

РОЗРОБКА КЛАСИФІКАТОРУ ЗОБРАЖЕНЬ ДЛЯ ШВИДКОГО ПОШУКУ У ВЕЛИКИХ БАЗАХ ДАНИХ

Філатов В.В., Філатова Г.Є., Поворознюк А.І.

*Національний технічний університет
«Харківський політехнічний інститут», м. Харків*

У даний час збору та обробці інформації приділяється велика увага. Істотний обсяг інформації містять стрічки новин електронних засобів масової інформації. Одним із способів обробки такої інформації є її кластеризація за смисловим навантаженням. Кластеризація новинної інформації проводиться як з урахуванням морфологічного аналізу текстів, так і графічного контенту. Таким чином, актуальним завданням є кластеризація зображень, що супроводжують текстову інформацію на різних веб-ресурсах, включаючи портали новин. Предмет дослідження: класифікатор зображень, що малочутливий до зростання кількості інформації в базах даних. Метою дослідження є підвищення продуктивності пошуку однакових зображень у базах даних, у яких швидкість додавання інформації досягає 10-12 тисяч зображень на добу, шляхом розробки класифікатора зображень. Розроблено загальну модель класифікатора зображень на основі контенту, в основі якої лежить перетворення зображень до його сигнатури (вектор значень ознак), що використовуються для однозначної класифікації зображення. У роботі пропонується використовувати двовимірне дискретне косинусне (ДКП) перетворення для обчислення сигнатури зображення. На основі дослідження властивостей ДКП при аналізі напівтонової та колірних складових зображень відповідно (колірна модель RGB) отримані математичні вирази обчислення компонентів сигнатури, інваріантної до поворотів та дзеркальних відображень зображення по будь-якій осі. Розроблено класифікатор зображень, що малочутливий до зростання кількості інформації в базах даних. Виконано дослідження властивостей розробленого класифікатора зображень, у результаті якого обґрунтовано перетворення всіх зображень до одного розміру 64×64 пікселів. При цьому їх сигнатури зберігають свою здатність ідентифікувати конкретні зображення. Проведені експерименти показали, що кластеризація інформації за зображеннями за допомогою розробленого класифікатора виявилася досить швидкою та маловитратною з погляду обсягів інформації та вимог до обчислювальної потужності. Подальші дослідження спрямовані на пошук оптимальних параметрів запропонованого класифікатора, а також вивчення можливості використання класифікатора для нечіткої кластеризації та нечіткого пошуку у базі даних.