Assume we have a universal family of hash functions. If we pick a random function from this family, it is universal if for any pair of keys $x \neq y$:

$$\Pr[h(x) = h(y)] \leq \frac{1}{m}.$$
load factor $\lambda := \frac{n}{m}$

search take $\Theta(1)$ time in expectation, but there will be keys for which using $h$ causes longer search times.

What if $\lambda$ is really small?

expected $\neq$ colliding pairs:

$\mathbb{E}[h(x) = h(y)] \leq \binom{n}{2} \frac{1}{m} = \frac{n(n-1)}{2m}$
So, if \( m = n^2 \), expected # coll. is \( \leq \frac{1}{2} \).

Could keep rebuilding until no collisions. Expect \( < 2 \) builds, but size is too big.
Perfect hashing:

Top level hash table of size $n$

If entry $u$ has $n_u$ keys, make it point to table of size $n_{u^2}$.

It has its own hash function.

Rebuild each little table until no collisions
To get a worst-case $\Theta(1)$ time search.

Total size of little tables:

$$\sum_{\omega} n_\omega^2 = \sum_{\omega} \left( \sum_{\omega} \left[ h(x) = \omega \right] \right)^2$$

$$= \sum_{\omega} \left( \sum_{\omega} \left[ h(x) = \omega \right] \left[ h(y) = \omega \right] \right)^2$$

$$+ 2 \sum_{\omega} \sum_{\omega} \left[ h(x) = \omega \right] \left[ h(y) = \omega \right]$$

$$= \left( \sum_{\omega} \left[ h(x) = \omega \right] \right)^2 +$$

$$2 \sum_{x \neq y} \omega \left[ h(x) = \omega \right] \left[ h(y) = \omega \right]$$
\[ \sum_{x} h(x) = \sum_{y} h(y) + 2 \sum_{x \neq y} \mathbb{E} \left[ h(x) = h(y) \right] \]

\[ \leq n + 2 \sum_{x \neq y} \mathbb{E} \left[ h(x) = h(y) \right] \]

\[ \leq n + 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{n} \]

\[ = 2(n-1) \]
Multiplicative Hashing

Use some prime $p$ that's larger than our universe of possible keys.

$(p > 2^{32}$ if keys are Java ints)

For any $a \in \{1, 2, \ldots, p-1\} \land b \in \{0, 1, \ldots, p-1\}$

$h_a(x) := ((ax + b) \mod p) \mod m$
It is the set $p(p-1)$ possible $h_{aj}$. Pick one uniformly at random.

**Lemma:** Let $a \in \{1, \ldots, p-1\}$. The set $\{az \pmod{p} \mid z \in \{1, \ldots, p-1\}\}$ has exactly $p-1$ distinct elements, all non-zero.

$\Rightarrow$ multiplication by a modulo $p$ defines a permutation of $\{0, \ldots, p-1\}$.

$\Rightarrow$ multiplicative inverse
of a mod p is well-defined.

Proof: For any \( z \in \{ 1, \ldots, p-1 \} \), neither \( a \) nor \( z \) is divisible by \( p \) (both < \( p \)).

\( p \) is prime, so \( az \) is not divisible by \( p \) either.

So \( az \mod p \neq 0 \).

Suppose \( az \mod p = az' \mod p \) for some \( z, z' \in \{ 1, \ldots, p-1 \} \).

Assume \( z \equiv z' \mod p \).

Algebra to get \( a(z-z') \mod p = 0 \).

So \( z-z' = 0 \), so \( z = z' \).
\[(\text{For each } r = \{1, \ldots, p-1\}, \text{ at most one integer } z \in \{1, \ldots, p-1\} \text{ such that } az \mod p = r.\)\]
Thm: It is universal.
Fix any keys $x \neq y$. All arithmetic mod $p$, so assume $x, y \in \{0, \ldots, p-1\}$.

For $h_{a,b}(x) = h_{a,b}(y)$, we have
\[(ax + b \mod p = ay + b \mod p) \mod a \mid 0 \mod m\]

Previous lemma implies $ax \mod p \neq ay \mod p$. 

Fix any \( r, s \in \{0, \ldots, p-1 \} \) s.t. \( r \neq s \). What is probability that \( ax + b \mod p = r \) and \( ay + b \mod p = s \)?

\[
\iff \quad a(x-y) \equiv (r-s) (\mod p)
\]

Has one solution in \( a \):

\[
a \equiv (r-s) (x-y)^{-1} (\mod p)
\]

Also, \( 6 \equiv r - ax (\mod p) \)

So \( a + b \) are fixed by our choice of \( r + s \).
So...

\[ p \] 
\[ \{ (ax+b) \mod p = r + \] 
\[ a, b \} \] 
\[ (ay+b) \mod p = s \} \] 
\[ \frac{1}{p(p-1)} \] 

If this happens, 

\[ h_{a,b}(x) = h_{a,b}(y) \] 

if \( r \mod m = s \mod m \).

Need to know how many pairs \((r, s)\) with \(r \neq s\) but 
\[ r \mod m = s \mod m \].

Fix just \(r\). There are \(|p!/m|\) choices for \(s\).
(floor because \( r \neq s \))

\[ p \text{ is prime, so } \frac{p}{m} \text{ is not an integer, so } \]

\[ \left\lfloor \frac{p}{m} \right\rfloor = \frac{p}{m} - \frac{1}{m} = \frac{(p-1)}{p(p-1)} \]

\[ \frac{p-1}{m} = \frac{m}{m} \text{ pairs } (r,s) \]

\[ p \text{ choices for } r, \text{ so } \frac{p(p-1)}{m} \text{ pairs } (r,s) \]

\[ \Pr_{a,b} \left[ h_{a,b}(x) = h_{a,b}(y) \right] \leq \frac{p(p-1)}{m} \cdot \frac{1}{p(p-1)} \]

\[ = \frac{1}{m} \]
Carter–Wegman trick to avoid $t \mod$ operations.

$p$ must be a Mersenne prime, $p = 2^k - 1$ for some $k$.

Suppose keys a small enough so we can use $p = 2^{31} - 1$.

Consider any $r + y$ s.t.

$$r = y \pmod p$$

$$y = q'(p + 1) + r'$$

↑ quotient ← remainder
public static final int DIGITS = 31;
public static final int mersennep = (1 << DIGITS) - 1;

public static int universalHash(int x, int a, int b, int m) {
    long hashVal = (long) a * x + b;

    hashVal = ((hashVal >> DIGITS) + (hashVal & mersennep));
    if (hashVal >= mersennep) {
        hashVal -= mersennep;
    }

    return (int) hashVal % M;
}