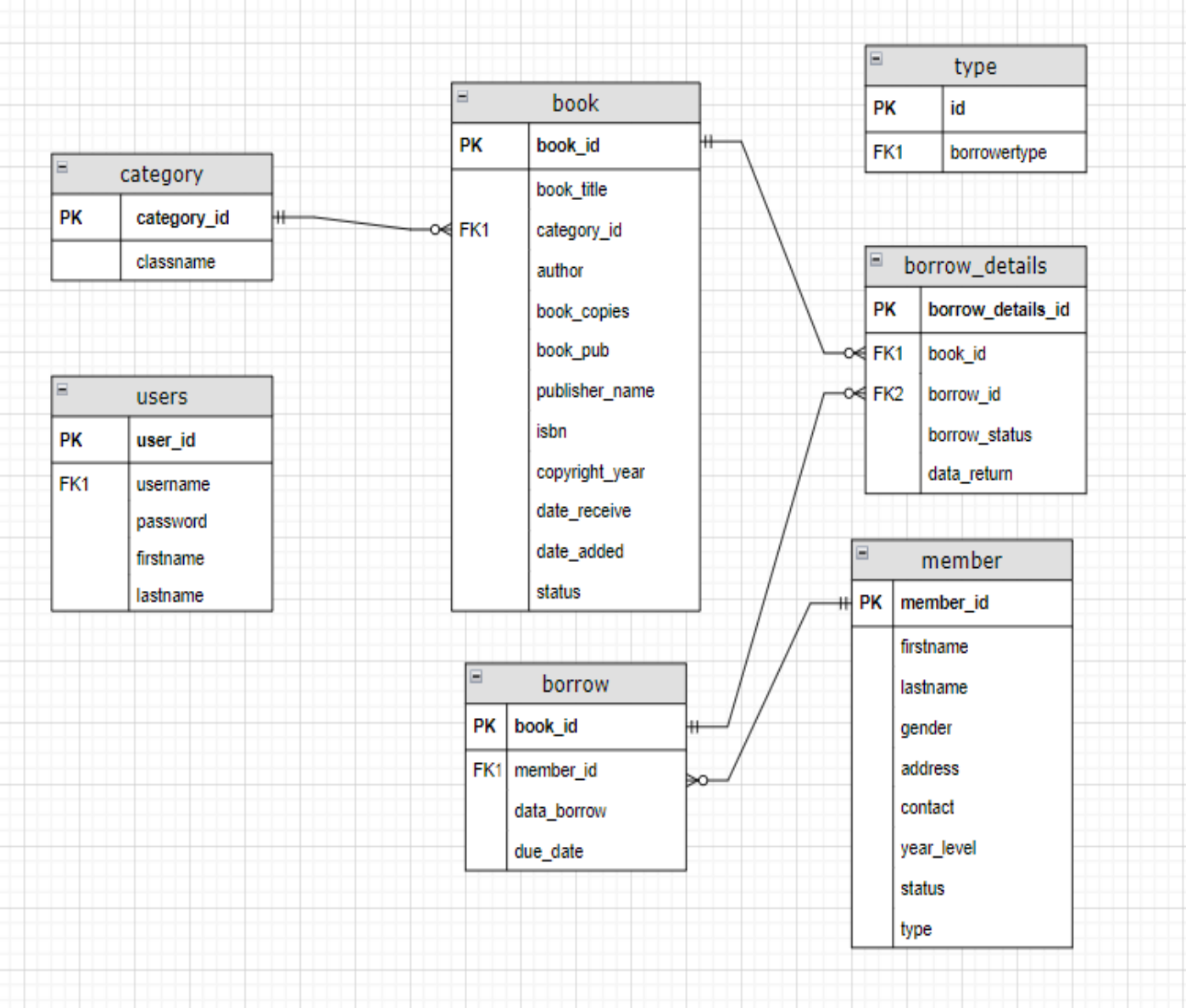
COMP1044 REPORT DOCUMENT

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 we 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?

Missing data:
We decided to search up the informations from internet, then fill them in.

- Author- finding information online
- Book publisher - finding information online
- Book publisher - finding information online
- Publisher name - finding information online
- Date_receive -removed

Member.csv
Contact -non existent