

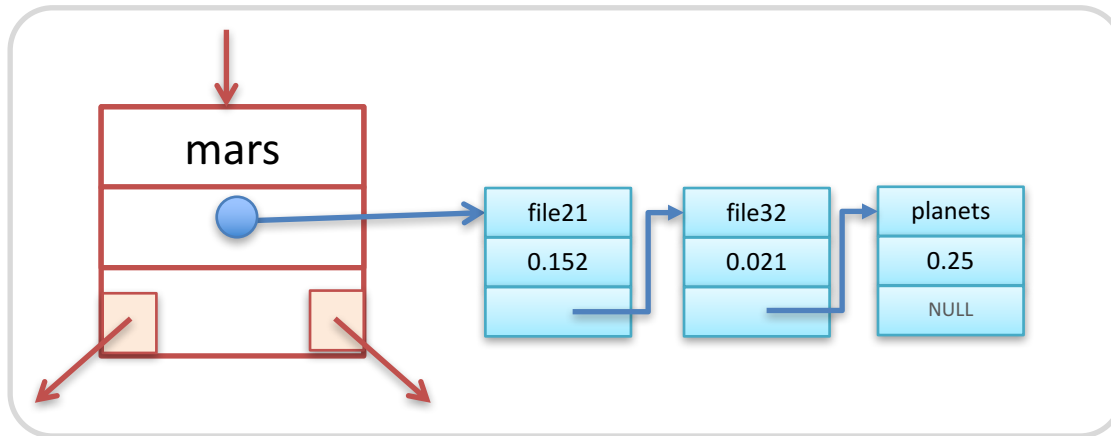
COMP2521 (20T1)

Ass1 : How to Implement?

Notes:

- The document offers some **suggestions only**, with incomplete pseudo code
- You can use code from labs/lecture material, however, must acknowledge it and provide a reference. For example you can use BST ADT implementation from one of the labs, and adapt it for this assignment.

InvertedIndexBST



```
struct InvertedIndexNode {  
    char *word; // key  
    struct FileListNode *fileList;  
    struct InvertedIndexNode *left;  
    struct InvertedIndexNode *right;  
};  
typedef struct InvertedIndexNode *InvertedIndexBST;
```

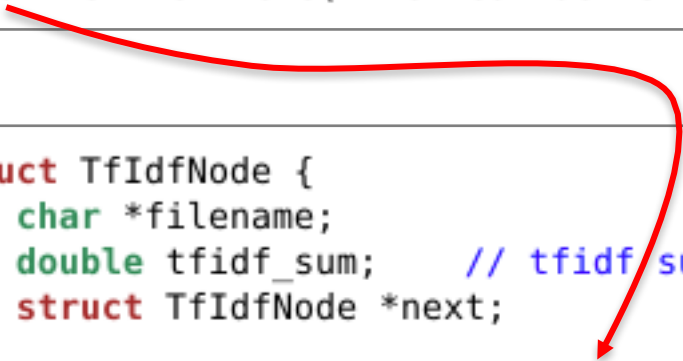
```
struct FileListNode {  
    char *filename;  
    double tf; // relative tf  
    struct FileListNode *next;  
};  
typedef struct FileListNode *FileList;
```

- You **MUST** use a binary search tree where key is a string and value is a list as shown in the figure. You can use **strcmp** to compare key strings.

```
// Functions for Part-1
char * normaliseWord(char *str);
InvertedIndexBST generateInvertedIndex(char *collectionFilename);
void printInvertedIndex(InvertedIndexBST tree);

// Functions for Part-2
TfIdfList calculateTfIdf(InvertedIndexBST tree, char *searchWord , int D);
TfIdfList retrieve(InvertedIndexBST tree, char* searchWords[] , int D);
```

```
struct TfIdfNode {
    char *filename;
    double tfidf_sum;    // tfidf sum value
    struct TfIdfNode *next;
};
typedef struct TfIdfNode *TfIdfList;
```



collection.txt

```
collection.txt x |
nasa.txt      news1.txt
file11.txt   mixed.txt   planets.txt
file21.txt
info31.txt
```

nasa.txt – one of the documents

```
nasa.txt x |
experiments NASA EU Japan robot
planet volcano moon weather
mars jupiter moon planets
```