



РОЗРОБКА СЛОВНИКА СИНОНІМІВ ДЛЯ СИНОНІМАЙЗЕРА НОВИНИХ ПОВІДОМЛЕНЬ

Тупікова Н.С.

*Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 707–63–60,
e-mail: elpys@rambler.ru*

В останні роки спостерігається динамічне зростання Інтернет-ресурсів, найважливішим видом яких є веб-сайти. Оптимізуючи сайт під пошукові системи великої уваги слід приділяти унікальності контенту. Унікальність контенту, в першу чергу, потрібна для позитивної оцінки сайту пошуковими машинами.

Одним із способів отримати унікальний контент для сайту новин є рерайтинг готових статей. Рерайтинг полягає в переписуванні вихідного тексту таким чином, щоб у результаті вийшов унікальний текст без втрати сенсу. Рерайтинг тексту проводиться, як правило, підбором синонімів до слів, або заміною фраз і цілих речень на смислові аналоги, в результаті чого пошукові системи сприймуть новий текст як унікальний [1].

У ході роботи були проаналізовані різні програми й он-лайн сервіси, які вже існують для реалізації рерайтингу, це програми підбору синонімів [2].

Синонімайзер – програма або сервіс для заміни слів у тексті на синоніми, що знаходяться в базі даних з метою видозміни тексту. Розроблялися синонімайзери виключно для допомоги тим, хто в тій чи іншій мірі пов'язаний з розробкою, управлінням та наповненням інтернет – ресурсів.

Синонімайзери можна розділити на автоматичні і ручні. Ручні пропонують користувачеві самому вибирати зі списку синоніми, які підходять. Автоматичні проробляють всю роботу без участі користувача.

Синонімайзери поділяються на серверні або он-лайн, тобто ті, що працюють виключно в Інтернеті, та десктопні, ті, що встановлюються на комп'ютер користувача [3].

Дія синонімайзера заснована на механізації роботи з текстами. Якість роботи синонімайзера безпосередньо залежить від словника синонімів, який використовує програма. Таким чином, словник синонімів для синонімайзера повинен бути створений коректно.

Словник синонімів для синонімайзера має вид $X | Y$, де X і Y можуть бути як будь-якою частиною мови, так і словосполученням або цілим реченням [4].

Аналіз електронних словників для синонімайзерів показав, що більшість слів мають не один синонім, а декілька. Це позначається в електронних синонімічних словниках наступним чином:

автомобиль|авто
автомобиль|автомашина
автомобиль|машина



Крім того, словники для синонімайзерів містять у собі усі форми слова. Так, наприклад, дієслово «краснеть» у словнику має вид:

краснев|багровев
краснел|багровел
краснела|багровела
краснеть|багроветь
краснея|багровея

Таким чином, можна помітити, що одне слово вживається у словнику декілька разів. Через це словник виходить об'ємним.

В наш час все прагне до спрощення і зменшення. Тому необхідно удосконалювати електронні словники для синонімайзерів, а саме зменшувати їх об'єм.

Якщо слово має декілька синонімів, ми пропонуємо писати їх через кому (,) після заголовного слова та вертикальної лінії (|). Це допоможе зменшити об'єм тому, що ми не будемо писати одне і те ж заголовне слово декілька разів, а також надасть користувачеві синонімайзера вибір синонімів. Це буде мати наступний вид: компанія | холдинг, підприємство

Деякі частини мови, наприклад, дієслова та прикметники, мають різні закінчення під час відмінювання. Надання всіх можливих форм слова робить словник об'ємним і громіздким. Для уникнення подібної ситуації, ми пропонуємо надавати в словнику лише слово у називному відмінку або інфінітив слова. Для того, щоб синонімайзер знаходив словоформи і замінював їх на синоніми, ми пропонуємо невраховувати закінчення слів. Це підвищить рівень релевантності.

Після користування синонімайзером, текст обов'язково повинен перевірятися рерайтером. Він повинен обрати потрібний синонім, якщо їх декілька, змінити закінчення обраного синоніма, а всі інші варіанти видалити.

При створенні електронного синонімічного словника було відібрано багато текстів з новинних сайтів з тематики саме інформаційних технологій. Крім слів, які стосуються тематики інформаційних технологій, словник містить слова загальної вживаності. Це такі слова як: «бесплатный» – «неоплачиваемый» тощо [5]. Об'єм створеного електронного синонімічного словника для синонімайзера складає близько 400 слів.

Список літератури

1 Рерайтинг: [Електронний ресурс]. – Режим доступу: <http://igroup.com.ua/seo-articles/rerajtynh/>

2 Програми для рерайтера: [Електронний ресурс]. – Режим доступу: http://mastersloga.ru/news/programmy_dlja_rerajtera/2013-04-14-47

3 Синонімайзер: [Електронний ресурс]. – Режим доступу: <http://animatika.ru/info/gloss/synonymizer.html>

4 Особливості побудови словника: [Електронний ресурс]. – Режим доступу: http://ua-referat.com/особливості_побудови_словника

5 Караванський С.Й. Практичний словник синонімів української мови: близько 15 000 синонімічних рядів / С.Й. Караванський. – К. : Кобза, 1995. – 472 с.