Python Data Set Description

This is a data set crawled from Coursera with the permission from Coursera. Each line describes a corresponded post. This discussion forum has several subforums. The course of this dataset is, https://class.coursera.org/programming1-002/forum. We crawled 24963 posts from this forum, and the attributes of each column are explained as follows.

badge: *whether the person who makes this post is instructor, community TA or others. If this entry is blank, it means this person is a normal student.*

forum_path: *this post belongs to which subforums*

parent_id: *the parent of current post*

parent_link: *the URL linking to parent post (might be invalid in some cases)*

post_author: *who makes this post*

post_id: *ID of current post*

post_link: *the URL linking to current post*

post_text: *the content text in this post*

post_time: *time when the post is posted (ignore the -04:00)*

thread_id: *the thread id of the current post. (A thread contains many posts.)*

thread_link: *the URL linking to the corresponded pos*

thread_resolve: *a label indicating whether this thread is resolved or not. If this entry is blank, it means this thread has no label; otherwise, you need to check the entry to see whether it is resolved or unresolved.*

thread_title: *the title of this thread*

timestamp: *a Unix format timestamp*

topic_title: *the topic title of current post (Threads are posted under different subforums, and each subforum has its own different topics. For example, in the Lecture subforum, you have week 1 subsubforum, and in week 1, threads relating to similar topics are categorized into corresponded topic areas)*

topic_link: *the URL linking to the topic title*

votes: *the total votes this post has received*