String matching

Given two strings

$P[1..m]$ "pattern"

$T[1..n]$ "text"

Is P a substring of T?

Idea: scan text in order, stop when we find P

Brute force: $O((n-m)m)$ time

$P = AAAA..AAB$

$T = AAAA\dots AAAA\dots AAA$

English

"shift"

```
AlmostBruteForce(T[1..n], P[1..m]):
  for s ← 1 to n − m + 1
      equal ← True
      i ← 1
      while equal and i ≤ m
          if T[s + i − 1] ≠ P[i]
              equal ← False
          else
              i ← i + 1
      if equal
          return s
  return None
```

$O(mn)$ worst

$O(n)$ in practice (usually)

WLOG, strings are over $\{0,1,2,3,4,5,6,7,8,9\}$

interpret P and T as numbers

$$p = \sum_{i=1}^{m} 10^{m-i} P[i] \qquad t_s = \sum_{i=1}^{m} 10^{m-i} T[s+i-1]$$

31415

$$t_{s+1} = 10\left(t_s - 10^{m-1} T[s]\right) + T[s+m]$$

31415$\boxed{92653}$58927182892 64514

Pick a prime # q

$\textsc{NumberSearch}(T[1..n], P[1..m])$:
  $\sigma \leftarrow 10^{m-1}$ mod q
  $p \leftarrow 0$
  $t_1 \leftarrow 0$
  for $i \leftarrow 1$ to m
    $p \leftarrow 10 \cdot p + P[i]$ mod q
    $t_1 \leftarrow 10 \cdot t_1 + T[i]$ mod q
  for $s \leftarrow 1$ to $n - m + 1$
    if $p = t_s$ ← compare P and T[s...s+m-1]
      return s
    $t_{s+1} \leftarrow 10 \cdot (t_s - \sigma \cdot T[s]) + T[s+m]$ mod q
  return None

O(n) aritmetic operations

O(mn) time

with mod q → O(n) time!

with mod q + brute force check

time = $O(n + Fm)$

where F = # false matches

We'd really like $F < \frac{n}{m}$

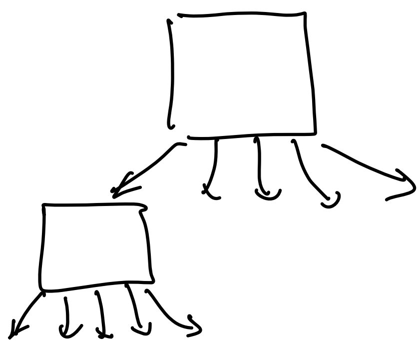
We can get $E[F] = O\left(\frac{n}{m}\right)$ $Pr[\text{false match}] \leq O\left(\frac{1}{m}\right)$

Zobrist hashing

hash value for every (piece, square)

hash value for board

$\bigoplus_{\text{piece}}$ hash(piece)



Update hash(board) for 1 move with 2-4 xors.

KarpRabin($T[1..n], P[1..m]$):

$q \leftarrow$ a random prime number between 2 and $\underbrace{\lceil m^2 \lg m \rceil}_{M}$

$\sigma \leftarrow 10^{m-1} \bmod q$

$\tilde{p} \leftarrow 0$

$\tilde{t}_1 \leftarrow 0$

for $i \leftarrow 1$ to m

  $\tilde{p} \leftarrow (10 \cdot \tilde{p} \bmod q) + P[i] \bmod q$

  $\tilde{t}_1 \leftarrow (10 \cdot \tilde{t}_1 \bmod q) + T[i] \bmod q$

for $s \leftarrow 1$ to $n - m + 1$

  if $\tilde{p} = \tilde{t}_s$

    if $P = T_s$ ⟨⟨brute-force $O(m)$-time comparison⟩⟩

      return s

  $\tilde{t}_{s+1} \leftarrow \big(10 \cdot (\tilde{t}_s - (\sigma \cdot T[s] \bmod q) \bmod q) \bmod q\big) + T[s+m] \bmod q$

return None

Lemma: Every integer x has $\le \lceil \lg x \rceil$ prime factors.

Proof: prime ≥ 2 □

assuming $p \ne t_s$

#primes $\le M$

$$\Pr[\tilde{p} = \tilde{t}_s] = \Pr[q \text{ divides } \underbrace{|p - t_s|}_{<10^m}] \le \frac{m}{\pi(M)}$$

$<10^m \rightarrow < O(m)$ prime factors.

Prime Number Theorem: $\pi(M) = \Theta\left(\frac{M}{\log M}\right)$

If we choose $M \ge m^2 \log m$

$$\Rightarrow \Pr[\tilde{p} = \tilde{t}_s] \le \frac{m}{\Theta(m^2)} = \Theta\left(\frac{1}{m}\right)$$

Carter Wegman

$$h_{a,b}(x) = ((ax + b) \bmod p) \bmod m$$

$$h_x(A) = h_A(x) = \left(\sum_{i=0}^{k-1} a_i x^i \bmod p\right) \bmod m$$

$$\boxed{h_b(P) = \sum_{i=0}^{m-1} b^i \cdot P[m-i] \bmod q}$$

interpret pattern as number in base b

random salt

FIXED

CarterWegmanKarpRabin($T[1..n], P[1..m]$):

    $q \leftarrow$ ***an arbitrary prime number larger than*** $\boldsymbol{m^2}$ M

    $\boldsymbol{b} \leftarrow$ **Random**$(\boldsymbol{q}) - \boldsymbol{1}$ ⟨⟨*uniform between* $\cancel{0}$ *and* $q-1$⟩⟩

    $\sigma \leftarrow \boldsymbol{b}^{m-1} \bmod q$

    $\tilde{p} \leftarrow 0$

    $\tilde{t}_1 \leftarrow 0$

    for $i \leftarrow 1$ to m

        $\tilde{p} \leftarrow (\boldsymbol{b} \cdot \tilde{p} \bmod q) + P[i] \bmod q$

        $\tilde{t}_1 \leftarrow (\boldsymbol{b} \cdot \tilde{t}_1 \bmod q) + T[i] \bmod q$

    for $s \leftarrow 1$ to $n - m + 1$

        if $\tilde{p} = \tilde{t}_s$

            if $P = T_s$ ⟨⟨*brute-force* $O(m)$*-time comparison*⟩⟩

                return s

        $\tilde{t}_{s+1} \leftarrow \big(\boldsymbol{b} \cdot \big(\tilde{t}_s - (\sigma \cdot T[s] \bmod q) \bmod q\big) \bmod q\big) + T[s+m] \bmod q$

    return None

q is prime $\Rightarrow$ division mod q works $\boxed{\mathbb{Z}/q\mathbb{Z} \text{ is a field}}$

$\Rightarrow$ any polynomial of degree $m-1$ has $\le m-1$ roots

$$\Pr[\tilde{p} = \tilde{t}_s] = \Pr_b\Big[\underbrace{P(b) - T_s(b)}_{\text{poly deg } m-1} = 0\Big] \le \frac{m-1}{q} = \frac{m-1}{m^2} < \frac{1}{m}$$