АНАЛІЗ ІСНУЮЧИХ СИНОНІМАЙЗЕРІВ ТА ОГЛЯД ПРОБЛЕМ ЩОДО ЇХ СТВОРЕННЯ

Юрьєва К.О.
Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська 79/2, (057) 707-63-60

Лексика мови постійно збагачується. Одне й те ж саме поняття характеризується з різних сторін, виділяються нові його ознаки. В лексемах може відбиватися ставлення мовців до предметів реального світу, їхні емоції та волевиявлення. Між словами, що позначають одне розгалужене поняття чи його компоненти або різні відтінки й оцінки, встановлюються близькі зв'язки. їх називають синонімічними, а слова — синонімами [1].

Синоніми – це слова, які мають близьке або однакове лексичне значення, але різні за написанням і звучанням. Вони допомагають уникати одноманітності мови, невиправданих повторень однакових слів. Синоніми також дають змогу більш точно сформулювати думку, роблять мову виразною. Такі слова виникають у мові постійно. Це обумовлено рядом причин. Однією з основних є прагнення людини знайти у вже відомих предметах, явища навколишнього світу якісь нові риси й відтінки, тобто поглибити і розширити існуючі поняття навколишньої дійсності. Нова додаткова ознака поняття називається новим словом, подібним або тотожним за значенням з уже наявними найменуваннями.

Нерідко виникнення синонімів зумовлено проникненням і освоєнням запозичених слів, близьких або тотожних за значенням українським, наприклад, літак - аероплан. Спостерігалося і зворотне явище, коли предмет, названий спочатку запозиченим словом, потім отримував синонімічне українське найменування, наприклад: гелікоптер - вертоліт.

Таким чином, можна сказати, що в українській мові багато синонімів, але процес появи нових синонімів не переривається. Синонімічні ряди постійно поповнюються [1].

Словник — упорядкований в алфавітному чи тематичному порядку список заголовних слів, лексикографічно опрацьованих. Таке довідкове видання містить впорядкований перелік мовних одиниць (слів, словосполучень тощо) з короткими їх характеристиками або характеристиками позначених ним понять або з перекладом на іншу мову. Словник є одним із засобів накопичення результатів пізнавальної діяльності людства, показником культури народу. Вони виконують інформативну та нормативну функції: являють собою універсальні інформаційні джерела для розуміння того чи іншого явища та найточніша консультація щодо мовних норм [2].

Словники синонімів мають досить цікаву функцію, яка здатна полегшити написання того чи іншого тексту з можливістю уникання одноманітності словникових форм. Гарним прикладом роботи з словниками синонімів є також синонімайзери.

Синонімайзер – це програма, яка дозволяє зробити з одного вихідного тексту велику кількість однакових за змістом текстів, які написані різними словами. Його зазвичай використовують для розмноження статей при просуванні сайту за допомогою бірж статей або каталогів статей, а також для розмноження опису сайту при реєстрації у великій кількості каталогів.

Попередньо, перед генерацією статті, кожній фразі тексту, де це тільки можливо, задається група синонімів і допустимих перестановок. Те ж саме робиться і зі словами усередині фрази.

Якщо провести хорошу роботу з синонімами в тексті статті і потім включити випадкову генерацію заданої кількості статей, то можна отримати значну кількість текстів, які з певною ймовірністю будуть унікальними для пошукових систем.

Для того, щоб запобігти появі занадто схожих статей внаслідок випадкового генерування, в синонімайзерах зазвичай передбачена функція видалення схожих статей за методом шинглів. Метод шинглів використовується пошуковими системами для визначення унікальності тексту сторінок сайтів. Шингл – це набір з певної кількості слів, що стоять один за одним у тексті [3].

Основні проблеми й складності використання синонімайзера укладається в наступному:

1) здійснюваний добір синонімів повинен проводитися в рамках контексту словосполучення;

2) в українській мові відсутній порядок слів у реченні, що суттєво ускладнює синтаксичний аналіз тексту;

3) для полегшення читання тексту в ньому не повинно бути використана безліч однакових слів;

4) повинна виконуватися закономірність Зіфра.

Для подальшого аналізу роботи з текстом, отриманого за допомогою синонімайзера необхідно здійснювати порівняння текстів на унікальність частинами й вибирати найбільш унікальний текст, який бажане вичитувати для виправлення помилок у закінченнях слів.

Отже, актуальними залишаються проблеми розробки сучасних комп'ютерних технологій у лексикографії. Особливістю електронних словників є те, що вони неможливі без технічної підтримки. Зокрема, завдання які виконують такі видання формують їх переваги, через що вони відповідають читацьким потребам.

Список літератури:

1. *Королева Н.В.* Проблема синонимии как одна из важных проблем лексикологии / Н.В. Королева, Е.В. Артамонова // Вестник КАСУ. – 2009. – №2. – с. 91-97

2. *Реформатский А.А.* Введение в языковедение: Учебник для вузов/ А.А.Реформатский; Под ред. В.А. Виноградова - 5-е изд., испр. – М.: Аспект Пресс. – 2006. – 536с.

3. *Кронгауз М.А.* Семантика. Учебник для студ. лингв. фак. высш. учеб. Заведений / М. А. Кронгауз. - 2-е изд., испр. и доп. – М.: Академия. – 2005. –352 с.