

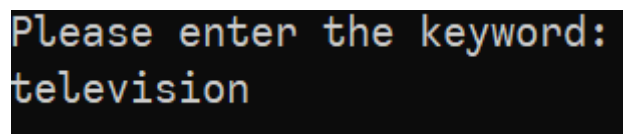
INTRODUCTION

Binomial Heaps are highly advantageous data structures invented by French computer scientist Jean Vuillemin in 1977. The most noteworthy aspect of this structure is that it allows different heaps to merge together, whereas providing $O(\log n)$ or $O(1)$ efficiency for all operations.

In this project, we are expected to read 42 files according to the input taken. Then, we should be holding our records in a binomial heap file. **Here, I preferred a max-heap property.** Since we would like to access the files with the most frequency of the given keyword, holding the ones with the highest similarity score at the topmost part is essential.

SCREENSHOTS AND IMPLEMENTATION

Before anything else, we should be taking the input from the user. I will be using “television” as an example.



```
Please enter the keyword:
television
```

Here, we take this keyword and start calculating similarity score for all 42 files. If a match is found in a document, its similarity score is incremented by one.

We have many documents containing the word “television”. Hence, let’s display the number of relevant documents:

please see next page

```
List of all documents with the keyword "television":

File number #2. Similarity score is 6.
File number #3. Similarity score is 1.
File number #4. Similarity score is 4.
File number #6. Similarity score is 1.
File number #13. Similarity score is 1.
File number #16. Similarity score is 5.
File number #19. Similarity score is 13.
File number #20. Similarity score is 1.
File number #31. Similarity score is 2.
File number #33. Similarity score is 3.
File number #34. Similarity score is 3.
File number #35. Similarity score is 2.
File number #38. Similarity score is 1.
File number #39. Similarity score is 7.
File number #40. Similarity score is 2.

~~~~~
```

From this point, we should build our priority queue using a binomial heap implementation to hold these records. Then, we will extract the maximums one by one.

We will use “INSERT” operation for all of these files. Then, our binomial heap will be constructed.

Inserting in a binomial heap actually means creating a new heap with the current node and then using UNION operation with the main heap that we already got.

My source code is as follows:

please see next page

a. INSERT operation

```
document* insertNodeToHeap(document* H, document* insertThis) { /*  
Inserting to a binomial heap actually means creating a new, empty heap and  
unifying it with what we already have. */  
  
document* H1 = initializeHeap();  
  
/* Set all fields as empty...*/  
  
insertThis->parent = NULL;  
  
insertThis->child = NULL;  
  
insertThis->sibling = NULL;  
  
insertThis->degree = 0;  
  
H1 = insertThis;  
  
/* Now, we have a heap from our node. */  
  
/* And unify it with the main heap we have! Also, upgrade it. */  
  
H = unionHeaps(H, H1);  
  
return H; /* Return our main heap. */  
  
}
```

During the insertion, we use UNION to connect our newly-created null heap to the main heap we have

b. UNION operation

c.

```
document* unionHeaps(document* H1, document* H2) { /* This function  
here is identical to what we've seen in our lecture slides. We iterate  
through our nodes, checking all  
possible conditions regarding the degrees and unifying them  
eventually.*/  
  
  
document* previous;  
document* next;
```

```

document* current;
document* HEAD = initializeHeap();
HEAD = mergeHeaps(H1, H2); /* merging the root lists, as we seen in
our lectures. */

if (HEAD == NULL) /* Taking precaution against empty heaps. */
    return HEAD;
previous = NULL;
current = HEAD;
next = current->sibling;
while (next != NULL) {
    if ((current->degree != next->degree) || ((next->sibling != NULL) /*
CASE ONE */
        && (next->sibling)->degree == current->degree)) {
        previous = current;
        current = next; /* MOVE ONE RIGHT */
    } else { /* CASE TWO */
        if (current->similarityScore > next->similarityScore) {
            current->sibling = next->sibling;
            Get_a_BT_k(next, current);
        } else { /* CASE THREE */
            if (previous == NULL)
                HEAD = next;
            else /* CASE FOUR */
                previous->sibling = next;
            Get_a_BT_k(current, next);
            current = next;
        }
    }
    next = current->sibling;
}
return HEAD;
}

```

This was the enqueue part. For dequeue operation, I search the linked list of roots and find the maximum node as shown in our lectures. Then, I set its flag to -1. After this, I set all of my heap to NULL and construct the binomial heap WITHOUT that extracted node (if the flag value is equal to -1 for a node, then it will not be included in the construction of our binomial heap). I use a for loop to extract top 5 documents. If we have less than this, a break statement finalizes all found results. My implementation is below:

```

for(m=0; m<5; m++) { /* For extracting possible top five documents, do the
following: */

    document* maxNode = findMaximumNode(H); /* Finding the maximum
number towards the linked list of roots, */

    if(maxNode != NULL) {

        if(m==4){

            printf("Doc%d(%d).\n\n", maxNode->documentNumber, maxNode-
>similarityScore); /* giving the results */

        } else {

            printf("Doc%d(%d), ", maxNode->documentNumber, maxNode-
>similarityScore);} /* giving the results */

            printResults[m] = maxNode; /* This array will be used for printing later */

            numberInserter++; /* Iterate through the next */

            H = NULL; /* Now, this is the important part! After the extracted our
maximum number, it's flag value is set to -1. Also now, I completely resetted my
heap. It is now null.

            It will be rebuilt WITHOUT the extracted max.*/

            maxNode->flag = -1;

            for(i=0; i<sizeOfArr; i++) {

                if(relevantDocuments[i]->flag != -1) /* Do not take the already-extracted
maximum nodes into consideration */{

                    H = insertNodeToHeap(H, relevantDocuments[i]); /* Build the heap */

                } else {

                    continue;}

            }} else {

                break;}

        }

```

These two parts conclude both constructing the binomial heap (i.e. enqueue) and obtaining maximum nodes from it.

GETTING THE RESULTS

please see next page

```
Please enter the keyword:
television
```

```
List of all documents with the keyword "television":
```

```
File number #2. Similarity score is 6.
File number #3. Similarity score is 1.
File number #4. Similarity score is 4.
File number #6. Similarity score is 1.
File number #13. Similarity score is 1.
File number #16. Similarity score is 5.
File number #19. Similarity score is 13.
File number #20. Similarity score is 1.
File number #31. Similarity score is 2.
File number #33. Similarity score is 3.
File number #34. Similarity score is 3.
File number #35. Similarity score is 2.
File number #38. Similarity score is 1.
File number #39. Similarity score is 7.
File number #40. Similarity score is 2.
```

```
~~~~~
```

```
The relevance order is: Doc19(13), Doc39(7), Doc2(6), Doc16(5), Doc4(4).
```

```
Doc19(13): Pros:Great picture, feature loaded Cons:Breaks, unable to be repaired by authorized repair repeated
vision with wonderful picture quality and form factor but lasted only 3 months then was unrepairable. It was
We purchased the Samsung UN55ES8000 as our primary TV in our family room. We decided on this set for its stunnin
s features. The TV itself is an LED TV with a 240MHZ refresh rate and very good color out of the box that does
rst, the connectivity of the television is typical. Three HDMI ports with one capable of ARC, one for CEC, and
onent inputs as well as a direct coaxial input. Finally there is USB inputs and the TV has an expansion slot
all formfactor was one of the reasons that we purchased the TV. It has a very thin bezel and overall thickness
n is a small Samsung symbol that glows when the television is turned on. At the top center of the television
the set. The television also comes with two remote controls. A standard button remote and a touch interface
n the rear of the television is a control pad for cycling through the menus for use without a remote. The soft
It can be directly connected to the internet by ethernet or wireless. Setup is simple through the menus. The
ty to access youtube, netflix, hulu, etc directly through the television as well as the ability to use Samsung
of this television is the motion control and voice command control. You can activate the motion control and
olume, etc. The voice feature is similar except it responds to your voice commands. The first 3 months we
. The voice control was convenient especially if we didn't want to hunt down the remote. The motion control w
he position of the television and the brightness of the room the set was in. We were extremely pleased with the
into owning the set is when the gremlins began appearing. At first, there was the occasional freezing of the
d off spontaneously. Over the course of two weeks, the television began doing this continuously and was unwate
was to unplug it. Samsung sent a repair service to fix the set and they replaced the main board. The set wor
aving the same issues again. Again Samsung sent the TV back for repairs and they again replaced the mainboard
orked for 8 hours before the problems began again. Samsung customer service finally issued a full refund for t
rom the onset of the problem took about two months. We have never owned a Samsung TV set before and unlikely
ust received a bad set but, it was not a good sign when the return freight shipper stated this is the 30th Sam
p and send back to Samsung.
```

```
Doc39(7): Pros:Easy navigation, good quality sound, quality resolution. Cons:Not too great remote. The Bottom
```

(The results continue and it prints out all documents with the desired output format.)

Let's show a different example, this time using the word "computer"...

```
Please enter the keyword:  
computer
```

```
List of all documents with the keyword "computer":
```

```
File number #3. Similarity score is 2.  
File number #4. Similarity score is 1.  
File number #6. Similarity score is 3.  
File number #8. Similarity score is 4.  
File number #13. Similarity score is 4.  
File number #16. Similarity score is 1.  
File number #25. Similarity score is 1.  
File number #32. Similarity score is 2.  
File number #34. Similarity score is 1.  
File number #35. Similarity score is 1.  
File number #36. Similarity score is 1.  
File number #40. Similarity score is 1.
```

```
~~~~~
```

```
The relevance order is: Doc8(4), Doc13(4), Doc6(3), Doc3(2), Doc32(2).
```

```
Doc8(4): Pros: Cheap and you don't need to buy a computer monitor. Cons: The stand in the back doesn't deta  
Dude this TV rocks!! Not only does it work as a tv but you can hook it up to your computer and wamo a 26  
uch if your looking to save space and you don't want a "giant tablet" (those new all in one computers are  
ng as you don't mind not watching TV while doing your homework (a travesty I know) its a great 2 for one.  
image missing (you can change this with the remote don't worry) but its only on the start up screen and
```

```
Doc13(4): Pros: Inexpensive, good picture, slim profile Cons: Audio not so great No audio output The Bottom  
Our kitchen computer died last month. It was a nice Sony all-in-one wall mount that we used to watch TV a  
o fix (new motherboard needed) so it has been moved to the electronic recycle lab. We realized that we mi  
ls or listen to Pandora radio while we eat. I didn't want to do another all-in-one computer so I looked i  
r size that would overshadow our kitchen bar table so I was leaning toward 22 to 26 inches. After a lot o  
as 1080 resolution an antenna input along with two HDMI inputs. It also has a VESA 75mm hole pattern on t  
unt we had for the Sony computer. The picture looks great if you feed it a nice signal. When it gets stan  
TV used to look like. Our cable in the house has a lot of SD channels over the antenna cable (no cable bo  
d up when I ran the auto channel setup. Those channels look awesome. My only complaint is that the audio  
amsung is getting compared to that (the Sony also cost 10 times more). The Samsung has downward facing sp  
being said but there is no ambience to speak of. Since there is no audio output (headphone or optical) th
```

Also, let's demonstrate a situation where we have less than 5 relevant documents.


```
Please enter the keyword:  
husband
```

```
List of all documents with the keyword "husband":
```

```
File number #3. Similarity score is 1.
```

```
File number #29. Similarity score is 1.
```

```
~~~~~
```

```
The relevance order is: Doc29(1), Doc3(1),
```

```
Doc29(1): Pros:Picture and sound quality is amazing. Cons:It is little bit expensive as compared to  
On my wedding my father gifted me with Samsung UN60EH6000 60" 1080p HD LED LCD Television. It was th  
bulous and the channels of cable are also crystal clear on it. It hang it on my wall and has enhance  
at our home instead of cinema. Watching movies on Samsung UN60EH6000 60" 1080p HD LED LCD Televisio  
. Our entire family love to watch our wedding video on LCD. If you have this at home then there is n  
player with it and children can also play games on it with their play station as well.
```

```
Doc3(1): Pros:Good quality product, good screen, easy to use. Cons:Hard to set up and get working es  
al skills.
```

```
I received this from my husband for Christmas last year and I love it. It took some fiddling to make  
ow connected to my computer and the range wasn't great. Now I can go anywhere with it and it doesn't  
I am in the bath. I put it in the slot on my treadmill so that I can watch and walk. I use a stretch  
as I drift off. Love it! We have only two televisions in the house. This one which I drag around wit  
n the top all feel the same and you can push the wrong one easily. Volume up/down is next to the off  
gs from your computer but it can be complicated to set up. Web-The browser is pretty basic and klunk  
er a site I wanted to check out while I am watching tv. I wouldn't use it regularly for work. Same g  
less laptop. Capture is cool. Just push the capture button and it does a screen capture and saves th  
re skating that I can then either send via email or print. All in all I am very pleased with the pro
```

```
Process returned 0 (0x0)   execution time : 1.058 s
```

```
Press any key to continue.
```

The word “husband” had occurrences in only 2 files, each with a similarity score of 1.
Our programme successfully terminated the loop before reaching 5.

Finally, let’s show what happens if no relevant documents can be found:

```
Please enter the keyword:
abcdefg

List of all documents with the keyword "abcdefg":

No documents could be found!
Terminating the programme...

Process returned 0 (0x0)   execution time : 5.099 s
Press any key to continue.
```

These screenshots conclude our results section.

DISCUSSION: ADVANTAGES OF PRIORITY QUEUE

In this project, we are given the task of ranking documents. This means that some of the records we hold are more *important* than others and means more for our purpose. Yes, they can be hold in arrays, linked lists or something else: however, if we hold them in a data structure which we can reach the most prioritized one with the easiest effort, scaling of our project for large inputs will be much more efficient. Here what we give importance to is the similarity scores. Therefore, documents with higher similarity score should be kept in a max-heap structure for obtaining them even faster.

There are fibonacci heaps, classical binary heaps and binomial heaps that can provide this priority queue property. Binomial heaps are a *forest* of trees and merging them is lot faster, $O(\log n)$. This is the reason why we've chosen priority queues instead of other data structures.

This explanation concludes my project report.

Thank you for the term, Sir/Madam