

CS 7260

Lecture on Nov 26th

Scriber : Kamal Shadi

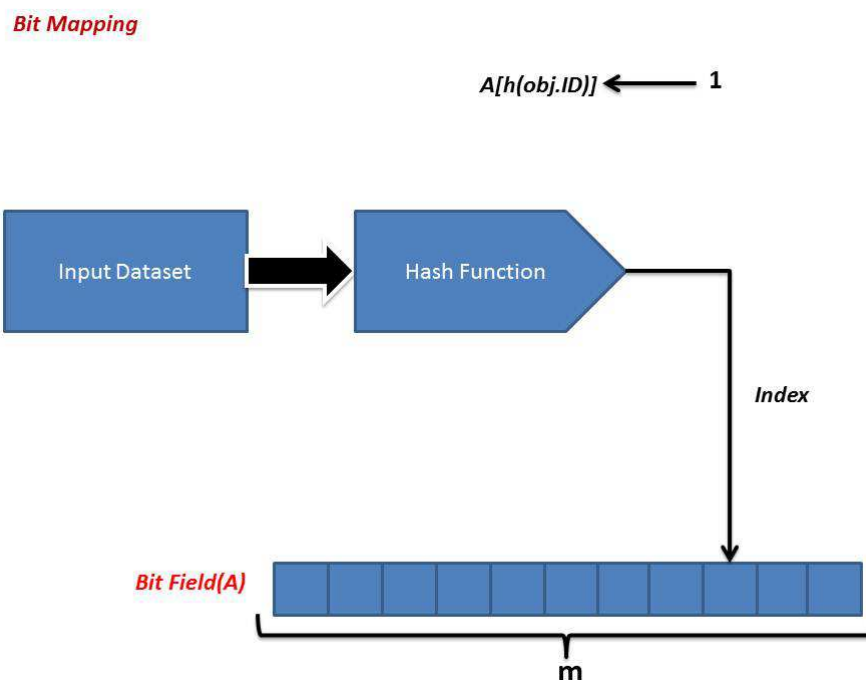
We have gone through two algorithms to count the number of distinct elements in two sets.

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Number of elements in each set is already known, so we need an algorithm for $|A \cap B|$. Here are the two visited algorithms:

- Mean Registers
- Bit map

Bit mapping algorithm is recapped in following figure:



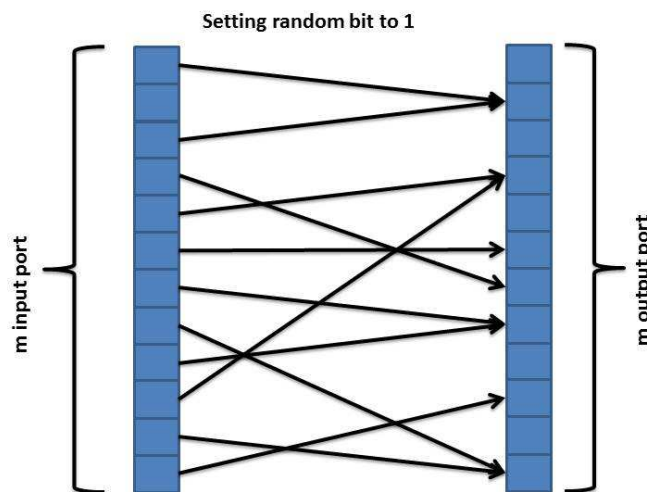
The probability that an arbitrary bit remains zero after m random bits are set to 1 is:

$$P_0 = \left(1 - \frac{1}{m}\right)^m$$

In limiting scenario when m gets larger: $\lim_{m \rightarrow \infty} P_0 = e^{-1}$

Therefore number of bits which are set to 1 is $m(1-e^{-1})$.

This number ($1-e^{-1}=63.2\%$) reminds us of head of line blocking in input queuing problem and actually this is the same problem with minor difference. Setting random bit to 1 is visualized in figure below. One could easily map it to HOL problem. The minor difference is that throughput of system shown in figure below would decrease to 58.6% after some cycles due to continuous competing.



Bit mapping update cost is very low, thus the computational complexity is negligible but it requires more space in linear fashion with number of distinct element.

The estimator for number of distinct element in bitmapping is:

$$\hat{d} = m \ln \frac{m}{c}$$

Where c is number of zeros. There is a limit on the number of distinct element detected by estimator because when $c=0$, the estimator output is infinity. Therefore, $m \ln(m)$ for $c=1$, is our guess for maximum number of distinct element. It could be proved with **Coupon Collector Problem**.

Coupon Collector Problem

There are m types of coupon with equal probability of collecting each. How many times you should buy a coupon to collect all types?

This is a famous problem with answer (for n coupon):

$$\frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1} = n \ln(n)$$

If the probability of collecting coupon i is P_i then the answer would be:

$$\text{number of times} = \int_0^{\infty} [1 - \prod_{i=1}^m (1 - e^{-P_i t})] dt$$

In next session we will prove that with Poisson process.