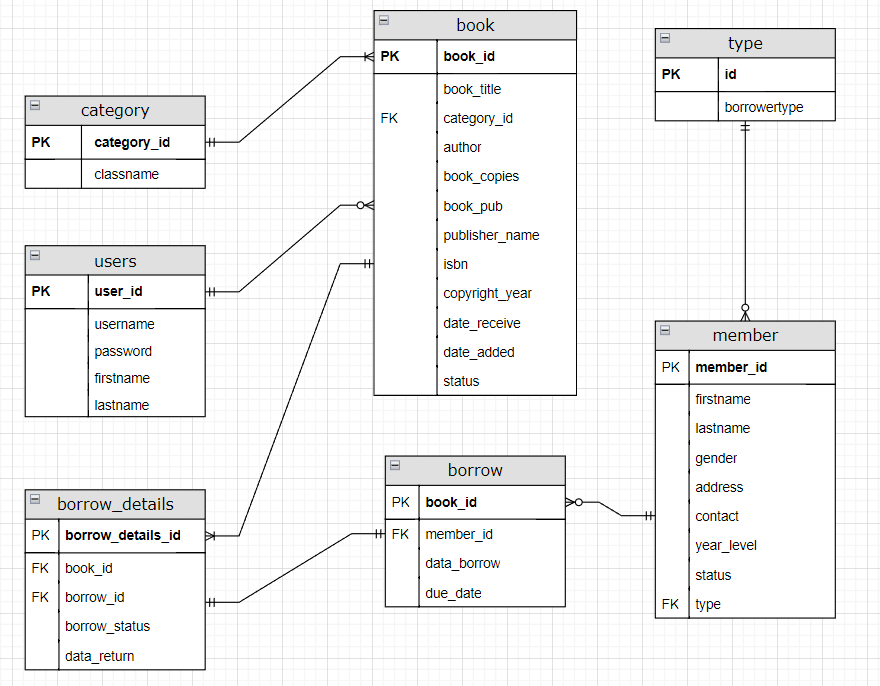
COMP1044 REPORT DOCUMENT

GROUP NAME: GROUP 11

| STUDENT ID | STUDENT NAME |
| --- | --- |
| 20412776 | LEO TAN HAI KEN |
| 20282376 | YEP JOHN WAYNE |
| 20312846 | RULAND MUHAMMAD FURQAN |
| 20303049 | NEOH JIE REN |
| 20398899 | LAU SOOK MUN |

# TASK A

**ER DIAGRAM**



**SUMMARISING YOUR OPINION ABOUT THE DATASET**

**Question 1 : Why did you choose the datatypes of each field?**

In the book table, we decided to give integer data type to ‘book\_id’, ‘category\_id’, and ‘book\_copies’ because the data set only contains a number, specifically an integer. As for ‘book\_title’, ‘author’, ‘book\_pub’, ‘publisher\_name’, ‘isbn’, and ‘status’, they are all a varchar data type because they contain a character. For a clarification, although ‘isbn’ seems like an integer data type, it has a dash which makes it invalid to be an integer. To continue, the ‘date\_received’ and ‘date\_added’ are a datetime data type because the data type consist of the date, therefore, it is categorize in the datetime data type. Lastly, ‘copyright\_year’ categorize as a year data type since it only has a year in the data set.

In the member table, ‘member\_id’ is the only entity to have the integer data type. Meanwhile, ‘firstname’, ‘lastname’, ‘gender’, ‘address’, ‘contact’, ‘type’, ‘year\_level’ as well as ‘status’ have a varchar data type because they all contain a character in the data set.

In the borrowdetails, ‘borrow\_details\_id’, ‘book\_id’, ‘borrow\_id’ have an integer data type. While ‘borrow\_status’ and ‘date\_return’ have a data type of varchar and datetime respectively.

In the users table, only the ‘user\_id’ has an integer data type while the others such as ‘username’, ‘password’, ‘firstname’, ‘lastname’ have a varchar data type.

In the borrow table, both ‘borrow\_id’ and ‘member\_id’ have integer data type while ‘date\_borrow’ and ‘due\_date’ have datetime data type.

In the type table, ‘id’ has an integer data type and ‘borrower\_type’ has a varchar data type.

In the category table, ‘catergory\_id’ and ‘classname’ have integer and varchar data type respectively.

**Question 2 : Reason behind the design of the tables and connections between the tables including PK, FK?**

The connection from category table to book table is one to many with category\_id and book\_id as the primary key. The connection of the users table to the book table is one to many with user\_id as the primary key and category\_id as the foreign key. The connection from book to to borrow\_details is one to many also with borrow\_details\_id as the primary key and book\_id and borrow\_id as the foreign key. The table borrow and borrow\_details is a one to one with borrow\_id as the primary key and member\_id as the foreign key. The connection from member table to borrow table is one to many with member\_id as the primary key and type as the foreign key. Lastly, table type has a one to many connection with id as the primary key.

**Question 3 : What kinds of errors did you find in the dataset?**

* + Database file “type” seems to be completely irrelevant from the other datasets
  + Column of “Contact” in file “member” looks unusual
  + “Book\_id” in file “book” starts from 15 rather than 1 (previous books?)
  + [Similar issue occurs in “borrow\_id”,”borrow\_details\_id”, and “member\_id”]
  + Why do we have “book\_pub” and “publisher\_name” at the same time in file “book”, aren’t they the same thing? (some publisher\_name even contains address, sus)
  + If there exists multiple copies of a book, each book should have a unique id, else there’s no way to know exactly which book is borrowed out and the current status of those individual books.
  + What is “Date\_recieve” in the file “book” and why is it blank?

**Question 4 : How did you find the completeness of the dataset?**

The completeness of the dataset seems low, there should be more data that was once recorded in the database, as “Book\_id” in file “book” starts from 15 rather than 1, so I would assume that there should be 14 more books that wasn't included in the database, and similar issue occurs in “borrow\_id” starting from 482,”borrow\_details\_id” starting from 162, and “member\_id” starting from 52. Moreover, there are several boxes that are filled with .. or even left blank (date\_recieved in “book”), contact (number?) in the “member” database is also a complete mess with random numbers. Lastly I notice that there are multiple copies of the same book, therefore If there exists multiple copies of a book, each book should have a unique id, else there’s no way to know exactly which book is borrowed out and the current status of those individual books.

**Question 5 : How did you address those problems?**

5.————Let's discuss this in meeting————

Missing data:

Book.csv

Author- finding information online

Book publisher - finding information online

Publisher name - finding information online

Date\_receive -removed

Member.csv

Contact -non existent