**Programming Assignment #1 report**

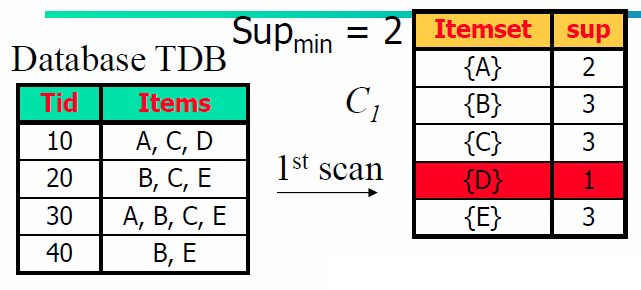
한양대학교 컴퓨터전공

2012004087

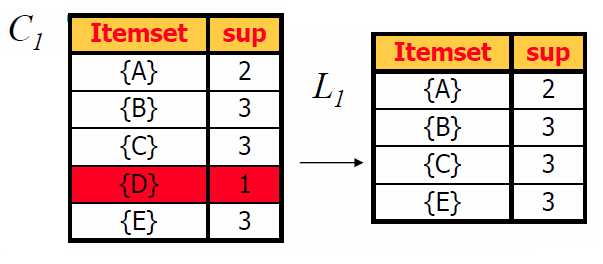
이기준

**Summary of your algorithm**

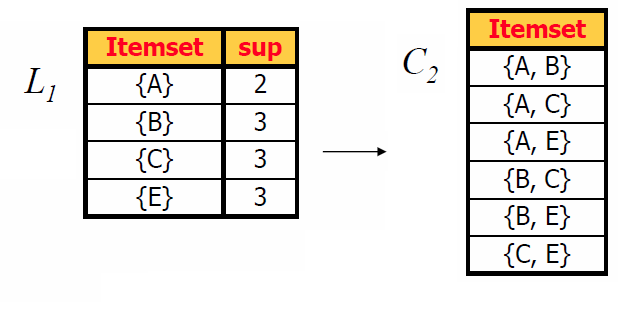
Apriori pruning principle : frequnet하지 않은 itemset의 superset은 만들어지지 않고 테스트하지 않는다.



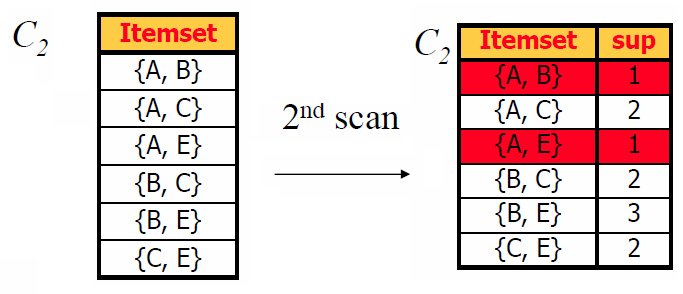
Transaction DB를 1-itemset의 목록과 support 값을 스캔한다.



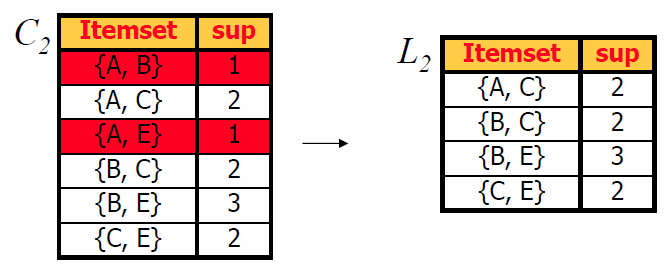
1-itemset 목록 중에서 support값이 minimum support을 만족 시키지 못하면 목록에서 제외시킨다.



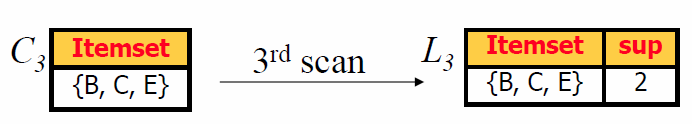
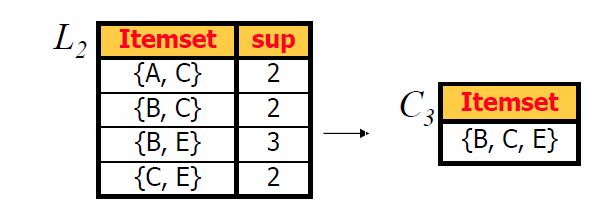
제외 시킨 1-itemset목록을 참고하여 만들어 질 수 있는 모든 2-itemset candidates를 만든다.



2-itemset candidates의 support 값을 DB를 참조하여 스캔한다.

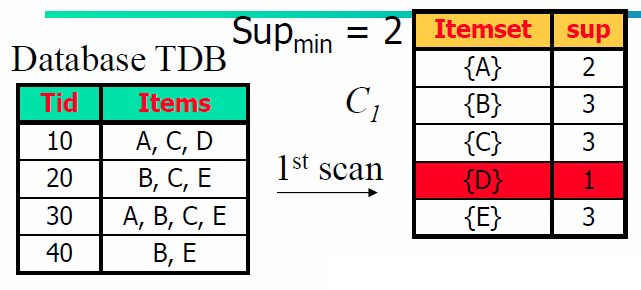


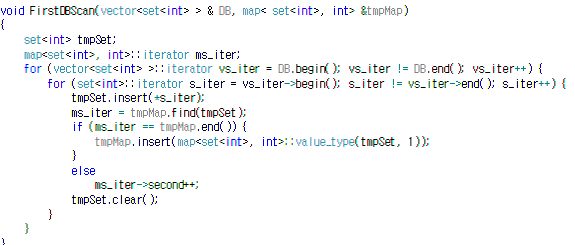
2-itemset candidates에서 minimum support을 만족 시키지 못하면 목록에서 제외시킨다.



위 방법과 동일하게 frequent candidates가 없거나 candidates set이 만들어지지 않을 때까지 귀납적으로 n-candidates 를 찾아낸다.

**Detailed description of your codes (for each function)**

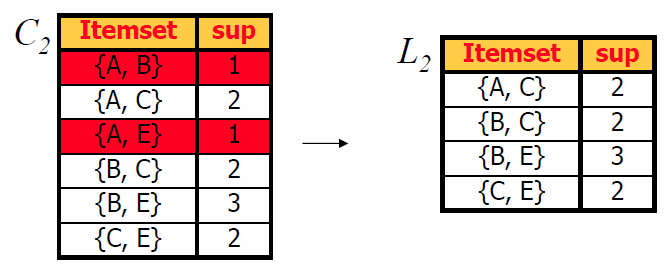
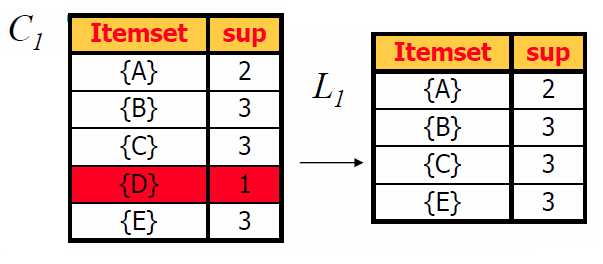


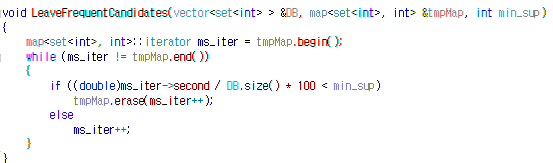


Transaction DB를 tid순으로 내부 items를 순차적으로 스캔한다. (2중 for문)

Tid의 items를 하나씩 보면서 1-itemset candidates에 있으면 support값을 높여주고 없으면 새로 갱신해주는 방식으로 구현했다.

(사실, 1-itemset candidates는 map<int, int> (itemset, sup)으로 구현하는게 적절하지만 모든 n개의 itemset candidates를 모두 담는 변수가 있기 때문에 통일성을 위하여 map<set<int>, int>로 구현하였다.)

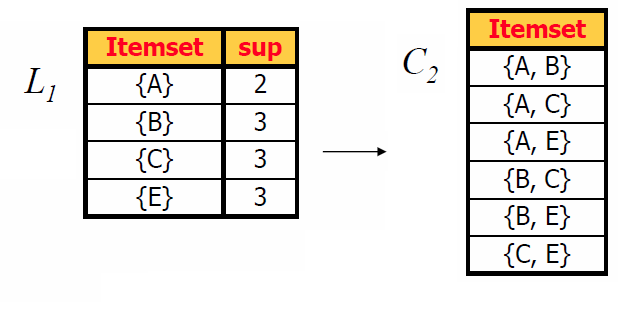
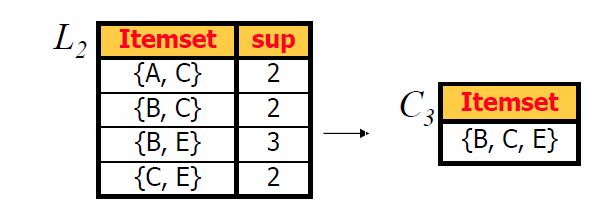


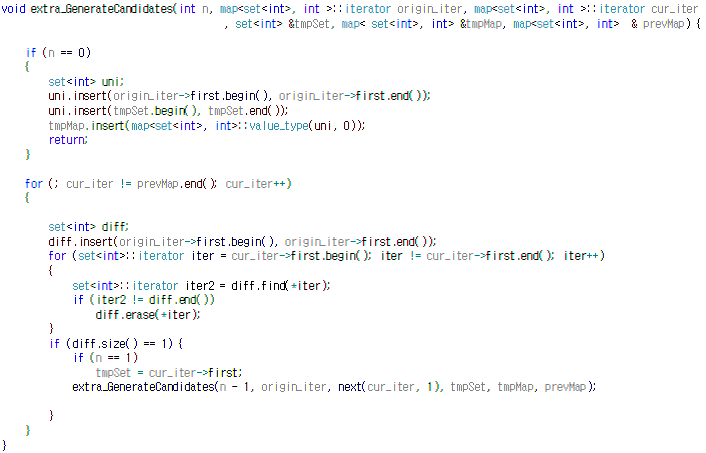


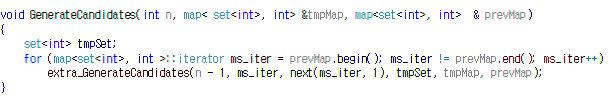
Apriori algorithm의 원리에 따라서,

Support / Transaction id의 개수 \* 100 < minimum support 일경우는 pruning의 대상이 된다.

지워지지 않은 itemset candidates는 frequent하다.

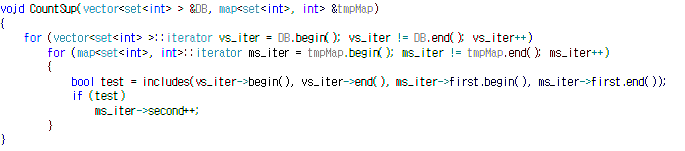
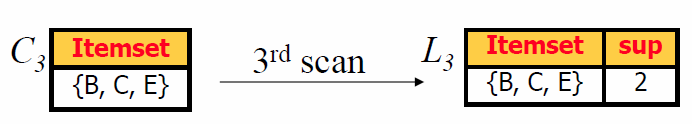
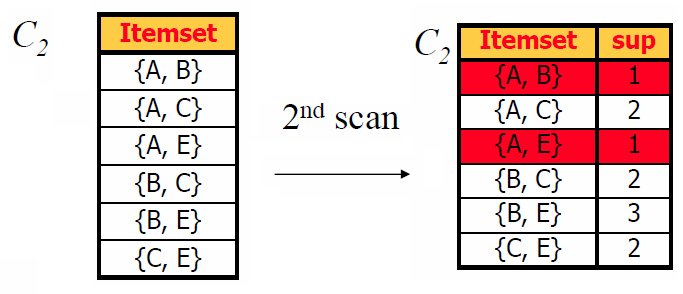




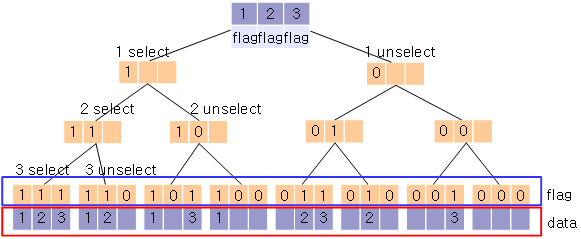
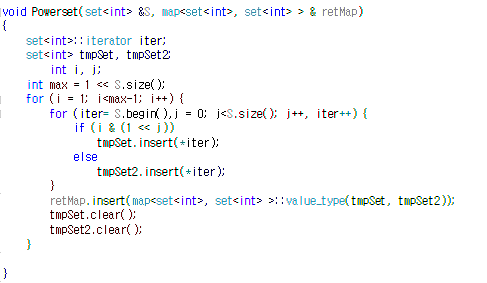


이전 n-1 frequent itemset candidates 중 하나를 고정한 뒤 차집합의 원소 개수가 1인 또다른 candidates들이 n개가 있으면 n개의 itemset을 하나로 합쳐서 새로운 n itemset candidates를 구성한다. GenerateCandidates에서는 고정된 n-1 frequent itemset candidates를 지정하며, Extra\_generateCandidates() 는 재귀형식으로 고정된 candidates와 차집합의 원소개수를 확인한다. n개를 충족하면 재귀 탈출조건이 만족되어 n-itemset candidates에 추가시킨다.

이 함수에서는 support값은 세지 않는다. 오직 n-itemset candidates만 구성한다.



Transaction DB의 tid를 하나씩 스캔하면서 만들어진 n-itemset candidates에 해당되면 support 값을 count한다.

****

output단계에서 candidates를 itemset과 associate itemset으로 나누는 작업이다.

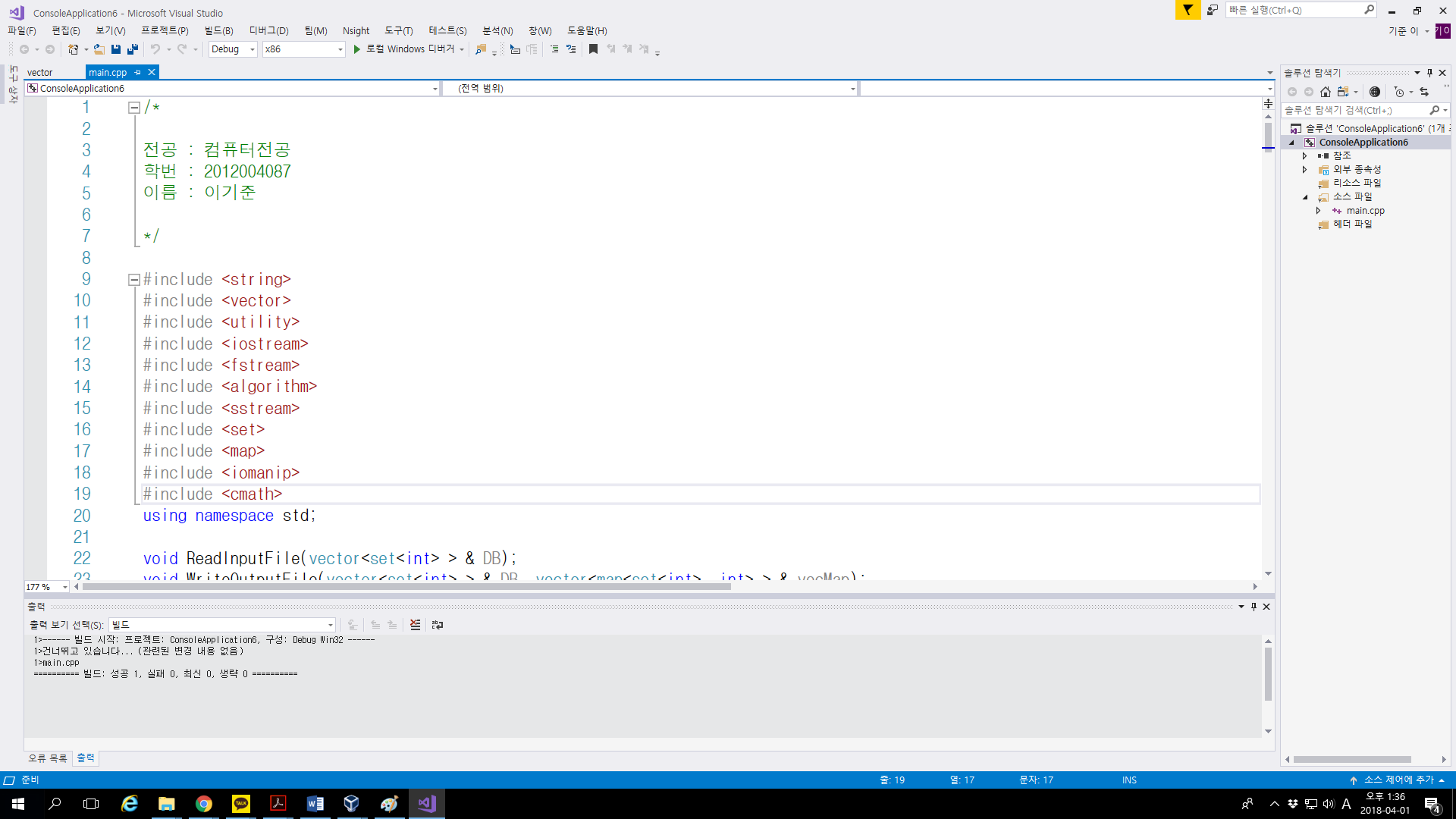
Candidates의 부분집합과 그 부분집합의 여집합도 같이 뽑아낸다.

공집합과 전체집합은 itemset이나 associate itemset이 없어지므로 생략한다**.**

**Instructions for compiling your source codes at TA's computer (e.g. screenshot)**

OS : Windows10, 개발환경: Visual Studio 2017 사용언어:C++

Visual Studio에서 Ctrl + F7 로 컴파일



Ubuntu(linux)에서 g++로 컴파일해도 실행가능, 파일은 main.cpp 하나이기 때문에 따로 makefile은 없음



**Any other specification of your implementation and testing**

윈도우10에서 실행

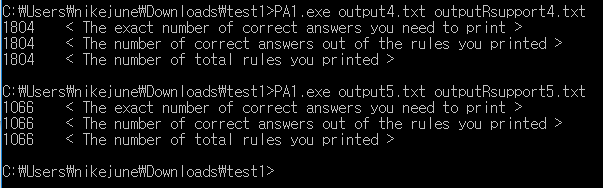
****

****



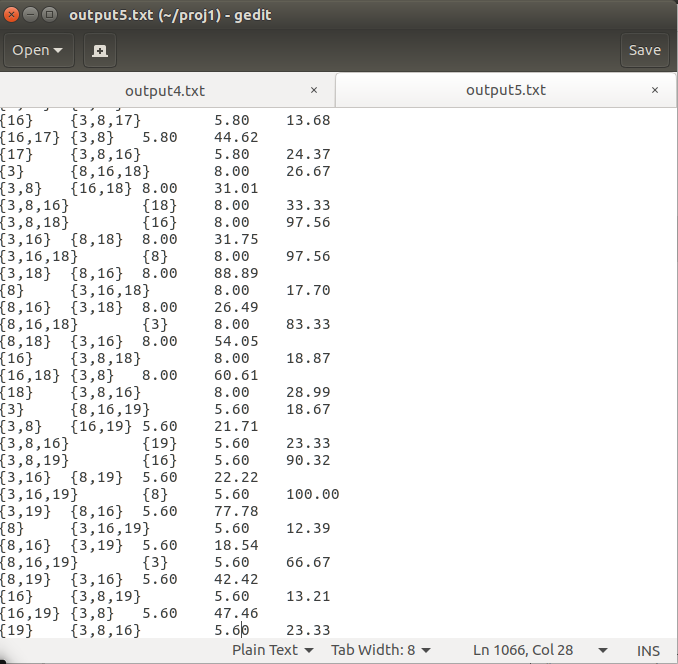
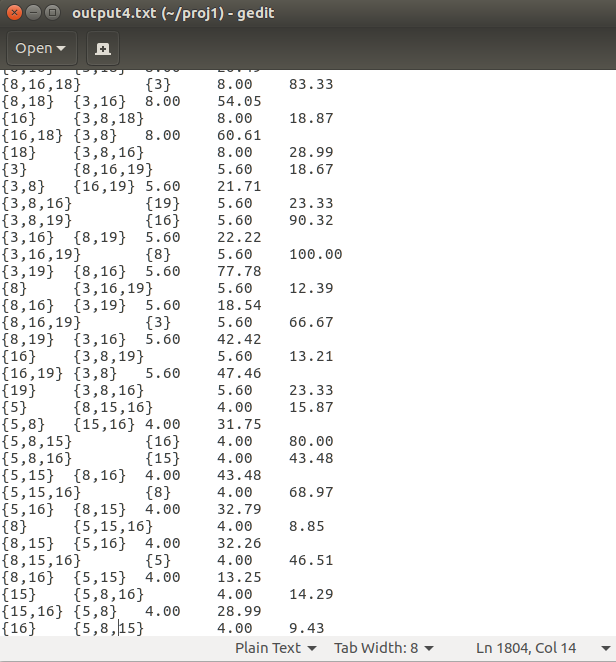


PA1 테스트 프로그램으로 min\_sup 4일떄, 5일 때 모두 테스트

****

**우분투에서 실행**

****

원래 output.txt로 나오는데 output4.txt, output5.txt로 이름을 변경한 것 임