DATABASE MANAGEMENT SYSTEMS
Spring 2015 Lab Test-1 (Even Machines)

You are provided a comma separated variable (CSV) file containing the publication data of PubMed, where each row has the following format:

PubID, title, year, author1 | author2 | author3 | ... | authorN |

Note that author is a multivalued attribute. In the CSV file the author names are delimited by the '|' symbol.

Create a database with three tables, namely Publications, Authors and Relationship. The schema of the tables should be as follows:

```
Publications( PubID, title, year )
Authors( AuthorID, AuthorName )
Publish( PubID, AuthorID )
```

1. Create the tables with appropriate integrity constraints.
2. Parse the csv file and populate the tables accordingly.
3. Write one or more SQL queries for obtaining each of the following:
   (a) Find the year which has seen the most publications by a single author. The resulting tuples should be in the format: <year, AuthorID, count>. Order the tuples chronologically.
   (b) Find the authors who have strictly increasing number of publications between 1989 and 2000 (including these years). The result should be in the format: <AuthorID>
   (c) Find all the pairs of authors who have always published with no one except each other. The result should in this format: <AuthorID, AuthorID>
4. You are allowed to use a programming language of your choice to find the result of this part of the question. Query the database to extract whatever information is required, process that information according to the problem and output the result.

   Consider a co-authorship network where nodes are the authors and an edge connects two authors if they have ever collaborated. We wish to allocate dinner table seats to delegates attending a conference in a way that co-authors are seated on different tables. Your task is to find the minimum number of dinner tables needed to achieve this objective.
   [Greedy Algorithm for Graph Colouring: *Order the nodes in the network. Select each node in that order and use the minimum colour not already used in its coloured neighbours*.]