CISC637 PROJECT: Lizhou Yuan, Yiqun Jia, Ramin Ramazi

SDB Design

1. Overviw

The goal of this project is to model the user posts and comments in a social network. SDB is short for Social DataBase. We plan to model the posts that each user made, and the comments that users write below each post. For gathering the data we are going to use the real users posts and comments in the Wechat(similar with whatsApp). In Wechat, the users' post can be either text, image, or combination of both. Other users can submit comments for each post.

2. Data

As described above, we employ the social data, posts and comments, of Wechat as our data source. Currently we have 2 plans to gain the data we need. PLAN A: Generate the instances for each entity manually, at least 20; PLAN B: Develop a module to export data to JSON files by crawler[1].

In SDB, we have non-relational data such as icon, long text, etc. Currently, we have 3 entities and 3 relations. For each entity, there are more than 3 essential attributes. The **schema** is listed below:

User:

Field	Type	Null	Key	Default	Extra
user_id	int(11)	NO	PRI	NULL	
nickname	varchar(30)	YES		NULL	
icon	int(11)	YES		NULL	

Message:

Field	Type	Null	Key	Default	Extra
mes_index	int(11)	NO	PRI	NULL	
user_id	int(11)	YES	MUL	NULL	
content	varchar(3000)	YES		NULL	
post_time	datetime	YES		NULL	

Comment:

Field	Type	Null	Key	Default	Extra
com_index	int(11)	NO	PRI	NULL	
mes_index	int(11)	YES	MUL	NULL	
user_id	int(11)	YES	MUL	NULL	
content	varchar(500)	YES		NULL	
comment_time	datetime	YES		NULL	

1. A sample project: https://github.com/Chion82/WeChatMomentExport

CISC637 PROJECT: Lizhou Yuan, Yiqun Jia, Ramin Ramazi

Briefly speaking, it's designed for the data management of the user generated data in Wechat. We care about which user post which message/moment and which user comment on which message, and even what the comment do the user make under certain message. Besides, we are also interested in the time related properties. For example, how many user post messages in this month? Who is the most active user last week? Etc. In this phase, we don't care about the content of the user's publication, but only the behavior of the user.

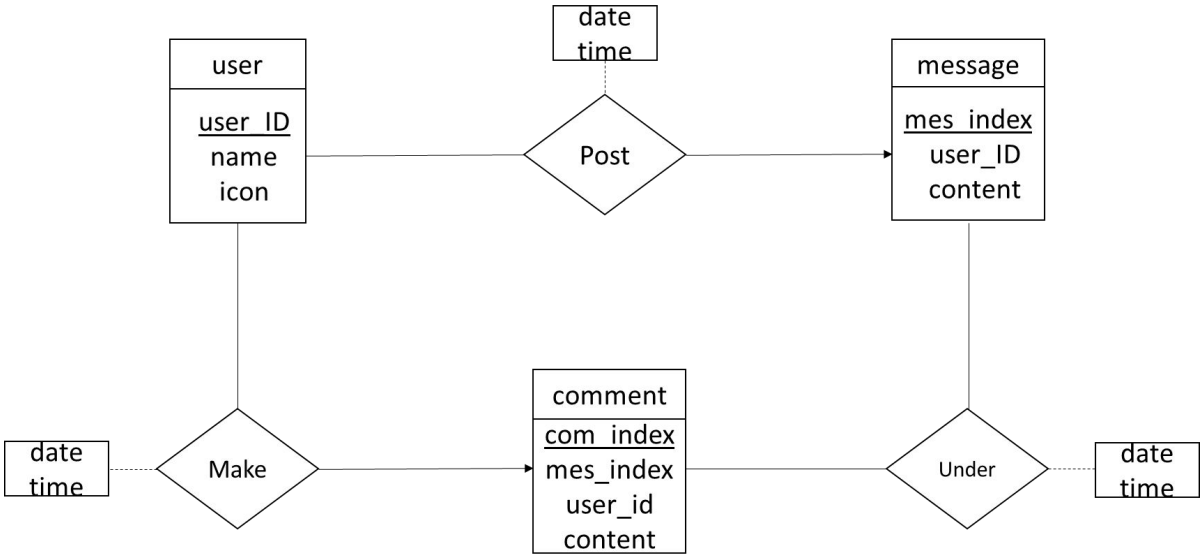
Our database will include non-relational data such as pictures and long text and they are all practical data types in the real social network. For example the Wechat user's profile include avatar which will be included in our user table. The long text will be included both in message table and comment table. Also the message might include text and pictures. We take all these non-relational data into consideration.

We have 3 tables: user, message, comment.
In the user table, the *user_id* is distinct for every user and each user must have an id. *Nickname* and *avatar(icon)* is also a part of the user profile in Wechat. In a word, id is for inner use, nickname and avatar are shown the real user.
The message table includes *mes_index*, which is the index of the message, again this index is used for inner demand, to identify each message and each message should have a unique id. The *user_id* is same as the user table, which points out which user post this message/moment in Wechat. *Content* is the text or picture the user post, in this phase we don't consider the picture part, but we will take it into account later. *Post_time* is when the user post this moment/message.
The comment table is much similar with the message table. It also includes the user id who make the comment and comment time.

The data snapshot we shown is just about the raw data we are likely to use.

1. A sample project: https://github.com/Chion82/WeChatMomentExport

3. E-R Diagram



Instance of crawled data:

1. A sample project: https://github.com/Chion82/WeChatMomentExport