

1 Gii thiu

2 Phng php gom cm

2.1 Thut ton tm lng ging gn nht (Nearest Neighbor Search)

Algorithm 1 Nearest neighbor search

document d trong ton b d liu term t trong document d document d' c cha t cp nht khong cch
 $distance(d, d')$ tn timer score: $dis_{min}(d) = \min\{distance(d, d')\}$ thm document d vo index

2.2 Thut ton Locality Sensitive Hashing

Algorithm 2 Locality Sensitive Hashing

```

1: Khi to L hash tables
2: foreach document  $d$  mi do
3:   foreach hash table  $l$  trong  $L$  do
4:     Tnh hash code cho document  $d$ :  $LSH_l(d)$ 
5:     Thm tt c cc documents  $d'$  c  $LSH_l(d') = LSH_l(d)$  vo tp  $S$ 
6:      $dis_{min}(d) = 1$ 
7:     foreach document  $d'$  trong  $S$  do  $distance(d, d') < dis_{min}(d)$   $dis_{min}(d) = distance(d, d')$  thm
      document  $d$  vo index

```

Algorithm 3 Locality Sensitive Hashing ci tin

```

1: Khi to L hash tables
2: foreach document  $d$  mi do
3:   foreach hash table  $l$  trong  $L$  do
4:     Tnh hash code cho document  $d$ :  $LSH_l(d)$ 
5:     Thm tt c cc documents  $d'$  c  $LSH_l(d') = LSH_l(d)$  vo tp  $S$ 
6:      $dis_{min}(d) = 1$ 
7:     foreach document  $d'$  trong  $S$  do  $distance(d, d') < dis_{min}(d)$   $dis_{min}(d) = distance(d, d')$ 
       $dis_{min}(d) \geq threshold$  thm document  $d$  vo index

```

3 Phng php phn lp

4 Kt chng