Advanced Computing Project

From :-
Fabio santiago(104934797)
Sarath Pulikkal(105099194)
Munish (105146209)

Search Engine

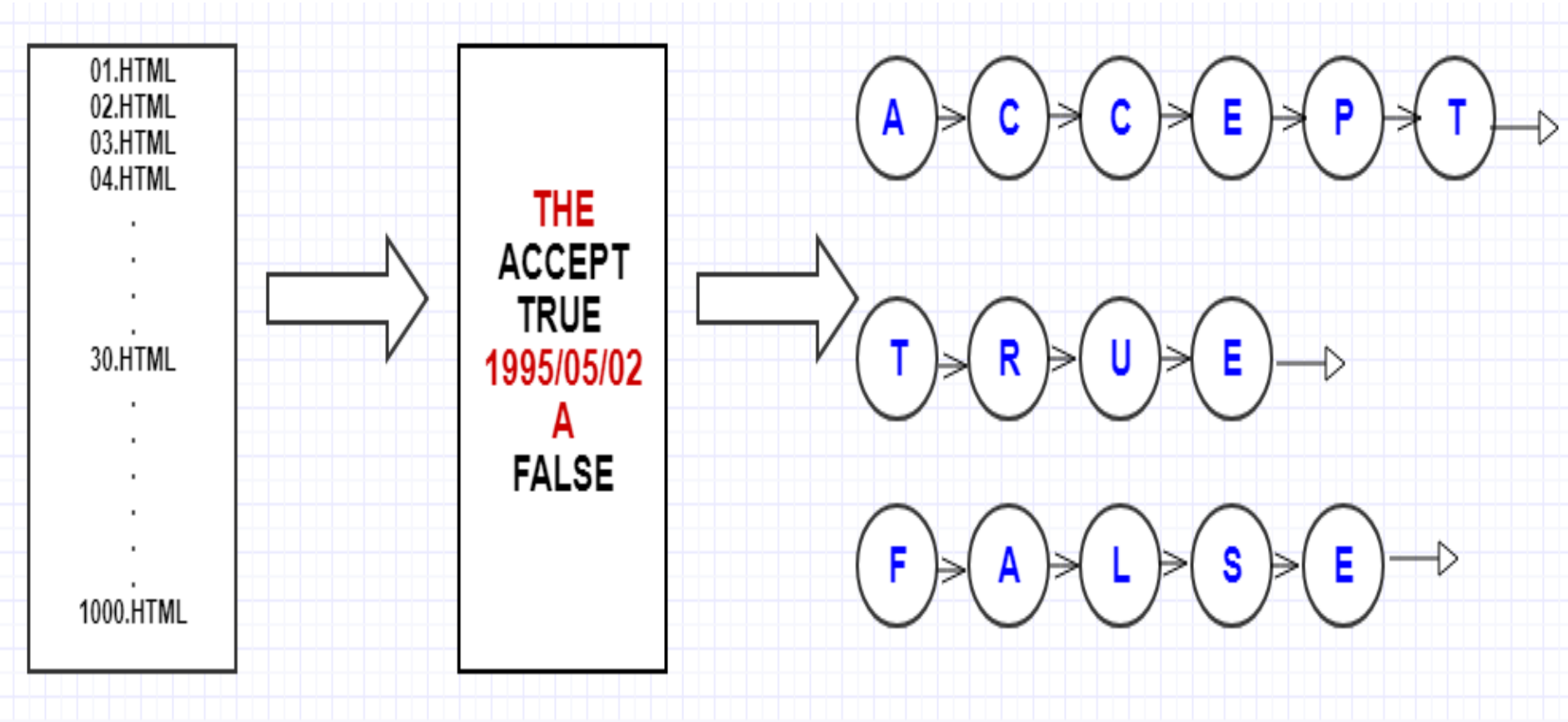
Problem Scope

- Snapshot of Pages
- Maximum number of pages in search result.
- Valid search words.

Search Engine – Problems

- Build database time
- Search time

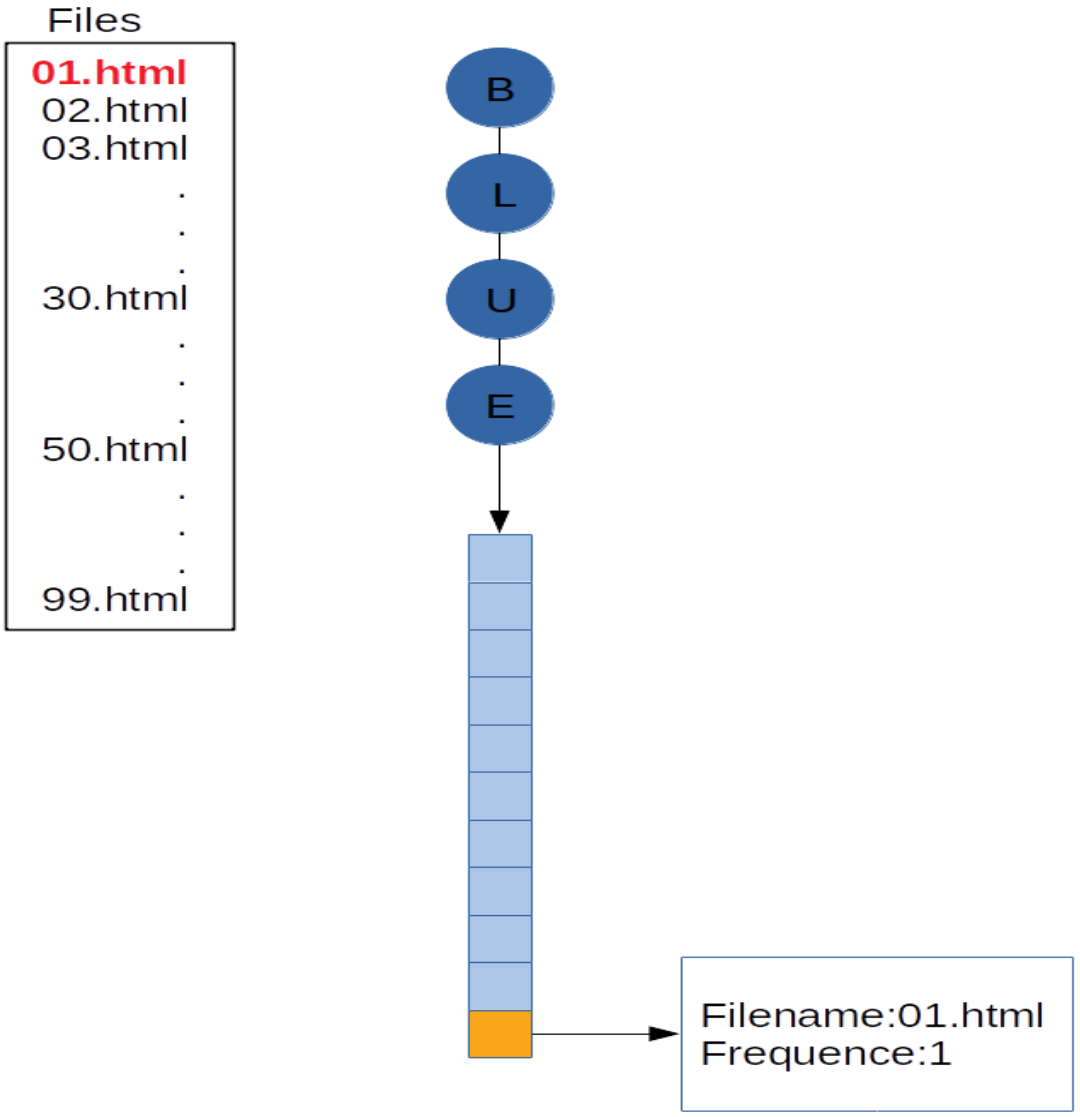
Build the database

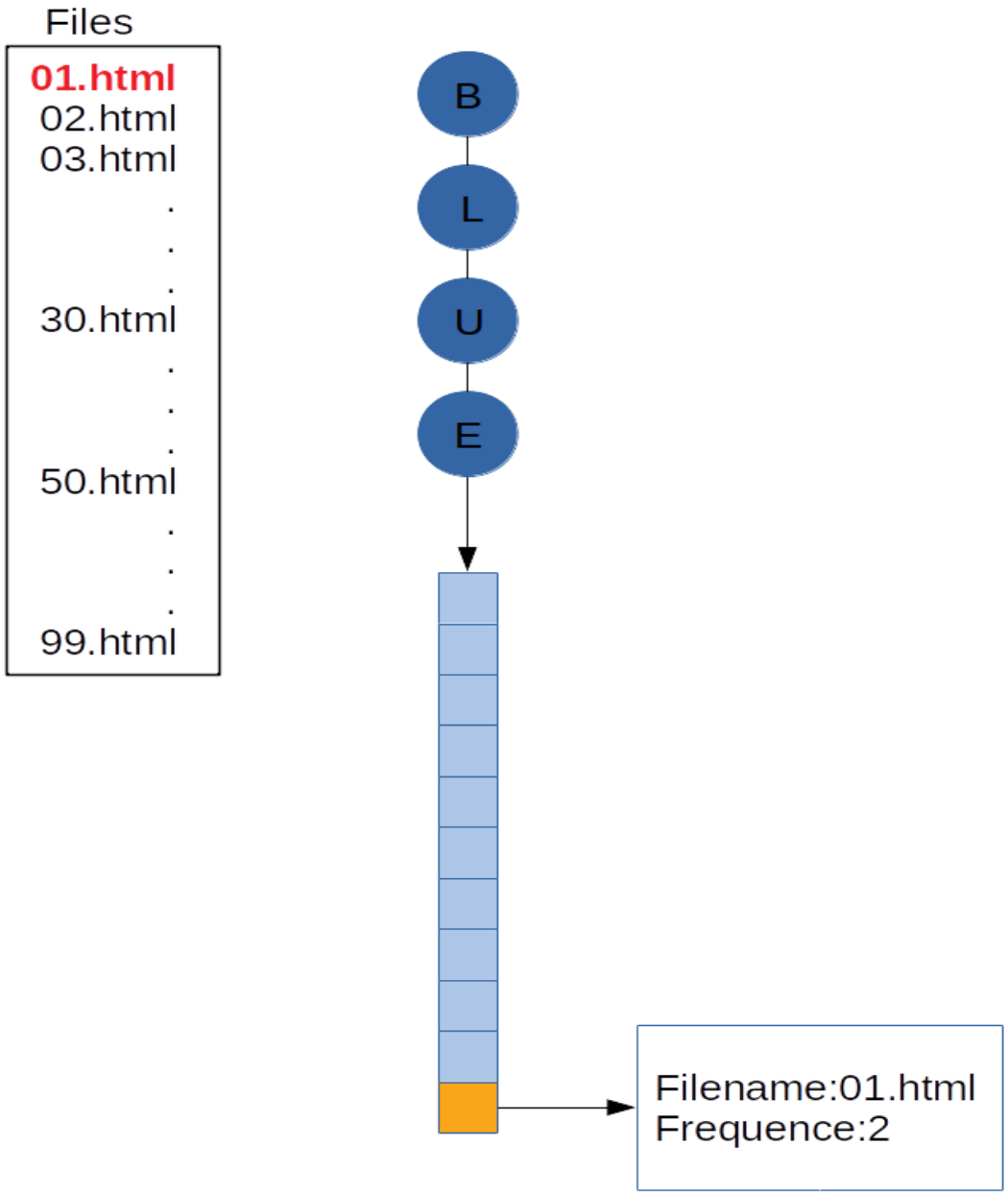


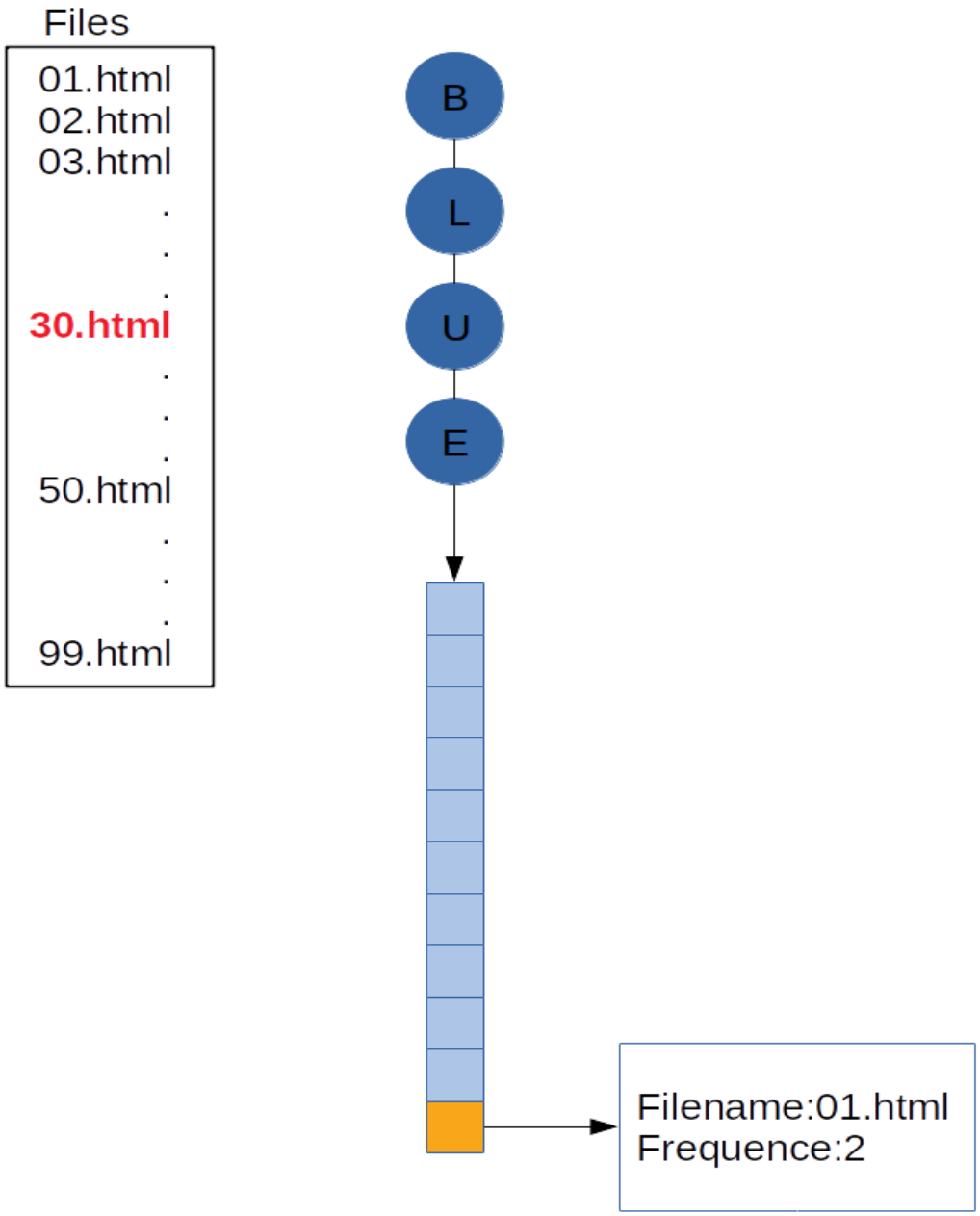
The Algorithm

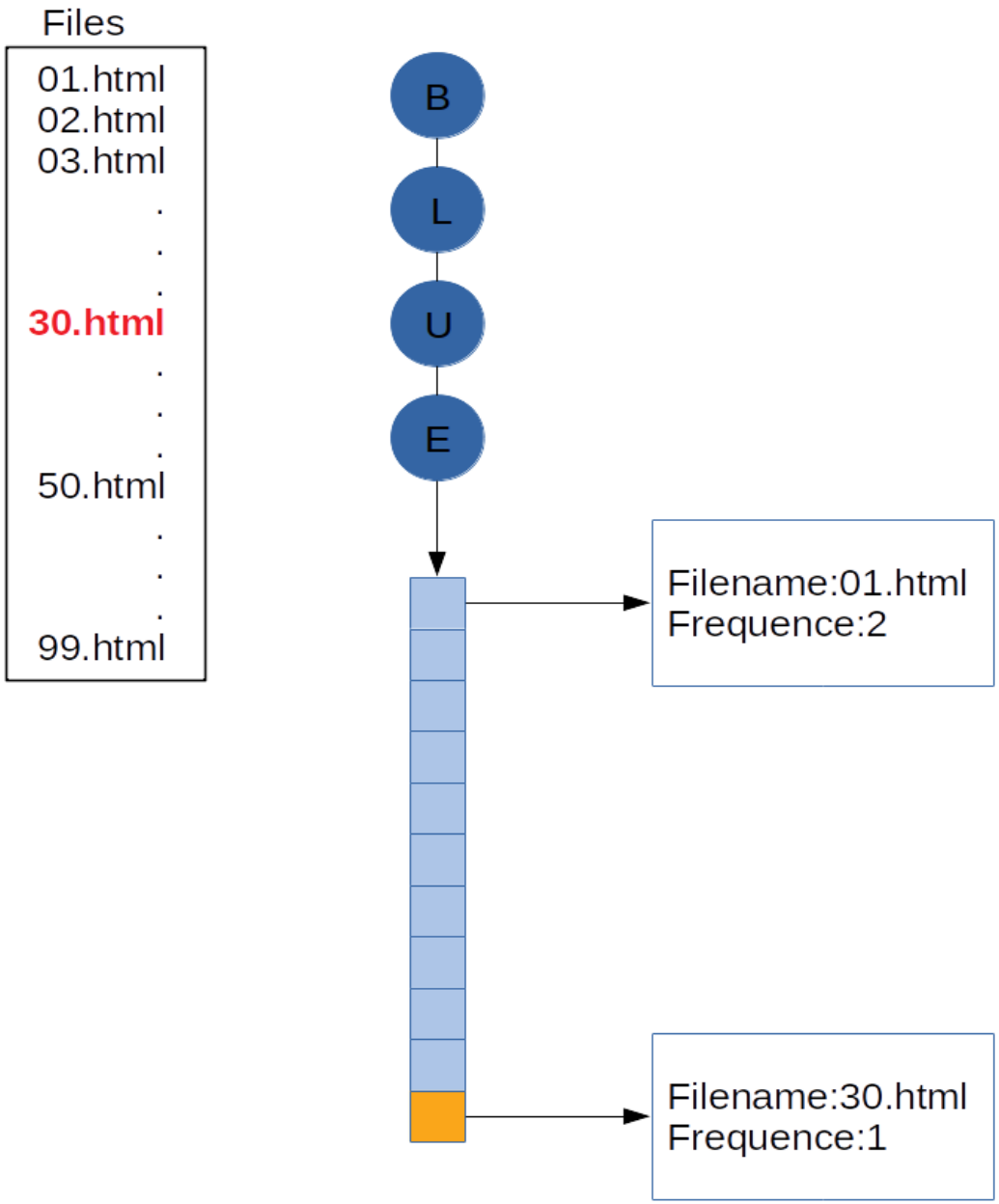
```
1  for each file F in FILES
2      for each word W in F
3          if W does not exist in TST
4              create an empty arrary A of size 11
5              A[10].filename = F.filename
6              A[10].frequence++
7          else
8              if A[10].filename == F.filename
9                  A[10].frequence++
10             else
11                 sort array A
12                 A[10].filename = F.filename
13                 A[10].frequence++
14
15 for each word W in TST
16     get array A of W
17     sort A
18     put array A
19
20
```

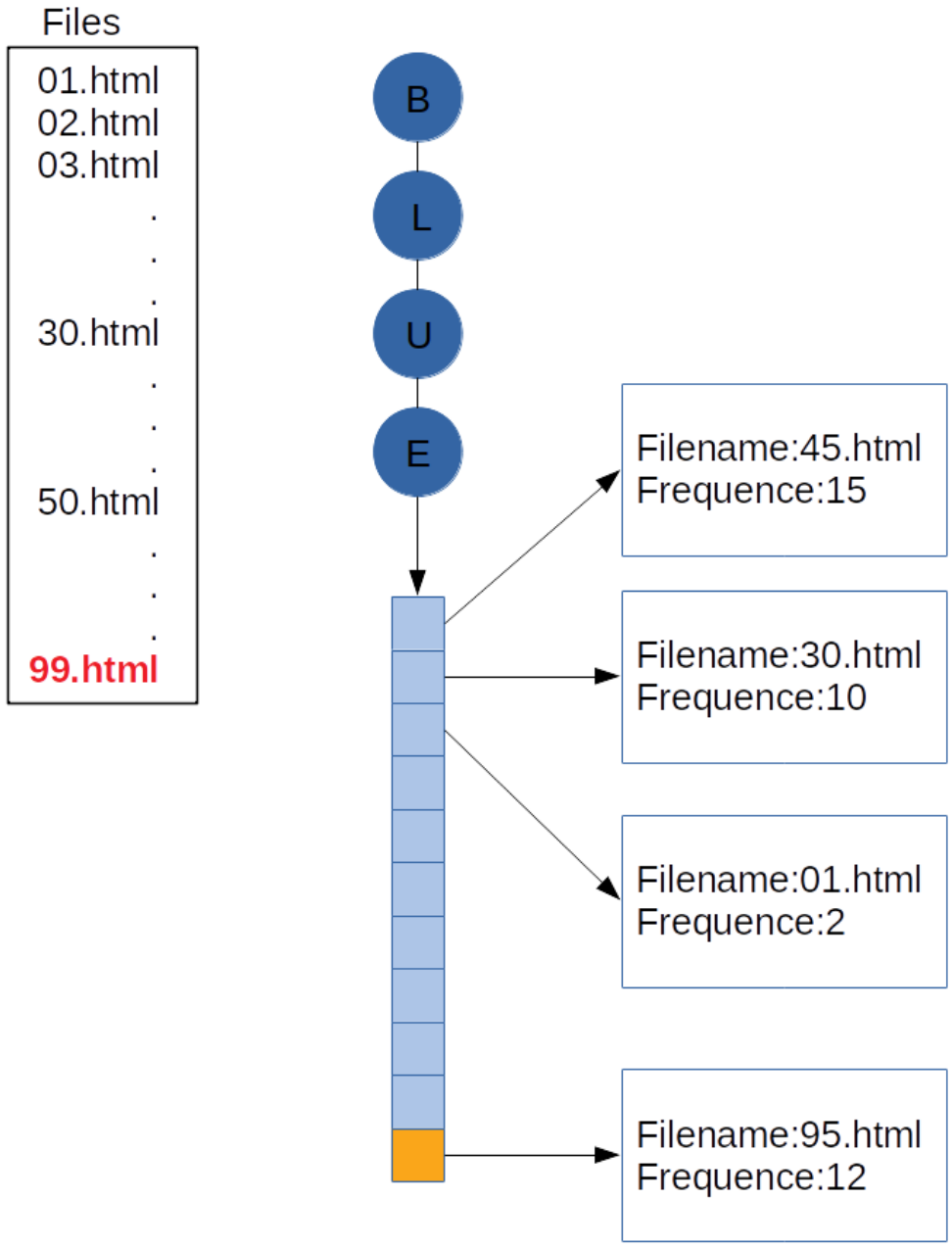
Build the database

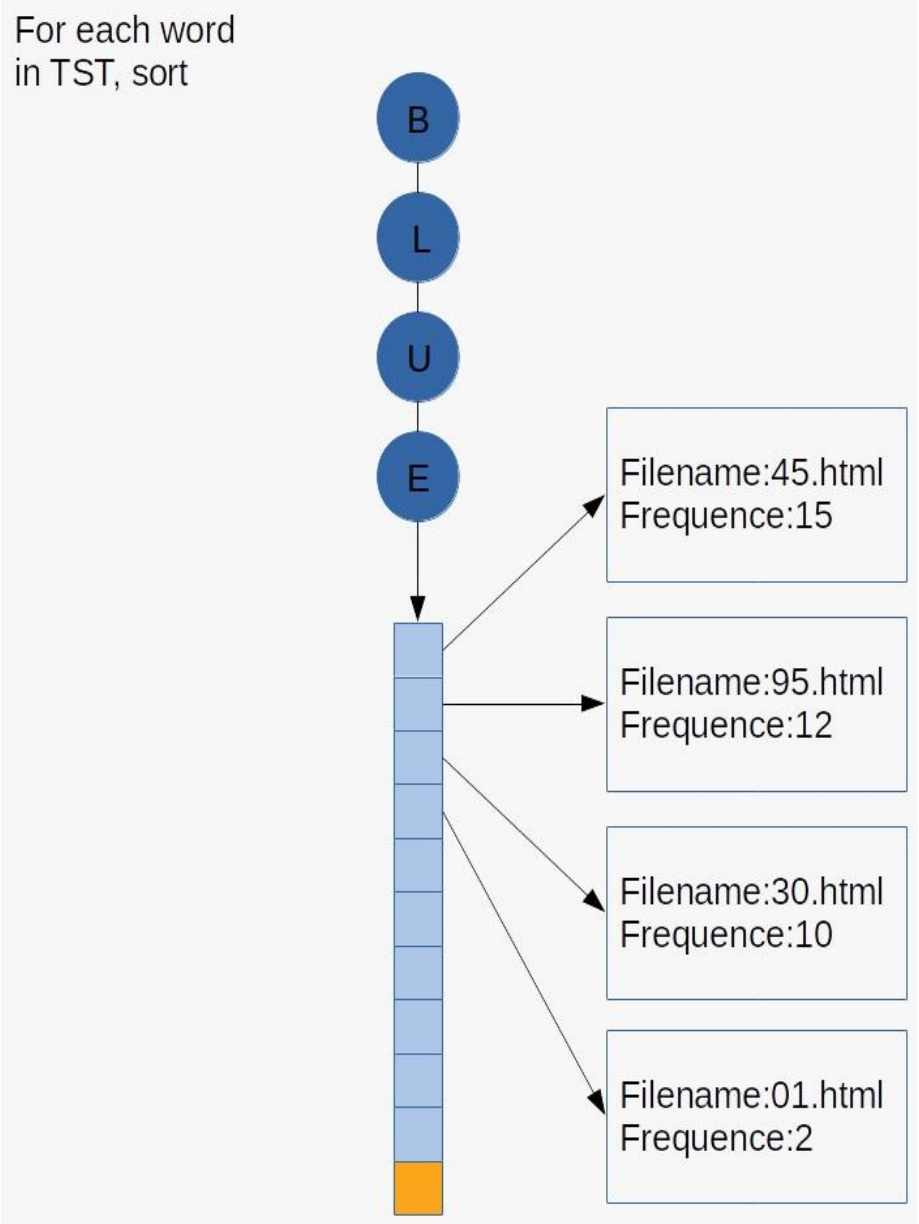



Files
01.html
02.html
03.html
30.html
50.html
99.html
B
L
U
E
Filename:01.html
Frequence:2


Files
01.html
02.html
03.html
30.html
50.html
99.html
B
L
U
E
Filename:01.html
Frequence:2


Files
01.html
02.html
03.html
30.html
50.html
99.html
B
L
U
E
Filename:01.html
Frequence:2
Filename:30.html
Frequence:1


Files
01.html
02.html
03.html
30.html
50.html
99.html
B
L
U
E
Filename:45.html
Frequence:15
Filename:30.html
Frequence:10
Filename:01.html
Frequence:2
Filename:95.html
Frequence:12

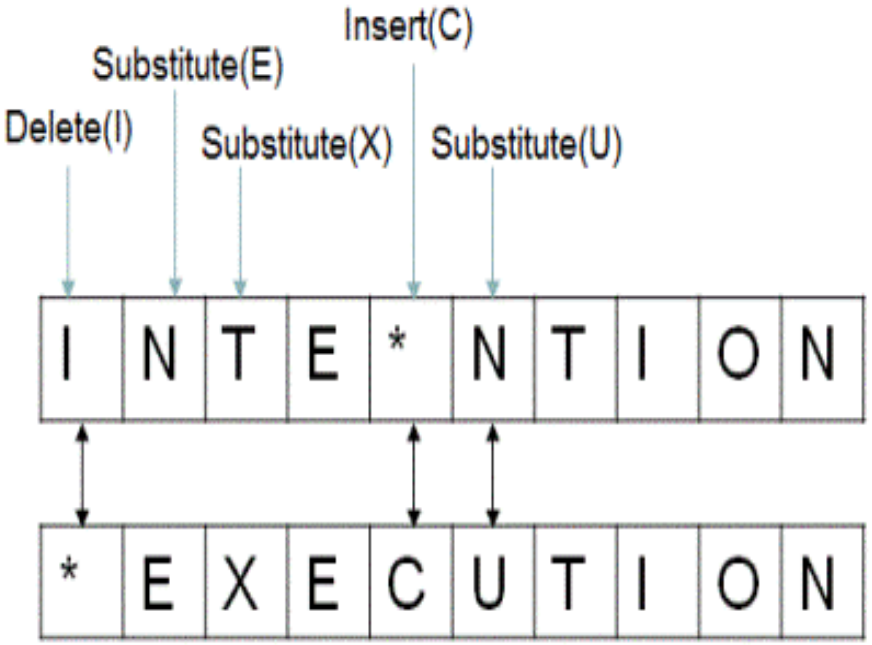

For each word
in TST, sort
B
L
U
E
Filename:45.html
Frequence:15
Filename:95.html
Frequence:12
Filename:30.html
Frequence:10
Filename:01.html
Frequence:2

Edit Distance

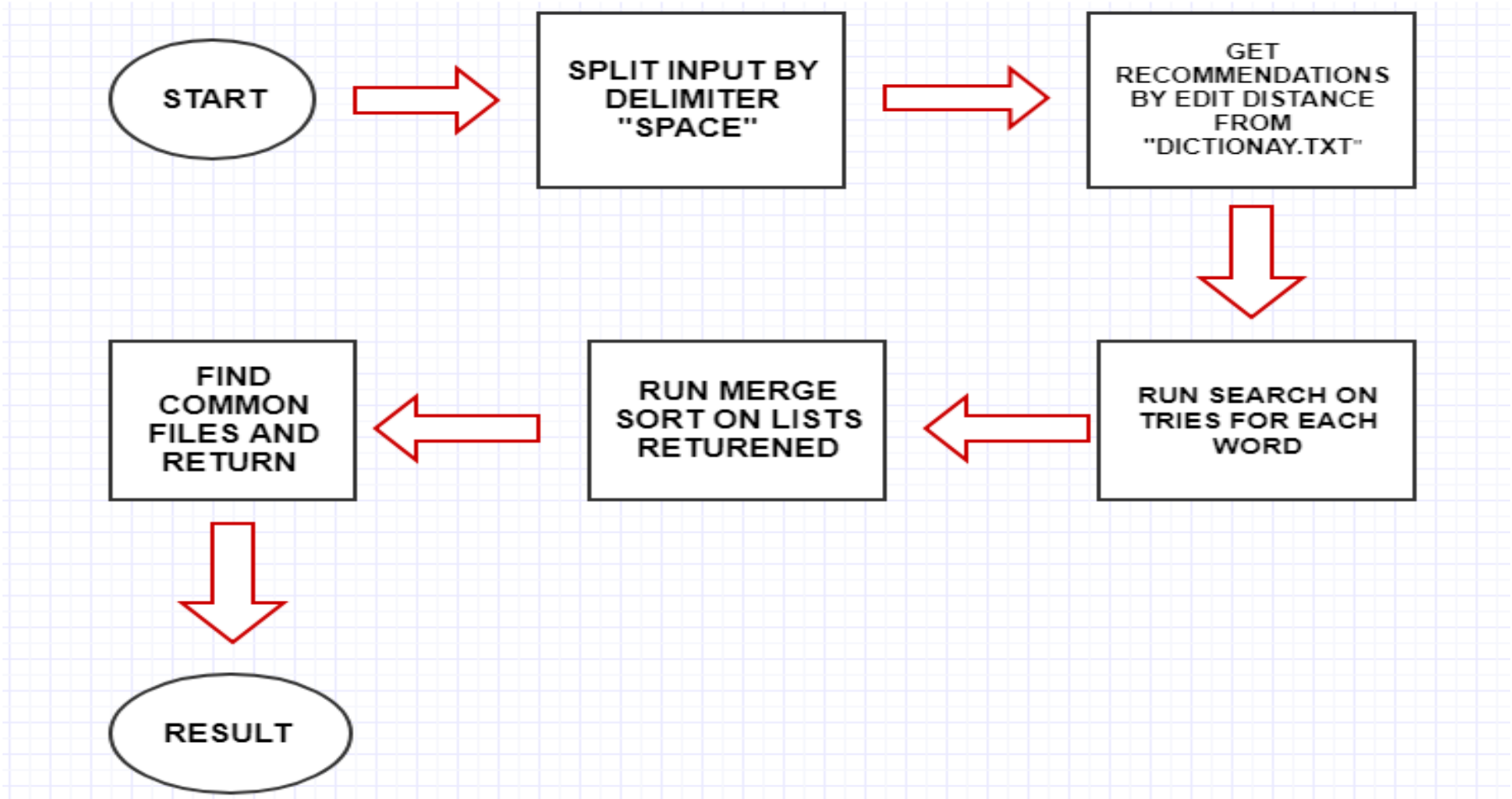
- The Edit Distance is defined as the minimum number of edits needed to transform one string into the other.
- Given an **initial string s**, and a **target string t**.

The list of valid changes are:

1) Inserting a character
2) Deleting a character
3) Changing a character to another character.



OVERVIEW


START
SPLIT INPUT BY DELIMITER "SPACE"
GET RECOMMENDATIONS BY EDIT DISTANCE FROM "DICTIONAY.TXT"
RUN SEARCH ON TRIES FOR EACH WORD
RUN MERGE SORT ON LISTS RETURENED
FIND COMMON FILES AND RETURN
RESULT

Concepts Used

- Regular Expression
- Ternary search trie(TST)
- Merge Sort
- Edit Distance

LIVE DEMO OF SEARCH ENGINE

Thank
You

Q
A