Hash Table and Generic ADT

Project 3

Hash Table (1st week)

- After download and untar the file, change the directory to “strings”
- Create table.c to implement the set operations with hash table

```
SET *createSet(int maxElts);
void destroySet(SET *sp);
int numElements(SET *sp);
void addElement(SET *sp, char *elt);
void removeElement(SET *sp, char *elt);
char *findElement(SET *sp, char *elt);
char **getElements(SET *sp);
```

ADT

- An additional array of flags (EMPTY, FILLED, DELETED)

```
struct set{
   int count;
   int length;
   char **data;
   int *flags;
};
```

```
struct set{
   int count;
   int length;
   char **data;
   ~~int~~ char *flags;
};
```

Hash Table (1st week)

- Use the hash function from the lecture:

```
unsigned strhash(char *s)
{
    unsigned hash = 0;
    while (*s != '\0')
        hash = 31 * hash + *s ++;
    return hash;
}
```

(Hash value may exceed the max length of the set)

addElement in unsorted.c

```
void addElement(SET *sp, char *elt) {
      //declare variables
      assert (sp ≠ NULL && elt ≠ NULL);
      locn = search(sp, elt, &found);
      if (!found){
            assert(sp→count < sp→length);
            copy = strdup(elt);
            assert(copy ≠ NULL);
            [logic]
            sp→count++;
      }
}
```

```
sp→data[locn] = copy;
sp→flags[locn] = FILLED;
```

- What about removeElement?

Probing Logic and Search Function

- What happens when there is a conflict during insertion?
- What happens when we remove an element from the hash table?
- What do we return when the element is not found?
- 3 Cases:
  - FILLED
  - DELETED
  - EMPTY

A Simple Workflow

sp->data:							
sp->flag:	E	E	E	E	E	E	E

- (index+hash)%length (assuming length = 7):
- [elt1,elt2,elt3,elt4,elt5] -> hash function -> [7, 8, 14, 15, 21]
- Activities:
  Step 1: insert elt1
  Step 2: insert elt2
  Step 3: insert elt3
  Step 4: insert elt4
  Step 5: delete elt1
  Step 6: delete elt3
  Step 7: insert elt5
  Step 8: get all elements

Workflow Beggin

- Insert elt1 (hash value 7):

sp->data:	elt1						
sp->flag:	F	E	E	E	E	E	E

- Insert elt2 (hash value 8):

sp->data:	elt1	elt2					
sp->flag:	F	F	E	E	E	E	E

- Insert elt3 (hash value 14):

sp->data:	elt1	elt2	elt3				
sp->flag:	F	F	F	E	E	E	E

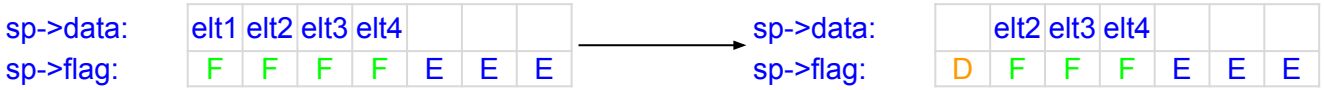
- Insert elt4 (hash value 15):

sp->data:	elt1	elt2	elt3	elt4			
sp->flag:	F	F	F	F	E	E	E

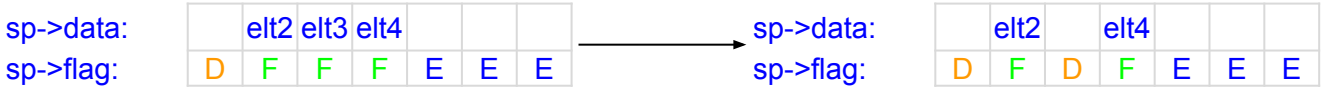
(We insert right when we see a 'E'.)

Workflow Continue

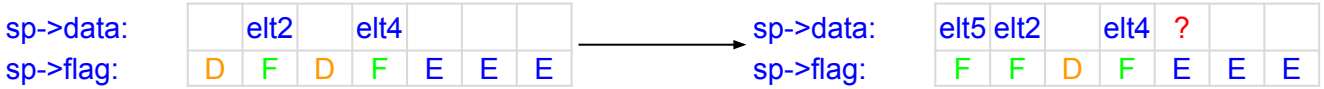
- Delete elt1 (hash value 7):



- Delete elt3 (hash value 14):
  - Do we stop when we see a 'D'?



- Insert elt5 (hash value 21):
  - Do we insert when we see a 'D'?
  - Where do we insert?



Workflow Finish

- Get all Elements:
  - Memcpy or copy every slot?
  - Two separate indexes

sp->data:	elt5	elt2		elt4			
sp->flag:	F	F	D	F	E	E	E
index:	0	1	2	3	4	5	6

elts:	elt5	elt2	elt4
index	0	1	2

Test Cases

- Same as Lab 2:

  ./unique /scratch/coen12/Macbeth.txt

  ./unique /scratch/coen12/Macbeth.txt /scratch/coen12/Bible.txt

  ./unique -l /scratch/coen12/Macbeth.txt /scratch/coen12/Bible.txt

  ./parity /scratch/coen12/Macbeth.txt

report.txt

- Copy the text file from project2 and add a new column for “hashing”

time ./unique /scratch/coen12/Macbeth.txt

```
unique
------
                              unsorted  sorted  hashing
GreenEggsAndHam.txt             ????    ????    ????
Macbeth.txt                     ????    ????    ????
Genesis.txt                     ????    ????    ????
HoundOfTheBaskervilles.txt      ????    ????    ????
TheWarOfTheWorlds.txt           ????    ????    ????
TreasureIsland.txt              ????    ????    ????
TheSecretGarden.txt             ????    ????    ????
TwentyThousandLeagues.txt       ????    ????    ????
TheCountOfMonteCristo.txt       ????    ????    ????
Bible.txt                       ????    ????    ????


parity
------
                              unsorted  sorted  hashing
GreenEggsAndHam.txt             ????    ????    ????
Macbeth.txt                     ????    ????    ????
Genesis.txt                     ????    ????    ????
HoundOfTheBaskervilles.txt      ????    ????    ????
TheWarOfTheWorlds.txt           ????    ????    ????
TreasureIsland.txt              ????    ????    ????
TheSecretGarden.txt             ????    ????    ????
TwentyThousandLeagues.txt       ????    ????    ????
TheCountOfMonteCristo.txt       ????    ????    ????
Bible.txt                       ????    ????    ????
```