

데이터베이스 2차 과제

- 조인 알고리즘의 구현

여러 Column의 값이 저장된 2개의 파일을 2가지 방법을 이용해 Join 한 후 결과를 출력하는 프로그램(파이썬)을 작성하시오

구현해야 하는 조인 알고리즘 (수업 시간에 배운 내용대로 구현)

- Nested Loop Join :
블록 내의 한 튜플과 상대 블록의 모든 튜플을 비교 한 후 다음 상대 블록 진행
- Block Nested Loop Join :
블록 내의 모든 튜플과 상대 블록의 모든 튜플을 비교 한 후 다음 상대 블록 진행

입력 파일 형태 (제공)

파일은 text 형태이며, 첫번째 줄은 Column의 개수와 row의 개수를 알려준다.

이후의 line에는 newline 문자로 구분되는 한 line 마다 지정한 개수의 column 값이 있다.

각 Column의 값은 4 Byte integer이다.

Ex)

```
5 3
1234 2938 20394 2938 293938
293929 293949 28344 1059 2323
12 232 47 540 3059
```

프로그램 실행

- 입력값으로 두 개의 input file 이름, join 방법(nested, block_nested), join column, 한 block 내의 튜플 개수, 출력 파일명등을 입력 받는다.

Ex) 프로그램 실행 화면 예

```
조인할 파일 2개를 입력하세요: input1.txt input2.txt
조인 방법을 선택하세요 (1. Nested loop 2. Block Nested loop): 1
조인할 각 컬럼 값들을 입력하세요: 1 2
한 블록 내의 튜플 개수를 입력하세요: 40
출력할 파일명을 입력하세요: output.txt
```

*조인할 각 컬럼은 2개의 파일에서 어떤 컬럼들이 조인되는지 알려줌

예) 1 2 이면 첫 번째 파일의 첫 번째 컬럼과 두 번째 파일의 두 번째 컬럼 값이 조인 대상이 됨
*블록 내의 튜플 개수는 파일에서 한 번에 읽어들이 수 있는 튜플의 개수를 의미

출력

두 개 파일의 join 결과를 output file에 출력한다.

출력은 두 파일의 column을 concat해서 출력하며 각 file column 순서를 지킨다. (자연 조인 아니어도 됨)

조인 후에는 join 결과 row의 수와 조인에 사용된 블록의 접근 횟수 출력한다. (화면에 출력)

Ex) 화면 출력 예

Result rows: 10000

블록의 접근횟수 : 200

제출 마감

- 12/7일까지 제출 (10점)
- Jupyter 소스 파일 : 학번.ipynb
- Copy 시 모두 0점 처리 (유사 코드 분석기 사용 예정)