(1+\left(\frac{1}{1-\alpha}\right)^2\right)$
V ÚSPEŠNOM PRÍPADE CCA: $\frac{1}{2}\left(1+\frac{1}{1-\alpha}\right)$.

# DVOJITÉ HAŠOVANIE

VÝBER ĎALŠÍCH RIADKOV JE ZÁVISLÝ NA VKLADANOM PRVKU
(IDEÁL = PRE RÔZNE PRVKY PRI KOLÍZII ABY BOLI POSTUPN. ADRESY, NA KT. SA HĽADÁ VOĽNÝ RIADOK, RÔZNE)

$\boxed{\text{DVE HASH. FCIE } h_1, h_2}$

INSERT(x): AK x NIE JE V TABUĽKE, NÁJDE SA NAJMENŠIE $i = 0, 1, \ldots$ TAKÉ, ŽE $\boxed{(h_1(x) + i\, h_2(x)) \bmod m}$ JE PRÁZDNY RIADOK, A TAM SA PRVOK x VLOŽÍ.
$\forall x$ $h_2(x)$ A m MUSIA BYŤ NESÚDELITEĽNÉ.

PREDPOKLAD: $\forall x \in U$: $\{h_1(x) + i\, h_2(x) \bmod m\}_{i=0}^{m-1}$
JE NÁHODNÁ PERMUTÁCIA RIADKOV TABUĽKY
(V PRAXI NIE JE MOŽNÉ ÚPLNE ZREALIZOVAŤ)

METÓDA NEPODPORUJE OPERÁCIU DELETE.

PROBLÉM: DEFINOVANIE FUNKCIE $h_2$

11.02.2009 We
PETER ČERNO

# NTINO66 DATOVÉ STRUKTÚRY I

## VYHĽADÁVANIE V NEÚSPEŠNOM PRÍPADE

$\boxed{q_i(n,m)}$ := PRAVDEPODOBNOSŤ, ŽE KEĎ MÁ TAB. $m$ RIADKOV, $n$ JE OBSADENÝCH, POTOM $h_1(x)+j\,h_2(x)$ PRE $j=0,\dots,i-1$ SÚ OBSADENÉ

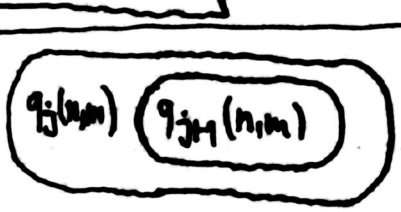
!

$q_0(n,m)=1$, $q_1(n,m)=\frac{n}{m}$, $q_2(n,m)=\frac{n(n-1)}{m(m-1)}$

TD. $q_i(n,m)=\prod_{j=0}^{i-1}\frac{n-j}{m-j}$ $\quad i\geq 0,\ n,m\geq i$

$\Rightarrow q_j(n,m)=\frac{n}{m}\,q_{j-1}(n-1,m-1)$ $\quad j,n>0,\ m>1$

$\boxed{C(n,m)}$ := OČAKÁVANÝ POČET TESTOV PRI NEÚSP. VYHĽ.




$q_j(n,m)-q_{j+1}(n,m)=q_j(n,m)\left(1-\frac{n-j}{m-j}\right)=$
PRAVD., ŽE $h_1(x)+i\,h_2(x)$ SÚ OBSADENÉ PRE $i=0,1,\dots,j-1$ A PRE $i=j$ JE VOĽNÉ.

$\Rightarrow$ ZREJME $C(n,m)=\sum_{j=0}^{n}(j+1)\left(q_j(n,m)-q_{j+1}(n,m)\right)=$

$=\sum_{j=0}^{n}q_j(n,m)$

ZREJME $C(0,m)=1$

$C(n,m)=\sum_{j=0}^{n}q_j(n,m)=q_0(n,m)+\sum_{j=1}^{n}\frac{n}{m}\,q_{j-1}(n-1,m-1)=$

$=1+\frac{n}{m}\sum_{j=0}^{n-1}q_j(n-1,m-1)=1+\frac{n}{m}\,C(n-1,m-1)$

INDUKCIOU: $\boxed{C(n,m)=\frac{m+1}{m-n+1}}$

# VYHĽADÁVANIE V ÚSPEŠNOM PRÍPADE

POČET TESTOV PRI VYHĽADANÍ $x \in S$ =
POČET TESTOV PRI VKLADANÍ $x \in S$
($x$ BOLO VLOŽENÉ V $i$-TOM KROKU S PRST. $\frac{1}{n}$, $i = 0, \dots, n-1$)

$\Rightarrow$ OČAKÁVANÝ POČET TESTOV PRI ÚSP. VYHĽ.

$$= \frac{1}{n} \sum_{i=0}^{n-1} C(i, m) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{m+1}{m-i+1} =$$

$$= \frac{m+1}{n} \left[ \sum_{j=1}^{m+1} \frac{1}{j} - \sum_{j=1}^{m-n+1} \frac{1}{j} \right] \approx$$

$$\frac{1}{\alpha} \ln\left(\frac{m+1}{m-n+1}\right) \approx \boxed{\frac{1}{\alpha} \ln\left(\frac{1}{1-\alpha}\right)}$$

## POROVNANIE

| NEÚSPEŠNÉ VYHĽ. | ÚSPEŠNÉ VYHĽ. |
|---|---|
| USP. SEPAR. RETAZCE | USP. SEP. R. = SEP. R. = PRENIEST. |
| SEPAR. RETAZCE = H.S PRENIEST. | S DVOMA UKAZATEĽMI |
| H. S DVOMA UKAZ. | VICH |
| VICH = LICH | LICH |
| EICH | EICH |
| LISCH = EISCH | EISCH |
| DVOJITÉ HASH. | LISCH |
| H. S LIN. PRIDÁVANÍM | DVOJITÉ HASH. |
| | H. S LIN. PRIDÁVANÍM |

PRE $\alpha < 0{,}7$ JE VICH LEPŠÍ AKO H. S DVOMA UKAZ.
H. S LIN. PRIDÁVANÍM VHODNÉ IBA PRE $\alpha < 0{,}7$
DVOJITÉ HASH. PRE $\alpha < 0{,}9$.
PRE ZRASTAJÚCE HASH. SA ODPORÚČA VZIAŤ $\beta = \frac{m}{m'} = 0{,}86$
DALŠIE OTÁZKY $\rightarrow$ VIĎ SKRIPTÁ !!!

11.02.2009 lk
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## UNIVERZÁLNE HAŠOVANIE

IDEA: MÁME MNOŽINU $H$ HAŠ. FCIÍ. $U \to \{0, \dots, m-1\}$ T.Ž. $\forall S \subseteq U, |S| \leq m$ SA VÄČŠINA FCIÍ Z $H$ CHOVÁ DOBRE VOČI $S$ (TJ. MÁLO KOLÍZIÍ)

FCIE. $h \in H$ RÝCHLO SPOČÍTATEĽNÉ $\rightsquigarrow$ ZADANÉ ANALYTICKY $\rightsquigarrow$ PROBLÉM PRI URČOVANÍ $|H|$

RIEŠENIE: NAMIESTO S FCIAMI PRACUJEME S ICH INDEXMI

$$H = \{h_i \mid i \in I\}$$

PRI APLIKÁCII METÓDY VYBERÁME PRVOK $i \in I$ (S ROVN. ROZD.)

FORMÁLNE: NECH $U$ JE UNIVERZUM. SYSTÉM FUNKCIÍ $H = \{h_i \mid i \in I\}$ Z $U$ DO $\{0, 1, \dots, m-1\}$ SA NAZÝVA C-UNIVERZÁLNY $\Leftrightarrow$

$$\boxed{\forall x, y \in U, x \neq y : \left|\{i \in I \mid h_i(x) = h_i(y)\}\right| \leq \frac{c|I|}{m}}$$

## EXISTENCIA UNIVERZÁLNYCH SYSTÉMOV

$U = \{0, 1, \dots, N-1\}$ $N$ JE VEĽKÉ PRVOČÍSLO

$H = \{h_{a,b} \mid (a,b) \in U \times U\}$

$$\boxed{h_{a,b}(x) := \big((ax+b) \bmod N\big) \bmod m}$$

NECH $x, y \in U$, $x \neq y$. CHCEME NÁJSŤ $\forall a, b \in U$: $h_{a,b}(x) = h_{a,b}(y)$

$h_{a,b}(x) = h_{a,b}(y) \Leftrightarrow \exists i \in \{0,\dots,m-1\},\ r,s \in \{0,\dots,\lceil \frac{N}{m} \rceil - 1\}$:

$$\boxed{\begin{aligned} ax+b &\equiv i + rm \pmod N \\ ay+b &\equiv i + sm \pmod N \end{aligned}}$$

POVAŽUJME $x, y, i, r, s$ ZA KONŠTANTY. POTOM RIEŠIME SYSTÉM LIN. ROVNÍC V TELESE $\mathbb{Z}_N$.

$\det \begin{pmatrix} x & 1 \\ y & 1 \end{pmatrix} = x - y \neq 0 \Rightarrow$ EXISTUJE JEDINÉ RIEŠ. $\binom{a}{b}$

PRE DANÉ $x, y$ MÔŽE $i$ NADOBÚDAŤ $m$ HODNÔT A $r$ A $s$ $\lceil \frac{N}{m} \rceil$ HODNOT

$\Rightarrow$ PRE $x, y \in U$ $x \neq y$ EXISTUJE $\boxed{m \left( \lceil \frac{N}{m} \rceil \right)^2}$ DVOJÍC $(a,b) \in U \times U$ : $h_{a,b}(x) = h_{a,b}(y)$.

$$m \left( \lceil \frac{N}{m} \rceil \right)^2 \leq \frac{c \cdot N^2}{m} \Rightarrow \underline{\underline{c = \left( \lceil \frac{N}{m} \rceil \right)^2 / \left( \frac{N}{m} \right)^2}}$$

## VLASTNOSTI UNIVERZÁLNEHO HAŠOVANIA

MAJME $c$-UNIVERZÁLNY SYSTÉM $H = \{h_i \mid i \in I\}$

PRE VŠETKY $i \in I$, $x, y \in U$ DEFINUJME:

$$\boxed{\delta_i(x,y) := \begin{cases} 1 \Leftrightarrow & x \neq y \ \& \ h_i(x) = h_i(y) \\ 0 \Leftrightarrow & x = y \ \vee \ h_i(x) \neq h_i(y) \end{cases}}$$

PRE $S \subseteq U$ DEF. $\delta_i(x,S) := \sum_{y \in S} \delta_i(x,y)$

PRE FIXOVANÚ $S \subseteq U$ A FIXOVANÉ $x \in U$:

$$\sum_{i \in I} \delta_i(x,S) = \sum_{i \in I} \sum_{y \in S} \delta_i(x,y) = \sum_{y \in S} \sum_{i \in I} \dots =$$

$$= \sum_{y \in S, y \neq x} \left| \{ i \in I \mid h_i(x) = h_i(y) \} \right| \leq \begin{cases} (|S|-1)\, c\, |I|/m & x \in S \\ |S|\, c\, |I|/m & x \notin S \end{cases}$$

$\Rightarrow$ HORNÝ ODHAD OČAKÁVANEJ HODNOTY $\delta_i(x,S)$ JE

$$\boxed{\frac{1}{|I|} \sum_{i \in I} \delta_i(x,S) \leq \begin{cases} c \frac{|S|-1}{m} & x \in S \\ c \frac{|S|}{m} & x \notin S \end{cases}}$$

= HORNÝ ODHAD OČAKÁV. DĽŽKY RETAZCA NA ADR. $h_i(x)$ ...

11.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

**VETA:** OČAKÁVANÝ ČAS OP. MEMBER, INSERT, DELETE V c-UNIV. HAŠ. $O(1+c\alpha)$, $\alpha = \frac{|S|}{m}$ (FAKT. NAPLNENIA)
PRE n OP. $O(n(1+\frac{c}{2}\alpha))$.

POZOR! ROZDIEL V PREDPOKLADOCH:
i JE VYBERANÝ Z I S ROVNOMERNÝM ROZDELENÍM
NIE JE KLADENÝ ŽIADNY PREDP. NA VSTUPNÉ DATA. !
VÝBER INDEXU i MÔŽEME OVPLYVNIŤ !! DATA NIE

PRE $t \geq 1$ AKÁ JE PRAVDEPODOBNOSŤ, ŽE
PRE $i \in I$ JE $\delta_i(x,S)$ ASPOŇ $t\mu$, KDE $\mu$ JE OČ. H. $\delta_i(x,S)$

OZNAČME $J = \{i \in I \mid \delta_i(x,S) \geq t\mu\}$

$$\mu = \frac{1}{|I|}\sum_{i\in I}\delta_i(x,S) > \frac{1}{|I|}\sum_{i\in J}\delta_i(x,S) \geq \sum_{i\in J}\frac{t\mu}{|I|} =$$

$$\mu > \frac{|J|}{|I|}t\mu \Rightarrow |J| < \frac{|I|}{t} \quad \text{A} \quad P(\delta_i(x,S) \geq t\mu) < \frac{1}{t}$$

TOTO TVRDENIE PLATÍ V TEORII PRAVD. OBECNE
A NAZÝVA SA MARKOVOVA NEROVNOSŤ

PROBLÉMY: ZAISTENIE ROVNOM. ROZD. PRI VÝBERE $i \in I$
$|I|$ JE ČASTO VEĽMI VEĽKÉ → GENEROVANIE POST. 0 A 1

# DOLNÝ ODHAD NA VEĽKOSŤ $|I|$

$U = \{0, \dots, N-1\}$, $H = \{h_i \mid i \in I\}$ JE c-UNIV., $I = \{0, \dots, |I|-1\}$

DEFINUJME $U_0, U_1, \dots$ : $U_0 := U$

$U_{i+1}$ JE NAJVÄČŠIA PODMNOŽINA $U_i$ TAKÁ, ŽE $h_i(U_{i+1})$ JE JEDNOPRVKOVÁ

PRE $U_i$, $j < i$ JE $h_j(U_i) \subseteq h_j(U_{j+1}) \Rightarrow |h_j(U_i)| = 1$

HASHUJEME DO TABUĽKY VEĽKOSTI $m$ $\Rightarrow |U_{i+1}| \geq \left\lceil \frac{|U_i|}{m} \right\rceil$.

$|U_0| = N \Rightarrow$ INDUKCIOU $|U_i| \geq \left\lceil \frac{N}{m^i} \right\rceil$

$i = \lceil \log_m N \rceil - 1$ JE NAJVÄČŠIE PRIR. ČÍSLO, T.Ž. $\frac{N}{m^i} > 1$

$\Rightarrow$ $U_i$ MÁ ASPOŇ DVA PRVKY $x \neq y$

POTOM $h_j(x) = h_j(y)$ PRE $j = 0, \dots, i-1$

$$\Rightarrow i \leq \left| \{ j \in I \mid h_j(x) = h_j(y) \} \right| \leq \frac{c\,|I|}{m}$$

$$\Rightarrow \boxed{|I| \geq \frac{m}{c} \left( \lceil \log_m N \rceil - 1 \right)}$$

$\Rightarrow$ VEĽKOSŤ c-UNIV. SYSTÉMU RASTIE ÚMERNE ASPOŇ S LOGARITMOM VEĽKOSTI UNIVERZA.

12.02.2009 Th
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

## MALÝ UNIVERZÁLNY SYSTÉM

SKONŠTRUUJEME C-UNIVERZÁLNY SYSTÉM TAKÝ, ŽE LOGARITMUS VEĽKOSTI JEHO INDEXOVEJ MNOŽINY PRE VEĽKÉ UNIVERZÁ JE AŽ NA ADITÍVNU KONŠTANTU MENŠÍ NEŽ $4(\log m + \log\log N)$.

$p_1, p_2, p_3 \dots$ ≡ POSTUPNOSŤ VŠETKÝCH PRVOČÍSIEL

$U = \{0, 1, \dots, N-1\}$ $N \in \mathbb{N}$, JE DANÉ $m$

NECH $t$ JE NAJMENŠIE PRIR. ČÍSLO T. Ž. $\boxed{t \ln p_t \geq m \ln N}$

DEFINUJME: $\boxed{H_1 = \{ g_{c,d}(h_\ell) \mid t < \ell \leq 2t,\ c,d \in \{0, \dots, p_{2t}-1\} \}}$

$$\boxed{h_\ell(x) = x \bmod p_\ell, \quad g_{c,d}(x) = ((cx+d) \bmod p_{2t}) \bmod m}$$

UKÁŽEME, ŽE $H_1$ JE 3,25-UNIVERZÁLNY SYSTÉM

$I = \{ (c,d,\ell) \mid c,d \in \{0, \dots, p_{2t}-1\},\ t < \ell \leq 2t \}$

$|I| = t\, p_{2t}^2$

VETA: $\boxed{i(\ln i + \ln\ln i) > p_i > i \ln i}$

$\Rightarrow \underline{\underline{p_i \leq 2 i \ln i}} \Rightarrow |I| \leq 16 t^3 \ln^2 2t \Rightarrow$

$\Rightarrow \log|I| \leq 4 + 3\log t + 2\log\log t$

PRE DOSŤ VEĽKÉ $t$, $(\log t \geq 2\log\log t)$ PLATÍ:

$\log(|I|) \leq 4(1 + \log t)$, Z DEFINÍCIE $t$: $t \leq m \ln N$

$\Rightarrow \underline{\underline{\log(|I|) \leq 4(1 + \log m + \log\log N)}}$ ✓

## 3,25 - UNIVERZALITA :

NECH $x, y \in U$, $x \neq y$. OZNAČME

$$G_1 = \{(c,d,\ell) \mid g_{c,d}(h_\ell(x)) = g_{c,d}(h_\ell(y)),\ h_\ell(x) \neq h_\ell(y)\}$$

$$G_2 = \{(c,d,\ell) \mid \quad \text{—}\,\|\,\text{—} \quad ,\ h_\ell(x) = h_\ell(y)\}$$

ODHADNEME VEĽKOSŤ $G_1$ A $G_2$.

(a) $(c,d,\ell) \in G_1$ AK $\exists i \in \{0,1,\dots,m-1\}$, $r,s \in \{0,\dots,\lceil \frac{p_{2t}}{m} \rceil - 1\}$

$$\boxed{c}\,(x \bmod p_\ell) + \boxed{d} \bmod p_{2t} = i + rm$$

$$\boxed{c}\,(y \bmod p_\ell) + \boxed{d} \bmod p_{2t} = i + sm$$

SÚSTAVA LIN. ROVNÍC S REG. MATICOU V TELESE $\mathbb{Z}_{p_{2t}}$

PRE KAŽDÉ $\ell, i, r, s$ PRÁVE 1 DVOJICA $(c,d)$

$$\Rightarrow |G_1| \leq tm\left(\lceil \tfrac{p_{2t}}{m} \rceil\right)^2 \leq t\frac{p_{2t}^2}{m}\left(1+\frac{m}{p_{2t}}\right)^2 = \frac{|U|}{m}\left(1+\frac{m}{p_{2t}}\right)^2$$

(b) OZNAČME $\boxed{L = \{\ell \mid t < \ell \leq 2t,\ x \bmod p_\ell = y \bmod p_\ell\}}$

A POLOŽME $\boxed{P = \prod_{\ell \in L} p_\ell}$ $\qquad x \equiv y \pmod{p_\ell} \Leftrightarrow p_\ell \mid |x-y|$

$$\Rightarrow P \mid |x-y| \Rightarrow \boxed{P < N}$$

$p_t < p_\ell$ PRE KAŽDÉ $\ell \in L$ $\Rightarrow \boxed{P > p_t^{|L|}}$

$$\Rightarrow |L| \leq \frac{\ln N}{\ln p_t} \leq \frac{t}{m} \quad (\text{Z DEFINÍCIE } t)$$

$(c,d,\ell) \in G_2 \Leftrightarrow \ell \in L$ A $c,d \in \{0,\dots,p_{2t}-1\}$

$$\Rightarrow |G_2| \leq |L|\,p_{2t}^2 \leq \frac{t\,p_{2t}^2}{m} = \frac{|U|}{m}$$

12.02.2009 Th
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

## ODHAD VÝRAZU $\left(1+\frac{m}{p_{2t}}\right)^2$

$m \geq p_{2t} \Rightarrow t \ln p_t < t \ln p_{2t} \leq m \ln m \leq m \ln N$ ↯

$\Rightarrow m < p_{2t} \Rightarrow \left(1+\frac{m}{p_{2t}}\right)^2 < 4$

LEPŠÍ ODHAD: PŘEDPOKLÁDÁME, ŽE $t \geq 6 \wedge \ln m + \ln\ln m + \ln\ln\ln m < \ln N$

UKÁŽEME, ŽE $m < \frac{p_t}{\ln t}$

AK $m \geq \frac{p_t}{\ln t}$, POTOM $m \geq \frac{p_t}{\ln t} > \frac{t \ln t}{\ln t} = t$

$t \ln p_t < t(\ln t + \ln\ln t + \ln\ln\ln t) <$

$< m(\ln m + \ln\ln m + \ln\ln\ln m) < m \ln N$ ↯

$\Rightarrow \frac{m}{p_{2t}} < \frac{\frac{p_t}{\ln t}}{2t \ln 2t} < \frac{t(\ln t + \ln\ln t)}{2t \ln t \ln 2t} < \frac{1}{2}$

$\Rightarrow \left(1+\frac{m}{p_{2t}}\right) < 2{,}25$

$|\{i \in I \mid h(x) = h(y)\}| = |G_1| + |G_2| \leq \frac{|I|}{m}\left(1+\frac{m}{p_{2t}}\right)^2 + \frac{|I|}{m} \leq$

$\leq \underline{\underline{\frac{|I|}{m} \cdot 3{,}25}}$

## DOLNÝ ODHAD NA VELKOSŤ $c$

MAJME UNIVERZUM $U$ O VELKOSTI $N$ A NECH $H$ JE $c$-UNIVERZÁLNY SYSTÉM HAŠOVACÍCH FCIÍ. DO TABUĽKY O VELKOSTI $m$. POTOM

$$c \geq \frac{N-m}{N-1} > 1 - \frac{m}{N}$$

**LEMMA:** $\sum_{i=0}^{m-1} b_i(b_i-1) \geq b\left(\frac{b}{m}-1\right)$, $b = \sum_{i=0}^{m-1} b_i$

**DÔKAZ:** $\left(\sum_{i=0}^{m-1} b_i\right)^2 = b^2 \leq m \sum_{i=0}^{m-1} b_i^2$ CAUCHY-SCHWARZ

$\Rightarrow \sum_{i=0}^{m-1} b_i(b_i-1) \geq \frac{b^2}{m} - b = b\left(\frac{b}{m}-1\right)$ $\square$

MAJME FUNKCIU $f: U \to S$, $|U| = N$, $|S| = m$

NECH $A = \{(u,v) \in U \times U \mid u \neq v, f(u) = f(v)\}$

OZNAČME PRE $s \in S$: $k_s = |f^{-1}(s)|$.

ZREJME $|A| = \sum_{s \in S} k_s(k_s-1) \geq N\left(\frac{N}{m}-1\right) = N\frac{N-m}{m}$

KEĎ $H = \{h_i \mid i \in I\}$ JE $c$-UNIV. SYSTÉM, POTOM

$$|I|\, N \frac{N-m}{m} \leq \sum_{i \in I} |\{(u,v) \in U \times U \mid u \neq v,\ h_i(u) = h_i(v)\}| =$$

$$= \sum_{(u,v) \in U \times U,\ u \neq v} |\{i \in I \mid h_i(u) = h_i(v)\}| \leq$$

$$\leq N.(N-1).c\,\frac{|I|}{m}$$

$$\Rightarrow N-m \leq c(N-1) \Rightarrow c \geq \frac{N-m}{N-1} > 1 - \frac{m}{N}.$$

12.02.2009 Th
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## PERFEKTNÉ HAŠOVANIE

IDEA: NÁJSŤ PRE DANÚ MNOŽINU HAŠ. FCIU., KT. NEPRIPUŠŤA KOLÍZIE

NEVÝHODA: NIE JE MOŽNÉ PRIR. SPÔSOBOM REAL. INSERT

$U = \{0, \dots, N-1\}$. SÚBOR FUNKCIÍ $H$ z $U$ DO $\{0, \dots, m-1\}$ SA NAZÝVA $\boxed{(N, m, n)\text{-PERFEKTNÝ}} \Leftrightarrow \forall S \subseteq U, |S| = n$ $\exists h \in H$ PERFEKTNÁ PRE $S$ $(\forall s, t \in S \; s \neq t \Rightarrow h(s) \neq h(t))$

$$\boxed{h \text{ JE PERFEKTNÁ PRE } S \subseteq U \Leftrightarrow \forall i = 0, \dots, m-1 : |h^{-1}(i) \cap S| \leq 1}$$

PRE DANÚ $h$ JE POČET $S \in \binom{U}{n}$ PRE KT. JE $h$ PERFEKTNÁ

$$= \sum \prod_{j=0}^{n-1} |h^{-1}(i_j)| \quad 0 \leq i_0 < \dots < i_{n-1} < m \qquad \nabla_0$$

Z CAUCHY-SCHWARZOVEJ NEROVNOSTI PLYNIE, ŽE VEĽKOSŤ TOHTO SÚČINU JE MAXIMÁLNA, KEĎ VŠETKY ČLENY $= \frac{N}{m}$ $\Rightarrow$ $h$ MÔŽE BYŤ PERFEKTNÁ PRE NAJVIAC $\boxed{\binom{m}{n}\left(\frac{N}{m}\right)^n}$ MNOŽÍN $S \in \binom{U}{n}$.

$$\Rightarrow \boxed{|H| \geq \frac{\binom{N}{n}}{\binom{m}{n}\left(\frac{N}{m}\right)^n}}$$

DOLNÝ ODHAD PRE $|H|$

INÝ POSTUP: NECH $H = \{h_1, \dots, h_t\}$ JE $(N, m, n)$-PERF.

POLOŽME $U_0 := U$

$U_{i+1}$ JE NAJVÄČŠIA PODMNOŽINA $U_i$ TAKÁ, ŽE $|h_{i+1}(U_{i+1})| = 1$

ZREJME $|U_{i+1}| \geq \frac{|U_i|}{m}$ $\Rightarrow$ INDUKCIOU $|U_i| \geq \frac{N}{m^i}$

PRE $j \leq i$ JE $h_j(U_i)$ JEDNOBODOVÁ MNOŽINA

$\Rightarrow$ ŽIADNA $h_j$, $j \leq i$ NIE JE PERFEKTNÁ PRE $S \subseteq U$ TAKÚ, ŽE $|S \cap U_i| \geq 2$.

$\Rightarrow$ MUSÍ BYŤ $|U_t| \leq 1$ TJ. $\frac{N}{m^t} \leq 1$ $\Rightarrow$ $\boxed{t \geq \frac{\log N}{\log m}}$

## EXISTENCIA $(N, m, n)$-PERFEKTNÉHO SÚBORU

SÚBOR FUNKCIÍ $H = \{h_1, \dots, h_t\}$ Z $U$ DO $\{0, \dots, m-1\}$ REPREZENTUJEME POMOCOU MATICE $M(H) \in \{0, \dots, m-1\}^{N \times t}$

PRE $x \in U$ A $i \in \{1, \dots, t\}$ JE NA $x$-TOM RIADKU A $i$-TOM STĹPCI HODNOTA $h_i(x)$.

PLATÍ: ŽIADNA $h \in H$ NIE JE PERFEKTNÁ PRE $S = \{s_1, \dots, s_n\} \subseteq U$ $\Leftrightarrow$ PODMATICA $M(H)$ TVORENÁ RIADKAMI $s_1, \dots, s_n$ NEMÁ PROSTÝ STĹPEC.

PRE DANÚ $S \in \binom{U}{n}$ JE POČET TAKÝCH MATÍC:

$$\boxed{\Big(\underbrace{m^n - \prod_{i=0}^{n-1}(m-i)}_{\text{POČET NEPROSTÝCH FCIÍ. Z } S \text{ DO } \{0, \dots, m-1\}}\Big)^t \underbrace{m^{(N-n)t}}_{\text{DOPLNÍME ZVYŠNÉ RIADKY}}}$$

KEĎŽE $\left|\binom{U}{n}\right| = \binom{N}{n}$, POČET $\forall$ MATÍC KTORÉ <u>NEREPREZENTUJÚ</u> $(N, m, n)$-PERF. SYSTÉM JE:

$$\leq \binom{N}{n} \cdot \dots$$

12.02.2009 Th
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

POČET ∀ MATIC $M(H)$ JE $m^{Nt}$.

AK $\binom{N}{n}\left(m^n - \prod_{i=0}^{n-1}(m-i)\right)^t m^{(N-n)t} < m^{Nt}$, $(*)$ $/ : m^{Nt}$

POTOM NUTNE EXISTUJE $(N,m,n)$ – PERFEKTNÝ SYSTÉM.

$$\Leftrightarrow \binom{N}{n}\left(1-\frac{\prod_{i=0}^{n-1}(m-i)}{m^n}\right)^t < 1 \Leftrightarrow t > \frac{\ln\binom{N}{n}}{-\ln\left(1-\frac{\prod_{i=0}^{n-1}(m-i)}{m^n}\right)}$$

$$\ln\binom{N}{n} \le n \ln N$$

$$-\ln\left(1-\frac{\prod_{i=0}^{n-1}(m-i)}{m^n}\right) \ge \prod_{i=0}^{n-1}\left(1-\frac{i}{m}\right) =$$

$$= e^{\sum_{i=0}^{n-1}\ln\left(1-\frac{i}{m}\right)} \ge e^{\int_0^{n-1}\ln\left(1-\frac{x}{m}\right)dx} \ge -\frac{n^2}{m}$$

$\Rightarrow$ AK $\underline{\underline{t > n(\ln N)e^{n^2/m}}}$, POTOM PLATÍ $(*)$.

## KONŠTRUKCIA PERF. HAJOVACEJ FCIE.

UNIVERZUM $U = \{0,1,\dots,N-1\}$, $N$ JE PRVOČÍSLO

MAJME $S \subseteq U$ O VEĽKOSTI $n$. BUDEME UVAŽOVAŤ FCIE.:

$$\boxed{h_k(x) = (kx \bmod N) \bmod m} \quad \text{PRE } k = 1,\dots,N-1$$

PRE $i = 0,1,\dots,m-1$, $k = 1,\dots,N-1$ OZNAČME

$$\boxed{b_i^k = \left|\{x \in S \mid h_k(x) = i\}\right|}$$

PLATÍ:

$$b_i^k \ge 2 \Rightarrow (b_i^k)^2 - b_i^k \ge 2$$

$$b_i^k \le 1 \Rightarrow (b_i^k)^2 - b_i^k = 0$$

!

**VETA:** $h_k$ JE PERFEKTNÁ $\Leftrightarrow \sum_{i=0}^{m-1} (b_i^k)^2 - n < 2$

$h_k$ PERFEKTNÁ $\Rightarrow \forall i = 0, \dots, m-1 : b_i^k \leq 1 \Rightarrow \sum_{i=0}^{m-1} (b_i^k)^2 = n < n+2$

$h_k$ NIE JE PERFEKTNÁ $\Rightarrow \exists i_0 \in \{0, \dots, m-1\} : b_{i_0}^k \geq 2$

$\forall i \ (b_i^k)^2 \geq b_i^k$, $(b_{i_0}^k)^2 \geq b_{i_0}^k + 2 \Rightarrow \sum_{i=0}^{m-1} (b_i^k)^2 \geq n+2$

**ODHAD** $\sum_{k=1}^{N-1} \left[ \sum_{i=0}^{m-1} (b_i^k)^2 - n \right]$

$$= \sum_{k=1}^{N-1} \left[ \sum_{i=0}^{m-1} \left| \{ x \in S \mid h_k(x) = i \} \right|^2 - n \right] =$$

$$= \sum_{k=1}^{N-1} \left| \{ (x,y) \mid x, y \in S,\ x \neq y,\ h_k(x) = h_k(y) \} \right| =$$

$$= \sum_{x,y \in S,\ x \neq y} \left| \{ k \mid 1 \leq k < N,\ h_k(x) = h_k(y) \} \right|$$

$h_k(x) = h_k(y) \Leftrightarrow \exists i \in \{0, \dots, m-1\},\ r, s \in \{0, \dots, \lfloor \frac{N}{m} \rfloor\}$

$$\left.\begin{aligned} kx \bmod N &= i + rm \\ ky \bmod N &= i + sm \end{aligned}\right\} \Rightarrow k(y-x) \equiv (s-r)m \pmod N$$

KDE $-\lfloor \frac{N}{m} \rfloor \leq s-r \leq \lfloor \frac{N}{m} \rfloor$, $s-r \neq 0$ PREDOŠE $x \neq y$

PRE KAŽDÉ $t \in \{-\lfloor \frac{N}{m} \rfloor, \dots, -1, 1, \dots, \lfloor \frac{N}{m} \rfloor\}$

EXISTUJE PRÁVE JEDNO $k \in \{1, \dots, N-1\} : k(y-x) \equiv tm$

$\Rightarrow$ EXISTUJE NAJVIAC $2\lfloor \frac{N}{m} \rfloor$ ČÍSIEL $\left(= 2\lfloor \frac{N-1}{m} \rfloor\right)$

$k \in \{1, \dots, N-1\}$ T.Z. $h_k(x) = h_k(y)$.

POZOR! JEDNÁ SA IBA O HORNÝ ODHAD,

NIE JE MOŽNÉ HO JEDNODUCHO VYLEPŠIŤ

12.02.2009 Th
PETER ČERNO

# NTINO66 DATOVÉ STRUKTÚRY I

$$\Rightarrow \sum_{x,y \in S,\, x \neq y} \left| \{ k \in \{1, \dots, N-1\} \mid h_k(x) = h_k(y) \} \right| \leq$$

$$\leq \sum_{x,y \in S,\, x \neq y} 2 \left\lfloor \frac{N-1}{m} \right\rfloor \leq 2(N-1) \frac{n(n-1)}{m}$$

$$\Rightarrow \boxed{\exists k : \sum_{i=0}^{m-1} (b_i^k)^2 \leq \frac{2n(n-1)}{m} + n} \quad !$$

EXISTUJE VIAC AKO $\boxed{\frac{N-1}{4}}$ TAKÝCH $k$, ŽE:

$\boxed{\sum_{i=0}^{m-1} (b_i^k)^2 < \frac{3n(n-1)}{m} + n}$ DK: AK EXISTUJE $\geq \frac{3(N-1)}{4}$ S OP. NEROVN.

$$\Rightarrow \sum_{k=1}^{N-1} \left( \sum_{i=0}^{m-1} (b_i^k)^2 - n \right) \geq \frac{3n(n-1)}{m} \cdot \frac{3(N-1)}{4} > 2(N-1)\frac{n(n-1)}{m} \quad ↯$$

$$\Rightarrow P\left( k \in \{1, \dots, N-1\} \mid \sum_{i=0}^{m-1} (b_i^k)^2 < \frac{3n(n-1)}{m} + n \right) \geq \frac{1}{4}.$$

VETA: $\boxed{n = m}$

(a) DET. ALGORITMUS V ČASE $O(nN)$ NÁJDE $k$: $\sum_{i=0}^{m-1} (b_i^k)^2 < 3n$

(b) PRAVD. ALG. V OČAK. ČASE $O(n)$ NÁJDE $k$: $\sum_{i=0}^{m-1} (b_i^k)^2 < 4n$ (4 ITER.)

(c) $\boxed{m = n(n-1)+1}$ DET. V ČASE $O(nN)$ NÁJDE PERFEKTNÚ $h_k$

(d) $\boxed{m = 2n(n-1)}$ V OČ. ČASE $O(n)$ NÁJDE PERF. $h_k$ (4 ITER. NA $k$)

PROBLÉM: $m = \Theta(n^2)$

(1) NÁJDEME $k$ TAK, ŽE PRE $m=n$ JE $\sum_{i=0}^{m-1}(b_i^k)^2 < 3n$
(V PRAVDEPODOBN. PRÍPADE .. $< 4n$)

PRE $i=0,\dots,m-1$ NÁJDEME $S_i = \{s \in S \mid h_k(s) = i\}$

(2) PRE KAŽDÉ $i=0,1,\dots,m-1$ TAKÉ, ŽE $S_i \neq \emptyset$ $\boxed{b_i^k = |S_i|}$
NÁJDEME PRE $c_i = |S_i|(|S_i|-1)+1$, (RESP. $c_i = 2|S_i|(|S_i|-1)$)
TAKÉ $k_i$, ŽE $h_{k_i}$ JE PERFEKTNÁ PRE $S_i$ DO TAB. VELKOSTI $c_i$
POLOŽME $c_i := 0$, KEĎ $S_i = \emptyset$.

(3) PRE $i=0,1,\dots,m$ DEFINUJME $d_i := \sum_{j=0}^{i-1} c_j$
A PRE $x \in U$ POLOŽME $g(x) := d_\ell + h_{k_\ell}(x)$, KDE $\ell = h_k(x)$.

<u>VETA</u>: $g$ JE PERFEKTNÁ, $O(1)$
DET. PRÍPAD: $g$ MAJÚC DO TABUĽKY $< 3n$, NÁJDENIE $g$ V $O(nN)$
PRAVD. PRÍPAD: —||— $< 6n$, NÁJDENIE $g$ V ČASE $O(n)$
HODNOTY $k, k_i$ ZABERAJÚ $O(n \log N)$ PAMÄTI. = <u>PROBLÉM</u>

D: $g(S_i)$ SÚ DISJUNKTNÉ A $h_{k_i}$ JE PERF. NA $S_i \Rightarrow g$ <u>JE PERFEKTNÁ</u>
$d_m$ JE HORNÝ ODHAD NA POČET RIADKOV V TABUĽKE
PRE $S_i \neq \emptyset$ JE $|S_i|(|S_i|-1)+1 \leq |S_i|^2 = (b_i^k)^2$
$\Rightarrow$ <u>V DET. PRÍPADE</u>: $\sum_{i=0}^{m-1} c_i \leq \sum_{i=0}^{m-1}(b_i^k)^2 < 3n$
$k$ NÁJDEME V ČASE $O(nN)$, $k_i$ V ČASE $O(|S_i|N)$
$\Rightarrow g$ SKONŠTRUUJEME V ČASE
$O(nN + \sum_{i=0}^{m-1}|S_i|N) = O(nN)$

<u>V PRAVD. PRÍPADE</u>: $d_m = \sum_{i=0}^{m-1} c_i = 2\Big(\overbrace{\sum (b_i^k)^2}^{<4n} - \overbrace{\sum (b_i^k)}^{=n}\Big) < 6n$
$k$ NÁJDEME V ČASE $O(n)$ A $k_i$ V ČASE $O(|S_i|)$
OČAKÁVANÝ POČET ITERÁCIÍ $= 4$, ... $\Rightarrow g$ NÁJ. V ČASE $O(n)$

$\square$

13.02.2009 Fr
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

PRE $m$ NECH $q = |\{p \text{ JE PRVOČÍSLO}: p \mid m\}|$

$$m \geq \prod_{i=1}^{q} p_i > q! = e^{\sum_{i=1}^{q} \ln i} \geq e^{\int_1^q \ln x \, dx} = e^{q \ln(\frac{q}{e})+1} \geq \left(\frac{q}{e}\right)^q$$

$\Rightarrow \ln m > q \ln(q) - q$ A PRE VHODNÉ $c > 0$:

$$\boxed{\ln m \geq c\, q \ln(q)}$$

$x \ln(x)$ JE SPOJITÁ, RASTÚCA $\Rightarrow \exists q_1 \geq q : \boxed{\ln m = c\, q_1 \ln(q_1)}$

$$\ln \ln m = \ln c + \ln q_1 + \ln \ln q_1 \leq \ln c + 2 \ln q_1$$

$$\Rightarrow q \leq q_1 = \frac{\ln m}{c \ln(q_1)} \leq \frac{2 \ln m}{c(\ln \ln m - \ln c)}$$

<u>VETA</u>: $\delta(m)$ = POČET PRVOČ., KT. DELIA $m$ $\boxed{= O\left(\frac{\log m}{\log \log m}\right)}$

MAJME $S = \{s_1 < s_2 < \dots < s_n\} \subseteq U$, $d_{i,j} := s_j - s_i$, $1 \leq i < j \leq n$

$$\boxed{D := \prod_{1 \leq i < j \leq n} d_{ij} \leq N^{n^2}} \qquad \boxed{s_i \not\equiv s_j \pmod p,\ i<j \Leftrightarrow d_{i,j} \not\equiv 0 \pmod p}$$

POČET PRVOČ. DELITEĽOV $D$ JE NAJVIAC $c \frac{\ln D}{\ln \ln D}$.

$\Rightarrow$ MEDZI PRVÝMI $1 + c \frac{\ln D}{\ln \ln D}$ PRVOČÍSLAMI EXISTUJE $p$

T.Z. $s_i \not\equiv s_j \pmod p$ PRE KAŽDÉ $1 \leq i < j \leq n$

A MEDZI PRVÝMI $2c \frac{\ln D}{\ln \ln D}$ MÁ ASPOŇ POLOVICA PRV. TÚTO VLASTN.

$$\boxed{\text{TJ. } \phi_p(x) = x \bmod p \text{ JE PERFEKTNÁ PRE } S}$$

!!!

PLATÍ $p_t \leq 2t \ln t$ PRE $t > 2$

$$\Rightarrow p \leq 2\left(1 + c\frac{\ln D}{\ln \ln D}\right) \cdot \ln\left(1 + c \frac{\ln D}{\ln \ln D}\right) \leq \dots = O(\ln D)$$

$$= O(n^2 \ln N)$$

<u>VETA</u>: PRE KAŽDÚ $n$-PRVKOVÚ $S \subseteq U$ EXISTUJE PRV. $p$ VEĽKÉ $O(n^2 \ln N)$ T.Ž. $\phi_p(x)$ JE PERFEKTNÁ PRE $S$.

MEDZI PRVÝMI $4cn^2 \ln N$ PRIRODZENÝMI ČÍSLAMI JE ASPOŇ $1/2$ PRVOČÍS. TAKÝCH, ŽE $\phi_p$ JE PERFEKTNÁ PRE $S$.

TEST, ČI $\phi_p$ JE PERFEKTNÁ PRE $S$ V ČASE $O(n \log n)$

NÁHODNE ZVOLENÉ $p \leq 4cn^2 \ln N$ JE PRVOČÍSLO S PRAVDEP. $\Theta\left(\frac{1}{\ln(4cn^2 \ln N)}\right)$

$\Rightarrow$ OČAKÁVANÝ POČET NEÚSP. TESTOV JE $O(\ln(4cn^2 \ln N))$

$\Rightarrow$ OČAKÁVANÝ ČAS ALGORITMU JE $\underline{\underline{O(n \log n\,(\log n + \log\log N))}}$

<u>METÓDA FREDMAN, KOMLÓS, SZEMERÉDI</u> NECH $S \in \binom{U}{n}$

(1) NÁJDEME PRVOČÍSLO $q_0$ VEĽKÉ $O(n^2 \log N)$ : $\phi_{q_0}$ PERF.

POLOŽME $S_1 := \{\phi_{q_0}(s),\ s \in S\} = \phi_{q_0}(S)$

(2) NÁJDEME PRVOČÍSLO $q_1$ TAKÉ, ŽE $n(n-1) < q_1 \leq 2n(n-1)+2$

$\Rightarrow$ EXISTUJE $\ell \in \{1, \dots, q_0 - 1\}$ : $h_\ell(x) = [(\ell x) \bmod q_0] \bmod q_1$ JE PERFEKTNÁ PRE $S_1 \subseteq \{0, \dots, q_0 - 1\}$

POLOŽME $S_2 := h_\ell(S_1)$

(3) PRE PODMN. $S_2$ UNIVERZA $\{0, \dots, q_1 - 1\}$ VIEME SKONŠTRUOVAŤ PERF. HAŠ. FUNKCIU $g$ DO TABUĽKY S MENEJ AKO $3n$ RIADKAMI V ČASE $O(n\, q_1)$

<u>VÝSLEDNÁ PERF. HAŠ. FCIA</u> JE $\boxed{f = g \circ h_\ell \circ \phi_{q_0}}$

$f$ JE PERFEKTNÁ HAŠ. FCIA ✓

TABUĽKA MÁ $< 3n$ RIADKOV ✓

VYČÍSLENIE $f$ V $O(1)$ ✓

PAMÄŤ: $\phi_{q_0}$ ... $q_0 \in O(n^2 \log N)$, $h_\ell$ ... $q_1 \in O(n^2)$, $\ell \in O(q_0)$

$g$ ... $n+1$ ČÍSIEL VEĽKOSTI $O(q_1)$

$\Rightarrow$ CELKOVÁ PAMÄŤ $O(\log n + \log\log N + n \log n)$ ✓

ZLOŽITOSŤ ALGORITMU: $O(n^3 \log n \log N)$

13.02.2009 Fr
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## DYNAMICKÉ PERFEKTNÉ HAŠOVANIE

$U = \{0, \dots, N-1\}$ N JE PRVOČÍSLO

PRE $t = 1, \dots, N-1$ NECH $\mathcal{H}_t = \{ h_k : U \to \{0, \dots, t-1\} \mid h_k(x) = (kx \bmod N) \bmod t \}$

PRE $S \in \binom{U}{n}$, $s \in \mathbb{N}$ AK NÁHODNE ZVOLÍME $k \in \{1, \dots, N-1\}$, POTOM $h_k \in \mathcal{H}_s$ S PRAVD. ASPOŇ $\frac{1}{2}$ SPĺŇA:

$$\boxed{\sum_{i=0}^{s-1} (b_i^k)^2 < \frac{4n^2}{s} + n} \qquad (*)$$

VIEME, ŽE: $\sum_{k=1}^{N-1} \left[ \sum_{i=0}^{s-1} (b_i^k)^2 - n \right] \leq 2(N-1)\frac{n(n-1)}{s}$

AK BY EXISTOVALO VIAC AKO $\frac{N-1}{2}$ TAKÝCH $k$, ŽE $\sum_{i=0}^{s-1} (b_i^k)^2 \geq \frac{4n^2}{s} + n$, POTOM

$$\sum_{k=1}^{N-1} \left[ \sum_{i=0}^{s-1} (b_i^k)^2 - n \right] \geq \frac{N-1}{2} \left( \frac{4n^2}{s} \right) > 2(N-1)\frac{n(n-1)}{s} \quad ↯$$

NECH $k$ SPĺŇA $(*)$

PRE KAŽDÉ $i = 0, \dots, s-1$ $\quad S_i = \{ t \in S \mid h_k(t) = i \}$, $|S_i| = b_i^k$

KEĎ NÁHODNE VOLÍME $j(i) \in \{1, \dots, N-1\}$,

POTOM S PRAVDEPODOBNOSTOU $\frac{1}{2}$ JE $h_{j(i)} \in \mathcal{H}_{2(b_i^k)^2}$ PERFEKTNÁ NA $S_i$.

KAŽDÁ MNOŽINA $S_i$ BUDE TERAZ REPREZENTOVANÁ POMOCOU $h_{j(i)}$ VO SVOJEJ TABUĽKE $T_i$ PRE $i = 0, 1, \dots, s-1$

DÔVOD: KEĎ SI AKTUALIZAČNÉ OP. INSERT ALEBO DELETE MENÍTA OPRAVU $T_i$, TAK SA NEMUSÍ PRACOVAŤ S HODNOTAMI V INÝCH TABUĽKÁCH.

$s$ VOLÍME AKO $\sigma(|S|)$, KDE $\sigma(n) = 4/3\sqrt{6}(1+c)n$

MEMBER(x)

$i \leftarrow h_k(x)$

if x JE NA $h_{j(i)}(x)$-TEJ POZÍCII V $T_i$ then $x \in S$ else

else $x \notin S$

INSERT(x)

if $x \in S$ then STOP

$++n$

if $n \leq m$ then

$i \leftarrow h_k(x)$, $|S_i| \leftarrow |S_i| + 1$

if $|S_i| \leq m(i)$ A POZÍCIA $h_{j(i)}(x)$ v $T_i$ JE PRÁZDNA then

VLOŽ x DO $T_i$ NA POZÍCIU $h_{j(i)}(x)$

else if $|S_i| \leq m(i)$ A POZÍCIA ... JE OBSADENÁ then

NÁJDI NOVÚ $h_{j(i)} \in \mathcal{H}_{2m(i)^2}$ A PREHASHUJ

else $m(i) \leftarrow 2 * m(i)$

if NIE JE DOSŤ PRIEJDORU PRE $T_i$ ALEBO

$\sum_{i=0}^{\sigma(m)-1} 2(m(i))^2 \geq \frac{8m^2}{\sigma(m)} + 2m$ then REHASHALL

else ALOKUJ NOVÚ $T_i$, NÁJDI $h_{j(i)} \in \mathcal{H}_{2m(i)^2}$ A PREHASHUJ

else REHASHALL

REHASHALL

$m := (1+c)|S|$

NÁJDI $h_k \in \mathcal{H}_{\sigma(m)}$ TAK, ABY $\sum_{i=0}^{\sigma(m)-1} 2(b_i^k)^2 < \frac{8m^2}{\sigma(m)} + 2m$

POLOŽ $m(i) := |S_i|$, $S_i := \emptyset$; ZVOL NÁHODNE $h_{j(i)} \in \mathcal{H}_{2(m(i))^2}$

foreach $x \in S$ do INSERT(x) +ALOKUJ T, $T_i$

DELETE(x)

if $x \in S$ then $i \leftarrow h_k(x)$, $--n$, $|S_i| \leftarrow |S_i| - 1$

OPSTRÁŇ x Z POZ. $h_{j(i)}(x)$ V TAB. $T_i$

if $n < \frac{m}{1+2c}$ then REHASHALL

13. 02. 2009 Fr
PETER ČERNO

# NTIN066 DÁTOVÉ ŠTRUKTÚRY I

## PERFEKTNÉ MAPOVANIE DO TABUĽKY VEL. $m = |S|$

EXISTENCIA $(N, n, n)$ - PERFEKTNÉHO SÚBORU
VYPLÝVA Z $n^N > e^{n + \ln(n)} \ln(N)$. (VIĎ $t > n(\ln N) e^{n^2/m}$)

r-HYPERGRAF $(X, E)$, $E \subseteq \binom{X}{r}$

CYKLUS $\equiv$ KAŽDÝ VRCHOL LEŽÍ V ASPOŇ 2 HRANÁCH
ACYKLICKÝ $\equiv$ ŽIADEN PODGRAF NIE JE CYKLUS

$\boxed{\text{KROK 1}}$: NECH $(V, E)$ JE r-HYPERGRAF, $|E| = n = m$
HĽADÁME $g: V \to \{0, \dots, m-1\}$ TAKÚ, ŽE
$h: E \to \{0, \dots, m-1\}$ $h(e) = \sum_{v \in e} g(v) \bmod m$ JE PROSTÁ
PRE ACYKLICKÝ $(V, E)$ FUNGUJE POSTUP:
(1) ZVOLÍME BIJEKCIU $h: E \to \{0, \dots, m-1\}$
(2) KEĎ $e = \{v_1, \dots, v_r\}$ A $g(v_i)$ JE DEF. PRE $i = 2, \dots, r$,
POLOŽÍME $g(v_1) := h(e) - \sum_{i=2}^{r} g(v_i) \bmod m$ (INDUKTÍVNE)

$\boxed{\text{KROK 2}}$: MÁME DANÚ $S \subseteq U$, $|S| = n$
NÁJDEME r FUNKCIÍ $f_1, \dots, f_r : U \to V$ TAKÝCH, ŽE
$(V, E)$, KDE $E = \{\{f_1(x), \dots, f_r(x)\} \mid x \in S\}$,
JE ACYKLICKÝ r-HYPERGRAF.

HAS. FCIA. JE POTOM DEFINOVANÁ AKO
$$f(x) := \sum_{i=1}^{r} g(f_i(x)) \quad \forall x \in U$$

$f$ JE PERFEKTNÁ NA $S$.

PROBLÉM: REPREZENTÁCIA $(V, E)$ A NÁHODNÁ VOĽBA $f_i$
(STAČÍ VRAJ Z NEJAKÉHO c-UNIV. SÚBORU FUNKCIÍ)

# EXTERNÉ HAŠOVANIE

$b > 1$

MINIMALIZÁCIA POČTU I/O OPERÁCIÍ, 1 STRÁNKA = $b$ PRVKOV

$h \to U \to \{0,1\}^k$ PROSTÁ

CHCEME ULOŽIŤ $S \subseteq U$

PRE $\alpha \in \{0,1\}^{<k}$ DEFINUJEME $\boxed{h_S^{-1}(\alpha) = \{s \in S \mid \alpha \text{ JE PREFIX } h(s)\}}$

$\alpha$ JE KRITICKÉ AK $0 < |h_S^{-1}(\alpha)| \leq b$ A $\forall \alpha' [\alpha$ JE $|h_S^{-1}(\alpha')| > b$

PRE KAŽDÉ $s \in S$ $\exists!$ KRITICKÉ $\alpha [ h(s)$. DEFINUJEME $d(s) := |\alpha|$.

REPREZENTÁCIA: JEDNOZNAČNÁ KORESPONDENCIA MEDZI KRITICKÝMI SLOVAMI A STRÁNKAMI

KU KRIT. SLOVU $\alpha$ KORESPONDUJE STRÁNKA, KDE JE $h_S^{-1}(\alpha)$

ADRESÁR: KAŽDÉMU SLOVU $\alpha \in \{0,1\}^{d(S)}$, $d(S) = \max_{s \in S} d(s)$

PRIRADÍ STRÁNKU ZODP. KRITICKÉMU SLOVU $\beta \sqsubseteq \alpha$.

AK $\neg\exists$ KRIT. $\beta \sqsubseteq \alpha$, PRIRADÍ NIL.

REPREZENTUJEME AKO ZOZNAM $2^{d(S)}$ ADRIES

OPERÁCIE MEMBER, INSERT A DELETE VYŽADUJÚ $O(1)$ OPERÁCIÍ S EXTERNOU PAMÄŤOU.

PRI REPREZENTÁCII MNOŽINY $S$ O VEĽKOSTI $n$ JE OČAKÁVANÝ POČET POUŽITÝCH STRÁNOK $\frac{n}{b \ln 2}$ A OČAKÁVANÁ VEĽKOSŤ ADRESÁRA $\frac{e}{b \ln 2} n^{1+1/b}$.

$\rightsquigarrow$ OČAKÁVANÝ POČET PRVKOV NA STRÁNKE JE $b \ln 2 \approx 0{,}69\, b$.

TO JE OK, HORŠIE JE TO S VEĽKOSŤOU ADRESÁRA...

EXPERIMENTY: POUŽITEĽNOSŤ DO $|S| = 10^{10}$ AK $b \approx 100$.

14.02.2009 Sa
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## STROMY

MEMBER, INSERT, DELETE, MIN, MAX

SPLIT(x) : KONŠTRUUJE $S_1 = \{s \in S \mid s < x\}$, $S_2 = \{s \in S \mid s > x\}$
+ OZNÁMI, ČI $x \in S$

JOIN2$(S_1, S_2)$ : $\max S_1 < \min S_2 \rightarrow$ ZOSTROJÍ $S_1 \cup S_2$

JOIN3 $(S_1, x, S_2)$ : $\max S_1 < x < \min S_2 \rightarrow$ ZOSTROJÍ $S_1 \cup \{x\} \cup S_2$

ORD(k) .. NÁJDE k-TÝ NAJMENŠÍ PRVOK $1 \le k \le |S|$

## (a,b)-STROMY

NECH $a \le b$, $a, b \in \mathbb{N}$. POTOM KOREŇOVÝ $\boxed{(a,b)\text{-STROM } (T, t)}$:

(1) KAŽDÝ VNÚTORNÝ VRCHOL $v \neq t$ MÁ $\ge a$, $\le b$ SYNOV

(2) VŠETKY CESTY Z KOREŇA DO LISTOV MAJÚ ROVNAKÚ DĹŽKU

+ $\boxed{2 \le a, \ 2a-1 \le b}$

KOREŇ t JE BUĎ LIST, ALEBO MÁ $\ge 2$, $\le b$ SYNOV !

$\forall a, b \in \mathbb{N}$, $2 \le a$, $2a-1 \le b$ $\forall n \in \mathbb{N}$: $\exists$ (a,b)-STROM S PRÁVE n LISTAMI A VÝŠKOU $\in [\log_b n ; 1 + \log_a (\frac{n}{2})]$.

STROM O VÝŠKE h MÁ $2a^{h-1} \le n \le b^h$ LISTOV

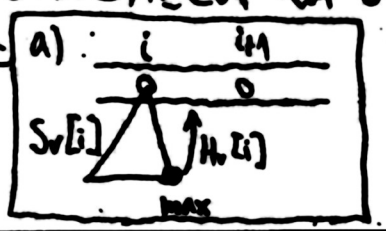
(a,b)-STROM (T,t) REPREZENTUJE MNOŽINU S, KEĎ MÁ PRESNE $|S|$ LISTOV A JE DANÝ IZOMORFIZMUS MEDZI LEXIKOGRAFICKÝM USP. LISTOV A USP. MN. S.

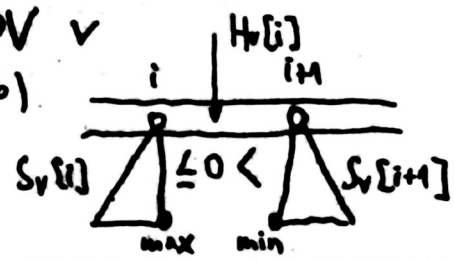
VNÚTORNÝ VRCHOL v : LIST : MÁ HODNOTU key

$p(v)$ = POČET SYNOV v

$S_v[1..p(v)]$ = UKAZATEĽA NA SYNOV v

$H_v[1..p(v)-1]$ =



PLATÍ: PRE KAŽDÉ $s \in S$ OKREM NAJVÄČŠIEHO EXISTUJE PRÁVE JEDEN VNÚTORNÝ VRCHOL $v$ A $i$ : $H_v(i) = s$.

PRE PORIADKOVÚ ŠTATISTIKU - OPERÁCIU $ord(k)$ MUSÍME ROZŠÍRIŤ ŠTRUKTÚRU VNÚTORNÉHO VRCHOLU $v$ O POLE $P_v(1..\rho(v))$

$P_v[i]$ = POČET PRVKOV $S$ REPREZENTOVANÝCH V PODSTROME $i$-TEHO SYNA VRCHOLU $v$.

IMPLEMENTÁCIU OP. POZRI V SKRIPTÁCH!

## STROMY S PARALELNÝM PRÍSTUPOM

PARALELNÁ IMPLEMENTÁCIA OP. INSERT A DELETE (TZV. VIAZOVANIE ZHORA - DOLU)

PREDPOKLAD: $b \geq 2a$

INSERT: VO VYHĽADÁVACEJ FÁZE SÚ UZAMKNUTÉ VRCHOLY $t$, $otec(t)$ A SYNOVIA VRCHOLU $t$. ALGORITMUS ZISTÍ, V KT. SYNOVI VRCHOLU $t$ MÁ POKRAČOVAŤ, A KEĎ $\rho(t) = b$, VYKONÁ ŠTIEPENIE ($\Rightarrow$ PRETO POTREBUJEME $b \geq 2a$)

DELETE: VO VYHĽ. FÁZE SÚ UZAMKNUTÝ $t$, $otec(t)$, BEZPROSTREDNÝ BRAT $y$ VRCHOLU $t$ A ICH SYNOVIA

AK $\rho(t) = a$, POTOM PO NÁJDENÍ VRCHOLU, KDE SA BUDE POKRAČOVAŤ, SA VYKONÁ BUĎ PREJOM (AK $\rho(y) > a$) ALEBO SPOJENIE (AK $\rho(y) = a$).

14.02.2009 Sa
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## A-SORT

ZALOŽENÝ NA (a,b)-STROMOCH, SCHOPNÝ VYUŽIŤ PREDTRIEDENOSŤ...

+ UKAZATEĽ PTV NA PRVÝ LIST + CESTA Z PTV DO KORENA
LISTY (a,b)-STROMU PREPOJENÉ DO ZOZNAMU (Nasl)

MAJME $(x_1, \dots, x_n)$ Z $U$. DEFINUJME

$$\boxed{f_i = |\{ j \mid i < j, x_i > x_j \}|}$$

$$\boxed{F = |\{ (i,j) \mid i < j, x_i > x_j \}| = \sum_{i=1}^{n} f_i}$$

POČÍTAME ZLOŽITOSŤ A-SORTU V ZÁVISLOSTI NA $n$ A $F$.

A-INSERT(x) VYŽADUJE ČAS POTREBNÝ NA NÁJDENIE MIESTA, KAM VLOŽIŤ $x$ + $O$(POČET VOLANÍ Štiepenie)

KAŽDÝ BEH Štiepenie VYTVORÍ 1 NOVÝ VNÚTORNÝ VRCHOL

KEĎŽE $a \geq 2$ A FINÁLNY (a,b)-STROM MÁ $n$ LISTOV, JE POČET VN. VRCHOLOV MENŠÍ AKO $n$ $\Rightarrow$

$\Rightarrow$ ČAS, KT. VYŽADUJÚ V BEHU PODPROCEDÚRY A-INSERT = ČAS NA NÁJDENIE MIEST JEDNOTLIVÝCH PRVKOV POSTUPNOSTI + $O(n)$.

KEĎ A-INSERT PRI HĽADANÍ MIESTA PRE PRVOK $x$ SKONČÍ VO VÝŠKE $h$, POTOM NÁJDENIE MIESTA PRE PRVOK $x$ VYŽADUJE ČAS $O(h)$.

VŠETKY PRVKY REPREZENTOVANÉ (a,b)-STROMOM POD PRVÝM VRCHOLOM VO VÝŠKE $h-1$ SÚ MENŠIE NEŽ $x$ A JE ICH ASPOŇ $a^{h-2}$

KEĎ $x = x_i$, POTOM POČET PRVKOV $< x$ JE $f_i$

$\Rightarrow a^{h-2} \leq f_i \Rightarrow h-2 \leq \log_a f_i$

CELKOVÝ ČAS ALGORITMU A-SORT JE PRETO

$$O\left(\sum_{i=1}^{n} \log f_i + n\right)$$

$$\sum_{i=1}^{n} \log f_i = n \log \left(\prod_{i=1}^{n} f_i\right)^{1/n} \leq n \log \frac{\sum_{i=1}^{n} f_i}{n} = n \log \frac{F}{n}$$

$\Rightarrow$ ZLOŽITOSŤ A-SORT JE $\underline{\underline{O\left(n + n \log \frac{F}{n}\right)}}$

ODPORÚČA SA POUŽIŤ (2,3)-STROMY.

## HLADINOVO PREPOJENÉ (a,b)-STROMY S PRSTOM

(a,b)-STROM S PREPOJENÝMI HLADINAMI

NAVYAC - UKAZATEĽ <u>PRST</u> NA NEJAKÝ LIST

OPERÁCIA PRST(x) - NASTAVÍ UKAZATEĽ Prst NA LIST, KTORÝ REPREZENTUJE NAJMENŠÍ PRVOK $\geq x$

(AK $x > \max S$, Prst BUDE UKAZOVAŤ NA NAJVÄČŠÍ LIST)

TIETO STROMY SÚ VHODNÉ PRE ÚLOHY, KDE ARGUMENTY RADY PO SEBE IDÚCICH OPERÁCIÍ SÚ V BLÍZKOM OKOLÍ NEJAKÉHO PRVKU $x \in U$.

14. 02. 2009 Sa
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## AMORTIZOVANÁ ZLOŽITOSŤ OPERÁCIÍ

$a \geq 2$, $b \geq 2a$ SÚ PEVNE DANÉ — ZREJME $c \geq 1$

$$\boxed{c := \min\left\{ \min\left(2a-1, \left\lceil \frac{b+1}{2} \right\rceil\right) - a,\ b - \max\left(2a-1, \left\lfloor \frac{b+1}{2} \right\rfloor\right) \right\}}$$

$\mathcal{P}$ JE POSTUPNOSŤ OPERÁCIÍ INSERT, DELETE.

$St_h$ = # ŠTIEPENÍ VO VÝŠKE $h$, $St = \sum_h St_h$

$Sp_h$ = # SPOJENÍ VO VÝŠKE $h$, $Sp = \sum_h Sp_h$

$P_h$ = # PRESUNOV VO VÝŠKE $h$, $P = \sum_h P_h$

POTOM PLATÍ **VETA**:

(1) $P \leq n$, $(2c-1)St + c\,Sp \leq n + c + \frac{c(n-2)}{a+c-1}$

(2) $St_h + Sp_h + P_h \leq \frac{2(c+2)n}{(c+1)^h}$

z (1) $St + Sp \leq \frac{n}{c} + 1 + \frac{n-2}{a} \leq \frac{3}{2}n + 1$

$\Rightarrow \lim_{n \to \infty} \frac{P + St + Sp}{n} \leq \frac{5}{2}$

---

DEFINUJME PRE VNÚTORNÝ $v \neq r$

$$\boxed{\begin{array}{l} b(v) := \min\{\rho(v) - a,\ b - \rho(v),\ c\} \\ \text{A PRE KOREŇ } b(r) = \min\{\rho(r) - 2,\ b - \rho(r),\ c\} \end{array}}$$

PRE VNÚTORNÝ $v \neq r$ PLATÍ:

(1) $b(v) \leq c$

(2) $\rho(v) = a \vee \rho(v) = b \Rightarrow b(v) = 0$

(3) $\rho(v) = a-1 \vee \rho(v) = b+1 \Rightarrow b(v) = -1$

(4) $\rho(v) = 2a-1 \Rightarrow b(v) = c$

!

AK $v'$ A $v''$ SÚ DVA RÔZNE VRCHOLY $\neq r$
$\rho(v') = \lceil \frac{b+1}{2} \rceil$, $\rho(v'') = \lfloor \frac{b+1}{2} \rfloor \Rightarrow b(v') + \underline{\underline{b(v'') \geq 2c-1}}$
$b(r) \leq c$

!

STROM $(T, r)$ OHODNOTÍME FCIOU $\boxed{b(T):}$
$b_h(T) := \sum \{ b(v) \mid v \neq r$, VNÚTORNÝ VRCHOL VO VÝŠKE $h \}$
$\underline{\underline{b(T) := \sum_{h=1}^{\infty} b_h(T)}} + b(r)$

$(T, r, v)$ JE <u>PARCIÁLNY $(a,b)$-STROM</u>, KEĎ $r$ JE KOREŇ STROMU, $v$ JE VNÚTORNÝ VRCHOL $T$ RÔZNY OD $r$ A PLATÍ: $a-1 \leq \rho(v) \leq b+1$, $2 \leq \rho(r) \leq b$
KEĎ $t$ JE VNÚTORNÝ VRCHOL $T$ RÔZNY OD $v$ A $r$, POTOM $a \leq \rho(t) \leq b$.
VŠETKY CESTY OD KOREŇA $r$ K LÍSTOM MAJÚ ROVNAKÚ DĹŽKU

◁ $v, v'$ VN. VRCHOLY $T, T'$ RÔZNE OD KOREŇOV
(1) $\rho(v) = \rho(v') \Rightarrow b_T(v) = b_{T'}(v')$
(2) $|\rho(v) - \rho(v')| = 1 \Rightarrow b_{T'}(v') \geq b_T(v) - 1$

!

<u>LEMMA</u>: $(T, r)$ $(a,b)$-STROM, $T'$ VZNIKNE Z $T$ ODBRANÍM / PRIDANÍM <u>VRCHOLU VO VÝŠKE 1</u> (TJ. $\pm 1$ LIST) POTOM $(T', r, v)$ JE PARCIÁLNY $(a,b)$-STROM A
$\boxed{b_1(T') \geq b_1(T) - 1}$, $b_h(T') = b_h(T)$ PRE $h > 1$, $\Rightarrow b(T') \geq b(T) - 1$

<u>LEMMA</u>: $(T, r, v)$ PARCIÁLNY $(a,b)$-STROM, $v$ VRCHOL VO VÝŠKE $\ell \geq 1$, $\rho(v) = b+1$, $T'$ VZNIKNE Z $T$ OP. $\boxed{\text{ŠTIEPENIE } (v)}$, POTOM $(T', r, \text{OTEC}(v))$ JE PARC. $(a,b)$-S.
$\boxed{b_\ell(T') \geq b_\ell(T) + 2c, \quad b_{\ell+1}(T') \geq b_{\ell+1}(T) - 1}$, $b_h(T') = b_h(T)$, $h \neq \ell, \ell+1$
$\Rightarrow b(T') \geq b(T) + 2c - 1$

14.02.2009 Sa
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

**LEMMA**: $(T, r, v)$ PARCIÁLNY $(a,b)$-STROM, $v$ VRCHOL VO VÝŠKE $\ell \geq 1$, $\rho(v) = a-1$, $y$ BEZPROSTREDNÝ BRAT $v$ TAKÝ, ŽE $\rho(y) = a$, $T'$ VZNIKNE Z $T$ OP. $\boxed{\text{SPOJENIE}(v,y)}$, POTOM $(T', r, \text{OTEC}(v))$ JE P. $(a,b)$-S.

$\boxed{b_\ell(T') \geq b_\ell(T) + c+1,\ b_{\ell+1}(T') \geq b_{\ell+1}(T) - 1}$ $b_h(T') = b_h(T)$ $h \neq \ell, \ell+1$

$\Rightarrow b(T') \geq b(T) + c$

**LEMMA**: $(T, r, v)$ PARCIÁLNY $(a,b)$-STROM, $v$ VRCHOL VO VÝŠKE $\ell \geq 1$, $\rho(v) = a-1$, $y$ BEZPROSTREDNÝ BRAT $v$ TAKÝ, ŽE $\rho(y) > a$, $T'$ VZNIKNE Z $T$ OP. $\boxed{\text{PRESUN}(v,y)}$, POTOM $(T', r)$ JE $(a,b)$-STROM

$\boxed{b_\ell(T') \geq b_\ell(T),}$ $b_h(T') = b_h(T)$ $h \neq \ell$, $b(T') \geq b(T)$

**DÔSLEDOK !!**

NECH $T_k$ VZNIKNE VYKONANÍM POSTUPNOSTI $\mathcal{P}$ NA PÔVODNE PRÁZDNOM STROM. POTOM, [AK OZNAČÍME $St_0 + Sp_0 = \#$LISTOV V $T_k$]

$$\boxed{b_h(T_k) \geq 2c\, St_h + (c+1)\, Sp_h - St_{h-1} - Sp_{h-1}} \quad h \geq 1$$

ĎALEJ $\boxed{b(T_k) \geq (2c-1)\, St + c\, Sp - n}$, KDE $n = |\mathcal{P}|$

KEĎŽE $c \geq 1$, TO $2c \geq c+1$, NUTNE

$$St_h + Sp_h \leq \frac{b_h(T_k)}{c+1} + \frac{St_{h-1} + Sp_{h-1}}{c+1} \leq \dots \leq \sum_{i=0}^{h-1} \frac{b_{h-i}(T_k)}{(c+1)^{i+1}} + \frac{n}{(c+1)^h}$$

$$= \frac{n}{(c+1)^h} + \sum_{\ell=1}^{h} b_\ell(T_k) \frac{(c+1)^\ell}{(c+1)^{h+1}}$$

## ODHAD $b(T_k)$ ZHORA

LEMMA: KEĎ $T$ JE $(a,b)$-STROM S $m$ LISTAMI, POTOM $0 \le \boxed{b(T) \le c + (m-2)\frac{c}{a+c-1}}$ !

$$\boxed{\begin{array}{l} 0 \le j < c \quad m_j = \text{POČET VN. VRCHOLOV} \neq \text{KOREŇA S } a+j \text{ SYNMI} \\ m_c = \text{POČET VN. VRCHOLOV} \neq \text{KOREŇA S ASPOŇ } a+c \text{ SYNMI} \end{array}}$$

KEĎ MÁ $v$ $\underline{a+j \text{ SYNOV}}$, POTOM $\underline{b_T(v) \le j}$, $\forall$ VN. $v$ $\underline{b_T(v) \le c}$

$\Rightarrow b(T) \le c + \sum_{j=0}^{c} j\, m_j$ (PRVÉ $c$ JE ODHAD KOREŇA)

$$\boxed{2 + \sum_{j=0}^{c}(a+j)m_j \le \sum\{\rho(v) \mid v \text{ JE VN. VRCHOL } T\} = m + \sum_{j=0}^{c} m_j}$$

$\Rightarrow \sum_{j=0}^{c}(a+j-1)m_j \le m-2$

$$b(T) \le c + \sum_{j=0}^{c} j \cdot m_j = c + \sum_{j=0}^{c} \underbrace{\frac{j}{a+j-1}}_{\le \frac{c}{a+c-1} \quad 0 \le j \le c}(a+j-1)m_j$$

$$\le c + \frac{c}{a+c-1}(m-2).$$

TERAZ UŽ DOKÁŽEME (1) Z VETY:

KAŽDÁ OP. DELETE POUŽIJE NAJVIAC 1 OP. PRESUN

$\Rightarrow P \le$ POČET OP. DELETE $\le n$ ✓

$$(2c-1)St + c\,Sp - n \le b(T_k) \le c + (n-2)\frac{c}{a+c-1} \quad ✓$$

LEMMA: $\forall h \ge 1$ A PRE KAŽDÝ $(a,b)$-STROM $T$ S $m$ LISTAMI PLATÍ: $\sum_{\ell=1}^{h} b_\ell(T)(c+1)^\ell \le (c+1)m$

$$\Rightarrow St_h + Sp_h \le \frac{n}{(c+1)^h} + \sum_{\ell=1}^{h} b_\ell(T_k)\frac{(c+1)^\ell}{(c+1)^{h+1}} \le \frac{2n}{(c+1)^h}$$

$$P_h \le Sp_{h-1} - Sp_h \le St_{h-1} + Sp_{h-1}$$

$$\Rightarrow \boxed{St_h + Sp_h + P_h \le \frac{2n(c+2)}{(c+1)^h}} \quad (2)$$

15.02.2009 Su
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## VYHĽADÁVANIE V USP. POLI

```
d:=1, h:=|S|
...
while d+1 < h do
  n:=next(d,h)
  ...
```

UNÁRNE VYHĽADÁVANIE: $next(d,h) = d+1$ $O(|S|)$

BINÁRNE VYHĽADÁVANIE: $next(d,h) = \lceil \frac{d+h}{2} \rceil$ $O(\log |S|)$

INTERPOLAČNÉ VYHĽADÁVANIE:

$$next(d,h) = d + \left\lceil \frac{x - A(d)}{A(h) - A(d)} (h-d) \right\rceil$$

NAJHORŠÍ PRÍPAD: VIAC AKO $|S|/2$ DOTAZOV
PRI ROVNOMERNOM ROZLOŽENÍ VSTUPU JE
OČAKÁVANÝ ČAS $O(\log \log |S|)$

ZOVŠEOBECNENÉ KVADRATICKÉ VYHĽADÁVANIE

PROCEDÚRA PRACUJE V BLOKOCH A DOTAZY ZADANÉ
VRÁTKI TOMO ISTÉHO BLOKU SÚ MEDZI SEBOU KORELOVANÉ
PRVÝ DOTAZ JE INTERPOLAČNÝ, STANOVÍ VELKOSŤ
KROKU PRE ĎALŠIE PRIBLÍŽENIE, POTOM STRIEDA
UNÁRNE A BINÁRNE DOTAZY, KEĎ NÁJDE HĽADANÝ
PRVOK ALEBO KEĎ SKONČÍ PRÁCU V BLOKU

BLOK KONČÍ, KEĎ ROZDIEL MEDZI $h$ A $d$ JE NAJVIAC
ROVNÝ VELKOSTI KROKU

KROK V KAŽDOM NASL. BLOKU KLESNE VŽDY PRIBLIŽNE
NA ODMOCNINU VELKOSTI KROKU V PREDCH. BLOKU ...

```
next(d, h)
  if BLOK then
    if TYP then TYP := FALSE
      if SMER then
        next(d,h) := h - KROK
        if A(..) < x then BLOK := FALSE
      else
        next(d,h) := d + KROK
        if A(...) > x then BLOK := FALSE
    else TYP := TRUE
      if KROK > min { ⌈(d+h)/2⌉ - d, h - ⌈(d+h)/2⌉ }
        then BLOK := FALSE
      next(d,h) := ⌈(d+h)/2⌉
  else
    KROK := ⌊√(h-d)⌋
    next(d,h) := d + ⌈ (x - A(d)) / (A(h) - A(d)) (h-d) ⌉
    if A(..) > x then SMER := TRUE
    else SMER := FALSE
    BLOK := TRUE
    TYP := TRUE
```

ZLOŽITOST next JE $O(1)$

15.02.2009 S4
PETER ČERNO

# NTIN066 DÁTOVÉ ŠTRUKTÚRY I

## OČAKÁVANÝ POČET DOTAZOV V 1 BLOKU

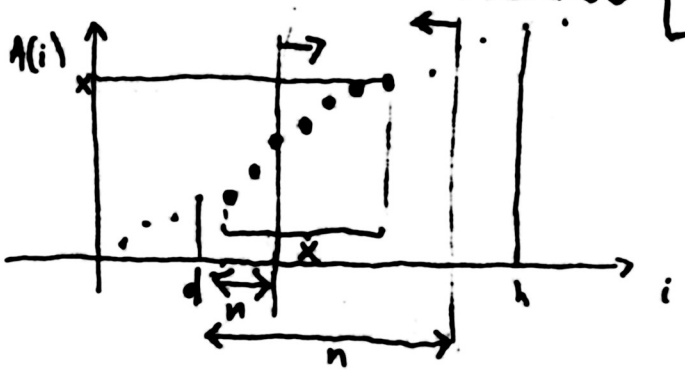
$p_i$ = PRAVDEPODOBNOSŤ, ŽE VRÁTI BLOKU SA POLOŽÍ ASPOŇ $i$ DOTAZOV

$\Rightarrow$ OČAKÁVANÝ POČET DOTAZOV $C$ VRÁTI BLOKU JE

$$C = \sum_{i \geq 1} i(p_i - p_{i+1}) = \sum_{i \geq 1} p_i$$

OZNAČME:

$h+d$ ARGUMENT PRVÉHO (INTERPOLAČNÉHO) DOTAZU
$k$ VEĽKOSŤ KROKU
$X$ NÁHODNÚ PREMENNÚ $\boxed{X = |\{i \mid i > d, A(i) \leq x\}|}$




KEĎ SA V BLOKU POLOŽÍ ASPOŇ $i$ DOTAZOV, $i > 2$,
POTOM $|X - n| \geq \lfloor \frac{i-2}{2} k \rfloor$

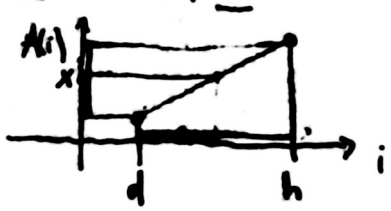
KAŽDÝ UNÁRNY DOTAZ, PO KTOROM NESKONČIL BLOK, NAŠIEL ĎALŠÍCH $k$ HODNÔT V ROZDIELI $|X-n|$

$$p_i \leq P\left(|X-n| \geq \lfloor \tfrac{i-2}{2} k \rfloor\right)$$

ČEBYŠEVOVA NEROVNOSŤ: $P(|Y - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \quad \forall \varepsilon > 0$

PRAVDEPODOBNOSŤ, ŽE $A(i) \leq x$ PRE $d < i < h$ JE

$$p = \frac{x - A(d)}{A(h) - A(d)}$$




$$\Rightarrow P(X=j) = \binom{h-d}{j} p^j (1-p)^{h-d-j} \qquad X \sim Bi(h-d, p)$$

$$\Rightarrow \mu = p(h-d), \quad \sigma^2 = p(1-p)(h-d)$$

Z ČEBYŠEVOVEJ NEROVNOSTI, KDE $k = \lfloor \sqrt{h-d} \rfloor$, $n = p(h-d) = \mu$

$$P\left(|X-n| \geq \left\lfloor \frac{i-2}{2} k \right\rfloor\right) \leq \frac{4p(1-p)(h-d)}{(i-2)^2 k^2} \leq \frac{4p(1-p)}{(i-2)^2} \leq \frac{1}{(i-2)^2}$$

$$\Rightarrow C = \sum_{i \geq 1} p_i \leq 2 + 2\sum_{i \geq 3} \frac{1}{(i-2)^2} = 2 + 2\frac{\pi^2}{6} < 6$$

## ODHAD POČTU BLOKOV

$krok \leq \lfloor \sqrt{|S|} \rfloor$, $krok_{i+1} \leq \lfloor \sqrt{krok_i} \rfloor$

$\Rightarrow krok_i \leq \sqrt[2^i]{|S|}$

$2 = \sqrt[2^j]{|S|} \Rightarrow j = \log_2 \log_2 |S|$

$\Rightarrow$ POČET BLOKOV JE NAJVIAC $1 + \log \log |S|$

OČAKÁVANÝ POČET DOTAZOV JE $C(1 + \log \log |S|)$

NAJHORŠÍ PRÍPAD: BINÁRNYCH DOTAZOV JE NAJVIAC TOĽKO, KOĽKO ICH JE POLOŽENÝCH PRI BINÁRNOM VYHĽADÁVANÍ .. $\leq 3 + \log(|S| - 1)$

UNÁRNYCH JE NAJVIAC $3 + \log(|S|-1) + \log \log |S| + 1$

INTERPOLAČNÝCH $1 + \log \log |S|$

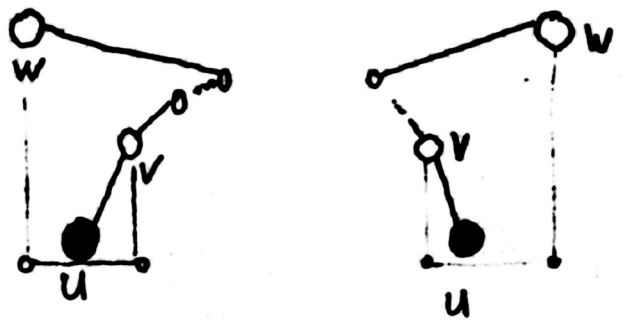
$\Rightarrow$ CELKOM $O(\log |S|)$ DOTAZOV.

16.02.2009 Mo
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## BINÁRNE VYHĽADÁVACIE STROMY

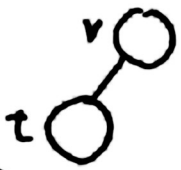
KAŽDÝ LIST REPREZENTUJE INTERVAL MEDZI DVOMA SUSEDNÝMI PRVKAMI:



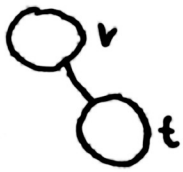

BUDEME PREDPOKLADAŤ, ŽE STROMY MAJÚ LISTY REPREZ. INTERVALY. HODNOTY BUDÚ ULOŽENÉ IBA VO VNÚTORNÝCH UZLOCH.

## KOREKTNOSŤ PROCEDÚRY VYHLADAJ

lhr(KOREŇ) = phr(KOREŇ) = NDF




phr(t) = v
lhr(t) = lhr(v)




lhr(t) = v
phr(t) = phr(v)

$$\lambda(t) = \begin{cases} -\infty & \text{lhr}(t) \text{ NDF} \\ \text{key}(\text{lhr}(t)) & \text{INAK} \end{cases} \qquad \pi(t) = \begin{cases} +\infty & \text{phr}(t) \text{ NDF} \\ \text{key}(\text{phr}(t)) & \text{INAK} \end{cases}$$

LEMMA: KEĎ JE T' PODSTROM T URČENÝ VRCHOLOM t', POTOM T' REPREZENTUJE $S \cap (\lambda(t), \pi(t))$ A $(\lambda(t), \pi(t))$ JE NAJVÄČŠÍ INTERVAL S TOTO VLASTNOSŤOU.

KOREKTNOSŤ VYHLADAJ PLYNIE Z NASL. INVARIANTU

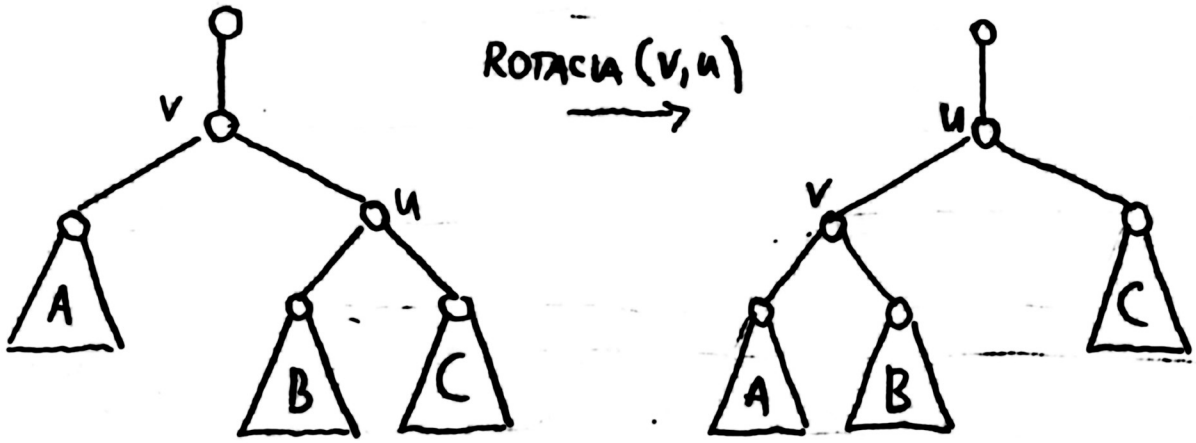
KEĎ PRI VYHĽADÁVANÍ x VSTÚPIME VRCHOL t, POTOM PLATÍ $\lambda(t) < x < \pi(t)$.

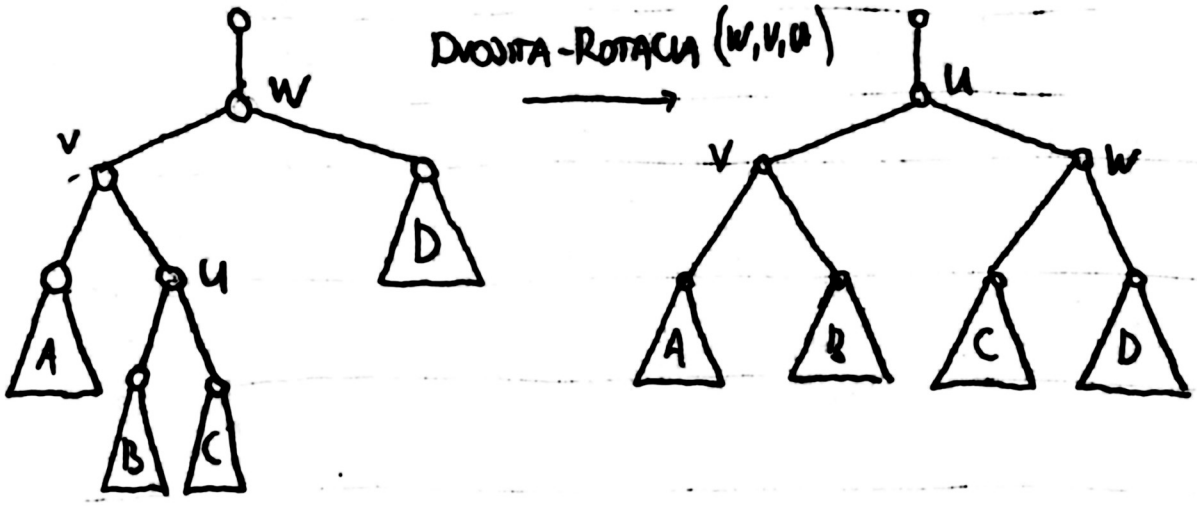
ČASOVÁ ZLOŽITOSŤ OPERÁCIÍ: $O(\text{výška}(T))$.

## PORIADKOVÉ ŠTATISTIKY : Ord(k)

ROZŠÍRENIE DAT. ŠTRUKTÚRY O $p(t)$ = POČET LISTOV V PODSTROME URČENOM VRCHOLOM $t$.

## VYVÁŽENÉ BINÁRNE VYHĽ. STROMY



16.02.2009 M.
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

## AVL STROMY

$\eta(v)$ = DĽŽKA NAJDLHŠEJ CESTY Z VN. VRCHOLU $v$ DO LISTOV (VÝŠKA PODSTROMU)

STRUKTÚRA VNÚTORNÝCH VRCHOLOV ROZŠÍRENÁ O $\omega$:

$\omega(v) := \eta(\text{ĽAVÝ SYN } v) - \eta(\text{PRAVÝ SYN } v)$

V AVL STROME VŽDY $|\omega(v)| \leq 1$.

OZNAČME:

$mn(i)$ = VEĽKOSŤ NAJMENŠEJ MNOŽINY REPREZENTOV. AVL-STROMOM $(T,t)$, PRE KTORÝ $\eta(t)=i$

$mx(i)$ = VEĽKOSŤ NAJVÄČŠEJ —||—

Z DEFINÍCIE: $mn(i) = mn(i-1) + mn(i-2) + 1$

$mx(i) = 2mx(i-1) + 1$

$mn(1) = 1 \qquad mx(1) = 1$

$mn(2) = 2 \qquad mx(2) = 3$

$\Rightarrow mx(i) = 2^i - 1, \quad mn(i) = F_{i+2} - 1$, KDE $F_i = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^i - \left(\frac{1-\sqrt{5}}{2}\right)^i\right]$

TJ. AK AVL STROM O VÝŠKE $i$ REPREZENTUJE $S$, $|S| = n$, POTOM: $F_{i+2} - 1 \leq n \leq 2^i - 1$

$\Rightarrow \log(n+1) \leq i \leq 1{,}44 \log(n+1)$ TJ. $\underline{\underline{i \in \Theta(\log n)}}$

PO INJERT(x) ZAVOLÁME KONTROLA-INSERT(t)

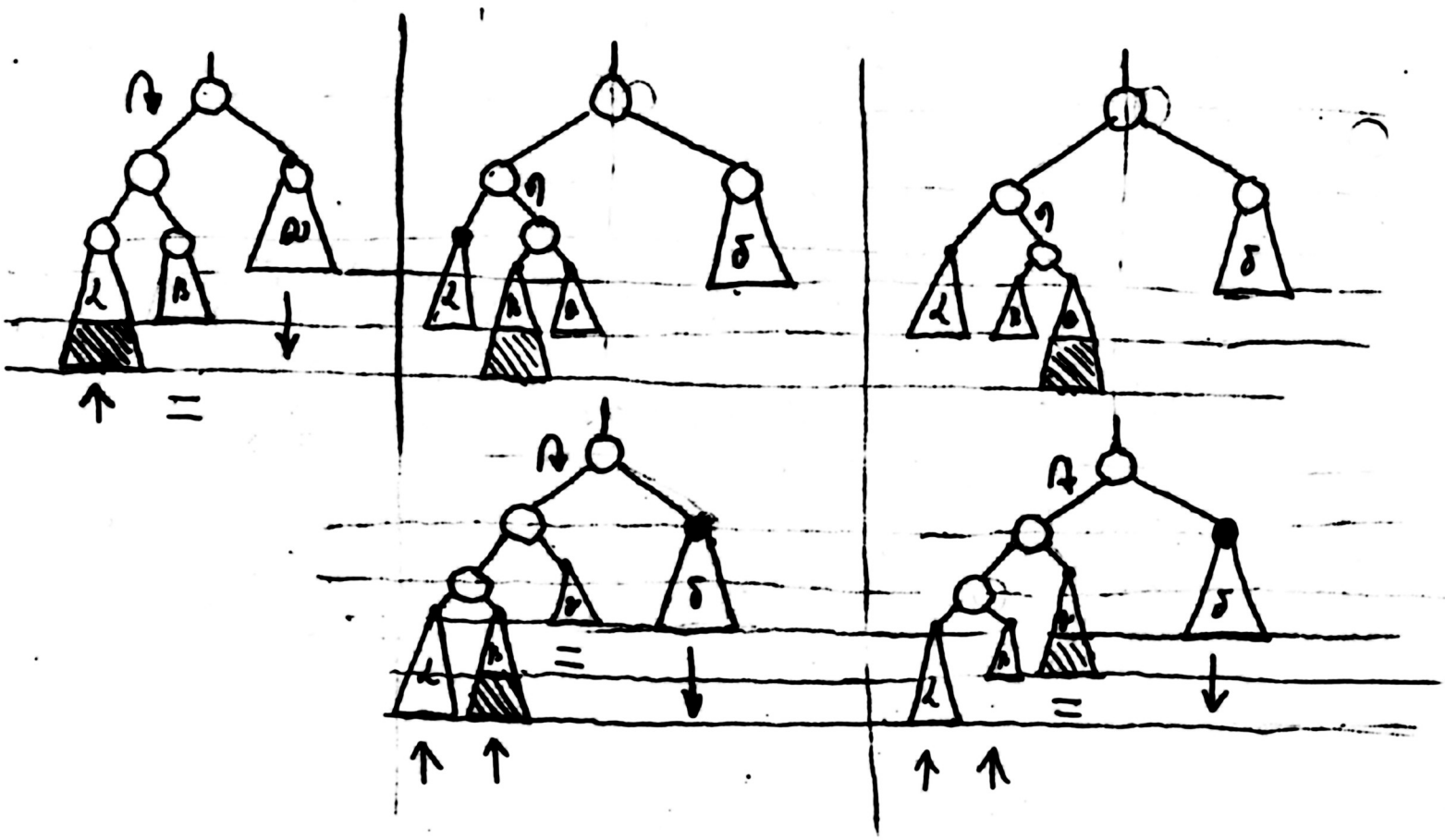
PO VYKONANÍ <u>ROTACIA</u> ALEBO <u>DVOJITÁ-ROTACIA</u> VYVÁŽOVANIE V OP. INSERT KONČÍ.

<u>LEMMA</u>: KEĎ SA HODNOTA $\eta(t)$ PRI OPERÁCII INSERT ZVÄČŠILA A $t$ NEBOL LIST, POTOM PO OPERÁCII JE $w(t) \neq 0$.

TJ. NENASTANE PRÍPAD $w(t)=0$, KT. BY SA ŤAŽKO OTESTOVAL.

DELETE: NECH $t$ JE VRCHOL, KT. OTEC BOL ODSTRÁNENÝ, TJ. BRAT $t$ BOL LIST, A HODNOTA $\eta(t)$ JE MENŠIA AKO $\eta(\text{OTEC}(t))$ ... ZAVOLÁME KONTROLA-DELETE(t).

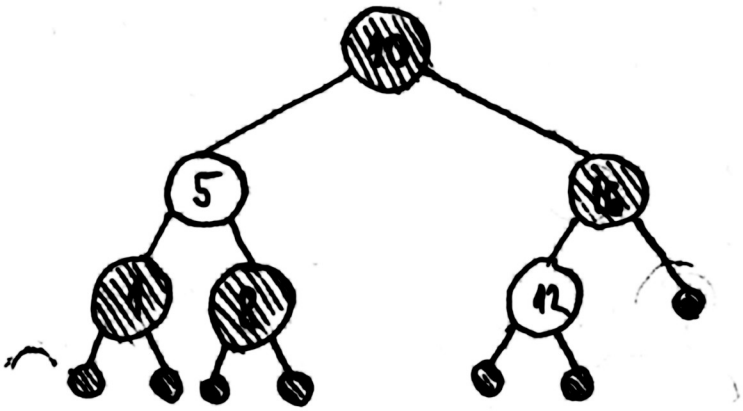
PROCEDÚRY ROTACIA ALEBO DVOJITÁ-ROTACIA MÔŽU BYŤ VOLANÉ $\log(|S|)$ KRÁT.



16.02.2009 M.
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## ČERVENO-ČIERNE STROMY




JE BIN. VYHĽ. STROM TAKÝ, ŽE:

(1) KAŽDÝ UZOL JE BUĎ ČERVENÝ ALEBO ČIERNY
(2) KOREŇ JE ČIERNY
(3) KAŽDÝ LIST JE ČIERNY.
(4) KEĎ JE UZOL ČERVENÝ, POTOM OBA JEHO POTOMKOVIA SÚ ČIERNI
(5) PRE KAŽDÝ UZOL $x$: $\forall$ CESTY Z UZLU $x$ DO LISTOV OBSAHUJÚ TEN ISTÝ POČET ČIERNYCH UZLOV

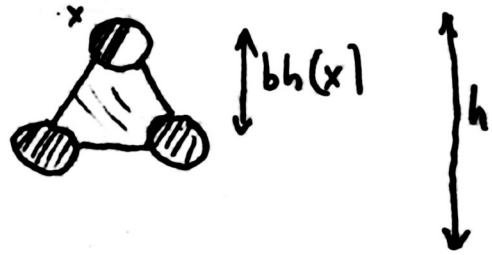
$bh(x)$ = # ČIERNYCH UZLOV NA CESTE Z $x$ DO LISTOV, OKREM UZLU $x$ !!!

<u>LEMMA</u>: KAŽDÝ ČERVENO-ČIERNY PODSTROM S KOREŇOM V $x$ OBSAHUJE ASPOŇ $2^{bh(x)}-1$ VNÚTORNÝCH UZLOV (INDUKCIA PODĽA $h(x)$)

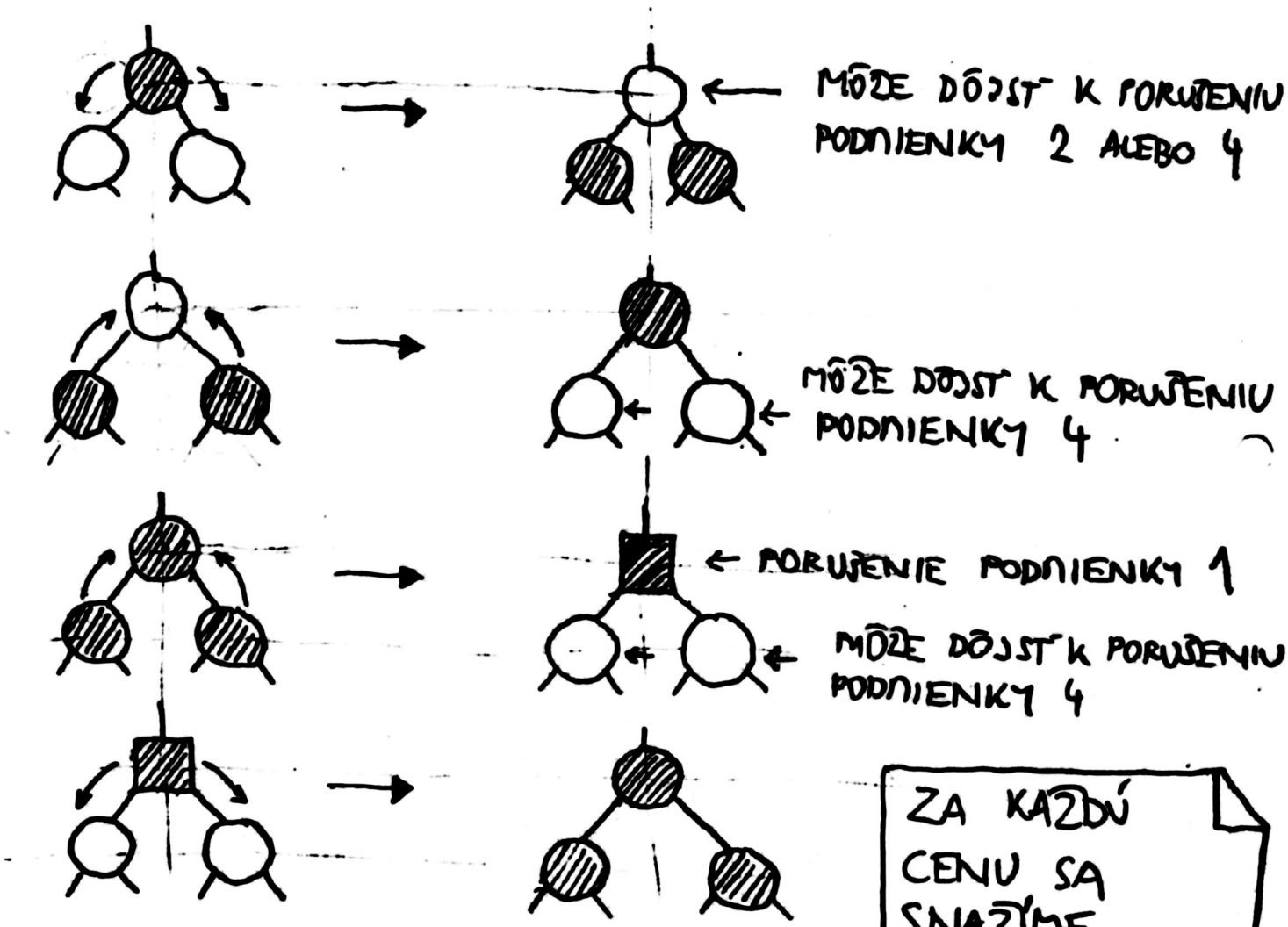
PLATÍ $\frac{\text{\# NAJDLHŠIA CESTA}}{\text{\# NAJKRATŠIA CESTA}} \leq 2$ TJ.

$bh(x) \geq \frac{1}{2} h(x) \Rightarrow n \geq 2^{h/2}-1$

$\Rightarrow \boxed{h \leq 2 \log_2 (n+1)}$



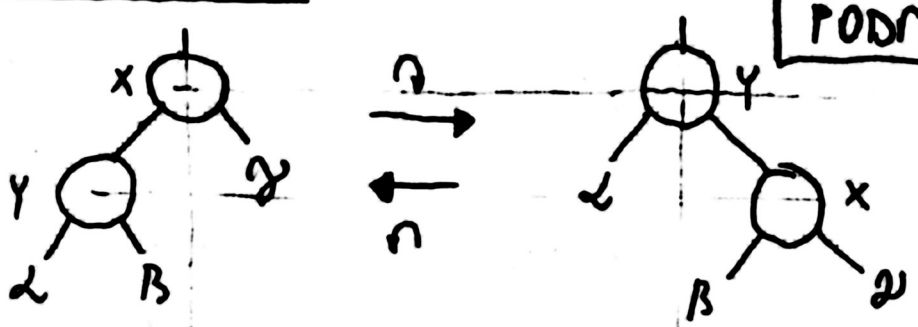
## POVOLENÉ FAREBNÉ ZÁSENY



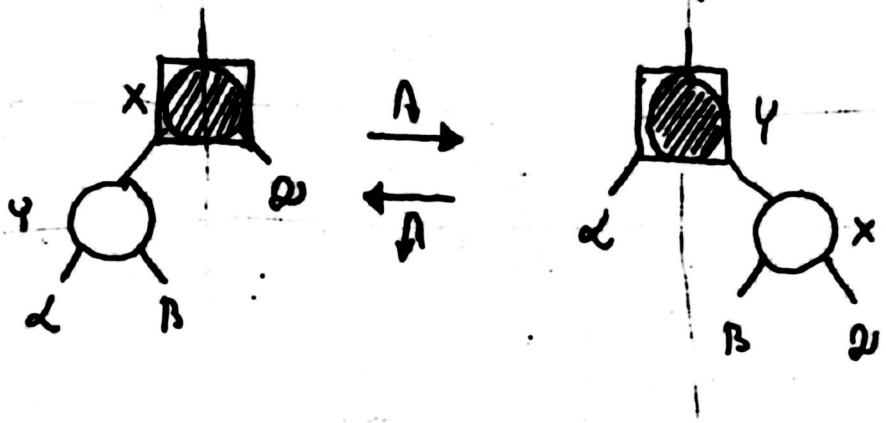

ZA KAŽDÚ CENU SA SNAŽÍME ZACHOVAŤ PODMIENKU 5

## POVOLENÉ ROTÁCIE

TYP 1:




TYP 2:



16.02.2009 M.
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

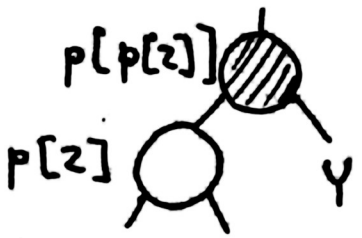
```
RB-INSERT(T, z)
    BST-INSERT(T, z)
    color[z] ← RED
    RB-INSERT-FIXUP(T, z)

RB-INSERT-FIXUP(T, z)
    while color[p[z]] = RED do
    [ ...
    color[root[T]] ← BLACK
```

**INVARIANT:** NA ZAČIATKU CYKLU JE color[z] = RED

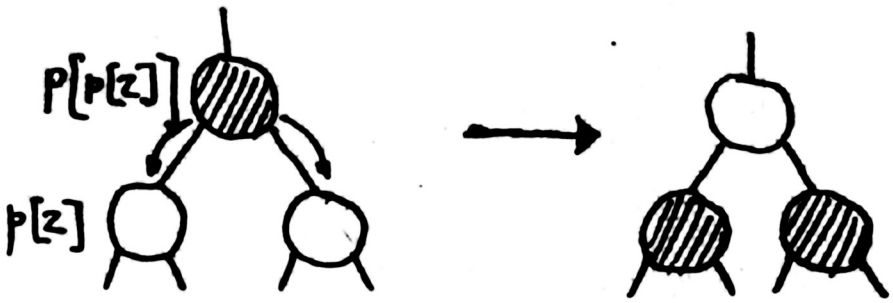
BÚNO NECH p[z] = left[p[p[z]]]
POLOŽME y ← right[p[p[z]]]

p[p[z]] URČITE EXISTUJE A JE ČERNÝ



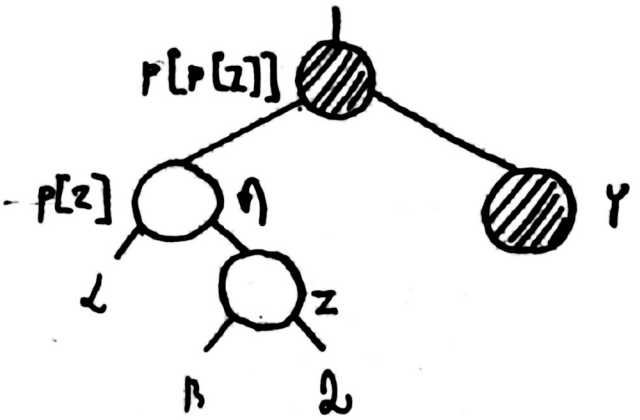

ROZLIŠUJEME 2 PRÍPADY:
1) color[y] = RED
2) color[y] = BLACK

(1) color[y] = RED



(2) <u>color[y] = BLACK</u>

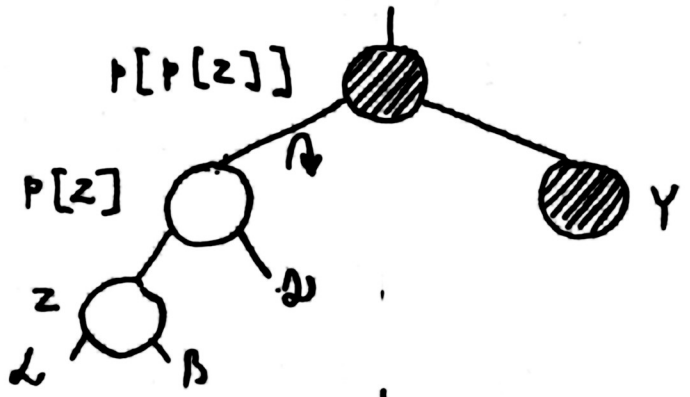
(a) z = right[p[z]].



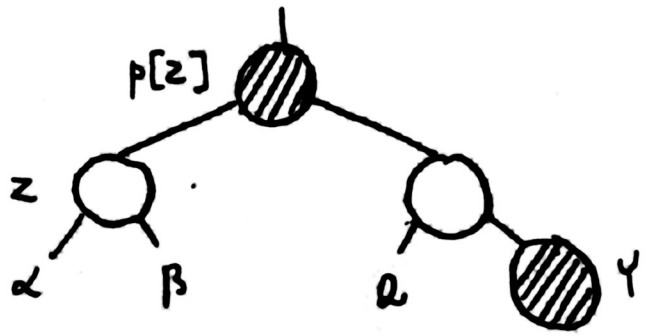

$z \leftarrow p[z]$
LEFT-ROTATE(T, z)

PREVEDIEME NA PRÍPAD (b)

(b) z = left[p[z]]




RIGHT-ROTÁCIA TYPU 2




VÝSLEDNÝ STROM UŽ JE KOREKTNÝ RB-TREE

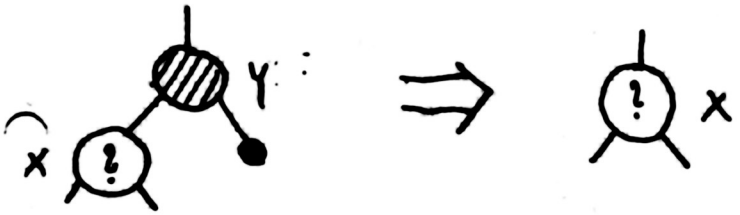
<u>POZNÁMKA</u> : MAX. 2 ROTÁCIE.

16.02.2009 M.
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## RB-DELETE (T, z)

POSTUPUJEME AKO PRI BST-DELETE (T, z)
POKIAĽ MAL VRCHOL, KTORÝ SME SKUTOČNE VYNECHALI, ČIERNU FARBU, ZAVOLÁME
RB-DELETE-FIXUP (T, x), KDE x JE AKO NA OBR.:




## RB-DELETE-FIXUP (T, x)

x JE DOUBLEBLACK !

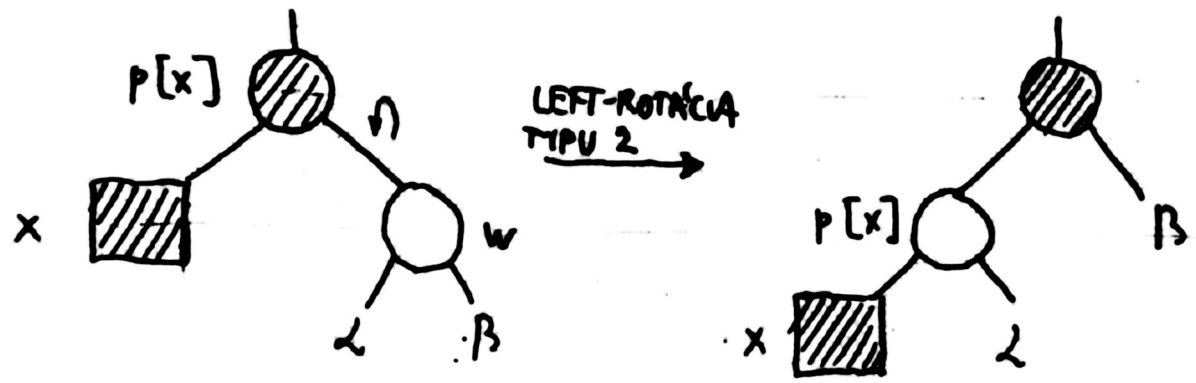
```
while x ≠ root[T] & color[x] = BLACK do
[ ...
color[x] ← BLACK
```

BÚNO x = left[p[x]]
POLOŽME w = right[p[x]]

p[x] URČITE EXISTUJE A JE ≠ NIL

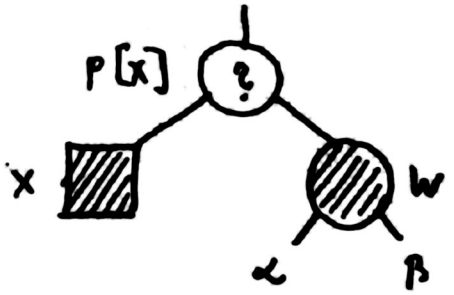
2 PRÍPADY:
1) color[w] = RED
2) color[w] = BLACK

(1) color[w] = RED



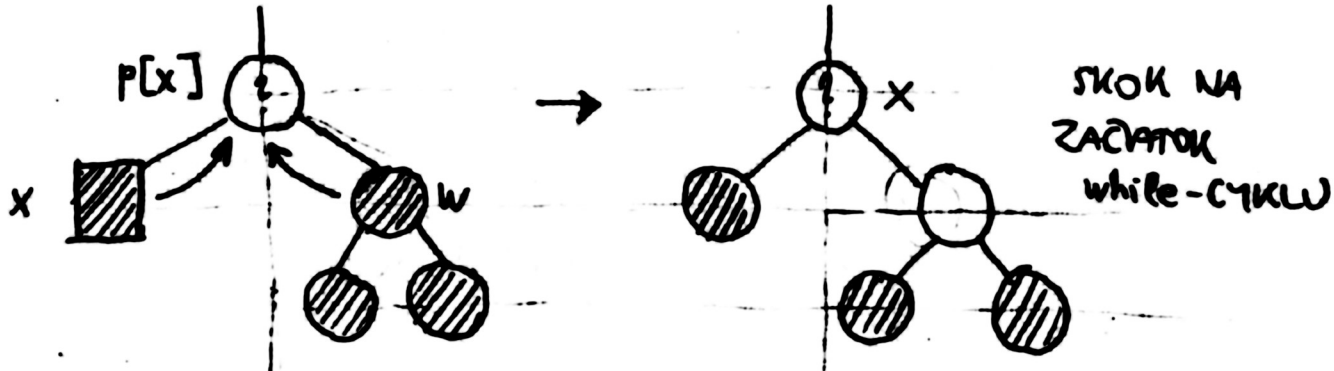

PO w ← right[p[x]] MÁME PRÍPAD 2 !

(2) <u>color[w] = BLACK</u>

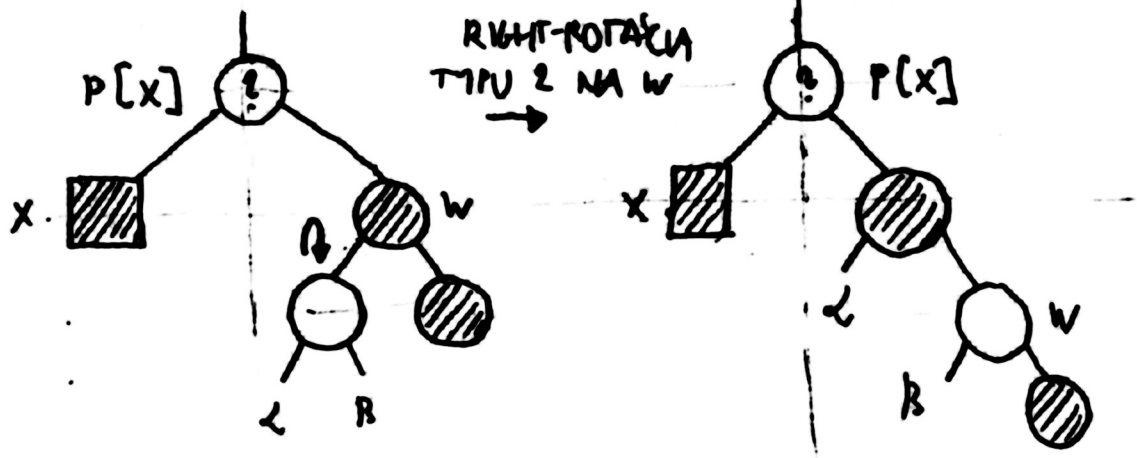



w NIE JE LIST, TJ. MUSÍ MAŤ NÁSLEDNÍKOV, PRETOŽE x JE DOUBLEBLACK !!

a) color[left[w]] = color[right[w]] = BLACK




b) color[right[w]] = BLACK

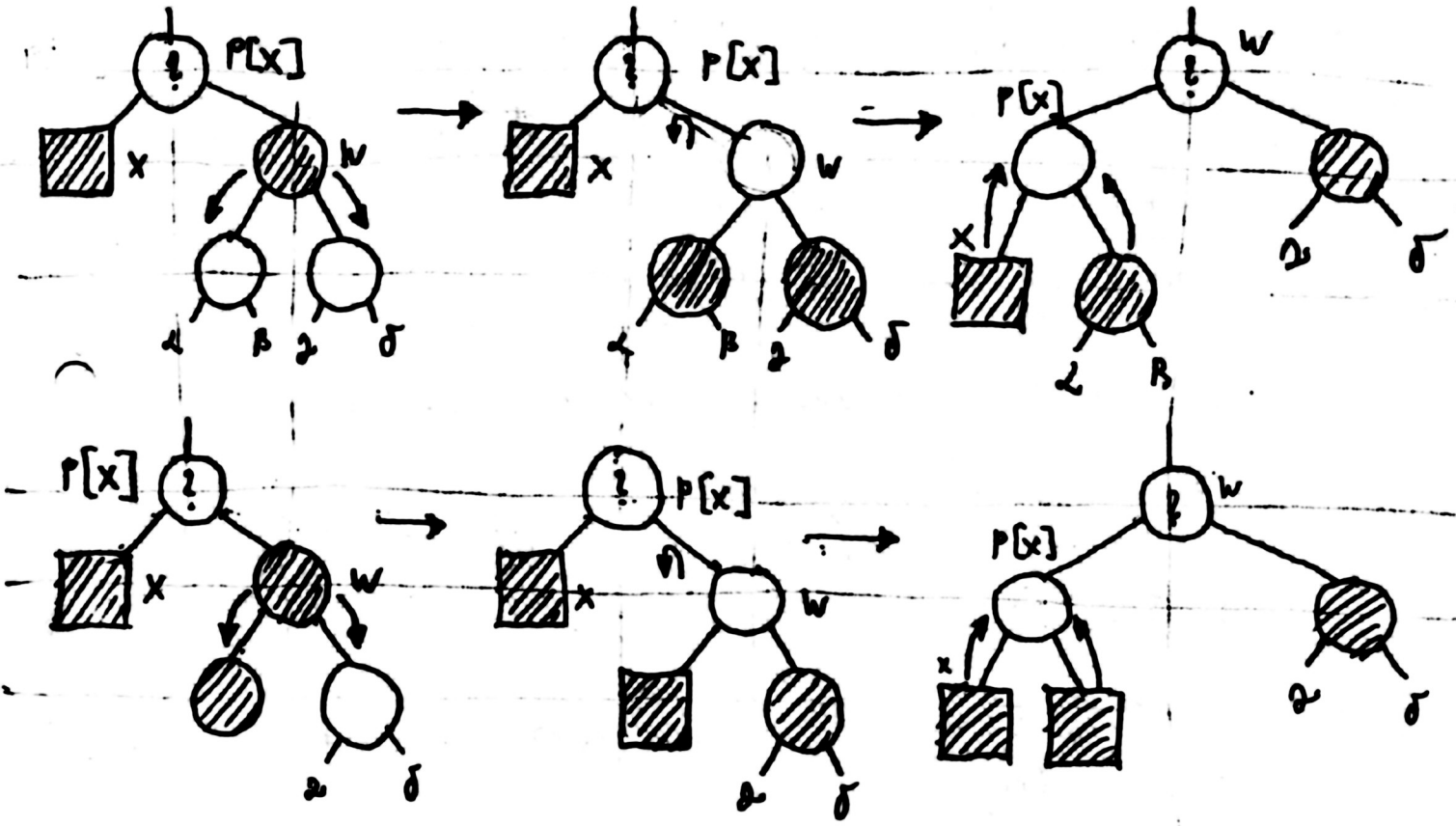



POLOŽÍME w ← right[p[x]]
A DOSTÁVAME PRÍPAD c)

16.02.2009 Mo
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

c) color [right[w]] = RED




POLOŽÍME $x \leftarrow root[T]$ A SME HOTOVÍ !

PRÍPAD 1 SA PREVEDIE NA PRÍPAD 2 A TAM TO SKONČÍ V BODE 2a)

POSÚVAME SA VYSSIE VŽDY LEN V 2a) (AK NEBOLA ROTÁCIA)

PRÍPAD 2b) SA PREVEDIE NA 2c) A TAM TO SKONČÍ

$\Rightarrow$ NAJVIAC 2 ROTÁCIE !

# VÁHOVO VYVÁŽENÉ STROMY

MAJME $\frac{1}{4} < \alpha \leq \frac{\sqrt{2}}{2}$ , $p(T)$ = POČET LISTOV V STROME $T$

BINÁRNY VYHĽADÁVACÍ STROM $T$ REPREZENTUJÚCI $S$ SA NAZÝVA $BB(\alpha)$-STROM $\Leftrightarrow$ $\forall$ VNÚTORNÝ VRCHOL $v$ :

$$\boxed{\alpha \leq \frac{p(T_{left(v)})}{p(T_v)} = 1 - \frac{p(T_{right(v)})}{p(T_v)} \leq 1 - \alpha}$$

PRE $BB(\alpha)$ REPR. $n$-PRVKOVÚ MNOŽINU PLATÍ :

$$\text{výška}(T) \leq 1 + \frac{\log(n+1) - 1}{\log \frac{1}{1-\alpha}}$$

TECHNICKÉ TVRDENIE : AKO VYVAŽOVAŤ $BB(\alpha)$ STROMY PRI AKTUALIZAČNÝCH OP.

VETA : PRE $1/4 < \alpha < 1 - \sqrt{2}/2$ EXISTUJE $c > 0$ POSTUPNOSŤ AKTUAL. OP. NA PRÁZDNOM STROM VOLÁ NAJVIAC $cn$ PROCEDÚR Rotacia A Dvojita-Rotacia.

17.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## HALDY A TRIEDIACE ALGORITMY

$S \subseteq U$, $f: S \to \mathbb{R}$ REÁLNE USPORIADANIE

INSERT$(s, a)$ PRIDÁ $s$ DO $S$, $f(s) = a$, PREDP. $s \notin S$

MIN, DELETEMIN, DELETE$(s)$, INCREASE / DECREASE $(s, a)$

HALDA - MINHEAP

MAKEHEAP$(S, f)$, MERGE$(H_1, H_2)$

### d-REGULÁRNY STROM

$n$ VRCHOLOV

VÝŠKA JE $\lceil \log_d(n(d-1)+1) \rceil$

VRCHOL $k$ MÁ SYNOV $(k-1)d+2, \dots, kd+1$

A OTCA $1 + \lfloor \frac{k-2}{d} \rfloor$

d-REG. STROM S VÝŠKOU $k$ MÁ $\sum_{i=0}^{k-1} d^i + 1 \le n \le \sum_{i=0}^{k} d^i$ VRCH.

$\Rightarrow d^k - 1 < n(d-1) \le d^{k+1} - 1 \Rightarrow k < \log_d(n(d-1)+1) \le k+1$

NIE JE ZNÁMA EFEKTÍVNA IMPL. MERGE

POMOCNÉ OPERÁCIE: UP A DOWN

### ZLOŽITOSŤ OPERÁCIÍ

UP: $O(\log_d |S|)$

DOWN: $O(d \log_d |S|)$

OPERÁCIA DOWN$(v)$ NA VRCHOL VO VÝŠKE $h$ VYŽADUJE V NAJHORŠIEM PRÍPADE ČAS $O(hd)$

VRCHOLOV V HĹBKE $i$ JE NAJVIAC $d^i$

AK MÁ STROM VÝŠKU $k$, POTOM VRCHOL V HĹBKE $i$ MÁ VÝŠKU NAJVIAC $k-i$.

$\Rightarrow$ OP. MAKEHEAP MŽADUJE ČAS

$O\left(\sum_{i=0}^{k-1} d^i (k-i) d\right)$, OZNAČME $\underline{A = \sum_{i=0}^{k-1} d^{i+1}(k-i)}$

$$\boxed{dA - A} = \sum_{i=0}^{k-1} d^{i+2}(k-i) - \sum_{i=0}^{k-1} d^{i+1}(k-i) =$$

$$= \sum_{i=2}^{k+1} d^i (k-i+2) - \sum_{i=1}^{k} d^i (k-i+1) =$$

$$= d^{k+1} + \sum_{i=2}^{k} d^i (k-i+2-k+i-1) - dk =$$

$$= \underline{\underline{d^{k+1} + d^2 \frac{d^{k-1}-1}{d-1} - dk}}$$

KEDŽE $k = \lceil \log_d (|S|(d-1)+1) \rceil \Rightarrow d^{k+1} \leq d^2 [(d-1)|S|+1]$

$\Rightarrow \boxed{A \leq 2d^2 |S|}$

<u>d-HEAPSORT</u> (PRE $n \leq 10^6$ ODPORUČENÉ $d = 6, 7$)

## DIJKSTROV ALG.

$\underline{G = (V, E, c), \quad c: E \to \mathbb{R}_0^+, \quad z \in V}$

```
d[z] ← 0,  Q ← {z}
foreach x ∈ V \ {z} do d[x] ← +∞
while Q ≠ ∅ do
    u ← EXTRACT-MIN (Q)
    foreach (u,v) ∈ E do
        if d[u] + c(u,v) < d[v] then
            if d[v] = +∞ then Q ← Q ∪ {v}
            d[v] ← d[u] + c(u,v)      (DECREASE ...)
```

AK $d = 2$
$\Rightarrow O((m+n)\log n)$

AK $\boxed{d = \max\{2, \frac{m}{n}\}}$
$\Rightarrow O(m \cdot \log_d n)$

AK $m > n^{1+\varepsilon}$ PRE $\varepsilon > 0$, POTOM $\log_d n = O(1)$

17.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## LEFTIST HALDY

BINÁRNY KOREŇOVÝ STROM $(T, r)$, KAŽDÝ VRCHOL MÁ NAJVIAC DVOCH SYNOV (ĽAVÝ / PRAVÝ)

$npl(v)$ = DĺŽKA NAJKRATŠEJ CESTY Z $v$ DO VRCHOLU, KTORÝ MÁ NAJVIAC JEDNÉHO SYNA, PRE LIST $\ell$: $npl(\ell) = 0$

LEFTIST HALDA REPREZENTUJÚCA $S$ A FCIU. $f: S \to \mathbb{R}$ :

(1) AK MÁ $v$ 1 SYNA, POTOM JE TO ĽAVÝ SYN

(2) AK MÁ $v$ 2 SYNOV, POTOM:

$npl(\text{PRAVÝ SYN } v) \leq npl(\text{ĽAVÝ SYN } v)$

(3) SPLNENÁ PODMIENKA USP. AKO PRI HALDÁCH (MIN-HEAP)

POSTUPNOSŤ $v_0, v_1, \dots, v_k$ SA NAZÝVA PRAVÁ CESTA Z VRCHOLU $v$, KEĎ $v = v_0$, $v_{i+1}$ JE PRAVÝ SYN $v_i$ $\forall i$ A $v_k$ NEMÁ PRAVÉHO SYNA.

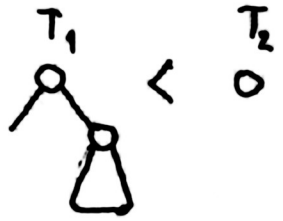
POTOM PODSTROM VRCHOLU $v$ DO HĺBKY $k$ JE ÚPLNÝ BINÁRNY STROM A MÁ ASPOŇ $2^{k+1} - 1$ VRCHOLOV

VETA: V LEFTIST HALDE JE DĺŽKA PRAVEJ CESTY Z KAŽDÉHO VRCHOLU $v$ NAJVIAC ROVNÁ

$\log(\text{VEĽKOSŤ PODSTROMU URČENÉHO VRCHOLOM } v)$

ALGORITMY:

MERGE - DEFINOVANÁ REKURZÍVNE, HĺBKA REKURZIE JE OHRANIČENÁ DĺŽKAMI PRAVÝCH CIEST.




MERGE $(T_1, T_2)$
if $T_1 = \emptyset$ then return $T_2$
if $T_2 = \emptyset$ then return $T_1$
if KEY [KOREŇ $T_1$] > KEY [KOREŇ $T_2$] then SWAP $T_1 \leftrightarrow T_2$
$T' :=$ MERGE ( pravy $[T_1]$, $T_2$)
pravy $[T_1] \leftarrow T'$
otec $[T'] \leftarrow T_1$
if npl [pravy $[T_1]$] > npl [lavy $[T_1]$] then
  SWAP SYNOV $T_1$
npl $[T_1] \leftarrow$ npl [pravy $[T_1]$] +1

DELETEMIN $(T)$
$T \leftarrow$ MERGE ( lavy $[T]$, pravy $[T]$)

MAKEHEAP $(S, \delta)$
$Q \leftarrow \emptyset$ <u>FRONTA</u>
foreach $s \in S$ do VLOŽ $T_s$ REPR. $\{s\}$ DO $Q$
while $|Q| > 1$ do
  ODSTRÁŇ $T_1, T_2$ z $Q$
  MERGE $(T_1, T_2)$ VLOŽ DO $Q$

ODHAD ZLOŽITOSTI: MERGE - POČET REKURZÍVNYCH VOLANÍ JE SÚČET DĹŽOK PRAVÝCH CIEST $\Rightarrow O(\log(|S_1|+|S_2|))$

MAKEHEAP - UVAŽUJME, ŽE NA ZAČIATKU ALGORITMU JE NA VRCHOLKU FRONTY ŠPECIÁLNY ZNAK, KTORÝ SA IBA PRENESIE NA KONIEC FRONTY. ODHADNEME ČAS, KT. SPOTREBUJÚ while-CYKLY MEDZI DVOMA PRENESENIAMI ŠPECIÁLNEHO ZNAKU.

17.02.2009 Tu
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

PREDPOKLADAJME, ŽE SA ŠPECIÁLNY ZNAK PRENIESOL UŽ k-TÝ KRÁT. V TOMTO OKAMIHU MAJÚ VŠETKY HALDY VO FRONTE (AŽ NA JEDNU) VEĽKOSŤ $2^{k-1}$.

V Q JE $\lceil \frac{|S|}{2^{k-1}} \rceil$ HÁLD, A KEĎŽE MERGE VYŽADUJE $O(k)$

$\Rightarrow$ MAKEHEAP POTREBUJE ČAS

$$O\left(\sum_{k=1}^{\infty} k \cdot \frac{|S|}{2^{k-1}}\right) = \underline{\underline{O(|S|)}} \qquad \sum_{k=1}^{\infty} \frac{k}{2^{k-1}} = 4$$

POZOR: DĽŽKA CESTY Z KOREŇA DO LISTU V LEFTIST HALDE MÔŽE BYŤ AŽ $|S|$ $\rightsquigarrow$ NEEFEKTÍVNE UP, DOWN!

OPRAV$(T, v)$ = VYTVORÍ LEFTIST HALDU Z BINÁRNEHO STROMU $T'$, KT. VZNIKOL Z LEFTIST HALDY $T$ ODTRHNUTÍM <u>PODSTROMU</u> S KOREŇOM VO VRCHOLE $v$.

```
OPRAV (T, v)
t ← OTEC[v], npl[t] ← 0
if prav[t] ≠ v then lav[t] ← prav[t]
prav[t] ← NIL
while SA ZMENŠILO npl[t] A t NIE JE KOREŇ do
    t ← OTEC[t]
    if npl[prav[t]] > npl[lav[t]] then
        SWAP lav[t] ↔ prav[t]
    npl[t] ← npl[prav[t]] + 1
```

ZLOŽITOSŤ: $O(\log |S|)$ TOTIŽ VŠETKY VRCHOLY, KDE SA ZMENŠILO npl, TVORIA PRAVÚ CESTU Z $t$.

DECREASE $(v, a)$

$T_1 \leftarrow$ PODSTROM URČENÝ VRCHOLOM $v$, $f[v] \leftarrow f[v] - a$

$T_2 \leftarrow$ OPRAV $(T, v)$

$T \leftarrow$ MERGE $(T_1, T_2)$

INCREASE $(v, a)$

$T_1 \leftarrow$ PODSTROM lavý $[v]$

$T_2 \leftarrow$ PODSTROM pravý $[v]$

$T_3 \leftarrow$ LEFTIST HALDA REPREZENTUJÚCA $v$, $f[v] \leftarrow f[v] + a$

$T_4 \leftarrow$ OPRAV $(T, v)$

$T_1 \leftarrow$ MERGE $(T_1, T_3)$

$T_2 \leftarrow$ MERGE $(T_2, T_4)$

$T \leftarrow$ MERGE $(T_1, T_2)$

DELETE $(v, a)$

$T_1 \leftarrow$ PODSTROM lavý $[v]$

$T_2 \leftarrow$ PODSTROM pravý $[v]$

$T_3 \leftarrow$ MERGE $(T_1, T_2)$

$T_4 \leftarrow$ OPRAV $(T, v)$

$T \leftarrow$ MERGE $(T_3, T_4)$

## AMORTIZOVANÁ ZLOŽITOSŤ

$h(D)$ = OHODNOTENIE KONFIGURÁCIE $D$

AMORTIZ. ZLOŽ. OP. $\boxed{am(o) := t(o) + h(D') - h(D)}$

$D_0 \xrightarrow{o_1} D_1 \xrightarrow{o_2} \dots \xrightarrow{o_n} D_n$, $h(D_0) = 0$, $h(D_i) \geq 0$

NECH MÁME ODHADY $am(o_i) \leq c(o_i)$. POTOM

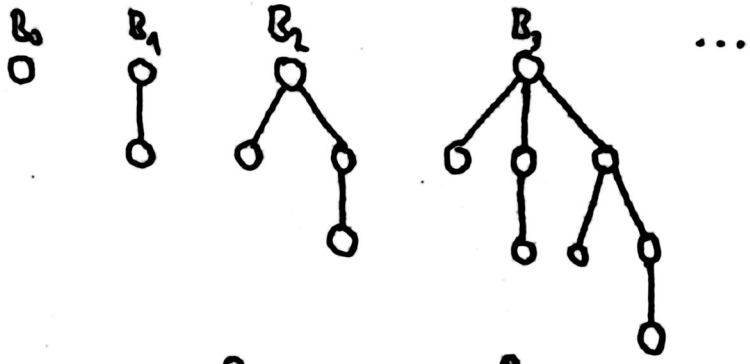
$\sum am(o_i) = h(D_n) - h(D_0) + \sum t(o_i) \leq \sum c(o_i)$

$\Rightarrow \sum t(o_i) \leq \sum c(o_i)$ !!!

17.02.2009 Tu
PETER ČERNO

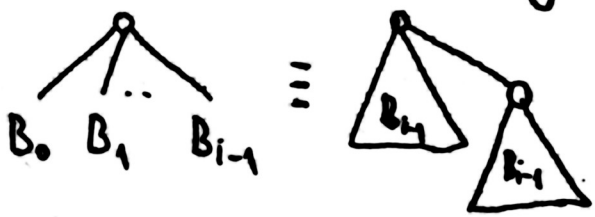
# NTIN066 DATOVÉ STRUKTÚRY I

## BINOMIÁLNE HALDY



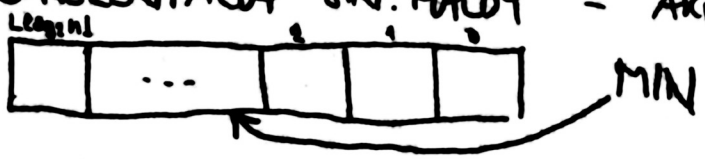

RÁD UZLU = #SYNOV
RÁD STROMU = RÁD KOREŇA

$B_i$ :



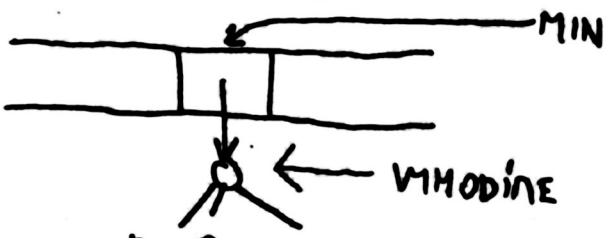

RÁD $B_i = i$
$|B_i| = 2^i$

REPREZENTÁCIA BIN. HALDY – AKO PRIR. ČÍSLO V DVOJKOVEJ SÚSTAVE




OPERÁCIA MERGE$(H_1, H_2)$ ... AKO PRI SČÍTANÍ BIN. ČÍSIEL

EXTRACT-MIN(H)




$B_0$ $B_1$ ... $B_{i-1}$ → VYTVORÍME NOVÚ HALDU
A SMERGUJEME S PÔVODNOU

INSERT$(H, i)$ = MERGE$(H$, MAKEHEAP$(i))$

AMORTIZOVANÝ ČAS : $O(1)$

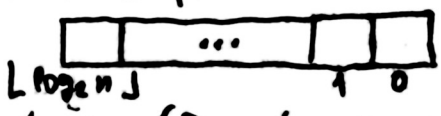
## LENIVÁ REPREZENTÁCIA BIN. HALDY

KORENE STROMOV ZVIAZANÉ DVOJSMERNÝM KRUH. ZOZNAMOM
ZOZNAM MÔŽE DOČASNE OBSAHOVAŤ VIAC STROMOV ROVN. RÁDU

MERGE (A INSERT) V $O(1)$

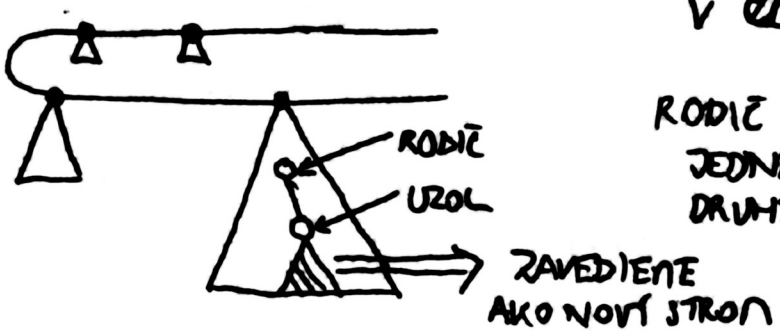
EXTRACT-MIN VOLÁ CONSOLIDATION

→ PLNÁ REPREZENTÁCIA




**INVARIANT**: KAŽDÝ STROM MÁ NA ÚČTE (VO SVOJOM KORENI) **2 €**

## FIBONACCIHO HALDY




PODPORA DECREASE KEY V AMORTIZOVANOM ČASE $O(1)$

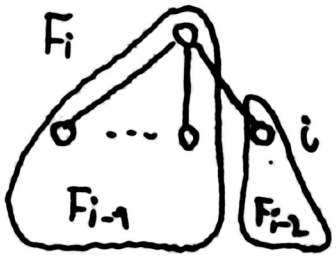
RODIČ MÔŽE STRATIŤ NAJVIAC JEDNÉHO SYNA; PRI STRATE DRUHÉHO → KASKÁDOVITÉ ODRÁZANIE

## DECREASE KEY : 6 €

1 € NA PRÁCU - ODREZANIE UZLU, ZAVEDENIE DO SPOJÁKU..
2 € NA ÚČET NOVÉHO STROMU
3 € POŠLEM RODIČOVI

**LEMMA**: NECH $x$ JE VRCHOL A $y_1, \dots, y_m$ JEHO SYNOVIA V POREDÍ, V KTOROM BOLI POD $x$ ZLIATI, POTOM PRE $\forall i \in \{1, \dots, m\}$ JE RÁD $y_i$ ASPOŇ $i-2$.

NECH $F_i$ JE RÁDU $i$ S MIN. POČTOM VRCHOLOV




DECREASE-KEY NA VRCHOL $i$ → LEGÁLNY STROM, Z MINIMALITY TO MUSÍ BYŤ $F_{i-1}$
VRCHOL $i$ MÁ RÁD $\geq i-2 \Rightarrow F_{i-2}$

$$\boxed{|F_i| = F_{i+1} \geq \left(\frac{1+\sqrt{5}}{2}\right)^i}$$

18.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

## TRIEDIACE ALGORITMY

### HEAPSORT

TRIEDENIE NA MIESTE (IN PLACE) POMOCOU d-MAXHEAP

### MERGESORT

ITERAČNÁ VERZIA - TZV. PRIRODZENÝ MERGESORT

RASTÚCE ÚSEKY NAHÁDZEME DO FRONTY Q
A POSTUPNE ICH MERGUJEME
PRI KONŠTANTNOM POČTE BEHOV MÁ ZLOŽITOSŤ $O(n)$

### QUICKSORT

```
QUICK(a_i, ..., a_j)
if i=j then return (a_i)
else if i < j then
  ZVOL i≤k≤j, SWAP a_i ↔ a_k, l ← i+1, q ← j, a ← a_k
  while true do
    | while a_l < a do ++l
    | while a < a_q do --q
    | if l ≥ q then break
    | else SWAP a_l ↔ a_q, ++l, --q
  return (QUICK(a_{i+1}, ..., a_{l-1}), a, QUICK(a_{q+1}, ..., a_j))
```

V PRAXI: PIVOTA VIBERÁME AKO NEDIAN Z TROCH ALEBO 5 ZVOLENÝCH PRVKOV...

NAJHORŠÍ ČAS: $O(n^2)$

OČAKÁVANÝ ČAS: $O(n \log n)$

## ANALÝZA OČAKÁVANÉHO PRÍPADU

(BUĎ NÁHODNE VOLENÝ PIVOT, ALEBO ROVNOM. ROZD. DÁT)

BUDEME POČÍTAŤ OČAKÁVANÝ POČET POROVNANÍ

**PRVÝ SPÔSOB**

KAŽDÉ DVA PRVKY $a_i, a_j$ QUICKSORT POROVNÁ NAJVIAC RAZ, PRIČOM KEĎ POROVNÁ $a_i$ A $a_j$, POTOM PRE NEJAKÝ BEH PODPROCEDÚRY QUICK JE $a_i$ ALEBO $a_j$ PIVOT, ALE V PREDCHÁDZAJÚCICH BEHOCH $a_i$ ANI $a_j$ NEBOLI PIVOTI

NECH $(b_1, \dots, b_n)$ JE VÝSLEDNÁ POSTUPNOSŤ

OZNAČME $X_{i,j}$ BOOLEOVSKÚ NÁH. VELIČINU, KTORÁ NÁ HODNOTU 1 $\Leftrightarrow$ QUICK POROVNAL $b_i$ A $b_j$

KEĎ $p_{i,j}$ JE PRAVDEP., ŽE $X_{i,j} = 1$, POTOM

$$E X_{i,j} = 0(1-p_{i,j}) + 1\, p_{i,j} = p_{i,j}$$

POČET POROVNANÍ PRI BEHU ALG. QUICKSORT JE

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{i,j}$$

$\Rightarrow$ OČAKÁVANÝ POČET POROVNANÍ JE $\sum_{i=1}^{n} \sum_{j=i+1}^{n} \overbrace{E(X_{i,j})}^{= p_{i,j}}$

**CHCEME SPOČÍTAŤ $p_{i,j}$**

UVAŽUJME BINÁRNY STROM, V KTOROM KAŽDÝ VRCHOL ZODPOVEDÁ JEDNÉMU BEHU QUICK

VNÚTORNÝ VRCHOL $v$ - OHODNOTENIE $v$ = ZVOLENÝ PIVOT

LIST = VOLANIE QUICK NAD JEDNOPRVKOVOU POSTUPNOSŤOU

OČÍSLUJME VRCHOLY TOHOTO STROMU PREHĽ. DO ŠÍRKY

NECH $(c_1, \dots, c_n)$ JE POSTUPNOSŤ PRVKOV $\{a_i \mid 1 \le i \le n\}$ V PORADÍ DANOM TÝMTO OČÍSLOVANÍM

18.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

PLATÍ $X_{i,j} = 1 \Leftrightarrow$ PRVÝ PRVOK V POSTUPNOSTI $(c_1, \dots, c_n)$ Z MNOŽINY $\{b_l \mid i \le l \le j\}$ JE BUĎ $b_i$ ALEBO $b_j$

!

PRAVDEPODOBNOSŤ TOHOTO JAV JE $\boxed{\frac{2}{j-i+1} = p_{i,j}}$

$$\Rightarrow \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n} \sum_{k=2}^{n-i+1} \frac{2}{k} \le 2n \sum_{k=2}^{n} \frac{1}{k} \le \underline{\underline{2n \ln n}}$$

**DRUHÝ SPÔSOB**

OZNAČME $QS(n)$ OČAKÁVANÝ POČET POROVNANÍ QUICKSORT-U PRI $n$-ČLENNEJ POSTUPNOSTI

POTOM $QS(0) = QS(1) = 0$ A

$$QS(n) = \frac{1}{n}\left(\sum_{k=0}^{n-1} n-1 + QS(k) + QS(n-k-1)\right) =$$

$$\boxed{= n-1 + \frac{2}{n} \sum_{k=0}^{n-1} QS(k)}$$

$$n\,QS(n) = n(n-1) + 2\sum_{k=0}^{n-1} QS(k)$$

$$(n+1)\,QS(n+1) = (n+1)n + 2\sum_{k=0}^{n} QS(k)$$

$$\underline{\underline{QS(n+1) = \frac{2n}{n+1} + \frac{n+2}{n+1} QS(n)}}$$

$$\Rightarrow QS(n) = \sum_{i=2}^{n} \frac{n+1}{i+1} \frac{2(i-1)}{i} \le 2(n+1) \sum_{i=2}^{n} \frac{1}{i+1} \le \dots$$

$$\dots \le 2n \ln(n+1) + 2\ln(n+1) - n - 1$$

$\Rightarrow QS(n) \le 2n \ln n$ PRE DOSŤ VEĽKÉ $n$.

# ZLIEVANIE RÔZNE DLHÝCH POSTUPNOSTÍ

VSTUP: MNOŽINA RASTÚCICH NAVZÁJOM DISJUNKT. POSTUPN.
VÝSTUP: POMOCOU MERGE ČO NAJRÝCHLEJŠIE SPOJIŤ TIETO POSTUPNOSTI DO JEDINEJ RASTÚCEJ POSTUPNOSTI

POSTUP JE URČENÝ ÚPLNÝM BINÁRNYM STROMOM $T$
LISTY = VSTUPNÉ POSTUPNOSTI, VN. UZLY = ZLÚČENIE SYNOV

NECH $\ell(P)$ JE DĽŽKA POSTUPNOSTI $P$.
POTOM ČASOVÁ ZLOŽITOSŤ JE

$$\sum\{\ell(P(v)) \mid v \text{ JE VNÚTORNÝ VRCHOL } T\} =$$
$$= \underline{\underline{\sum\{d(t).\ell(P(t)) \mid t \text{ JE LIST A } d(t) \text{ JEHO HĽBKA}\}}}$$

ÚLOHU MÔŽEME PREFORMULOVAŤ:

VSTUP: $n$ ČÍSIEL $x_1, \dots, x_n$
VÝSTUP: ÚPLNÝ BINÁRNY STROM S $n$ LISTAMI A BIJEKCIA $\phi$ Z $\{1, \dots, n\}$ DO LISTOV $T$ T.Ž. $\sum_{i=1}^{n} d(\phi(i))\, x_i$ JE MINIMÁLNE $=: \mathrm{Cost}(T, \phi)$

OPTM $(x_1, \dots, x_n)$
$V$ JE MNOŽINA $n$ JEDNOPRVKOVÝCH STROMOV S OHODNOTENIAMI $x_1, \dots, x_n$, $c(\phi(i)) = x_i$, $\phi$ JE BIJEKCIA
while $|V| > 1$ do
  VYBER Z $V$ DVA STROMY $v_1, v_2$ S NAJMENŠÍM OHODN.
  VYTVOR NOVÝ STROM $v$ SPOJENÍM $v_1$ A $v_2$
  $c(v) \leftarrow c(v_1) + c(v_2)$, $v$ VLOŽ DO $V$

<u>VETA</u>: OPTM NÁJDE OPTIMÁLNY STROM, AK JE $(x_1, \dots, x_n)$ NEKLESAJÚCA, POTOM VYŽADUJE ČAS $O(n)$.

18.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

KAŽDÝ BEH while-CYKLU ZMENŠÍ POČET STRÁNOK O 1, ALE NEZMENÍ MNOŽINU LÍSTOV

INDUKCIA PODĽA $n$: ZKONSTRUOV. $(T, \phi)$ JE OPT. K $(x_1, \dots, x_n)$ !

PREDPOKLADAJME, ŽE PLATÍ PRE KAŽDÚ POST. $(y_1, \dots, y_{n-1})$

NECH $x_1 \leq \dots \leq x_n$. BUNO NECH ZVOLIL STRÁNY $\phi(1), \phi(2)$

UVAŽUJME $(y_1, \dots, y_{n-1})$, KDE $y_i = x_{i+2}$, $y_{n-1} = x_1 + x_2$

A NECH OPTIM$(y_1, \dots, y_{n-1})$ SKONŠTRUOVAL $(T', \psi)$, $\boxed{T' \text{ VZNIKOL Z } T \text{ ODSTR. } \phi(1) \text{ A } \phi(2)}$

NECH $(U, \Theta)$ JE OPTIMÁLNY VÝHĽADOM K $(x_1, \dots, x_n)$.

ZVOĽME VNÚTORNÝ VRCHOL $u$ STROMU $U$ TAKÝ, ŽE DĹŽKA CESTY Z KOREŇA DO $u$ JE NAJVÄČŠIA MOŽNÁ MEDZI VŠETKÝMI VNÚTORNÝMI VRCHOLMI STROMU $U$,

NECH $u_1, u_2$ SÚ SYNOVIA $u$ (NUTNE LÍSTY)

A NECH $i, j \in \{1, \dots, n\}$: $\Theta(i) = u_1$, $\Theta(j) = u_2$, $i < j$

DEFINUJME $\eta(1) = \Theta(i) = u_1$, $\eta(i) = \Theta(1)$ $\quad \eta(k) = \Theta(k)$

$\eta(2) = \Theta(j) = u_2$, $\eta(j) = \Theta(2)$ $\quad$ PRE OSTATNÉ $k$

Z VOĽBY $u$ PLYNIE $d(u_1) \geq d(\Theta(1))$, $d(u_2) \geq d(\Theta(2))$, $x_1 \leq x_i$, $x_2 \leq x_j$.

$$\mathrm{Cont}(U, \eta) - \mathrm{Cont}(U, \Theta) = (d(u_1) - d(\Theta(1)))(x_1 - x_i) + (d(u_2) - d(\Theta(2)))(x_2 - x_j) \leq 0$$

A KEĎŽE $(U, \Theta)$ JE OPTIMÁLNY, NUTNE AJ $(U, \eta)$ JE OPT.

ODSTRÁNENÍM $u_1$ A $u_2$ Z $U$ DOSTANEME $U'$. DEFINUJME $\tau$ Z $\{1, \dots, n-1\}$ PREDPISOM $\tau(i) = \eta(i+2)$, $\tau(n-1) = u$

$\tau$ JE BIJEKCIA Z $\{1, \dots, n-1\}$ DO MNOŽINY LÍSTOV $U'$

KEĎŽE $(T', \psi)$ JE OPTIMÁLNY PRE $(y_1, \dots, y_{n-1})$, PLATÍ:

$$Cont(T', \psi) \leq Cont(U', \tau).$$

ĎALEJ $Cont(T, \phi) = Cont(T', \psi) + x_1 + x_2$

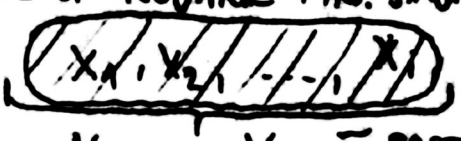
$Cont(U, \eta) = Cont(U', \tau) + x_1 + x_2$

$\Rightarrow Cont(T, \phi) \leq Cont(U, \eta)$, TJ. $(T, \phi)$ JE OPTIMÁLNY. $\square$

ČASOVÁ ZLOŽITOSŤ:

NECH $x_1 \leq \dots \leq x_n$ A V DANOM OKAMIHU SÚ $V_1, \dots, V_k$ POSTUPNE VYTVORENÉ VACPRVKOVÉ STROMY.

UŽ SPRACOVANÉ 1-PRV. STROMY — UKAZATEĽ NA NAJMENŠÍ LIST, KT. JE NESPR.




$x_{i+1}, \dots, x_n$ ≡ EŠTE NESPRACOVANÉ 1-PRV. S.

$V_1, \dots, V_k$ ≡ POSTUPNE VYTVORENÉ VACPRVKOVÉ STROMY

VYTVORME STROM $w$ SPOJENÍM $t_1$ A $t_2$ S NAJMENŠÍM OHODNOTENÍM.

KEĎ STROM $v_i$ VZNIKOL SPOJENÍM $u_1$ A $u_2$, POTOM

$\max\{c(u_1), c(u_2)\} \leq \min\{c(t_1), c(t_2)\}$

$\Rightarrow c(w) \geq c(v_i) \ \forall i$ A INDUKCIOU $c(v_1) \leq \dots \leq c(v_k)$

ZOTRIEDENIE POSTUPNOSTI PRIR. ČÍSIEL $x_1, \dots, x_n$, V ČASE $O(n+m)$ KDE $m = \max x_i$

$\Rightarrow P_1, \dots, P_n$ MOŽNO ZMERLOVAŤ V ČASE $O(\sum_{i=1}^{n} \ell(P_i))$. $\boxed{?}$

18.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## ROZHODOVACIE STROMY

JEDINÁ OP.: POROVNÁVANIE PRVKOV

BIN. STROM - VN. VRCHOLY OHODN. POROVNANÍM DVOJICE PRVKOV NA VSTUPE PERMUTÁCIA $\pi$

SKONČÍME V LISTE

KOREKTNOSŤ: MUSÍ PLATIŤ, ŽE DVE RÔZNE PERMUTÁCIE SKONČIA V RÔZNYCH LISTOCH

$\Rightarrow$ ASPOŇ $n!$ LISTOV

POLNÝ ODHAD POČTU POROVNANÍ = DĹŽKA NAJDLHŠEJ CESTY

DEFINUJME

$S(n) :=$ MINIMUM CEZ VŠETKY STROMY $T$ S ASPOŇ $n!$ LISTAMI Z DĹŽOK NAJDLHŠÍCH CIEST Z KORENA DO LISTU $T$

$A(n) :=$ MINIMUM CEZ VŠETKY STROMY $T$ S ASPOŇ $n!$ LISTAMI Z PRIEMERNÝCH DĹŽOK CIEST Z KORENA DO LISTU $T$

AK NAJDLHŠIA CESTA Z KORENA DO LISTU V BIN. STROME $T$ MÁ DĹŽKU $k$, POTOM $T$ MÁ NAJVIAC $2^k$ LISTOV

$$\Rightarrow n! \leq 2^{S(n)} \Rightarrow S(n) \geq \log_2 n!$$

STIRLINGOV KOREC: $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \overbrace{\left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right)}^{\text{MÔŽME PREDP. } \geq 1}$

$$\Rightarrow \log_2 n! \geq \frac{1}{2}\log_2 n + n(\log_2 n - \log_2 e) + \log_2 \sqrt{2\pi} \geq$$

$$\geq \left(n + \frac{1}{2}\right)\log_2 n - n\log_2 e = \left(n + \frac{1}{2}\right)\log_2 n - \frac{n}{\ln 2}$$

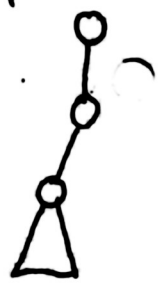
$B(T)$ = SÚČET V DĽŽOK CIEST Z KOREŇA DO LISTOV

$B(k) = \min\{B(T) \mid T$ JE BINÁRNY STROM S $k$ LISTAMI$\}$

KEĎ UKÁŽEME, ŽE $B(k) \geq k \log_2 k$, POTOM BUDE

$$A(n) \geq \frac{B(n!)}{n!} \geq \frac{n! \log n!}{n!} = \log_2 n! \geq \left(n+\frac{1}{2}\right)\log_2 n - \frac{n}{\ln 2}$$

DOKÁŽEME, ŽE $B(T) \geq k \log_2 k$ PRE KAŽDÝ BINÁRNY STROM $T$ S $k$ LISTAMI.



KEĎ V STROME $T$ VYNECHÁME KAŽDÝ VRCHOL, KTORÝ MÁ IBA JEDNÉHO SYNA, A TOHOTO SYNA SPOJÍME S JEHO PREDCHODCOM, DOSTANEME ÚPLNÝ BIN. STROM $T'$ S $k$ LISTAMI, ŽE $B(T') \leq B(T)$.

$\Rightarrow$ STAČÍ SA OBMEDZIŤ NA <u>ÚPLNÉ</u> BIN. STROMY

$T$ ÚPLNÝ BIN. STROM S 1 LISTOM $\Rightarrow B(T) = 0 = 1 \cdot \log_2 1$

—"— S 2 LISTAMI $\Rightarrow B(T) = 2 = 2 \cdot \log_2 2$

PREDPOKLADAJME, ŽE $B(i) \geq i \log_2 i$ PRE $i < k$

A NECH $T$ JE ÚPLNÝ BIN. STROM S $k$ LISTAMI

$T_1, T_2$ PODSTROMY KOREŇA $T$, $T_i$ MÁ $k_i$ LISTOV

$1 \leq k_1, k_2$, $k_1 + k_2 = k \Rightarrow k_1, k_2 < k$ A PODĽA IND. PREDP.

$B(k_i) \geq k_i \log_2 k_i$

$$\Rightarrow B(T) = k_1 + B(T_1) + k_2 + B(T_2) \geq k + B(k_1) + B(k_2) \geq$$

$$\geq k + k_1 \log_2 k_1 + k_2 \log_2 k_2 \overset{?}{\geq} k \log_2 k$$

STAČÍ UKÁZAŤ, ŽE PRE $k > 0$ A $x \in (0, k)$ JE

$$f(x) = x \log_2 x + (k-x)\log_2(k-x) + k - k\log_2 k \geq 0$$

$f\left(\frac{k}{2}\right) = 0$, $f'(x) < 0$ NA $\left(0, \frac{k}{2}\right)$ A $f'(x) > 0$ NA $\left(\frac{k}{2}, k\right)$.

19.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTÚRY I

## PRIEHRADKOVÉ TRIEDENIE

$Q_i$ SPOJ. ZOZNAMY, VKLADÁ SA VŽDY NA KONIEC
MÁME PRÍSTUP K PRVÉMU A POSLEDNÉMU PRVKU

$a_1, \dots, a_n \in \{0, \dots, m\}$

NEPRIAME ADRESOVANIE

**BUCKETSORT** $(a_1, \dots, a_n; m)$

STABILNÉ
$O(n+m)$

foreach $i = 0, \dots, m$ $\quad Q_i \leftarrow \emptyset$
foreach $i = 1, \dots, n$ $\quad Q_{a_i} \leftarrow Q_{a_i} \cdot a_i$
return $Q_0 \dots Q_m$

NECH $a_1, \dots, a_n$ SÚ RÔZNE Z $\langle 0, 1 \rangle$, $\alpha > 0$ PEVNÉ KLADNÉ

**HYBRIDSORT** $(a_1, \dots, a_n)$

$k \leftarrow \alpha n$
foreach $i = 0, \dots, k$ $\quad Q_i \leftarrow \emptyset$
foreach $i = 1, \dots, n$ $\quad Q_{\lceil k a_i \rceil} \ll a_i$
return KONKATENACIA HEAPSORT$(Q_i)$

VETA: $O(n \log n)$, AK $a_i$ SÚ NEZÁV., ROVN. ROZL., POTOM OČAKÁVANÝ ČAS JE $O(n)$.

OČAKÁVANÝ ČAS: POLOŽME $X_i = |Q_i|$
PRAVDEP. $x \in Q_i$ JE $\frac{1}{k}$ $\Rightarrow$ $P(X_i = q) = \binom{n}{q} \left(\frac{1}{k}\right)^q \left(1 - \frac{1}{k}\right)^{n-q}$
TJ. $X_i \sim Bi(n, \frac{1}{k})$

$\Rightarrow$ OČAKÁVANÝ ČAS POSL. RIADKU

$$E\left(\sum_{i=0}^{k} 1 + X_i \log X_i\right) \leq k + k E X^2 = k + k \cdot \left(\frac{n(n-1)}{k^2} + \frac{n}{k}\right) = O(n)$$

PRE $k = \alpha n$.

OČAKÁVANÁ ZLOŽITOSŤ BY ZOSTALA $O(n)$ AJ PRI KVADRAT. TRIEDENÍ.

$a = x_1 \dots x_n < b = y_1 \dots y_m \iff \exists i \in \{0, \dots, \min(n,m)\}$:
$x_j = y_j \ \forall j = 1, \dots, i$ A BUĎ $n = i < m$, ALEBO
$i < \min(n,m)$ A $x_{i+1} < y_{i+1}$.

PREDPOKLADAJME, ŽE $a_i = a_i^1 \dots a_i^{\ell(i)}$.

<u>WORDSORT</u> $(a_1, \dots, a_n)$

$\ell \leftarrow \max\{\ell(i) \mid i = 1, \dots, n\}$
foreach $i = 1, \dots, \ell$ $L_i \leftarrow \emptyset$
foreach $i = 1, \dots, n$ $L_{\ell(i)} << a_i$

$L_i$ OBSAHUJE SLOVÁ DĹŽKY $i$

$P \leftarrow \{(j, a_i^j) \mid i = 1, \dots, n;\ j = 1, \dots, \ell\}$
$P_1 \leftarrow$ BUCKETSORT$(P)$ PODĽA DRUHEJ KOMPONENTY
$P_2 \leftarrow$ BUCKETSORT$(P_1)$ PODĽA PRVEJ KOMPONENTY
for $i = 1, \dots, \ell$ $S_i \leftarrow \emptyset$
$(i, x) \leftarrow$ FIRST$[P2]$
while $(i, x) \neq$ NIL do
    $S_i << (i, x)$
    while $(i, x) =$ NEXT$[P2]$ do MOVE_NEXT$[P2]$
    $(i, x) \leftarrow$ NEXT$[P2]$

V $S_i$ SÚ $\forall$ DVOJICE $(i, x)$ TZ. $x$ JE $i$-TÝM PÍSMENKOM NEJAKÉHO SLOVA, ZOTRIEDENÉ PODĽA $x$

foreach $s \in \Sigma$ do $T_s \leftarrow \emptyset$
$T \leftarrow \emptyset$, $i \leftarrow \ell$
while $i > 0$ do
    $T \leftarrow L_i \cup T$
(*) DO $T_s$ NAMÁDZ SLOVÁ Z $T$ PODĽA PRVÉHO PÍSMENKA
    $T \leftarrow \emptyset$
(**) foreach $(i, x) \in S_i$ do $T \leftarrow T \cup T_x$, $T_x \leftarrow \emptyset$
    $--i$
return $T$

18.02.2009 We
PETER ČERNO

# NTIN066 DATOVÉ STRUKTURY I

ZLOŽITOST: OZNAČME

$\ell = \max\{\ell(i) \mid i = 1, \dots, n\}$, $L = \sum_{i=1}^{n} \ell(i)$

$m_i$ = POČET SLOV Z $\{a_1, \dots, a_n\}$ DÉLKY ALSPOŇ $i$

$\Rightarrow \underline{\underline{L = \sum_{i=1}^{\ell} m_i}}$

(*) VYŽADUJE ČAS $O(m_i)$

(**) VYŽADUJE ČAS $O(|s_i|) = O(m_i)$

$\Rightarrow$ CELKOVÝ ČAS ALGORITMU JE $O(L + m)$, $m = |\Sigma|$

## PORADKOVÉ STATISTIKY

<u>FIND</u>$(M = (a_1, \dots, a_n); i)$

ZVOL $a \in M$

$M_1 \leftarrow \{b \in M \mid b < a\}$

$M_2 \leftarrow \{b \in M \mid b > a\}$

if $|M_1| > i-1$ then return FIND$(M_1, i)$

else if $|M_1| < i-1$ then return FIND$(M_2, i - |M_1| - 1)$

else return $a$

ČAS. ZLOŽITOST V NEJHORŠÍM PŘÍPADĚ: $O(n^2)$

OČEKÁVANÝ ČAS ZA PŘEDPOKLADU NÁHODNÉ VOLBY $a \in M$

$a$ JE $k$-TÝ NEJMENŠÍ S PRAVD. $1/n$

OZNAČME $T(n, i)$ OČEKÁVANÝ ČAS ALG. FIND$(\dots, i)$

PLATÍ: $\underline{\underline{T(n,i) = n + \frac{1}{n}\left[\sum_{k=1}^{i-1} T(n-k, i-k) + \sum_{k=i+1}^{n} T(k, i)\right]}}$

(PLATÍ PRE $m=1$ AJ $m=2$)

PREDPOKLADAJME, ŽE: $T(m,i) \le 4m \quad \forall m<n \ \forall 1 \le i \le m$.
POTOM

$$T(n,i) = n + \frac{1}{n}\left[\sum_{k=1}^{i-1} T(n-k, i-k) + \sum_{k=i+1}^{n} T(k,i)\right] \le$$

$$\le n + \frac{1}{n}\left[\sum_{k=1}^{i-1} 4(n-k) + \sum_{k=i+1}^{n} 4k\right] =$$

$$= n + \frac{4}{n}\left[\frac{(2n-i)(i-1)}{2} + \frac{(n+i+1)(n-i)}{2}\right] = n + \frac{4}{n}\left(\frac{n^2+2ni-n-2i^2}{2}\right)$$

KDE ČITATEĽ NADOBÚDA MAXIMA PRE $i=\frac{n}{2}$ A $= \frac{3n^2-2n}{2}$

$$\Rightarrow T(n,i) \le n + \frac{4}{n}\left(\frac{3n^2-2n}{4}\right) = n+3n-2 < 4n \quad \square$$

## SELECT, LINEÁRNY V NAJHORŠOM PRÍPADE

```
SELECT(M, i)
if |M| ≤ 100 then
  ZOTRIEĎ M, return M[i]
else
  ROZDEĽ M DO DISJ. 5-PRVKOVÝCH
    A_1, A_2, ..., A_⌈n/5⌉ ; |A_⌈n/5⌉| ≤ 5
  NÁJDI m_j MEDIÁNY A_j
  m̄ ← SELECT({m_j | j=1,...,⌈n/5⌉}, ⌈n/10⌉)
  M_1 ← {m ∈ M | m < m̄}
  M_2 ← {m ∈ M | m > m̄}
  if |M_1| > i-1 then return SELECT(M_1, i)
  else if |M_1| < i-1 then return SELECT(M_2, i-|M_1|-1)
  else return m̄
```

KEĎ $n>100$, POTOM
$\lceil \frac{n}{5} \rceil \le \frac{21}{100} n$
$T(n) \le T(\frac{21}{100}n) + T(\frac{8}{11}n) + 6n$
NECH $c \ge \max\{a, 6\}$
DOKÁŽME IND. $T(n) \le cn$

LEMMA: AK $n \ge 100$, POTOM $|M_1|, |M_2| \le \frac{8}{11} n$

18.02.2009 We
PETER ČERNO

# NTIN066 ZHRNUTIE

## HASHOVANIE SO SEPAR. RETAZCAMI

$P(h(x)=i) = \frac{1}{m} \Rightarrow P(|S_i|=\ell) = \binom{n}{\ell}\left(\frac{1}{m}\right)^{\ell}\left(1-\frac{1}{m}\right)^{n-\ell}$

$\boxed{ENP}$ = OCAK. DĹŽKA NAJDLHŠIEHO RETAZCA

$P(\max_i \ell(i)=j) = P(\max_i \ell(i) \geq j) - P(\max_i \ell(i) \geq j+1)$

$ENP = \sum_j j P(\max_i \ell(i)=j) = \dots = \sum_j P(\max_i \ell(i) \geq j)$

$P(\max_i \ell(i) \geq j) = P(\ell(1) \geq j \vee \dots \vee \ell(m-1) \geq j) \leq$

$\leq \sum_{i=0}^{m-1} P(\ell(i) \geq j) \leq m \binom{n}{j}\left(\frac{1}{m}\right)^j \leq \underline{\underline{n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!}}}$ $\boxed{\alpha \leq 1}$ KLESÁ S $j$

! ODHADNEME $j_0 = \min_j \left(n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!} \leq 1\right)$

$\Rightarrow ENP \leq \sum_{j=1}^{j_0} 1 + \sum_{j=j_0+1}\left(n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!}\right) \dots \leq j_0 + \frac{1}{j_0} = O(j_0)$

ODHAD $j_0$:

$n\left(\frac{n}{m}\right)^{j-1}\frac{1}{j!} \leq \frac{n}{j!} \leq \frac{n}{(j/2)^{j/2}} \rightsquigarrow \underline{\underline{j_0 \leq \min_j \left(n \leq (j/2)^{j/2}\right)}}$ OZN. $\underline{\underline{k+1}}$

$\underline{\underline{\left(\frac{k}{2}\right)^{k/2} < n \leq \left(\frac{k+1}{2}\right)^{\frac{k+1}{2}}}} \rightsquigarrow k < 8\frac{\log n}{\log\log n}$

PRE $\alpha \leq 1$ JE OCAK. DĹŽKA MAX. RETAZCA $O\left(\frac{\log n}{\log\log n}\right)$

NEÚSPEŠNÉ VYHĽ.: $P(\ell(i)=0) + \sum_{\ell \geq 1} \ell P(\ell(i)=\ell) = \left(1-\frac{1}{m}\right)^n + \frac{n}{m} \approx e^{-\alpha}+\alpha$

ÚSP. VYHĽ.: $\frac{m}{n}\sum_{\ell=0}^{n}\binom{\ell+1}{2}p_{n\ell} = \dots = \frac{n-1}{2m}+1 \approx 1+\frac{\alpha}{2}$

ŠPECIÁLNA MEMÓRIA

HASHOVANIE S PREMIESTŇOVANÍM, S DVOMA UKAZATEĽMI

## ZRASTAJÚCE HASHOVANIE

ŠTANDARDNÉ: LISCH, EISCH

S POMOCNOU PAMÄŤOU: LICH, VICH, EICH

ANALÝZA LISCH $\boxed{C(a_1,\dots,a_n;a_{n+1})}$ = #TESTOV NEÚSP. VYHĽ.

RETAZEC DĹŽKY $\ell$: $(b_1,\dots,b_\ell)$

$\sum_{a_{n+1}=0}^{m-1} C(\vec{a};a_{n+1}) = c + \binom{\ell_1+1}{2} + \dots + \binom{\ell_k+1}{2} =$ $\boxed{\begin{array}{l}\ell_1+\dots+\ell_k=n\\ c=m-n\end{array}}$

$= c_{\vec{a}}(0) + \sum_{\ell=1}^{n}\binom{\ell+1}{2}c_{\vec{a}}(\ell)$ !

$$\sum C(\vec{a}; a_{n+1}) = c_n(0) + \sum_{\ell=1}^{n} \ell c_n(\ell) + \boxed{\sum_{\ell=1}^{n} \binom{\ell}{2} c_n(\ell)} \to S_n$$

$$c_n(0) = (m-n)\, m^n, \quad \sum_{\ell=1}^{n} \ell\, c_n(\ell) = n\, m^n$$

$$\boxed{\sum_{a_{n+1}=0}^{m-1} c_{\vec{a}, a_{n+1}}(\ell) = (m-\ell)\, c_{\vec{a}}(\ell) + (\ell-1)\, c_{\vec{a}}(\ell-1)}$$

$$\Rightarrow c_{n+1}(\ell) = (m-\ell)\, c_n(\ell) + (\ell-1)\, c_n(\ell-1) \quad !$$

$$S_n = \dots = \sum_{\ell=1}^{n-1} \Big[ \underbrace{\binom{\ell}{2}(m-\ell) + \binom{\ell+1}{2}\ell}_{\binom{\ell}{2}(m+2)+\ell} \Big] c_{n-1}(\ell) = (m+2) S_{n-1} + (n-1) m^{n-1}$$

SPOČÍTAJTE $T_c^n = \sum_{i=1}^{n} i c^i$, $(c-1) T_c^n$

ÚSP. VYHĽADÁVANIE: OČAKÁVANÝ POČET POROVN. KĽÚČOV:

$$\frac{1}{m^{n+1}} \sum_{\ell=1}^{n} \binom{\ell+1}{2} c_n(\ell) = \dots = \frac{1}{4}\left[\left(1+\frac{2}{m}\right)^n - 1 + \frac{2n}{m}\right]$$

$$\Rightarrow \frac{1}{n} \sum_{i=0}^{n-1} \left[ 1 + \frac{1}{4}\left[\left(1+\frac{2}{m}\right)^i - 1 + \frac{2i}{m}\right]\right] = \dots$$

## LICH, EICH, VICH

VICH VKLADÁ ZA POSLEDNÝ PRVOK, KT. JE EŠTE V POMOCNEJ ČASTI ...

$n = |S|$, $m = |\text{ADR. ČASŤ}|$, $m' = |\text{CELÁ TAB.}|$.

$\alpha = n/m'$ FAKT. ZAP., $\beta = m/m'$ ADR. FAKTOR, $\lambda : e^{-\lambda} + \lambda = \frac{1}{\beta}$

$\alpha \le \lambda\beta \Rightarrow$ NEÚSP.: $e^{-\alpha/\beta} + \frac{\alpha}{\beta}$, ÚSP.: $1 + \alpha/2\beta$ (AKO SEP. R.)

## LINEÁRNE PRIDÁVANIE

OČAK. POČET TESTOV:

NEÚSP. PRÍPAD: $\frac{1}{2}\left(1 + \left(\frac{1}{1-\alpha}\right)^2\right)$

ÚSP. PRÍPAD $\frac{1}{2}\left(1 + \frac{1}{1-\alpha}\right)$

18.02.2009 We
PETER ČERNO

# NTIN066 ZHRNUTIE

## DVOJITÉ HASHOVANIE

$h_1, h_2$, $\gcd(h_2(x), m) = 1 \ \forall x$

PREDPOKLAD: $\forall x \ \{h_1(x) + i\, h_2(x) \bmod m\}_{i=0}^{m-1}$ JE NÁHODNÁ PERMUTÁCIA RIADKOV

NEÚSP. PRÍPAD: $\boxed{q_i(n,m)}$ = PR. $h_1(x) + j h_2(x)$ SÚ OBS. PRE $j = 0, \dots, i-1$

$$= \frac{n}{m} \cdot \frac{n-1}{m-1} \cdots \frac{n-i+1}{m-i+1} = \underline{\underline{\frac{n}{m} q_j(n-1, m-1)}}$$

OČ. POČET TESTOV $C(n,m) = \sum_{j=0}^{n} (j+1)\left(q_j(n,m) - q_{j+1}(n,m)\right) = \boxed{\sum_{j=0}^{n} q_j(n,m)}$

$$C(n,m) = q_0(n,m) + \sum_{j=1}^{n} \frac{n}{m} q_j(n-1, m-1) = 1 + \frac{n}{m} C(n-1, m-1)$$

$$= \boxed{\frac{m+1}{m-n+1}} \quad \underline{\text{INDUKCIOU}} \quad \approx \frac{1}{1-\alpha}$$

ÚSP. PRÍPAD: $\frac{1}{n} \sum_{i=0}^{n-1} C(i,m) = \frac{m+1}{n}\left[\sum_{j=1}^{m+1} \frac{1}{j} - \sum_{j=1}^{m-n+1} \frac{1}{j}\right] \approx \frac{1}{\alpha} \ln\left(\frac{1}{1-\alpha}\right)$

## UNIVERZÁLNE HASHOVANIE

$H = \{h_i \mid i \in I\}$ c-UNIVERZÁLNY $|\{i \in I \mid h_i(x) = h_i(y)\}| \le \frac{c|I|}{m}$

EXISTENCIA: $h_{a,b}(x) = [(ax+b) \bmod N] \bmod m$

VLASTNOSTI: $\delta_i(x,y)$; $\sum_i \delta_i(x,S) \rightsquigarrow$ HORNÝ ODHAD OČ. DĺŽKY RET NA $h_i(x)$

MARKOVOVA VLASTNOSŤ: $P(\delta_i(x,S) \ge t\mu) < \frac{1}{t}$

DOLNÝ ODHAD NA VEĽKOSŤ $|I|$: $U_0 = U$, $U_{i+1} \subseteq U_i$, $|h(U_{i+1})| = 1$, $|U_{i+1}| \ge \lceil \frac{|U_i|}{m} \rceil$

MALÝ UNIVERZÁLNY SYSTÉM: $\log |I|$ JE $O(\log m + \log\log N)$

$t$ NAJMENŠIE PRIRODZENÉ: $\boxed{t \cdot \ln p_t \ge m \ln N}$ $\boxed{t \le m \ln N}$

$\boxed{H_1 = \{g_{c,d}(h_\ell) \mid t < \ell \le 2t,\ c,d \in \{0, \dots, p_{2t} - 1\}\}}$

$h_\ell(x) = x \bmod p_\ell$, $g_{c,d}(x) = ((cx+d) \bmod p_{2t}) \bmod m$

(1) VETA: $i(\ln i + \ln\ln i) > p_i > i \ln i \Rightarrow |I| = t p_{2t}^2 \le \dots \le 4(1 + \log t)$

(2) c-UNIVERZALITA: $G_1 = \{(c,d,\ell) \mid \dots = \dots,\ h_\ell(x) \neq h_\ell(y)\}$, $G_2 = \{\dots = \dots\}$

(a) $|G_1| \le t m \left(\lceil \frac{p_{2t}}{m} \rceil\right)^2 \le \underline{\underline{\frac{|I|}{m}\left(1 + \frac{m}{p_{2t}}\right)^2}}$

(b) $|G_2|$: $L = \{\ell \mid t < \ell \le 2t,\ x \bmod p_\ell = y \bmod p_\ell\}$

$P = \prod_{\ell \in L} p_\ell$, $P \mid |x-y| \Rightarrow P < N$, $P > p_t^{|L|} \Rightarrow |L| \le \frac{\ln N}{\ln p_t} \le \underline{\underline{\frac{t}{m}}}$

$|G_2| \le |L| \cdot p_{2t}^2 \le \underline{\underline{\frac{|I|}{m}}}$

ODHAD VÝRAZU $\left(1+\frac{m}{p_{2t}}\right)^2$

$m \geq p_{2t} \Rightarrow t \ln p_t < t \ln p_{2t} \leq m \ln m \leq m \ln N$ ↯ $\Rightarrow m < p_{2t}$

DÁ SA UKÁZAŤ $m < \frac{p_t}{\ln t} \rightsquigarrow \frac{m}{p_{2t}} < \frac{1}{2}$

<u>DOLNÝ ODHAD NA VEĽKOSŤ c</u>: $c \geq \frac{N-m}{N-1} > 1 - \frac{m}{N}$

<u>LEMMA</u>: $\sum_{i=0}^{m-1} b_i(b_i-1) \geq b\left(\frac{b}{m}-1\right)$

MAJME $f: U \to S$, $A = \{(u,v) \mid u \neq v,\ f(u)=f(v)\}$, $k_s = |f^{-1}(s)|$

$|A| = \sum_{s \in S} k_s(k_s-1) \geq N\left(\frac{N}{m}-1\right)$ ↯

$\Rightarrow |I| \cdot N\frac{N-m}{m} \leq \sum_i \{(u,v) \dots\} \leq \sum_{(u,v)} \{i \in I \dots\} \leq N(N-1)\, c\, \frac{|I|}{m}$

## <u>PERFEKTNÉ HASHOVANIE</u>

$H(N,m,n)$ – PERFEKTNÝ : $\forall S \in \binom{U}{n}\ \exists$ PERF. $h \in H$ $U \to \{0,\dots,m-1\}$

<u>DOLNÝ ODHAD $|H| \geq 1$</u>: $h$ PERF. PRE $S \Leftrightarrow |h^{-1}(i) \cap S| \leq 1\ \forall i$

PRE DANÚ $h$ : # PERF. $S \in \binom{U}{n}$ JE $\boxed{\sum \prod_{j=0}^{n-1} |h^{-1}(i_j)| \quad 0 \leq i_0 < \dots < i_{n-1} < m}$

$\leq \binom{m}{n}\left(\frac{N}{m}\right)^n \Rightarrow |H| \geq \binom{N}{n} \Big/ \binom{m}{n}\left(\frac{N}{m}\right)^n$

<u>DOLNÝ ODHAD $|H| \geq 2$</u>: $H = \{h_1, \dots, h_t\}$, $U_0 := U$, $U_{i+1} \subseteq U_i$, $|h_{i+1}(U_{i+1})| = 1$

MUSÍ BYŤ $|U_t| \leq 1 \Rightarrow t \geq \log N / \log m$

<u>EXISTENCIA</u>: $M(H) \in \{0,\dots,m-1\}^{N \times t}$, PRE $S \in \binom{U}{n}$ JE $\left(m^n - \prod_{i=0}^{n-1}(m-i)\right)^t m^{(N-n)t}$ NEPERF. MATIC

STAČÍ $\binom{N}{n} \cdot (\dots) < m^{Nt}$ STAČÍ $t > n(\ln N)\, e^{n^2/m}$

$-\ln\left(1 - \frac{\prod_{i=0}^{n-1}(m-i)}{m^n}\right) \geq \prod_{i=0}^{n-1}\left(1-\frac{i}{m}\right) \geq e^{-\frac{n^2}{m}}$

<u>KONŠTRUKCIA</u>: $h_k(x) = (kx \bmod N) \bmod m \quad k = 1, \dots, N-1$

$b_i^k = |S_i^k| = |\{x \in S \mid h_k(x) = i\}|$, $b_i^k \geq 2 \Rightarrow (b_i^k)^2 - b_i^k \geq 2$; $b_i^k \leq 1 \Rightarrow _ = 0$

<u>VETA</u>: $h_k$ JE PERF. $\Leftrightarrow \sum_{i=0}^{m-1}(b_i^k)^2 - n < 2$

ODHAD $\underline{\sum_{k=1}^{N-1}\left[\sum_{i=0}^{m-1}(b_i^k)^2 - n\right]} = \sum_{x,y \in S,\ x \neq y} \overbrace{|\{k \mid 1 \leq k \leq N-1\ \ h_k(x) = h_k(y)\}|}^{\leq 2\lfloor \frac{N-1}{2} \rfloor}$

$\Rightarrow \boxed{\exists k:\ \sum_{i=0}^{m-1}(b_i^k)^2 \leq \frac{2n(n-1)}{m} + n}$

VIAC AKO $\frac{N-1}{4}$ TAKÝCH $k$, ŽE $\sum (b_i^k)^2 < \frac{3n(n-1)}{m} + n$, $\frac{N-1}{2} \dots \frac{4n(n-1)}{m} + n$

$\Rightarrow$
- (a) $n = m$, DET. V ČASE $O(nN)$ NÁJDE $k$: $\sum (b_i^k)^2 < 3n$
- (b) $n = m$, OČ. ČAS $O(n)$ NÁJDE $k$: $\sum (b_i^k)^2 < 4n$
- (c) $m = \underline{n(n-1)+1}$ DET. V ČASE $O(nN)$ : PERF. $h_k$
- (d) $m = \underline{2n(n-1)}$ OČ. ČAS $O(n)$ : PERF. $h_k$

18.02.2009 We
PETER ČERNO

# NMNO66 ZHRNUTIE

PRE $m$ NECH $q$ JE # PRVOČÍSEL, KT. DELIA $m$

$m \geq q! \geq \left(\frac{q}{e}\right)^q \Rightarrow \ln m \geq c q \ln q$, $\exists q_1 \geq q$ $\ln m = c q_1 \ln q_1$

$\Rightarrow q \leq q_1 \leq \frac{2 \ln m}{c(\ln \ln m - \ln c)} \in O\left(\frac{\log m}{\log \log m}\right)$

MAJME $S = \{s_1 < \dots < s_n\} \subseteq U$ $d_{i,j} = s_j - s_i$ $\forall 1 \leq i < j \leq n$

$D := \prod d_{i,j} \leq N^{n^2}$

POČET PRV. DELITEĽOV JE NAJVIAC $c \frac{\ln D}{\ln \ln D}$

$\Rightarrow$ HĽADÁME $p$ : $\phi_p = x \bmod p$ JE PERF. PRE $S$

$p_t \leq 2t \ln t \Rightarrow p \leq \dots = O(\ln D) = O(n^2 \ln N)$

$\Rightarrow$ METÓDA FREDMAN, KOMLÓS, SZEMERÉDI

## DYNAMICKÉ PERFEKTNÉ HASHOVANIE

$h_k \in \mathcal{H}_s$ S PRAVD. ASPOŇ $\frac{1}{2}$ SPĽŇA

$\sum_{i=0}^{s-1} (b_i^k)^2 < \frac{4n^2}{s} + n$

$s$ VOLÍME AKO $\sigma(|S|)$, $\sigma(n) = \frac{4}{3}\sqrt{6}(1+\varepsilon)n$

## PERF. HASHOVANIE DO TAB. $m = |S|$

POMOCOU ACYKLICKÝCH $r$-HYPERGRAFOV, 2 KROKY

## EXTERNÉ HASHOVANIE

KRITICKÉ SLOVO, ADRESÁR, STRÁNKA, $b > 1$

$h : U \to \{0,1\}^k$, $\alpha \in \{0,1\}^{<k}$, $h_S^{-1}(\alpha) = \{s \in S \mid \alpha \sqsubseteq h(s)\}$

OČAKÁVANÝ POČET STRÁNOK: $\frac{n}{b \ln 2}$,

VELKOSŤ ADRESÁRA: $\frac{e}{b \ln 2} n^{1+1/b}$

## (a,b)-STROMY $2 \le a$, $2a-1 \le b$

STROM O VÝŠKE $h$ MÁ $2a^{h-1} \le n \le b^h$ LISTOV

PRE KAŽDÉ $s \in S$ OKREM NAJVÄČŠIEHO EXISTUJE PRÁVE JEDEN VN. VRCHOL $v$ A $i$ : $H_v(i) = s$

PARALELNÁ IMPLEMENTÁCIA, $b \ge 2a$

## A-SORT

$(x_1, \dots, x_n)$ z $U$, $f_i = |\{ j \mid i < j,\ x_i > x_j \}|$, $F = \sum f_i$

KEĎ A-INSERT PRI HĽADANÍ MIESTA PRE PRVOK $x$ ŠUDE DO VÝŠKY $h$, ČAS $O(h)$, V PRVKY POD 1. VRCHOLOM VO VÝŠKE $h-1$ SÚ $< x$ A JE ICH ASPOŇ $a^{h-2}$

PRE $x = x_i$ JE ICH $f_i$ TJ. $a^{h-2} \le f_i$

CELKOVÝ ČAS $O(\underbrace{\sum_{i=1}^{n} \log f_i}_{n \log F/n} + n)$ NA CELKOVÉ STIEPENIE

HLADINOVO PREPOJENÉ (a,b) STROMY

## AMORTIZOVANÁ ZLOŽITOSŤ $a \ge 2$, $b \ge 2a$

$$\boxed{c := \min\left\{ \min\left(2a-1, \left\lceil \frac{b+1}{2} \right\rceil\right) - a,\ b - \max\left(2a-1, \left\lfloor \frac{b+1}{2} \right\rfloor\right) \right\}} \quad c \ge 1$$

$St_k$, $St = \sum_k St_k$, $Sp_k$, $Sp$, $P_k$, $P$

(1) $P \le n$, $(2c-1)St + c\,Sp \le n + c + \frac{c(h-2)}{a+c-1}$

(2) $St_k + Sp_k + P_k \le \frac{2(c+2)n}{(c+1)^k}$

PRE VNÚT. $v \ne r$: $b(v) := \min\{\rho(v) - a,\ b - \rho(v),\ c\}$
PRE KOREŇ $r$: $b(r) := \min\{\rho(r) - 2,\ b - \rho(r),\ c\}$

!

PRE VNÚT. $v \ne r$ PLATÍ

(1) $b(v) \le c$
(2) $\rho(v) = a \vee \rho(v) = b \Rightarrow b(v) = 0$
(3) $\rho(v) = a-1 \vee \rho(v) = b+1 \Rightarrow b(v) = -1$
(4) $\rho(v) = 2a-1 \Rightarrow b(v) = c$

AK $v, v'$ SÚ VNÚT. V. $\ne r$
$\rho(v') = \left\lceil \frac{b+1}{2} \right\rceil$, $\rho(v'') = \left\lfloor \frac{b+1}{2} \right\rfloor$
$\Rightarrow b(v') + b(v'') \ge 2c - 1$
$b(r) \le c$

18.02.2009 We
PETER ČERNO

# NTIN066 ZHRNUTIE

## PARCIÁLNY $(T, r, v)$ STROM $\quad a-1 \le \rho(v) \le b+1$

(1) $\rho(v) = \rho(v') \Rightarrow b_T(v) = b_{T'}(v')$

(2) $|\rho(v) - \rho(v')| = 1 \Rightarrow b_{T'}(v') \ge b_T(v) - 1$

$$\boxed{\begin{array}{l} \text{PRIDANIE/UBRANIE VRCHOLU: } b_1(T') \ge b_1(T) - 1 \\ \text{ŠTIEPENIE: } b_\ell(T') \ge b_\ell(T) + 2c,\ b_{\ell+1}(T') \ge b_{\ell+1}(T) - 1 \\ \text{SPOJENIE: } b_\ell(T') \ge b_\ell(T) + c + 1,\ b_{\ell+1}(T') \ge b_{\ell+1}(T) - 1 \\ \text{PRESUN: } b_\ell(T') \ge b_\ell(T) \end{array}}$$

$$\boxed{b_h(T_k) \ge 2c\, St_h + (c+1)\, Sp_h - St_{h-1} - Sp_{h-1}}, \quad St_0 + Sp_0 = \#\text{LISTOV}$$

$$\Rightarrow \boxed{b(T_k) \ge (2c-1)\, St + c\, Sp - n}$$

## ODHAD $b(T_k)$ ZHORA

KEĎ $T$ JE $(a,b)$-STROM S $m$ LISTAMI, POTOM

$$\boxed{0 \le b(T) \le c + (m-2)\frac{c}{a+c-1}}$$

$\nabla$ $0 \le j < c$: $m_j = \#$ VN. VRCHOLOV $\neq$ KOREŇ S $a+j$ SYNMI $\longrightarrow b_T(v) \le j$

$m_c = \#$ —"— S ASPOŇ $a+c$ SYNMI $\quad b_T(v) \le c$

$$\Rightarrow \boxed{b(T) \le \overset{\text{KOREŇ}}{c} + \sum_{j=0}^{c} j \cdot m_j}$$

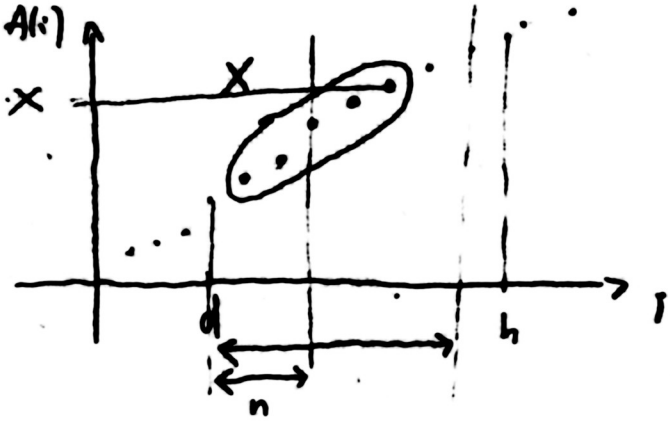
$$\boxed{2 + \sum_{j=0}^{c} (a+j)\, m_j \le \sum_{\text{W.VR.}v} \rho(v) = m + \sum_{j=0}^{c} m_j}$$

$$\Rightarrow \sum_{j=0}^{c} (a+j-1)\, m_j \le m-2 \Rightarrow \underline{\underline{b(T) \le c + \frac{c}{a+c-1}(m-2)}}$$

## ZOVŠEOBECNENÉ KVADR. VYHĽ.

$p_i$ = PRAVD., ŽE HRANICI BLOKU <u>ASPOŇ i DOTAZOV</u>

$\Rightarrow C = \sum_{i \geq 1} i(p_i - p_{i+1}) = \sum_{i \geq 1} p_i$

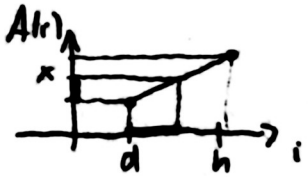



X NÁH. PREMENNÁ

$X = |\{i \mid i > d,\ A(i) \leq x\}|$

KEĎ SA V BLOKU POLOŽÍ ASPOŇ i DOTAZOV, POTOM

$\boxed{|X - n| \geq \lfloor \frac{i-2}{2} k \rfloor}$




$p_i \leq P\left(|X - n| \geq \lfloor \frac{i-2}{2} k \rfloor\right)$

PRAVD., ŽE $A(i) \leq x$ PRE $d < i < h$ JE $\frac{x - A(d)}{A(h) - A(d)} = p$

$\boxed{P(X = j) = \binom{h-d}{j} p^j (1-p)^{h-d-j}} \Rightarrow X \sim Bi(h-d, p)$

$\Rightarrow \mu = p(h-d) = n, \quad \sigma^2 = p(1-p)(h-d)$

$\Rightarrow P\left(|X - n| \geq \lfloor \frac{i-2}{2} k \rfloor\right) \overset{\text{ČEBYŠEV}}{\leq} \frac{\sigma^2}{(\lfloor \frac{i-2}{2} k \rfloor)^2} \leq \frac{1}{(i-2)^2}$

$\Rightarrow C = \sum p_i < 6$

$2 \leq \text{krok}_i \leq \sqrt[2]{|S|} \Rightarrow j = \log_2 \log_2 |S|$

NAJHORŠÍ PRÍPAD: BIN. DOTAZOV JE NAJVIAC TOĽKO, KOĽKO PRI BIN. VYHĽ. $\Rightarrow O(\log |S|)$