

Актуалізація проблеми боротьби з плагіатом

Олександр Діхтяренко

Кафедра основ інформатики, Київський національний університет будівництва
та архітектури, УКРАЇНА, м. Київ, Повітрофлотський пр-т. 31;
e-mail: aleks.wq@gmail.com

The necessity of an automated system for determining the content of the documents presented and the main areas of possible application.

Ключові слова – пошук співпадінь в текстах, автоматичне визначення змісту, плагіат.

І. Постановка задачі

Останні роки відзначилися швидким просуванням Інтернету у найрізноманітніші сфери життя людей. Це все заслуга технологічного прогресу, яку важко недооцінити, адже тепер обмін інформацією став доступний як ніколи раніше. А такий фактор як відстань між джерелом інформації та її споживачем взагалі перестав відігравати серйозну роль. Зараз будь хто може створити свою сторінку на безкоштовному сервісі блогів чи навіть цілий сайт на одній з безкоштовних систем конструкторів. Інформація розмішена таким способом стає доступною з усіх куточків світу де тільки є Інтернет. Але тут проявляється і зворотня сторона медалі. Інформація, яка розмішена на сайті, стає легкодоступною для використання не тільки у ознайомчих цілях. Звісно добре коли ми використовуємо чийсь розробку замість того щоб проходити весь шлях повторно, головне не забути вказати посилання на автора та дізнатися на яких умовах він розповсюджує результати своєї інтелектуальної діяльності. Використовуючи вже відоме рішення для однієї з поставлених задач ми зекономимо час для розробки чогось принципово нового, чи вирішення проблем які ще не були вирішені до нас. Це все сприяє більш швидкому розвитку науки. Проблема в тому, що чужі рішення можуть використовуватися без згоди на те автора або й взагалі видаватися за свої. Лінь хоч і рушій прогресу, але також являється його гальмом. Навіщо робити роботу, якщо у вільному доступі розмішена вже готова, яку можна змінити незначним чином і видати за свою? Таким чином з'являється проблема перевірки авторства тексту та виявлення плагіату. Один з способів вирішення цієї проблеми, це створення систем автоматичного виявлення співпадінь. Складність значним чином залежить від самої мови. Коли в англійській мові задача трохи простіша, адже в ній є правила послідовності слів у реченні, то в українській та російській від зміни порядку слів зміст може і не

змінитися або навпаки: змінитися на зовсім протилежний. Також у слов'янських мовах слова мають багато морфологічних словоформ. Крім того в будь якій мові існують синоніми, що дозволяє оперувати одним поняттям за допомогою різних слів. Це все дуже ускладнює процес машинного визначеного змісту тексту. Значний внесок у проблемі перетворення тексту в зміст і навпаки зробив Мельчук І.А., зокрема розроблена ним (за сприяння ряду інших вчених) теорія «Зміст \Leftrightarrow Текст» [1]. Ця теорія розроблялась з метою створення автоматичних систем перекладу тексту.

II. Детальніше про проблему

Запозичення результатів чужої діяльності може здійснюватися у різний спосіб. Може як бути повне копіювання тексту, таблиць, формул та графічної інформації так і часткове чи вибіркоче. Ключовим моментом такого виду плагіату являється те що запозичена інформація не зазнає ніяких змін. Але запозичений текст може зазнавати змін. Найпростіше – це перестановка слів у реченні, речень в абзаці чи абзаців у документі, звісно так щоб не втрачався початковий зміст. Деякі фрагменти можуть бути викинуті чи навпаки – своя вставка посеред скопійованого тексту. Даний спосіб модифікації допомагає надати більш унікальний вигляд роботі в очах людини що її перевіряє, але не допоможе при перевірці унікальності комп'ютерною програмою, адже найпростіший алгоритм все одно знайде точні співпадіння. Тому коли у виконавця роботи є підозри, що його робота перевірятиметься на унікальність, він постарается приховати факт плагіату і від програми. Одним з найпростіших способів – заміна символів кирилиці на символи з латиниці, що мають ідентичний вигляд. Також може здійснюватися заміна слів на їх синоніми, або взагалі – повний рерайт (переробка тексту таким чином, що він структурно повністю відрізняється від початкового, але має той же зміст). Іншими, «дрібнішими» варіантами унікалізації можуть бути заміна цифр літерами (і навпаки), повних назв – абрєвіатурами (і навпаки).

Висновок

Набула актуалізація створення автоматизованих систем пошуку співпадінь в наукових роботах та матеріалах які розміщуються для вільного користування в Інтернеті.

Література

1. Мельчук И.А Опыт теории лингвистических моделей «СМЫСЛ \Leftrightarrow ТЕКСТ». – М.: Школа «Языки русской культуры», 1999, 346 ст.