

# ISI 2016 Homework 1

*Given out: 11 October 2016*

*Due: 18 October 2016*

The database describes **an Internet forum** in which users can post questions and answers.

Posts have an Id and a text. They were created at some datetime by a user who is their owner. They have a title and may have tags, as well as an answer count, a comment count, and a vote count. Posts are either questions or answers. If they are answers, they have a parent post (which is a question). If they are questions, they can be “closed” at some point in time and then have a closing date; and they can have an accept-answer-Id. Posts have an original author and can be edited later by this author or other users. The complete version history is kept (i.e. who edited which post when to which new text).

Users have an Id, a display name, an age, a last access date, a website, and the number of upvotes as well as downvotes that they have given. The “community user” with Id=1 is a placeholder for wiki-style editing (the posts then belong to the “community”); this user may not vote on posts. Users who are administrative users can also receive badges, which have a name and were given to these users at a datetime.

Users can tag posts; the database stores the act of tagging and all information belonging to it: the user, the post, the tag. Tags are single words or multi-word expressions, and they have a definition (e.g. the tag “regression” could have the definition “regression is a statistical method that ...”). The database owner wants to analyse the tags later (count how many users gave the same tag, how many posts were tagged like this, build tag folksonomies, etc.). Users can also comment on posts, with a text and a score. Further, users can vote on posts, with the votes classified using different types. Common vote types include:

- 1 - AcceptedByOriginator - this was the answer that the questioner chose to accept.
- 2 - UpMod - the post was moderated up. That's good.
- 3 - DownMod - the post was moderated down. That's bad.

The date+time of all tagging, commenting and voting events are recorded.

*Model this information as an (E)ER diagram, and keep in mind that information that is given redundantly should only be modelled once, but can be modelled as derived attributes where it is repeated.*

*Please send your solution – ideally consisting of an ipython notebook, with the figure created by DIA (however, this is not compulsory and will not affect your grade) – to Bettina Berendt and Pedro Zuidberg.*