

프로그래밍 과제 01

1. 영어 사전 데이터 파일(dict.txt)이 제공된다. 이 파일은 170,000개 이상의 영어 단어와 그 단어의 의미를 포함하고 있다. 파일의 형식은 다음과 같다. 단어와 단어는 하나의 공백 라인으로 구분되어 있다. 각각의 단어는 하나 혹은 그 이상의 문자열로 구성된다. 즉, 어떤 단어 항목은 2개 이상의 단어로 구성되어 있다는 의미이다(예: Aaron's rod 등). 단어 다음에는 하나의 소괄호가 등장한다. 소괄호 안에는 일반적으로 그 단어의 품사가 명시되지만 빈 소괄호 일수도 있다. 소괄호 이후에는 그 단어에 대한 설명이 뒤따른다. 예를 들면 아래와 같다.

```
Abducted (imp. & p. p.) of Abduce
```

```
Abegge () Same as Aby.
```

```
Abraum salts (n.) A red ocher used to darken mahogany and for making chloride of potassium.
```

단어들은 알파벳순으로 정렬되어 있다. 영문 알파벳이 아닌 문자(예를 들어 '-'로 시작하는 단어들도 있다. 이 단어들의 경우 정렬을 할 때 영문 알파벳이 아닌 문자는 빼고 정렬하였다. 동일한 단어들이 여럿 있을 수 있다. 이 사전 파일을 읽어서 배열로 저장한 후 다음과 같은 기능을 제공하는 프로그램을 작성하라.

```
$ read dict.txt          // 데이터 파일을 읽어서 배열로 저장한다.
$ size                   // 저장된 단어의 개수를 출력한다.
176050
```

사전을 검색하는 기능을 다음과 같이 작동하도록 구현하라.

```
$ find yesterday          // 단어를 검색한다. 대소문자 구분은 하지 않는다.
Found 3 items.            // 3개를 찾았음을 알려준다.
Yesterday (n.) The day last past; the day next before the present.
Yesterday (n.) Fig.: A recent time; time not long past.
Yesterday (adv.) On the day last past; on the day preceding to-day; as, the
affair took place yesterday.
```

만약 찾는 단어가 존재하지 않을 경우 그 단어의 앞과 뒤의 각각 1단어씩을 다음과 같은 형식으로 출력한다. 앞 뒤의 단어들은 단어와 품사만 출력해야 한다.

```
$ find humoric            // 단어를 검색한다. 대소문자 구분은 하지 않는다.
Not found.                // 0개를 찾았음을 알려준다.
Humoralist (n.)
- - -                     // 찾는 단어가 없음을 이렇게 표시
Humorism (n.)
$ exit                    // 프로그램을 종료한다.
```

find 명령을 처리하는 함수는 반드시 순환함수(recursion)으로 구현해야 하며, 그 함수의 시간복잡도는 $O(\log N)$ 이 되어야 한다. find 함수는 찾는 단어가 존재할 경우 그 단어의 배열 인덱스를 반환하고, 존재하지 않을 경우 그 단어의 바로 앞 단어의 배열 인덱스를 반환해야 한다. 바로 앞 단어가 존재하지 않으면 -1을 반환한다.