Hashing

Hash functions → Hash code maps
→ compression maps

Open addressing → Linear Probing
→ Double hashing.

→ Good hash function

(i) quick to compute
(ii) distribute keys uniformly
(iii) good hash fns are very rare - Birthday paradox.

→ How to deal with non integer keys?

(i) find some way of turning keys into integers
eg. add ASCII values of a string.

(ii) then use standard hash function on integers.

From keys to indices

→ The mapping of keys to indices of a hash table is called a hash function.

→ A hash fn. is usually the composition of two maps, a hash code map and a compression map.

→ a good hash function minimises collisions.

- Hash code map : key → integer.
- compression map : integer → [0, N-1].

Popular hash code maps

Integer cast : For numeric types with 32 bits or less, we can interpret the bits of the number as an int.

Component sum : For numeric types with more than 32 bits, we can add the 32 bits components

component-sum hash code is bad for strings?

→ Polynomial accumulation : for strings of a natural language, combine the character values (ASCII) $a_0 a_1 a_2 \cdots a_{n-1}$ by viewing them as coefficients of a polynomial.

$$a_0 + a_1 x + a_2 x^2 \cdots x^{n-1} a_{n-1}$$

Polynomial is computed with Horner's Rule. at a fixed value x:

$$a_0 + x(a_1 + x(a_2 + \cdots x(a_{n-2} + x a_{n-1}))\cdots)$$

→ The choice $x = 33, 37, 39, 41$ gives almost 6 collisions on a vocabulary of 50,000 English words.

compression maps

→ Use the remainder.

$h(k) = k \bmod m$, k is key, m the size of the table

→ Need to choose m.

$m = b^e$ (bad)

→ If m is a power of 2, $h(k)$ gives the c least significant bits of k.

→ all keys with same ending go to same place

→ m prime (good)

helps ensure uniform distribution.

Example

n = 2000 character strings

m = 701 , a prime near $\frac{2000}{3}$

not near a power of 2.

→ use

$h(k) = \lfloor m(kA \bmod 1) \rfloor$

→ k is the key, m is the size of table, A is a const. $0 < A < 1$

→ steps (i) map $0 \ldots k_{max}$ into $0 \ldots k_{max} A$.

(ii) take fractional part

(iii) map into $0 \ldots m-1$.