

Есе по темі дипломної роботи для першої зустрічі з керівником

Татарин Микола, IT-92

8 травня 2023 р.

1 Тема

Застосунок для автоматичної категоризації художніх творів.

2 Проблема, що розглядається

На сьогоднішній день у світі існує величезна кількість текстової інформації (книг, статей, блогів, і т.д.). Тому знайти те, що тобі потрібно, іноді досить важко. Те, наскільки легко чи важко можна знайти той чи інший шматок інформації, залежить від декількох факторів. Один із таких факторів - це метадані, як от: заголовок, опис, ключові слова. Останні особливо корисні, бо дозволяють користувачу пошукової системи звузити пошук до тої теми, яка йому потрібна. В ідеалі ключові слова повинен надати автор твору, але іноді цього не відбувається. Або ж автор це зробив, але метадані були якимось чином втрачені. Тоді для їх відновлення хтось повинен буде перечитати цей твір, і знову скласти список ключових слів. Деякі платформи (наприклад, goodreads) дозволяють читачам колективно редагувати теги, проте було б непогано, якби це можна було робити автоматично.

3 Мета проекту

Розробити застосунок для автоматичної категоризації текстів (я планую обмежитись художньою літературою), і співставлення цих текстів з набором ключових слів (наприклад: «історія», «драма»). Використати для цього машинне навчання.

4 Що планується зробити

1. Розібрати вже існуючі моделі для вирішення цієї задачі і обрати оптимальну.

2. Підібрати існуючий датасет категоризованих за тегами текстів, або ж створити свій, використовуючи літературу з публічного домену.
3. Навчити на цьому датасеті обрану модель.
4. Створити простий для використання інтерфейс, що дозволить використовувати цю модель в довільних застосунках.
5. Створити веб або десктоп застосунок для автоматичної категоризації художніх творів.

5 Де це можна використати

Це можна використати для автоматичної категоризації і каталогування творів у онлайн каталогах (наприклад, [goodreads](#)) та бібліотеках.