BMME128 Group Project: Data Description

1   Notice

This is completely anonymised data from Sweek.com. By participating in this group project, you agree not to share this data, parts of it, or any results derived from it with anyone who is not enrolled in BMME128 in 2019. In particular, you agree not to make this data available publicly. This data is solely for the purpose of this group assignment and must be deleted once the course has ended. You obtain no rights to the data or the results of your analyses.

2   Data

The data includes all stories that were published in English, Indonesian, and Spanish, as well as the related interactions (reads, likes, follows, comments).

Green fields are unique in their table.

All times are in milliseconds since 1970 (UNIX time), apart from dates in the accounts table, which are in CET/CEST.

2.1 Stories

Data on stories published in English, Indonesian, and Spanish.

stories

STORY_ID	Anonymised unique ID for each story
CREATE_TIME	Server time when the story was created in milliseconds (UNIX time)
MODIFIED_TIME	Server time of the last edit (title, description, status, category etc.), but not chapter added; in milliseconds (UNIX time)
RATING	Authors can indicate adult content if it's not suitable for readers less than 18 years old
CATEGORY_REF	Refers to a category on Sweek. 0 = no category chosen. If the story is not published, that's possible. They all refer to a specific genre.
EXTRA_CATEGORY_REF	Users can also choose a second category, but it's not mandatory. If not chosen, it's NULL.
STORY_LANGUAGE	Identifies the language
NR_OF_CHAPTERS	Both published and unpublished chapters
IS_PUBLISHED	0 = draft (not visible to other users) 1 = published (visible to other users)
IS_FINISHED	This is something the user can indicate him- or herself to show the reader if 0 = unfinished or 1 = completed (no more new chapters)
HAS_COVER	FALSE = no cover, TRUE = the user upload their own cover
COPYRIGHT_REF	1 = all rights reserved, 2 = creative commons, 3 = open domain
IS_CLASSIC	0 = 'regular' story 1 = story uploaded by the Sweek team as a classical literature story that's free of rights and free to use
FIRST_PUBLISHED	Server time when the story went from draft to published
LAST_CHANGED	Story was modified (title, description etc.) or extra chapter added
ACCOUNT_ID	Anonymised unique account ID of the author.

2.2 Genres

A list of all genres.

genres

category_ref	Unique category/genre ID
name	Genre name

2.3 Accounts/Users

Users who created or interacted with stories.
- Information on some users is incomplete. In most cases, these are users from a different language background (English, Indonesian, or Spanish).
- The user with ID 15011 comprises all activities by anonymous users (those who did not create an account). At the moment, such users can only read one chapter. However, in the early days they could still read all chapters.

stories

ACCOUNT_ID	Anonymised unique account ID
CREATED_AT	Time the account was created
UPDATED_AT	Last time the e-mail, password or username was changed
GENDER	0 = not specified by user 1 = male 2 = female 3 = other
DOB	Date of birth as filled in by the user. Null = not filled in
READLANG	Preferred language.
FROMSOCIAL	facebook = sign up and log in with a Facebook account empty = e-mail based account

2.4 Comments

User can comment on stories and this is a list of all comments that have been left on the stories included in this data.

comments

COMMENT_ID	Anonymised unique comment ID
CREATE_TIME	Time the comment was posted
ACCOUNT_ID	User who posted the comment
STORY_ID	Story on which the comment was posted
PREVIOUS_COMMENT_ID	Refers to another comment, if this comment was made as a reply to another comment. Null means it's a first-level comment (i.e., not a reply).
DELETED	0 = not deleted, 1 = deleted by the author of the comment or the author of the story
WORD_COUNT	An approximation of the number of words in the comment.

2.5   Interactions

Apart from reading and commenting, users can also like and follow stories.

Interactions

INTERACTION_ID	Anonymised unique interaction ID
SERVER_EVENT_TIME	Server time in UNIX milliseconds of when the follow or like was registered on the server
ACCOUNT_ID	The user who interacted
INTERACTION_TYPE	2 = like 3 = follow (added to library and get notifications)
STORY_ID	The story the user interacted with

2.6   Reads

Event log of users reading stories.

reads

READ_ID	Anonymised unique ID
SERVER_EVENT_TIME	Time the read was finished
ACCOUNT_ID	The user who was reading
COUNT	Number of chapters of the story read
STORY_ID	The story the user was reading