

# 자료구조와 실습 설계과제

## 간단한 검색엔진 구현

- 2017112080 김혜수
- 2017112078 정지연

---

## contents

---

1 과제 개요

2 프로그램의 구성

3 수행 결과

4 결과 분석

# 1. 과제 개요

## ■ 설계과제 목적 및 내용 요약

txt파일을 읽어와 해당 파일의 영단어들을 색인하고,  
단어를 검색할 수 있는 검색엔진을 주어진 제한요소들을 고려하여 구현

### 1. 색인 과정

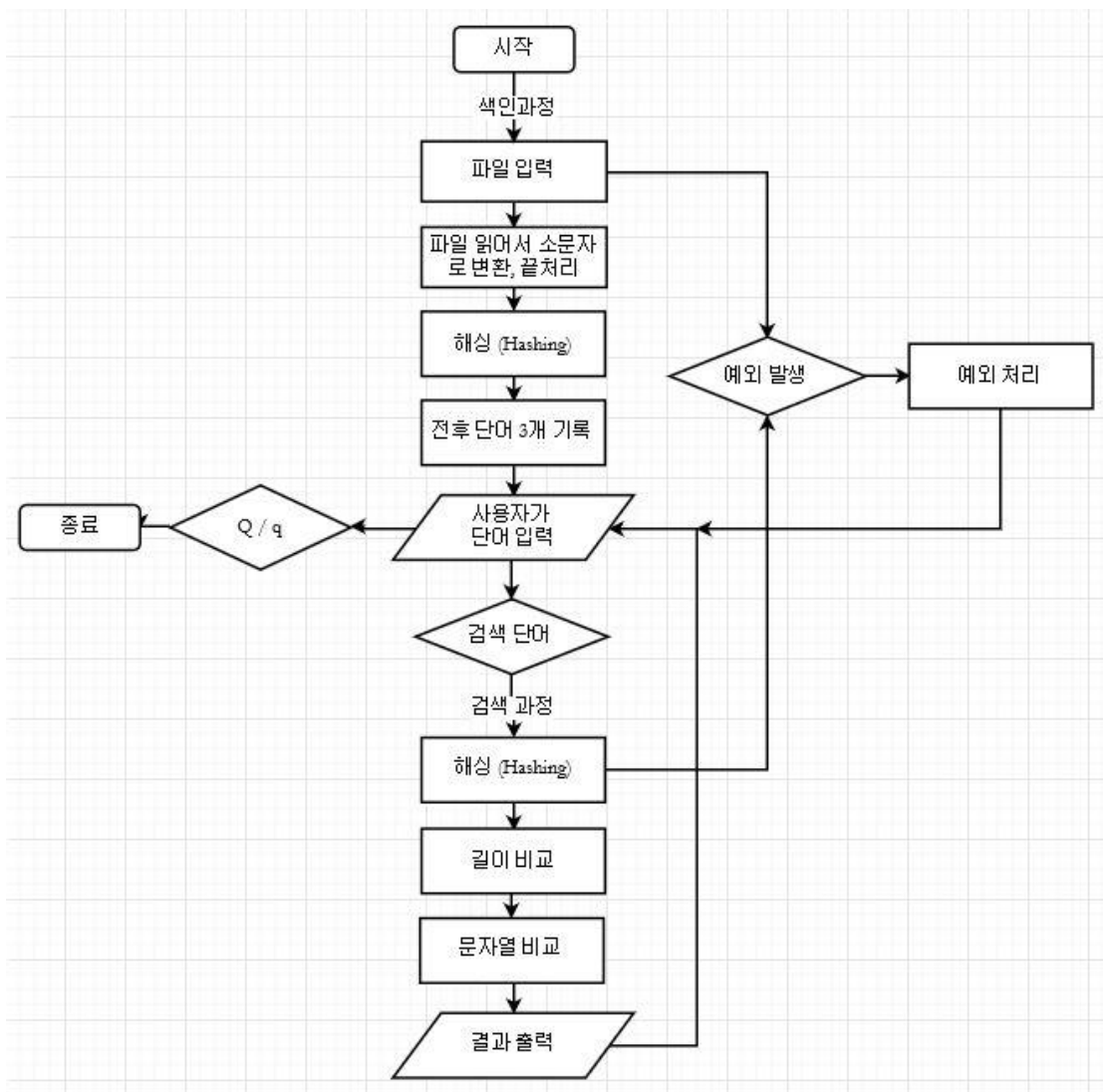
문서를 읽어 각 단어들이 어느 문서에 몇 번이나 나타나는지에 대한 정보를 자료구조를 이용하여 저장한다.

### 2. 검색 과정

사용자가 찾고자 하는 단어를 입력하면  
미리 저장해 놓은 자료구조를 활용 > 단어가 출현하는 문서 찾기 <  
문서명 / 단어 출현 빈도 / 그 단어가 출현한 문서 부분 (전후 3 단어씩) 출력

# 2. 프로그램의 구성

알고리즘 및 프로그램 세부 구성



## 2. 프로그램의 구성

알고리즘 및 프로그램 세부 구성

### 1. 색인

1. 파일을 불러와 파일에 있는 문장들을 한 단어 씩 끊음.
2. 해싱 과정 - 단어의 ASCII 코드 값을 더하고, 해시 테이블에 새로운 테이블을 만든다.
3. 단어 위치를 기록한 뒤 파일 내에 같은 단어가 있다면 해당 단어는 연속된 노드의 끝에 저장.

# 2. 프로그램의 구성

알고리즘 및 프로그램 세부 구성

## 2. 검색

1. 입력 받은 단어를 소문자로 변환한 뒤 단어의 끝에 NULL값 저장.
2. 글자 수가 맞는 경우 비교가 이루어지도록 길이가 같을 때까지 링크를 계속 뒤로 밀면서 글자 수가 같은 지 판단.
3. 같은 문자열이 있는 경우, 해당 노드부터 같은 결과의 노드를 모두 출력.

# 3. 수행결과

■ 프로그램 테스트 결과

## 1. 소문자 입력

```
C:\Windows\system32\cmd.exe
Total number of documents: 9
Total number of indexed words: 549
Total number of comparison: 379
Enter Your word. If you wanna terminate this program, Enter Q/q : computer

<document1.txt> computer: 1
In computer science, a data ...

<document5.txt> computer: 1
... ability of a computer to fetch and ...

Total number of comparison: 1

Enter Your word. If you wanna terminate this program, Enter Q/q : Q
계속하려면 아무 키나 누르십시오 . . .
```

# 3. 수행결과

■ 프로그램 테스트 결과

## 2. 대문자 입력

```
Enter Your word. If you wanna terminate this program, Enter Q/q : STRUCTURE

<document6.txt> STRUCTURE: 4
... of a data structure usually requires writing ...
... instances of that structure. The efficiency of ...
... of a data structure cannot be analyzed ...
... type, a data structure that is defined ...

<document1.txt> STRUCTURE: 2
... science, a data structure is a data ...
... precisely, a data structure is a collection ...

<document2.txt> STRUCTURE: 2
... on a data structure and the computational ...
... comparison, a data structure is a concrete ...

<document5.txt> STRUCTURE: 1
... items within the structure itself. Many data ...

<document8.txt> STRUCTURE: 1
... that allows data structure implementations to be ...

Total number of comparison: 1
```



# 3. 수행결과

■ 프로그램 테스트 결과

## 3. 존재하지 않는 단어 입력

```
Enter Your word. If you wanna terminate this program, Enter Q/q : AAAAAAA  
"aaaaaaa" does not exist.  
Enter Your word. If you wanna terminate this program, Enter Q/q :
```

# 4. 결과분석

■ 수행결과에 대한 토의

## 수행 결과에 대한 전반적인 분석

1. 한글과 같은 전각문자의 경우 확장성에 대한 한계가 존재한다.  
Hashing과 Chaining을 이용했기 때문에 시간 복잡도는  $O(n)$ .
2. 해시 테이블의 노드를 모두 비교하는 것이 아니라,  
먼저 글자수를 비교하므로 시간 복잡도가  $O(n)$ 보다 낮을 것이다.
3. 가독성을 위해 사용자가 입력하는 단어와  
최종 비교 연산횟수의 글자색을 달리하여 심미성을 증진시켰다.

**감사합니다**

- Thank you -