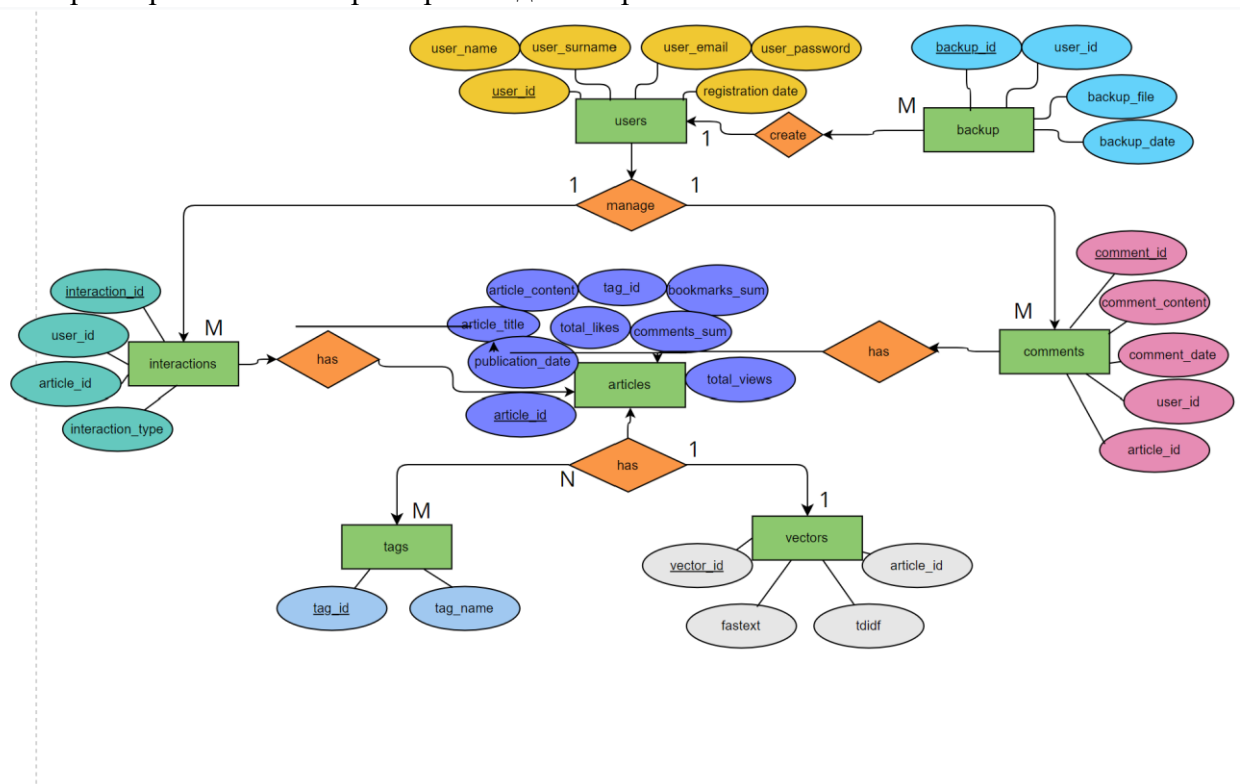


Планирую сделать проект, связанный с новостными постами, которые были получены результатом краулинга с новостного портала РИА Новости. Думаю, что для веб-интерфейса я выберу Flask для финального проекта, так как раньше я уже с ним работала. В самом приложении пользователи смогут регистрироваться, поэтому в базу данных будем сохранять имя, фамилию, почтовый адрес, пароль и дату регистрации. В базе данных также будут храниться новостные статьи и их атрибуты (название статьи, сам текст новости, дата публикации и теги). На самом сайте РИА Новости у каждой статьи имеется сразу несколько тегов, поэтому я в своей ER-диаграмме создала для них отдельную таблицу, которая будет соединена связью «многие ко многим» с таблицей новостей. Это позволит минимизировать избыточность данных и сделать систему более удобной для использования. Пользователь может делать поиск по новостным статьям и комментировать их, а также оставлять лайки, сохранять в закладки, учитывать просмотры. У каждой сущности есть свои индексы, которые помогут ускорить поиск. В ER-диаграмме я выбрала наиболее подходящие типы данных для каждого столбца, это сэкономит место в памяти и повысит производительность запросов.

В первом домашнем задании я сохраняла матрицы в pickle, и по ним происходил поиск, а в самом проекте думаю, что нужно будет создать отдельную таблицу, где будут храниться информационные векторные представления новостных постов. Это позволит избежать повторных расчетов векторов при каждом запросе.



Можно создать отдельную таблицу или в той же таблице с новостями можно добавить поля для хранения общего числа лайков, просмотров, комментариев и закладок. Это ускорит вывод при запросах. Также можно создать отдельную таблицу для резервного копирования, которая будет соединена с таблицей пользователей по индексу user_id.