

## **КОРПУС ЕКОНОМІЧНИХ ТЕКСТІВ ЯК ДЖЕРЕЛО УКЛАДАННЯ ТЕРМІНОЛОГІЧНОГО СЛОВНИКА**

**Бабурніч Д.Ю.**

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

Результати аналізу корпусів текстів, сформованих для окремих предметних областей, мають високу прикладну цінність. Спеціальні корпуси текстів і витягнуті з них дані затребувані як в науково-технічній лексикографії (при складанні термінологічних словників, класифікаторів, рубрикаторів), так і в сфері автоматичної обробки текстів (при автоматичному індексуванні і реферування документів, автоматичної класифікації і кластеризації документів, в інформаційному пошуку і машинному перекладі).

Проблему створення таких корпусів було розглянуто у роботах багатьох авторів, однак питання створення корпусу спеціальних текстів на даний момент досі залишається нерозв'язаним, що робить наше дослідження актуальним.

Корпус текстів розуміють як значний за обсягом, представлений в електронному вигляді, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, створений для вирішення конкретних лінгвістичних завдань.

Спеціалізований корпус – це жанрово специфічний корпус, який має на меті відобразити певну підмову

Для формування корпусу текстів спеціалізованої спрямованості було виділено окрему галузь економіки – екологічну економіку. Для наповнення корпусу екологічної економіки було використано статті новинних англomовних сайтів за категорією Ecological Economy.

Після вибору джерел всі тексти, отримані із різних джерел були представлені у вигляді txt-файлів, та пройшли обробку, в ході якої відбувалося видалення нетекстових елементів (рисунків), видалення із тексту переносів, забезпечення одноманітного написання тире та ін. Корпус нараховує 71 файл і 62293 слововживань.

Після вибору джерел тексти було розмічено за допомогою онлайн версії системи морфологічного тегування CLAWS.

Сам корпус менеджер було реалізовано за допомогою технології JavaFX.

Розроблений корпус вирішено використовувати для наповнення термінологічної бази словника. Розмічений корпус має такі переваги, як простота використання і багатофункціональність, що дозволяє значно розширити можливості корпусу при виконанні поставлених завдань. Корпусний менеджер виконує такі завдання, як пошук конкретних словоформ та вилучення термінів із текстів. Використання корпусу розмічених текстів у програмі вирішує завдання пошуку термінології в тексті та її вилучення для наповнення бази майбутнього словника.