

## 프로그래밍 과제 03

사전 파일(dict\_simplified.txt)을 읽어온다. 파일의 각 라인마다 하나의 “단어”와 그 단어에 대한 “설명”이 저장되어 있다. “단어”와 “설명”은 하나의 TAB 문자로 구분되어 있다. 사전은 오로지 영문 소문자로만 구성되어 있다. 쉼표, 마침표 등의 기호들은 모두 제거된 상태이다. 이 사전 파일을 읽어서 단어들 간의 연관성을 표현하는 그래프를 구성한다. 두 단어의 연관성은 다음과 같이 정의된다.

“임의의 두 단어 단어 A와 B에 대해서 단어 A의 설명에 단어 B가 등장하거나 혹은 단어 B의 설명에 단어 A가 등장하면 두 단어 A와 B는 연관성이 있다.”

예를 들어 아래의 두 줄은 각각 사전에서 단어 mountain과 volcano에 대한 항목이다.

```
mountain    a large mass of earth and rock rising above the common level of the earth or
adjacent land earth and rock forming an isolated peak or a ridge an eminence higher than a
hill a mount
...
volcano     a mountain or hill usually more or less conical in form from which lava
cinders steam sulphur gases and the like are ejected often popularly called a burning
mountain
...
```

단어 volcano에 대한 설명에 단어 mountain이 등장하므로 두 단어는 연관성이 있다. 이때 문제를 좀 쉽게 하기 위해서 두 단어가 완전히 일치할 경우에만 연관성을 인정하도록 구현한다. 예를 들어 mountain과 그것의 복수형인 mountains는 다른 단어로 간주하라,

사전에 있는 모든 단어를 정점(vertex)으로 하고(단어의 “설명” 부분에만 있는 단어는 포함하지 않는다.), 서로 연관성이 있는 단어 간에는 에지(edge)가 있는 그래프를 연결리스트로 표현하라. 그런 후 다음의 문제들을 순서대로 푸는 프로그램을 작성하라.

1. 만들어진 그래프의 정점의 개수와 에지의 개수를 출력한다.
2. 차수(degree)가 최대인 정점을 찾아 그 정점에 해당하는 단어와 차수를 출력한다.
3. 가장 큰 연결요소(connected component)를 찾아서 그 연결요소의 크기(정점의 개수)를 출력한다.
4. 하나의 단어  $x$ 와 탐색 깊이를 나타내는 정수  $k$ 를 입력받은 후 그래프에서 단어  $x$ 로 부터 떨어진 거리가  $k$ 이하인 모든 단어를 찾아 한 줄에 하나씩 출력한다. 이때 단어  $x$  자체를 맨 처음 출력하고, 나머지 단어들의 출력 순서는 마음대로 한다.

[프로그램의 출력 예 (출력의 예시일 뿐 정답이 아니다.)]

Answer1: 9422, 7504

Answer2: species 258

Answer3: 5039

Answer4: (단어 mountain과 k=2를 입력했을 시)

```
mountain
volcano
salse
sheeling
seaside
hut
saibling
...
```

[제출기한] 2022년 6월 12일 자정까지

[제출방법] 하나의 소스코드 파일을 LMS를 통해 제출한다.