
Tries

PV 062 ♦ Organizace souborů

Jan Staudek

<http://www.fi.muni.cz/usr/staudek/vyuka/>



Verze : jaro 2021

Trie, radix tree, prefix tree, . . .

- E.Fredkin, 1960, information retrieval
- trie (vyslov „try” jako „sky”) je m -ární strom
- datová struktura stromového typu používaná pro reprezentaci řetězců nad danou abecedou
- Na rozdíl od vyhledávacích stromů žádný uzel trie neuchovává klíč související s daným uzlem
- naopak, pozice uzlu v trie ukazuje, který klíč s uzlem souvisí
- každý uzel trie obsahuje m -prvkové pole ukazatelů, každý jeden prvek odpovídá znaku (prvku) abecedy, ve které jsou klíče vyjádřené
- všichni následníci uzlu trie mají společný prefix řetězu souvisejícího s daným uzlem

Trie, radix tree, prefix tree, . . .

- ❑ alfabetické klíče (v angličtině) požadují pro reprezentaci trie řádu 27
- ❑ trie řádu m může být prázdný
- ❑ pokud trie řádu m není prázdný, sestává z uspořádané posloupnosti m tries řádu m
- ❑ Kořen trie je asociován s prázdným řetězem
- ❑ hodnoty (data) se asociují se pouze s listy
- ❑ trie je strom, kde každý uzel reprezentuje jedno slovo nebo prefix alespoň jednoho slova
- ❑ Složitost hledání klíče v trie je daná délkou klíče nikoli počtem klíčů

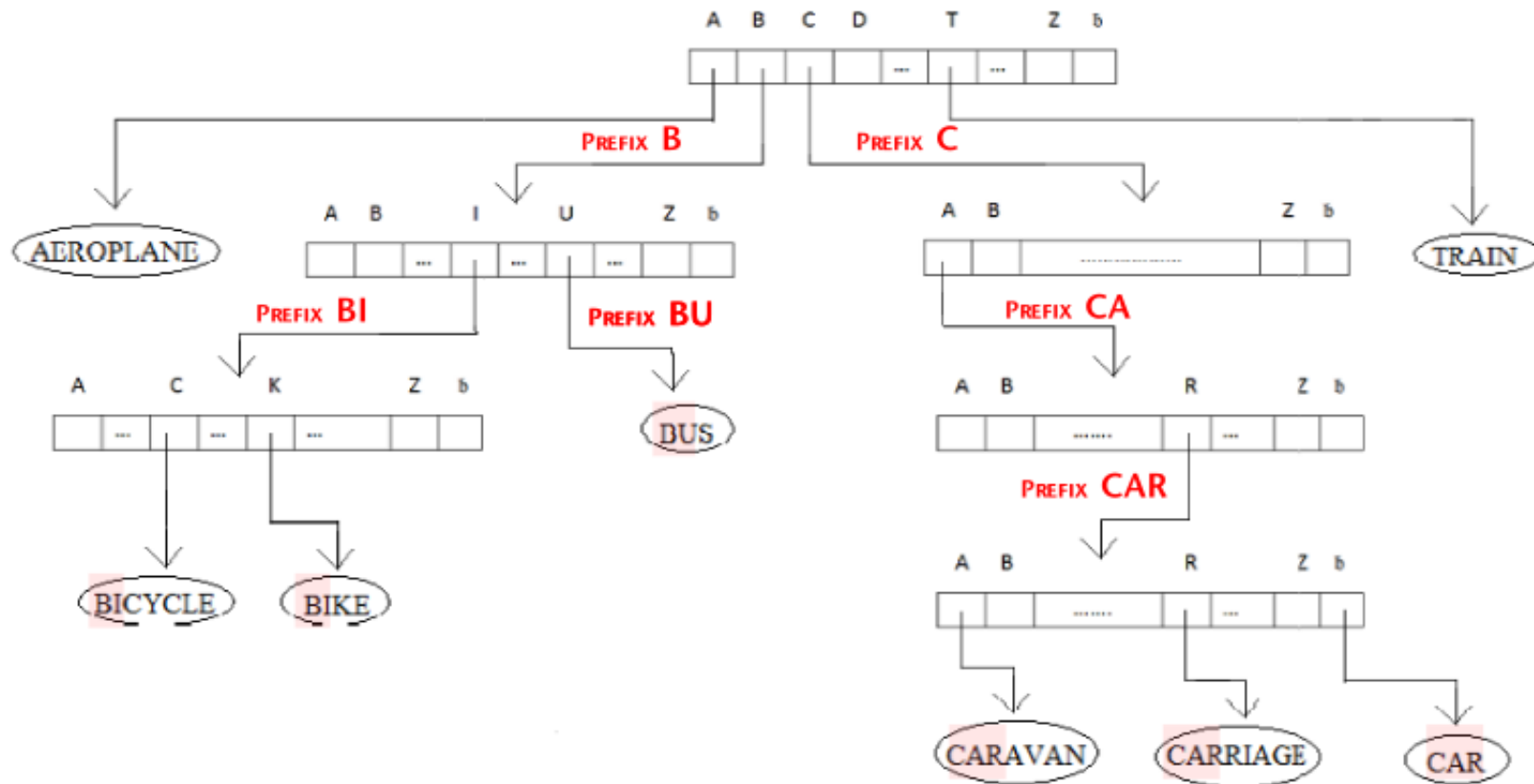
Trie, radix tree, prefix tree, . . .

- uzly – přímí následníci kořene – reprezentují prefixy délky 1
- uzly vzdálené 2 hrany od kořene reprezentují prefixy délky 2, . . .
- listový uzel vzdálený k hran od kořene reprezentuje slovo délky k
- každý uzel trie může být **rozvětovací** nebo **informační**
- rozvětovací uzel je pouze kolekce ukazatelů na rozvětovací uzly nebo na informační uzly
- informační uzel reprezentuje klíč uchovaný v trie
- trie **není m -ární vyhledávací strom**
 - ✓ řazení klíčových hodnot v uzlech nedodrží pravidla pro m -ární vyhledávací strom

Trie, příklad indexu slov

□ Index slov

aeroplane, bicycle, bike, bus, caravan, carriage, car, train



Trie

- jestliže je trie použitý jako index
 - ✓ pak listy obsahují adresy záznamů s odpovídajícími hodnotami klíčů
- jestliže je trie použitý jako reprezentátor hodnot
 - ✓ pak listy jsou indikátory existence záznamů s platnými hodnotami klíčů
- Oblasti použití, příklady
 - ✓ Rychlé prohledávání rozsáhlých textů, např. telefonních seznamů
 - ✓ *spell checking*
 - ✓ Síťový prohlížeč udržuje historii adres URL stránek, které uživatel navštívil. Organizováním této historie jako trie uživatel může zadat pouze prefix dříve použité adresy URL a prohlížeč může URL dokončit
 - ✓ Konstrukce adresářů pro rychlé procházení adresářů kapes s hašovanými záznamy při extenzibilním (rozšiřitelném) hašování

Trie, rekapitulace

- klíč v trie je uchovávaný v uzlech na cestě z kořene k vnějšímu uzlu a nikoliv jako celek v jednom uzlu
 - ✓ klíče musí být dělitelné na vhodné komponentní jednotky (symboly: znaky, cifry, ...)
- uzel m -árního trie
 - ✓ m -prvkové pole
 - ✓ každý prvek pole odpovídá jednomu „řádu“ klíče
 - ✓ prvek pole obsahuje ukazatel nebo indikátor prázdného místa
 - ✓ pozice prvku pole, obsahující ukazatel, určuje hodnotu řádu
- Jiný název – **radix searching**
 - ✓ základ je odvozený z abecedy použité pro kódování klíče
 - ✓ m je dáno základem použitým pro vyjádření klíče trie
 - 10, dekadické cifry,
 - 26/27, znaky/vč. mezery
 - ...

Co mohou být klíče v tries ?

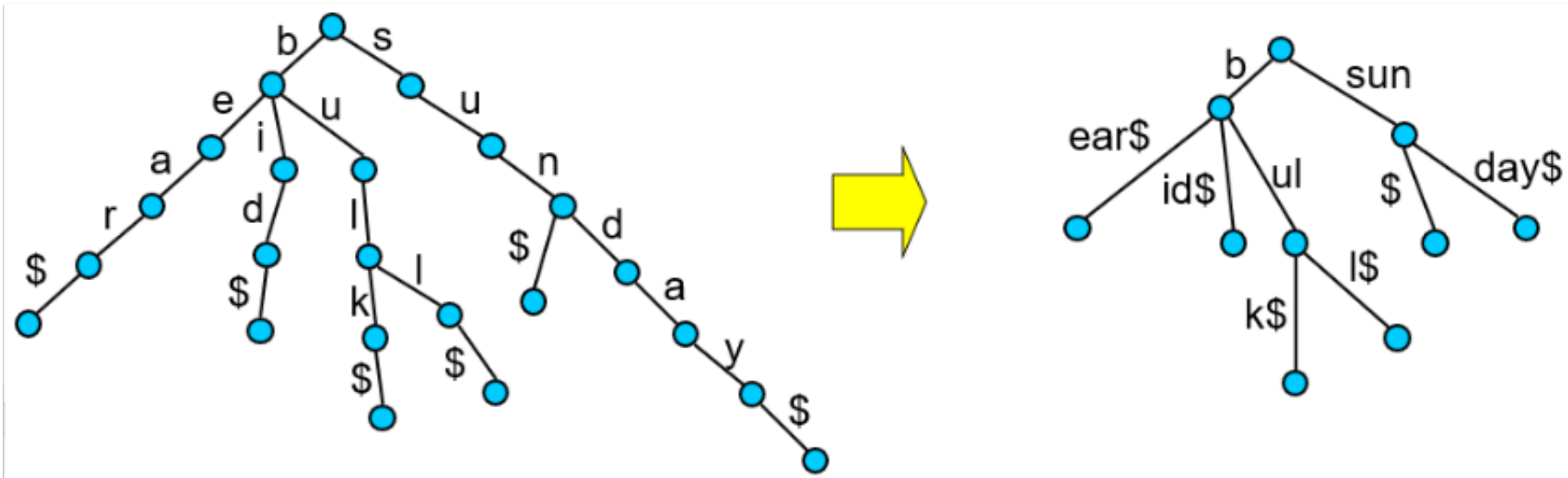
- řetězce (složené ze znaků)
- přirozená čísla (složená z číslic)
- celá čísla (složená z číslic, +, -)
- zlomky (složené z číslic, +, -, /)
- reálná čísla (složená z číslic, +, -, .)
- slova počítače (složená z 0, 1)
- objekty (složené z objektů)

Vlastnosti trie

- výška trie je dána délkou nejdelšího klíče
- výhody
 - ✓ délka hledání v trie je úměrná délce klíče
 - nikoli logaritmu počtu uzlů ve stromu
 - neúspěšné hledání může skončit na kterékoliv úrovni
 - ✓ pro případ velkého množství krátkých klíčů je prostorově efektivní, iniciální sekvence se sdílí
 - ✓ počet vnitřních uzlů je daný délkami klíčů, není nutné strom vyvažovat
 - ✓ při vkládání jsou rychlejší než hašované tabulky, nemusí se přepracovávat indexy při přeplnění
 - ✓ ...

Komprimovaný trie

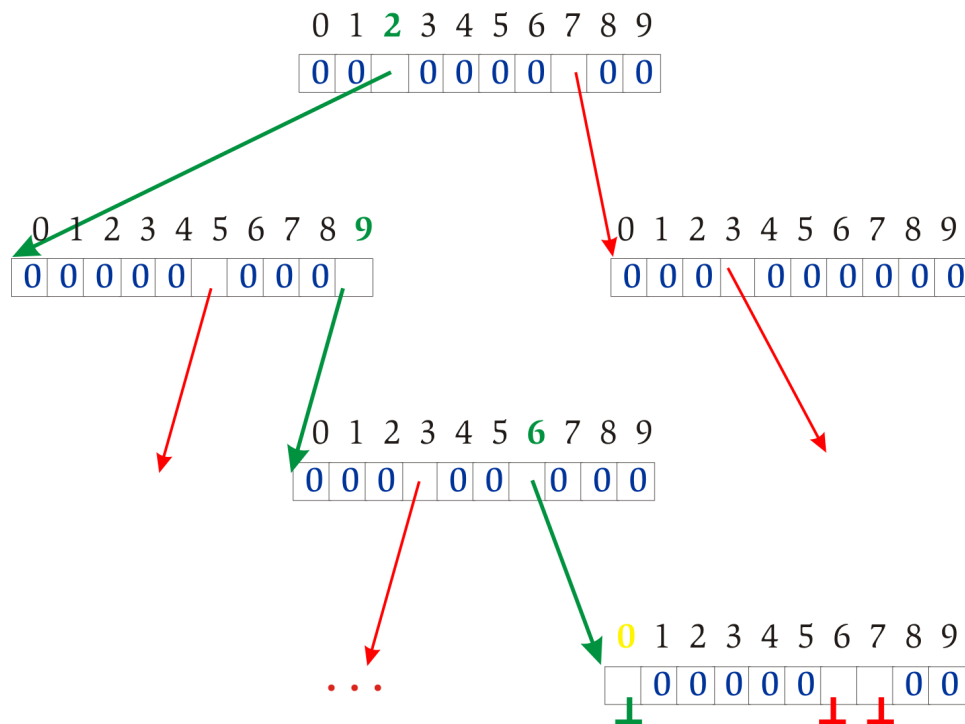
- Řetězec jednopotomkových uzlů se nahradí hranou ohodnocenou jedinečným řetězcem
- Každý nelistový uzel (kromě kořene) má alespoň dva potomky



Trie, příklad

□ množina hodnot klíčů

✓ 293..., 2960, 2966, 2967, 25..., 73...,



10. prvek pole na 2. úrovni
10-árního radixového trie
odkazuje na podstrom
reprezentující klíčové
hodnoty s druhou cifrou 9

Neexistence hodnoty/klíče 3897
se pozná už z 1. úrovně

0 ... Prázdný ukazatel

Trie, příklad aplikace v rozšiřitelném hašování

- implementace adresáře pomocí trie a pomocí tabulky

