

알고리즘 설계 HW #2

문자열 탐색 KMP 알고리즘

보고서 작성 서약서

1. 나는 타학생의 보고서를 베끼거나 여러 보고서의 내용을 짜집기하지 않겠습니다.

2. 나는 보고서의 주요 내용을 인터넷사이트 등을 통해 얻지 않겠습니다.

3. 나는 보고서의 내용을 조작하지 않겠습니다.

4. 나는 보고서 작성에 참고한 문헌의 출처를 밝히겠습니다.

5. 나는 나의 보고서를 제출 전에 타학생에게 보여주지 않겠습니다.

나는 보고서 작성시 윤리에 어긋난 행동을 하지 않고 정보통신공학인으로서 나의 명예를 지킬 것을 맹세합니다.

2020년 5월 22일

학부 정보통신공학과

학년 3

성명 심규환

학번 12181793



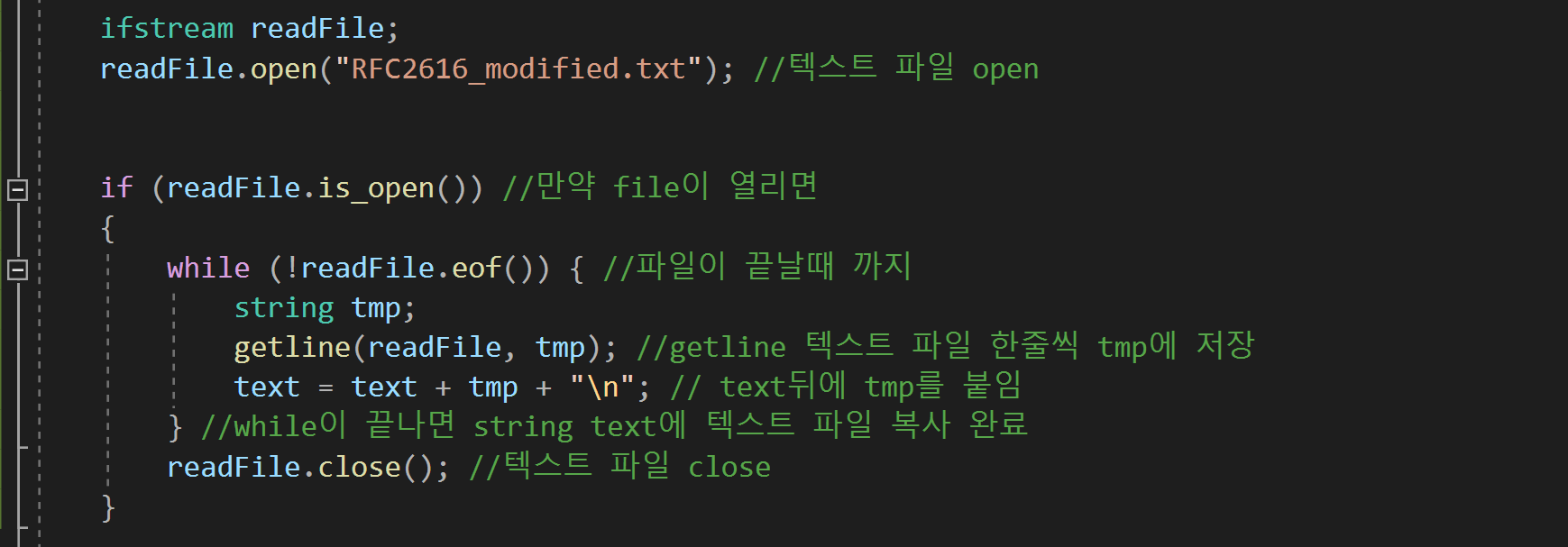
개요

* + - (1)
      * 구현상특징
      * 실행화면 결과
    - (2)
      * 구현상 특징
    - (3)
      * 실행화면 결과
    - (4)
      * 실행화면 결과
    - 전체 코드

(1) KMP알고리즘을 확장하여 숫자와 특수기호도 검색할 수 있도록 할 것.

\* 주어진 “RFC2616\_modified.txt”화일을 텍스트로 사용할 것.

* + - 구현상 특징



주어진 txt파일을 텍스트로 사용하기 위해 헤더파일 <fstream>의 함수들을 이용하였다. Open 함수를 통해 txt파일을 열고 is\_open함수로 파일이 열렸는지 확인하고eof함수로 파일의 끝을 확인한다. 그리고 tmp에 txt파일의 한 라인씩 받아오고 text에 tmp를 넣어준다. 마지막으로 close함수로 txt파일을 닫아준다. 그러면 string인 text에 txt파일이 모두 복사된다.

* + - 실행화면 결과

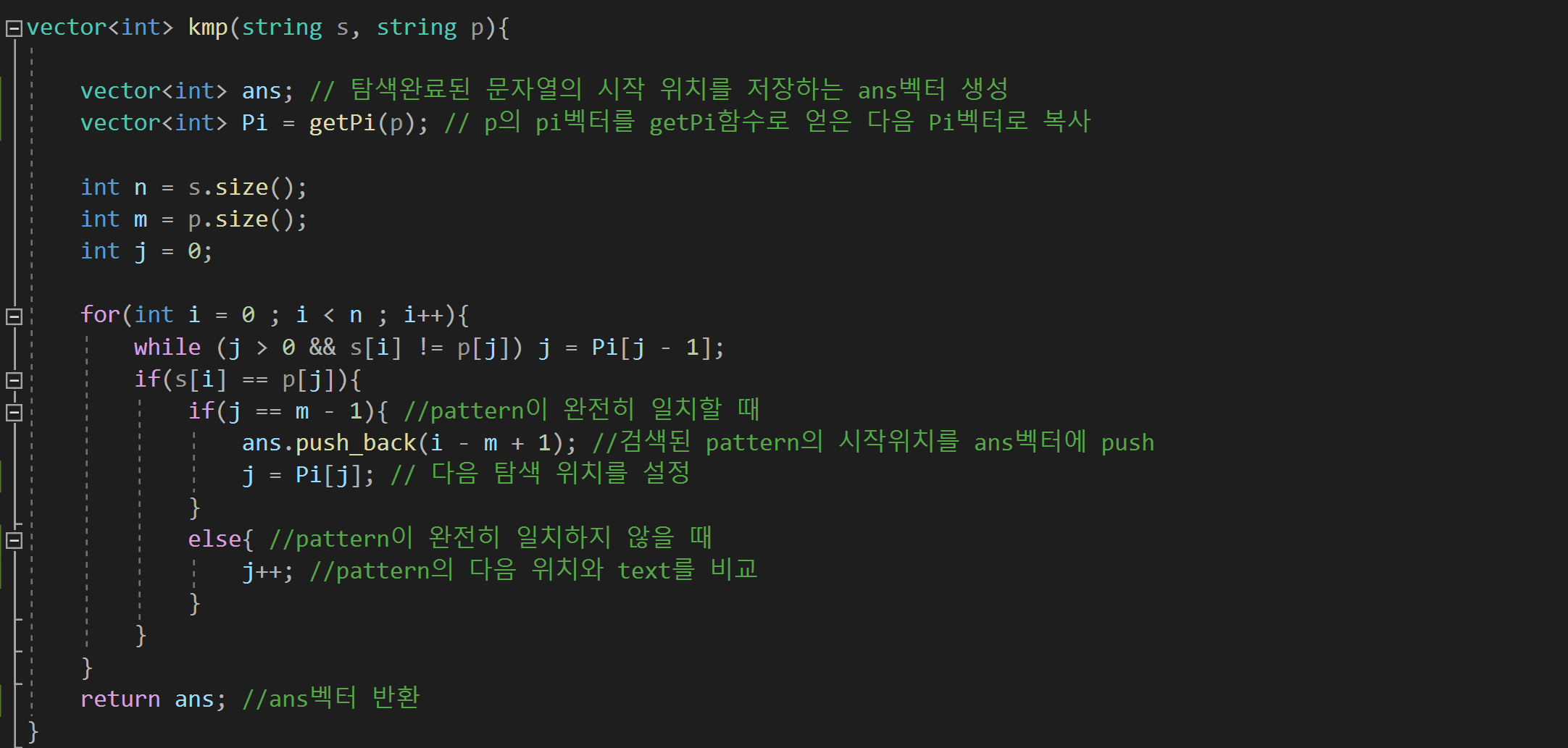
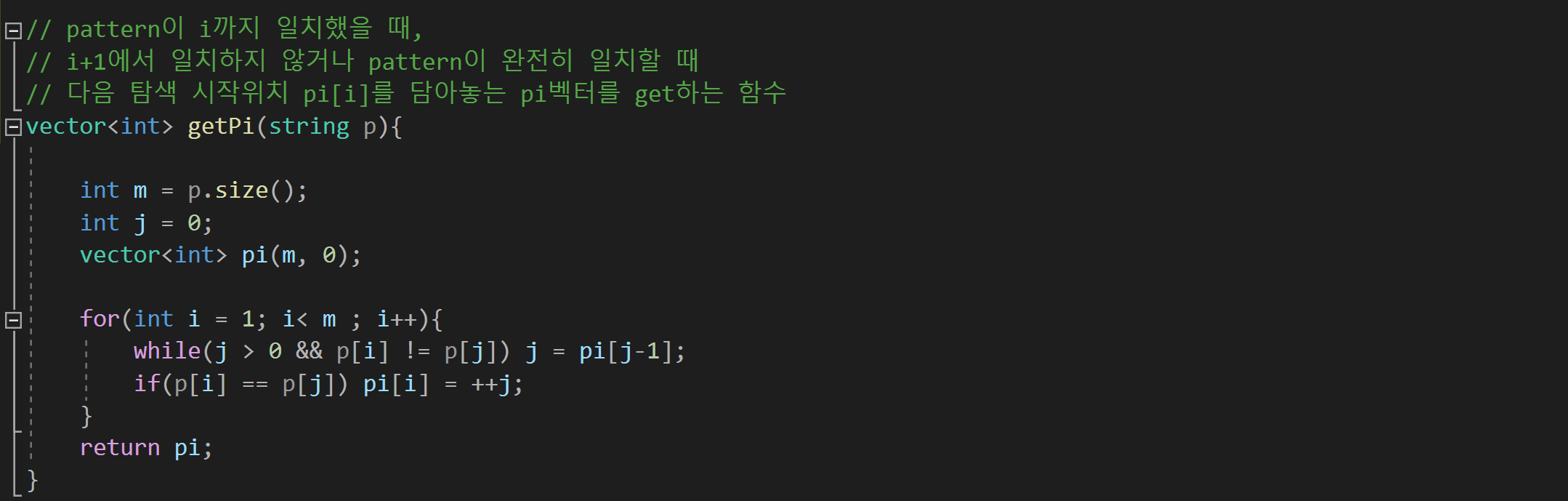


특수기호 **.** 과 / 그리고 숫자 1을 패턴으로 포함해도 정상적으로 탐색되는 것을 확인할 수 있다.

- 주어진 텍스트가 “ababab”이고 패턴인 “abab”인 경우 찾아진 위치가 1뿐만이 아니 라 3에서도 찾을 수 있는가? 또 다른 예로, “aaaaa”에서 “aa”를 찾으면 4번 찾아 지는가?

(2) 이런 탐색이 가능하도록 알고리즘을 수정한 뒤 어떻게 수정했는지 자세히 기술하라.

* + - 구현상 특징



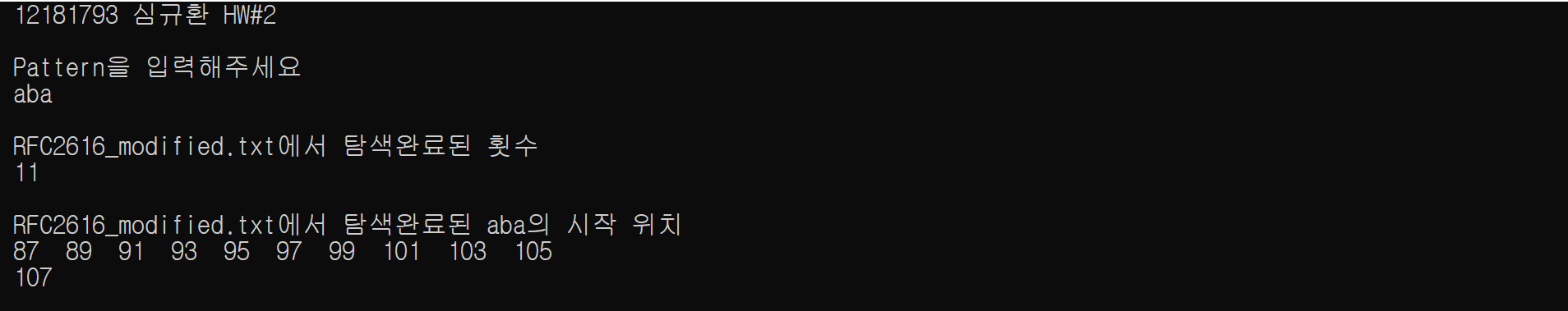
이전 강의 노트에 제시되어 있는 알고리즘에서는 text의 일부와 pattern이 완전히 일치한 후 다음 탐색 위치를 설정하지 않고 바로 탐색을 종료한다. 따라서 문제의 예에서 위치 1에서만 찾아지고 aa가 1번 찾아진 후 종료된다. 그리고 pi도 text의 일부와 pattern이 완전히 일치했을 때 다음 탐색 위치인 pi[m-1]값이 필요하지 않기 때문에 pi[0]~pi[m-2]까지만 설정한다.

이를 수정한 알고리즘에서는 pi벡터를 0~m-1까지 설정하고 text의 일부와 pattern이 완전히 일치한 경우에도 다음 탐색 위치를 설정하여 계속 탐색을 진행하도록 수정하였다.

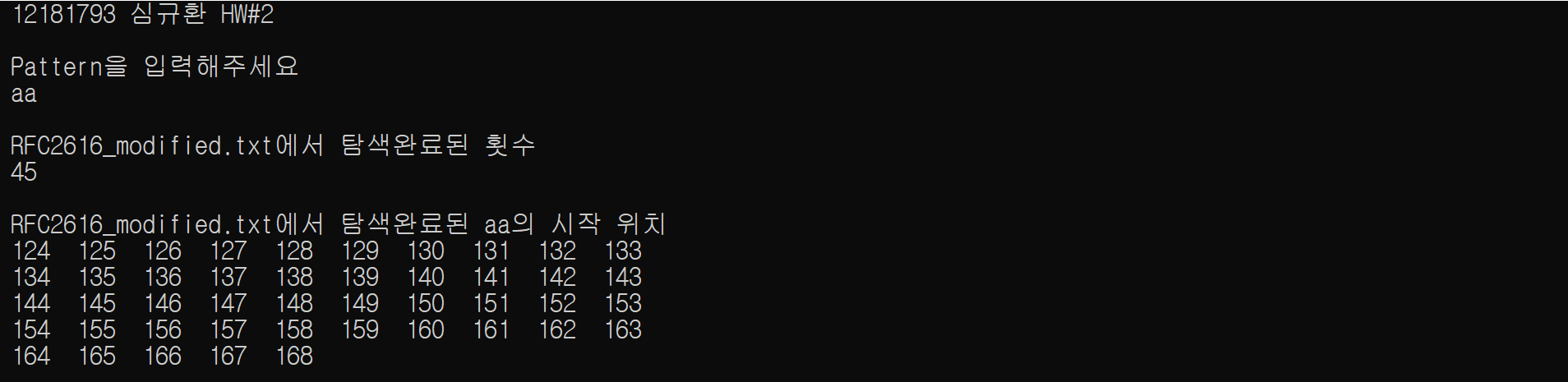
(3) “aba”, “aa” 이 검색되는 것을 보일 것

* + - 실행화면 결과

“aba” 검색



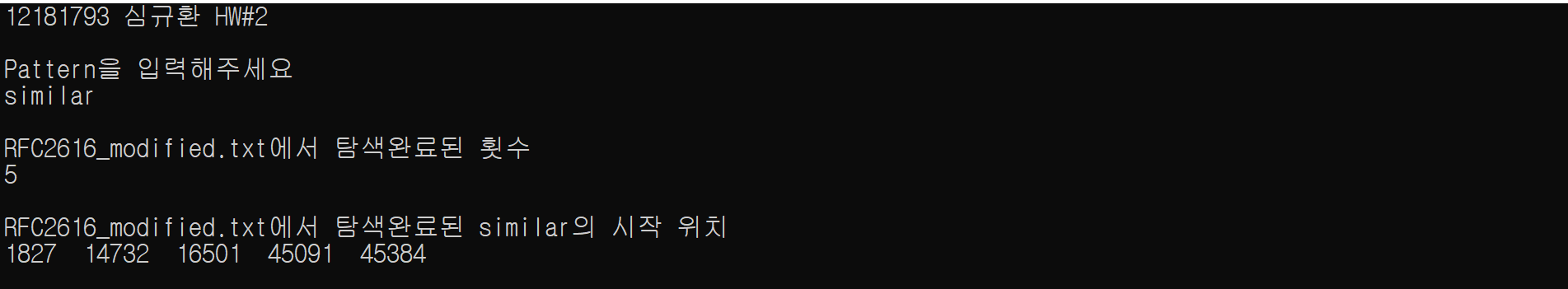
“aa” 검색



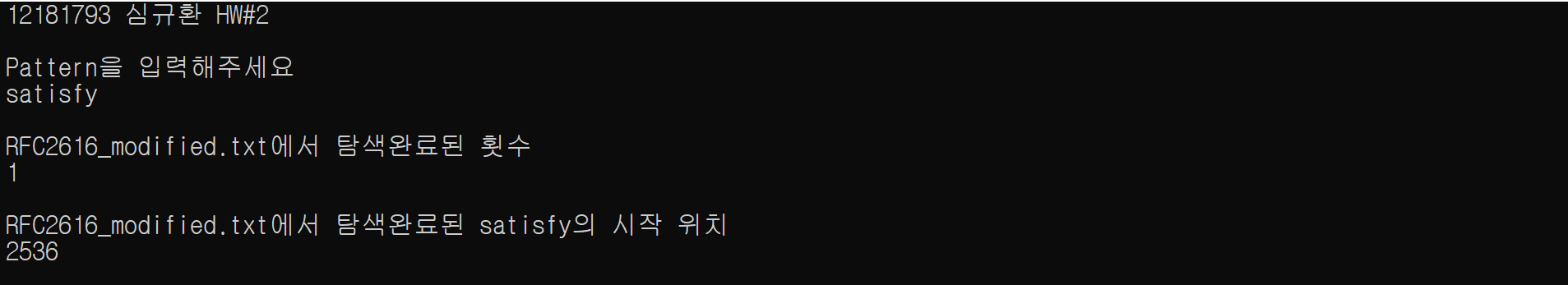
(4) “similar” “satisfy” “refer” “representation” “connections”이 검색되는 것을 보일 것

* + - 실행화면 결과

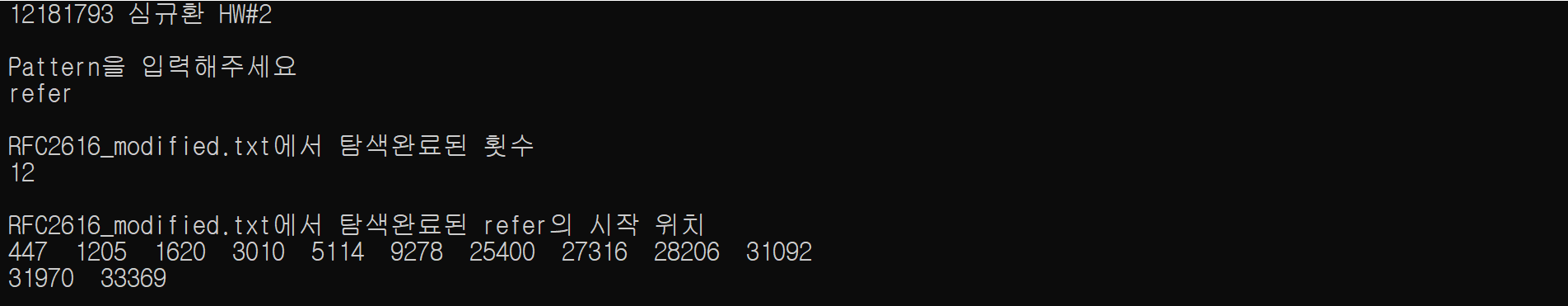
“similar”



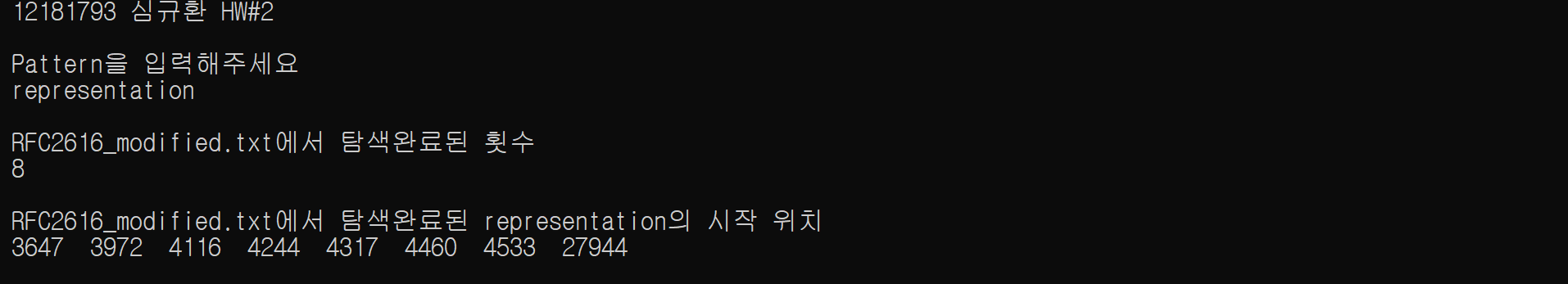
“satisfy” 검색



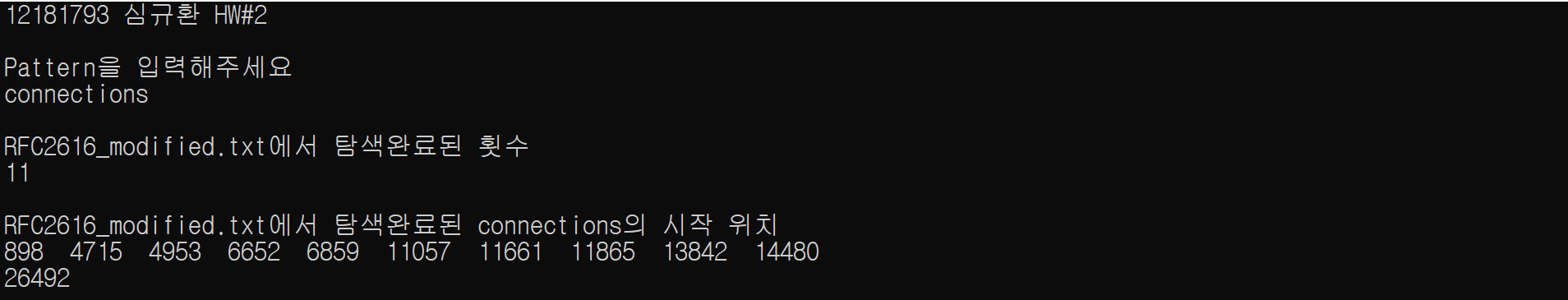
“refer” 검색



“representation” 검색



“connections” 검색



전체코드