User-Content on Online Boards

Ryan Chalmers, Joseph Farah, Etienne Guillot-Marquet, Huey Francois Nguyen
Group 60

I. Requirement Analysis

Introduction

Purpose

This application attempts to build an online social hub with the goal of creating strong, content focused, user-run communities. The application stores users' personal information for credentials validation and data analytics purposes, along with the content that they generate and/or share on online boards, their comments to other users' content and other users' comments to the former's content. Content takes the form of a registered user's submission.

It is inspired from reddit.com, the world's 6th most visited site according to Alexa.com and known as the "Front page of the Internet".

Scope and special requirements

In general, users from anywhere are allowed to navigate and view content on any online board. Registered users can do the same, and in addition, can subscribe to online boards to generate and/or share content. They are allowed a great degree of freedom of expression, limited only by the enforcement of established guidelines by the moderators of their subscribed online boards.

Terminology

Upvoting/downvoting = the act of approving/disapproving a submission by a registered user, including his own submission. This can occur for any reason, but common reasons include: liking/disliking the appeal, logic, agreeableness, timeliness, insight or usefulness of the submission. A registered user can be neutral, positive or negative towards a submission, and can demonstrate their positivity/negativity by upvoting/downvoting each submission only once. But, a submission can have an unlimited amount of "upvotes" or "downvotes" from registered users.

Subscribing = the act of joining an online board by a registered user. This can occur for any reason, but usually when the board aligns with his interests. The user may then receive updates of submission activity from the board.

Comment_chain = a relationship akin to the data structure of a tree, where the root submission is called post, and its nodes are children and descendant submissions called comments.

Moderates = a registered user with the status of "moderator" of an online board is said to moderate the board by maintaining the integrity of the board through his powers to ban registered users and to remove submissions, in accordance to the board's guidelines.

Banned = a registered user is banned if a moderator from an online board blocks that user from accessing the board for the purposes of submissions.

Data Requirements

Entities and Attributes

Users: Someone who created an account on the board is a user. A user is uniquely identified by their username. A user can enter an associated account email, password, and country. A user is automatically associated with a join date and an active status. Select users may or may not be site admins. A users score is a calculation described below.

Messages: Messages are a special type of user to user communication. Messages are uniquely identified by a message ID. The only other attribute is the text content of the message.

Boards: A board represents a central webpage that focuses on certain topic of interest, indicated by a board name. A board however is uniquely identified by the boards URL. Additional attributes include a brief description of the content and rules, and number of users subscribed.

Submissions: A submission is a user submission that is linked to a board, a post, or a comment. All submissions are uniquely identified with a generated submission ID. A submission has a date and time of posting associated with it. A submission may either be viewable or hidden from other users, indicated by the status attribute.

Post: A post is a type of submission that can only be made to the board. Each post should have a title and text content associated with it. A post may also be stickied at the top of the board by the is_Announcement attribute for ease of viewing.

User_posts: User posts are the regular types of post a user can make. User posts have no special attributes related to the post superclass. However, this set is mainly to make an important distinction against the other type of post, Ads.

Ads: Ads are a type of post that are used specially to advertise a product or service. An ad is special in that its related to company, and has a buyout price that they pay to the website for exposure priority.

Companies: Companies are customers of the site and advertise by buying an ad. Companies are uniquely identified with a generated company ID, and each have a company name attribute.

Links: Links are URLs that are associated to posts. In the schema, links are weak entities of posts. A link_url is not unique, rather a foreign key that requires post ID of the post it was derived from. Links are potentially useful for websites that wish to see what posts/boards their website has been linked to, as well as the volume reflecting the traffic their link has generated.

Comments: A comment is a type of submission that can only be made to posts and other comments. Comments only contain text comment content.

Relationships

Much of the main functionality of the website is focused around the user. The following relationships describe how the users are categorized and how they interact with the website.

Owns: A user can create a board, and automatically is given ownership to it. The relationship is needed to record users and the board they own, as board owners have special privileges and powers over said board. This relationship is many to one, as a user may own many boards, but a board is key and participation constrained to being owned by a single user.

Subscribes_to: A user may want to track and receive updates from boards they find interesting. This is a many to many relationship, as a user may subscribe to many boards, and a board may be subscribed by many users.

Submits: The main functionality of the website is to let the user make submissions to other users boards, posts, and comments. This is a many to one relationship. A user may submit many submissions, while each submission must be made by a single user.

Rates: Every submission has a score associated with it, based on ratings given by a user. It is useful to have posts that are liked by many users to have a score that reflects such popularity. A user may only add or subtract one to any submission, indicated by the up/down attribute. This is a many to many relationship, as a submission may be rated by many users and a user may rate many submissions. We can keep up/down as a attribute in the relationship as a user cannot simultaneously give a post a positive and negative rating.

Receives/Sends: User to user messages are handled by two relationships, receives and sends. Each unique message is tracked by these two relationships to the users that sent and the user that received it. Both are many to many, as many users can receive/send messages, and a message must be sent or received by a user. This model however does not capture the fact that both the sends and receives tables must be updated with concurrent information. The time that the message is received/sent is stored here.

Moderates: To keep boards from being flooded with junk/spam content, and removing hostile users from a board, moderators are users with privileges related to specific boards that they are associated with. Moderates is a many to many relationship, as a user may moderate many boards, and a board may be moderated by many users.

Banned: Users that are exceptionally hostile can be added to a ban list related to specific boards. This is a many to many relationship, as a user can be banned from many boards, and a board can have a banlist of many users.

Posted_on: As boards are a collection of posts, a relationship is needed to describe which posts belong on which board. This is a many to one relationship, as a board may have many posts, but a post must belong to one board.

Commented_on: Similarly to posted_on, users can give their input to a post. This is many to one, as a post may have many comments but a comment may only belong to one post.

Comment_chain: While it is important for users to interact with post content, users should also have the ability to interact with the comments of other users. Comment_chain is a relationship that describes comments as parent and child comments. This is a many to one, as a parent comment may have many child comments, but a child comment must have one parent comment.

Advertises: Companies can advertise by buying the topmost spot of the front page of any online board in the form of a submission. This spot is always above the most popular non-ad submission of that board. Advertises is a many to one relationship, as a company must make one or many advertisements, and an advertisement must be associated with a company.

Works_for: Companies who wish to advertise must have one or more of their employees register as a user. This allows the employees to make a submission on behalf of the company. This is a many to one relationship, as a company may have many associated users but a user may only be associated with a company.

Application description:

Overview
The application seeks to efficiently keep track of the ownership of registered users' submissions, their comments on other registered user's submissions and other registered users' comments on the

former's submissions. In addition, it keeps track of the score of each submission. Calculations are performed on their score to determine their rank within each post, each online board and on the main online board.

Preliminary calculations
The algorithm detects the most active posts within the last 24 hour based on their score and on their comments' score, all of which depend on users' upvotes/downvotes. The time frame is a 24 hours moving-average window, and the rank is updated every half-hour.

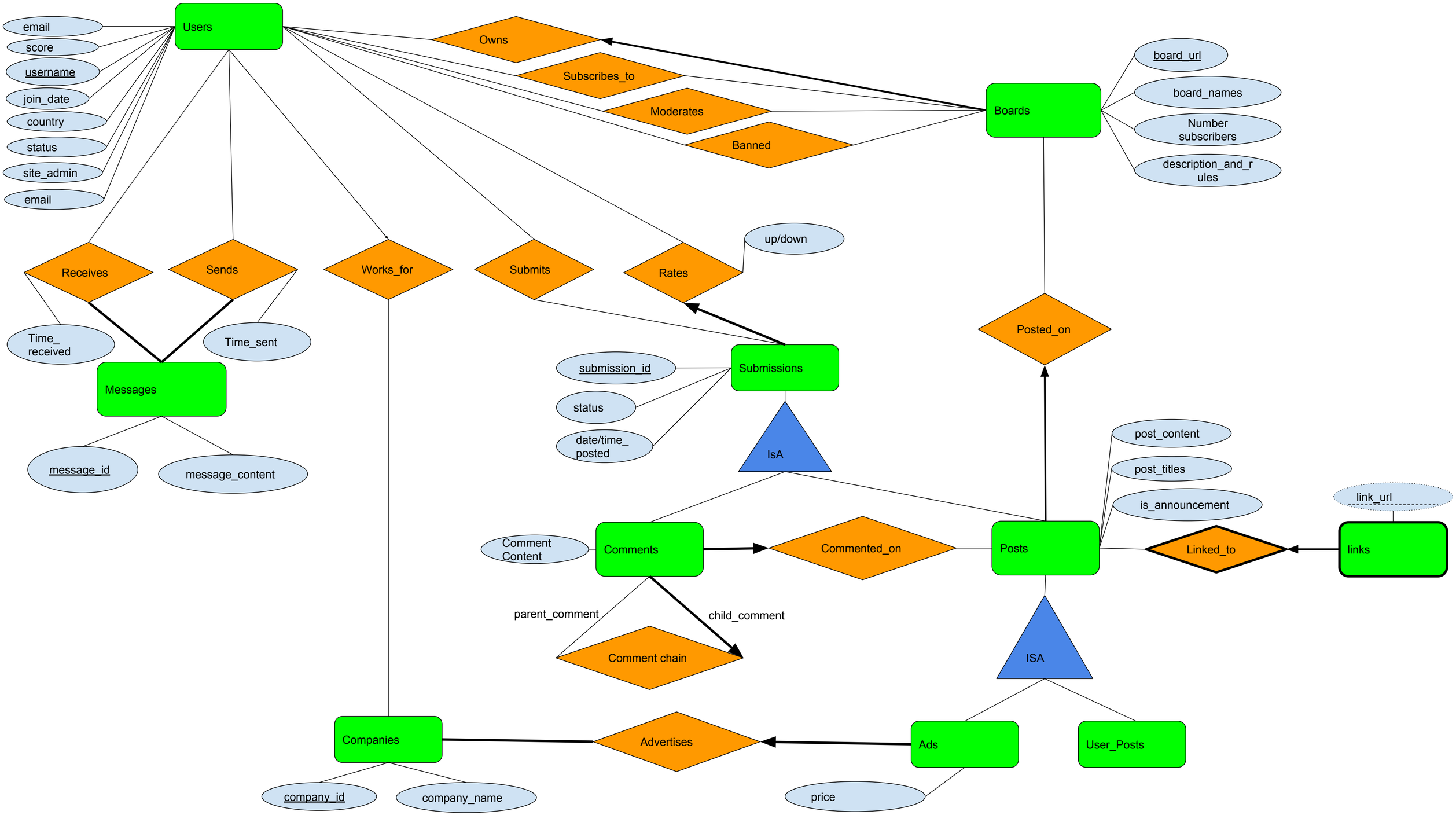
Algorithm description
The upvoting/downvoting mechanism assigns a score to each submission, and, along with the submissions datestamp, feeds it into an algorithm to rank each submission: popular submissions rise up to the front page of the website while less popular posts sink. Similarly, popular submissions rise to the top of their respective online board, while less popular posts sink. Similarly, popular comments within a submission rise to the top of the submission webpage, while less popular comments sink. Finally, each submission is exponentially penalized for its age to ensure the board stays current. The algorithm can be thought of as a modified sensitive moving-time average Google PageRank algorithm.

II. Relations
Entity sets
User(username,email,score,join_date,country,status,site_admin, email)
Messages(message_id, message_content)
Submissions(submission_id , status, date/time_posted)
Comments (submission_id, comment_content) (submission_id ref Submissions)
Posts (submission_id, post_content, post_title, is_announcement) (submission_id ref Submissions)
Boards (board_url, board_names,number_subscribers,description_and_rules)
Posts(submission_id, post_content, post_title, isAnnouncement) (submission_id ref Submissions)
User_posts(submission_id) (submission_id ref Submissions)
Ads(submission_id) (submission_id ref Submissions)
Companies(company_id , company_name)

Weak entity set
Links(submission_id, link_url) (submission_id ref Submissions)

Relationships
Banned(username, board_url) (username ref Users) (board_url ref Boards)
Moderates(username, board_url) (username ref Users) (board_url ref Boards)
Receives(username, message_id, time_received) (username ref Users) (message_id ref Messages)
Sends(username, message_id, time_sent) (username ref Users) (message_id ref Messages)
Submits(username, submission_id) (username refs Users) (submission_id ref Submissions)
Rates(username, submission_id , up/down) (username ref Users) (submission_id ref Submissions)
Works_for (username, company_id) (username ref Users) (company_id ref Companies)
Owns(board_url, username) (board_url ref Boards, username ref Users)
Subscribes_to(board_id, username) (board_id ref Boards, username ref Users)
Comment_chain(child_comment, parent_comment)
Commented_on(submission_id, submission_id) (submission_id ref Submissions, submission_id ref Submissions)
Posted_on(submission_id, board_url) (submission_id ref Submissions)

Advertises(submission_id, company_id) (submission_id ref Submissions)


email
score
username
join_date
country
status
site_admin
email
Users
Owns
Subscribes_to
Moderates
Banned
Boards
board_url
board_names
Number subscribers
description_and_rules
Receives
Sends
Works_for
Submits
Rates
up/down
Time_ received
Time_sent
Messages
message_id
message_content
submission_id
status
date/time_ posted
Submissions
IsA
Posted_on
post_content
post_titles
is_announcement
link_url
Comment Content
Comments
Commented_on
Posts
Linked_to
links
parent_comment
child_comment
Comment chain
ISA
Companies
Advertises
Ads
User_Posts
company_id
company_name
price