Kapitola

Pořádek jako zvýšení struktury

úsek | text | datum | Shrnutí odstavce

Představme si úklid kolem rodinného domu: Tu je opřen smeták, pod schody je lopatka na smetí, na řezivu je odložen smetáček, u studny je pohozen rýžový kartáč, ….Člověk, který se v prostředí pohybuje, si vytváří mimoděk v hlavě seznam těchto věcí. Ve chvíli, kdy se manželka zeptá: "Neviděl jsi rýžový kartáč?", proběhne v hlavě seznam a hlásí: "Počkej, někde jsem jej zahlédl. Zkusím si vzpomenout." a pak si buď vzpomene, že leží u studny, nebo rezignuje. V určitém okamžiku se rozhodne, že takový chaos je náročnější na čas, než když teď uklidí. Určí místo v koutku pod schodištěm, kam se budou po použití vracet všechny úklidové předměty. Snese sem výše uvedené předměty, dohledá kbelíky, stěrku na okna, hadry na podlahu a prachovku.

Tím vznikl strukturovaný podseznam "Úklidový kout" a napříště je definována jasná akce, kde začít hledat rýžový kartáč, i kam uklidit sáčky k vysavači.
V úklidovém koutě mohou být předměty volně poházeny, ale v určitém stádiu zaplněnosti prostoru zase začíná být procházení jednoduchého netříděného seznamu časově náročnější, než přeskupení do seznamu strukturovaného.
Řekněme, že nádoby budou vlevo, předměty s dlouhou násadou pověšené vpravo na zdi nahoře, uprostřed ostatní.

Hypotéza **Nutnost pořádku souvisí s narůstající délkou seznamu**

Důvod k vytvoření a udržování strukturovaného seznamu souvisí s délkou původního nestrukturovaného seznamu, takže procházení všech položek se stává časově náročnější, než jednorázová investice do vytvoření struktury. Ta se vyplatí vůči počtu zkrácených vyhledávání.
Matematika časů pro vyhledání v nestrukturovaném seznamu:

Prohlédnutí jedné položky seznamu = p
zatřídění do tematické podskupiny = t
Průměrná doba nalezení položky v lineárním netříděném seznamu délky "n" DN(n)=n*p/2
(Ta 2 ve jmenovateli říká, že položku najdeme průměrně po projití poloviny seznamu)
Doba do zjištění, že položka v seznamu není NN(n)=n*p
Doba do rozhodnutí o položce (tj. nalezena/nenalezena) je zase závislá na na pravděpodobnosti nenalezení. Triviálně předpokládejme, že čím je seznam delší, tím je pravděpodobnější, že v něm položka bude. Při hypotetickém seznamu všech známých věcí v něm bude určitě. Naopak, čím je seznam kratší, tím je pravděpodobnější, že v něm položka nebude.

Máme tedy funkci pravděpodobnosti ní(0)=1, ní(oo)=0, ní je ryze klesající.

K ní spřažená funkce pravděpodobnosti nalezení :
mí(0)=0, mí(oo)=1, mí ryze rostoucí.
Platí ní(i) + mí(i) = 1
pro jakoukoli délku seznamu "n".
Nyní jsme schopni zapsat dobu do rozhodnutí o položce v seznamu délky "n":
TR(n) = n*p/2 * mí(n) + n * p * ní(n) = n * p *(1/2 * (1- ní(n) + ní(n)) = n*p/2 (1+ní(n)) (1)

lim TR(n);n->oo= p/2 * lim[n* ní(n)];n->oo = oo
protože je zřejmé, že hledání v nekonečném seznamu musí trvat nekonečně dlouho, tedy z toho máme charakteristiku funkce ní, že O(ní) > O(1/n). (Vlastnost O() charakterizuje podobnost průběhu funkcí)
Docela by se nám hodilo, abychom uměli tuto funkci nějak matematicky popsat. Pokud by byla řešením běžné diferenciální rovnice, mohl by být výsledek pro velká n např.
ní(n)= 1/ln(n+1) (2)

, která požadavky splňuje.
Při hlubším zkoumání v matematických teoriích by jistě bylo možné najít přesnější tvar funkce ní. Nicméně věřím, že po dosazení přesnější funkce do následujících výpočtů se níže odvozená matematika strukturovaného seznamu zásadním způsobem nezmění.
TR(n) = n*p/2 * (1+1/ln(n+1)) (3)

Porovnejme mezi sebou lineární netříděný seznam délky 63 se strukturovaným:
TR(63)= 39,1 * p
Nejprve za konkurenta zvolme dichotomický seznam o přibližně stejném počtu položek, a který má 6 úrovní:
Počet hledání bude: (2+2+2+2+2+2)=12 , tedy TR(12)=8,2 * p

což je 4,77 x kratší, než u seznamu netříděného.
Porovnejme mezi sebou ještě další chotomie:
Řádu 1 (lineární) = 39
Řádu 2 (dichotomický) = 8,34

V chotomickém seznamu ve skutečnosti nehledáme např. v seznamu délky 15, ale 3 x v seznamu délky 5. Odvodím vzorec pro délku vyhledávání v k-chotomickém seznamu délky "n"

Tk(n)=ln(n)/ln(k) * (k*p/2 * (1+1/(ln(1+k))
Pokud nyní kvůli vyšetření průběhu těchto funkcí sestrojím podíl k-chotomického hledání vůči dichotomickému hledání, dostanu:

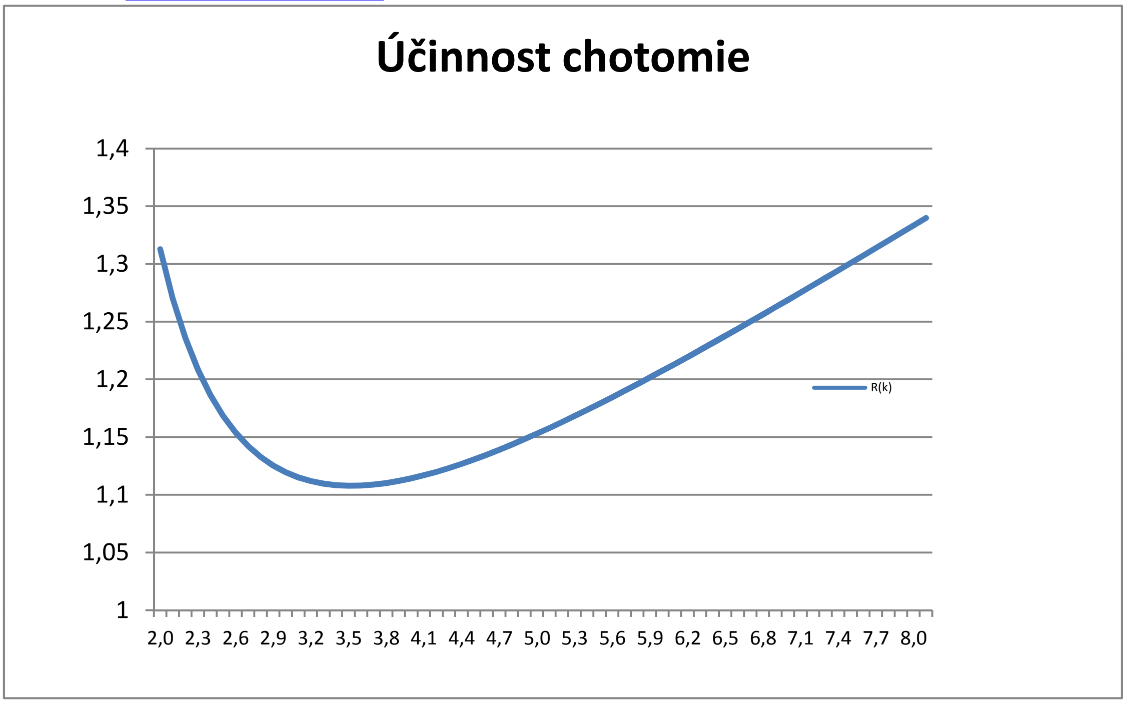
Rk(n)= Tk(n)/T2(n)=ln(2)/ln(k) * k/2 * (1+1/(ln(1+k))/(1+1/ln3)
(4)

a ta je již nezávislá na "n". Můžeme proto definovat funkci

R(k)= K* k * (1+1/(ln(1+k))/ln(k) (5)

kde K=ln2/(2+1/ln3) = 0,238

Graf viz Účinnost chotomie



Tuto funkci můžeme vyšetřovat na existenci minima. Šlo by samozřejmě použit metodu první derivace rovnu 0, ovšem s ohledem na to, že je tato funkce založena na pouhém odhadu průběhu pravděpodobnostní funkce výrazem (2), nemá smysl usilovat o takovou přesnost. Vystačíme si s pohledem do grafu. Minimum se zde vyskytuje na hodnotě přibližně 3,5, což znamená, že ideální strukturování netříděného seznamu je na 3 až 4 podpoložky.

Jako připomínadlo tohoto poznatku můžeme sestrojit 14-stěn tak, že krychli odřízneme všech 8 rohů. Tím máme 6 čtverců a 8 trojúhelníků. Odříznutí zvolíme tak, aby strany čtverců i trojúhelníků byly stejně dlouhé, a těleso bylo co nejpravidelnější. (Viz obálka knihy).