



ДОСВІД ЗАСТОСУВАННЯ ЗНАНЬ З ПРИКЛАДНОЇ ЛІНГВІСТИКИ У СТВОРЕННІ СИСТЕМИ, ЩО ВИКОРИСТОВУЄ МЕТОДИ OPINION MINING TA COREFERENCE RESOLUTION

Терещенко В.І.

Search Concepts AG

4052 Basel, Switzerland, 59 Gartenstrasse,

phone: +41 61 715 95 06

e-mail: contact@searchconcepts.io

В епоху розвитку інформаційних і комунікаційних технологій та їх впровадження в усі сфери людської діяльності, вирішення лінгвістичних проблем та завдань набуває особливої актуальності. Це пов'язано у більшій мірі зі збільшенням кількості інформації, яку можна знайти в мережі Internet, а також з прагненням людства до прискорення якості та швидкості обробки цієї інформації. Результатом таких прагнень є створення лінгвістичних систем обробки та аналізу інформації. І одними з перших проблем, на які натрапляють науковці під час створення таких систем, є Opinion Mining та Coreference Resolution [1].

Coreference – відношення між займенником (анафором) та його антецедентом, за якого і займенник, і його антецедент співвідносяться з одним і тим же предметом об'єктивної дійсності [2]. Нижче наведено приклад речення з кореферентним зв'язком:

«Чистий прибуток Visa[antecedent] склав 1,27 млрд дол, йдеться в повідомленні компанії. За результатами II кварталу 2012 року вона[anaphor] повідомила про прибуток в 1,29 млрд дол.»

Coreference Resolution – процес встановлення кореферентних зв'язків в реченні або між реченнями [3].

Opinion Mining – (Sentiment Analysis, аналіз тональності тексту) метод контент-аналізу в прикладній лінгвістиці, призначений для автоматизованої обробки емоційно забарвленої лексики або емоційної оцінки автора, яка безпосередньо відноситься до об'єктів мови, про які йдеться в тексті.

Результатом розв'язання обох завдань може бути система, здатна до якісної обробки та аналізу природної мови. Створенням саме такої системи займається компанія «Search Concepts AG», з центральним офісом в м. Базель, Швейцарія. Створюваний компанією програмний продукт під назвою «News Board» має на меті автоматизований аналіз та обробку поширюваних в мережі internet новин фінансово-економічної тематики. Гордістю компанії є те, що вже сьогодні «News Board» може робити аналіз текстів на англійській, французькій, німецькій, російській та іспанській мовах. Найбільшу користь від використання програми отримують люди, що цікавляться станом справ в економіці певної країни чи компанії, або ж ті, що бажають інвестувати капітал в ту чи іншу галузь чи компанію. Засновник компанії, Hans-Peter Gier як ніхто інший знайомий

з проблемою автоматичної обробки фінансово-економічної інформації і говорить про те що зазвичай, для того щоб дізнатися реальний стан справ компанії, людина може витратити дуже багато часу та сил на пошук та аналіз новин. Проте, з цією системою той же результат можна отримати за декілька хвилин.



Рис. 1. – Скріншот програми News Board by Search Concepts

Незважаючи на те що центральний офіс компанії знаходиться в Швейцарії, розробка програми починалася в Україні зі співпраці кафедри ІКС та новоствореної компанії Search Concepts AG. Зокрема, великий вклад в розробку лінгвістичної основи програми зробили професор кафедри інтелектуальних комп'ютерних систем НТУ «ХП», доктор технічних наук Хайрова Ніна Феліксівна, а також студенти кафедри: Тетяна Булатнікова, Сільветте Луда, Альона Ковальова, Світлана Петрасова та Віталій Терещенко.

В ході співпраці було створено декілька програмних продуктів для вирішення окремих лінгвістичних задач. Такими є програма для встановлення кореферентних зв'язків в російськомовних текстах економічної тематики та програма для швидкого та повного пошуку перекладних еквівалентів в мережі Internet. Всі вони розроблялись як окремі модулі для основної програми, проте можуть використовуватись і самостійно.

Таким чином, результатом плідної співпраці між компанією Search Concepts AG та кафедрою інтелектуальних комп'ютерних систем НТУ «ХП» став програмний продукт, що дозволяє проводити сентімент-аналіз новин фінансово-економічної тематики на декількох мовах. Створений програмний продукт користується досить великою популярністю у користувача. Розроблене в ході співпраці програмно-алгоритмічне забезпечення є винятковим прикладом використання теоретичних навичок та знань з прикладної лінгвістики на практиці.

Список літератури

1. Clark J.H., Gonzalez-Brenes J.P. "Coreference Resolution: Current Trends and Future Directions", 2008. – pp. 11-16.
2. Hobbs J. R. "Pronoun resolution" – California, 1976. – pp. 10-18.
3. Скатов Д. "Разрешение кореференции: обзорная экскурсия" – Н. Новгород, ДИКТУМ, 2012.