**Given -**

Input file with user's fav artists

Each line corresponds to a user

Each user can have up to 50 fav artists

**To find -**

Pairs of artists who appear together in atleast 50 lists ( 50 lines / 50 user's choices )

**Motivation -**

Largely motivated from what I had done in my IR(Information Retrieval ) project in the Fall semester.

I had to build a complete search engine ( that included – Parser, Indexer and Search components ).

* The Parsing component scans each document ( of the data set provided – Reuters collection – approximately 10,000 documents ) and extracts tokens.
* The Indexer builds an index ( which states if a particular token is present in the document).
* The Search component finds all the documents containing the query term.

To support conjunctive and disjunctive queries ( finding - documents containing all the query terms and documents containing atleast 1 query term ) I had to come up with a solution to a similar problem.

**Solution -**

* + Start by reading the input file.
  + Read 1 line at a time.
  + Obtain the artists in the list ( line ) - by splitting the contents of the line on the ',' character.
  + For each artist thus extracted, indicate (by setting the appropriate index to true in the BitSet) the occurance at a given line.
  + Perform a logical AND operation - to find the number of times 2 artists occur in the same line.
  + Check if the 2 artists occur together in more than 50 lines.

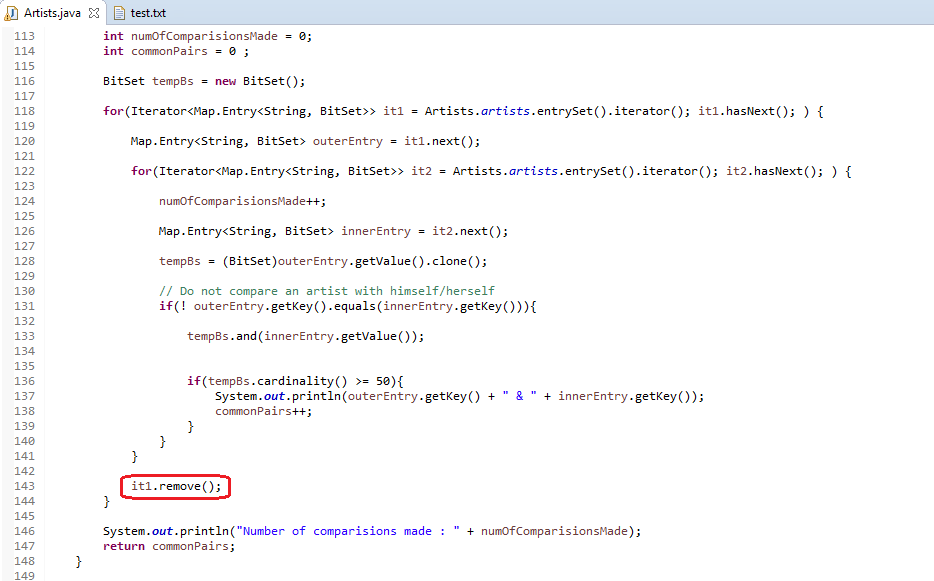
===================================================================================

**Improving performance –**

1. For 2 artists to occur together in atleast 50 lines, both of them must each occur atleast 50 times in the file.

Checking for this condition alone reduces number of candidates from 11,290 (total number of artists) to 124 ( those occuring in atleast 50 lines).

1. By removing the element from the first iterator after comparision, number of comparisions is reduced from n2 to [ (n X (n+1)) / 2 ] ( almost halved ).



Number of comparisions reduced from 15,376 to 7750.

n2 = 1242 = 15,376

[ (n X (n+1)) / 2 ] = ( 124 \* 125 ) / 2 = 7750

**Running :**

* Import the project as an existing project in Eclipse.
* Runt the Artists.java file.