

відноситься висока експресивність (вживання експресивних прикметників, іменників, прислівників, дієслів), посилення на своїх знайомих/ родичів при вихвалюванні себе, вживання риторичних фігур та тропів, вихвалювання “другої половинки”, улюблених речей/ місць/ роботи тощо, єдина для жінок та чоловіків цінність – почуття гумору. Оголошення жінок більш насичені яскравими образами та детальним описом.

Подальшим напрямом дослідження лаудативного дискурсу може бути вивчення впливу інших соціолінгвістичних чинників на лаудацію.

Література

Бігунова Н.О. Рівнева структура мовленнєвих актів похвали і схвалення [Електронний ресурс] / Н. О. Бігунова // Вісник Житомирського державного університету імені Івана Франка. – 2007. - № 32. – С. 117-119. – Режим доступу до статті: <http://eprints.zu.edu.ua/1564/> *Демьянков В. З.* «Теория речевых актов» в контексте современной лингвистической литературы (Обзор направлений) [Электронный ресурс] / В. З. Демьянков. – Режим доступа к статье: <http://www.infolex.ru/P036.html>. *Фоменко О. С.* Лаудативный дискурс: прагматичний та лінгвокультурологічний аспекти (на матеріалі англійської мови) / Олена Степанівна Фоменко, Марина Дмитрівна Рижикова // Мова і культура. (Науковий журнал). – К.: Видавничий дім Дмитра Бураго, 2009. – Вип. 12. – Т. VII (132). – 360 с. – С. 5-12. *Fox, K.* Watching the English: The Hidden Rules of English Behavior / Kate Fox. – Nicholas Brealey Publishing, 2008. – 432p. *Holmes J.* Women, Men and Politeness. – L.: Longman, 1995. – 254 p.

РІПЕЦЬКА І.Є.

(*Національний університет «Львівська політехніка»*)

ЕФЕКТИВНІСТЬ ЗАСТОСУВАННЯ ІНТЕРТЕКСТУАЛЬНОГО ПІДХОДУ ДО АВТОМАТИЗОВАНОЇ КЛАСИФІКАЦІЇ НАУКОВИХ ТЕКСТІВ

У статті окреслено сучасні підходи до автоматизованої класифікації наукових текстів. Пропонується інтертекстуальний підхід до класифікації досліджуваних текстів та підтверджується ефективність його застосування.

Ключові слова: (автоматизована) класифікація текстів, контент-аналіз, інтертекст, маркованість, цитата, універсальна десяткова класифікація.

Рипецкая И.Е. Эффективность применения интертекстуального подхода к автоматизированной классификации научных текстов. В статье рассмотрены современные подходы к автоматизированной классификации научных текстов и доказываются эффективность применения интертекстуального подхода.

Ключевые слова: (автоматизированная) классификация текстов, контент -анализ, интертекст, маркирование, цитата, универсальная десятичная классификация.

Ripetska I.E. Efficacy of intertextual approach applied to automated classification of scientific texts. The paper suggests an insight into up-to-date approaches to automated text classification. An intertextual approach has been suggested and its efficacy has been proved. *Key words:* (automated) texts classification, content analysis, intertext, labeling quote, Universal Decimal Classification.

В умовах сучасного технологічного поступу та глобалізації змінюються способи отримання, пошуку та представлення інформації. Сьогодні важко уявити без електронних книг та бібліотек, навіть світова наукова думка побоюється витіснення паперових видань електронними. В Україні Національна бібліотека імені В.І. Вернадського розміщує збірники фахових видань у електронному варіанті і надає до них вільний доступ без оплати та реєстрації [<http://www.nbuv.gov.ua/portal/>]. Така вседоступність до наукових матеріалів сприяє поширенню знань та збільшує інтерес до науки. Проблематичним є величезне нагромадження наукових текстів без відповідної їх класифікації, що ускладнює ефективний пошук потрібного.

Мета цього дослідження – здійснити підбір таких способів класифікації електронних текстів, щоб кількість лексичних одиниць, які потрібно аналізувати, звести до мінімуму і при цьому забезпечити високу точність визначення класу тексту. **Об’єкт дослідження** – статті українських фахових видань в галузі комп’ютерних наук та технологій. Ця галузь в універсальній десятковій класифікаційній системі (УДК) має кодове представлення 004 і ділиться на підкласи. **Завдання дослідження** – розроблення алгоритмів ефективною автоматизованою класифікації наукових (АКТ) статей для фахових видань та створення програмного продукту АКТ.

На початку ХХ століття виникли перші системи класифікації текстів, як от класифікаційна система Дьюї, універсальна класифікаційна система тощо. Кожна з них має власні, ефективні для паперових видань підходи до класифікації текстів [Bliss 1939], та електронні тексти потребують дещо інших принципів систематизації.

Традиційні підходи до класифікації текстів передбачають аналіз всього тексту, анотації чи заголовку. Наявні програми автоматизованої класифікації текстів використовують пострічковий аналіз тексту/анотації/заголовку, тобто кожна лексична одиниця перевіряється у тезаурусі та «чорному списку» (Рис. 1) і, таким чином, встановлюється клас всього тексту [Svenonius 2000]. У цій статті тезаурус

визначається як перелік значимих термінів із приміткою, до якого класу вони належать. «Чорний список» – перелік лексичних одиниць, які не є значимими при встановленні класу тексту і можуть бути не враховані.

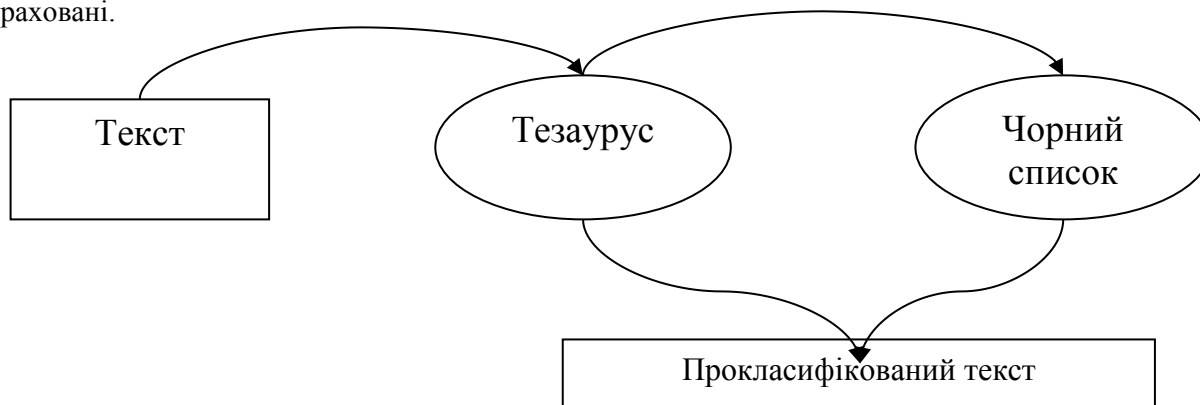


Рис.1. Алгоритм автоматизованої класифікації текстів

Для дослідження обрано науковий текст, а саме 25 статей фахових видань України (УДК 004) за 2008-2010 роки. Науковий текст – це вербалізоване наукове знання, а первинний науковий текст (монографія, стаття) націлений, в першу чергу, на представлення, наукового знання як результату пізнавально-творчої діяльності науковця, це нове знання виникає завжди як продовження уже відомих, «чужих» текстів як реакція на них [Чернявская 2009, с. 202-203].

Першим етапом дослідження був автоматизований контент - аналіз всієї статті, при цьому було встановлено найчастотніші слова і їх відсоток у текстах. У процесі дослідження виявилось, що кількість лексичних одиниць у заголовку і анотації недостатня для ефективної автоматизованої класифікації наукового тексту, тому програмі АКТ потрібно задати маркери для виявлення окремих значимих частин тексту. Серед значимих частин тексту особливу роль відіграє інтертекст. Інтертекст - це вся сукупність текстів, що репрезентуються в певному творі, незалежно від того, чи співвідносяться ці тексти з твором *in absentia* (наприклад, у випадку алюзії) або проникає в *in praesentia* (як у випадку цитати) [Пьєге-Гро 2008, с. 48].

У межах наукової комунікації потрібно дотримуватися чіткості і однозначності при поясненні теорій, гіпотез, а границя між своїм і «чужим» текстом повинна бути чітко маркована. В науковому дискурсі інтертекст легко виявити через експліцитні маркери.

Маркованість розглядається як наявність на фонетичному, лексичному, синтаксичному, стилістичному, композиційному рівні лінгвістичних сигналів міжтекстового діалогу [Чернявская 2009, с. 197]. Прикладами експліцитних маркерів інтертексту в науковому дискурсі є цитати, посилання, зноски, ремарки. Тому програма автоматизованого виявлення інтертексту у науковому дискурсі шукатиме графічні символи його представлення.

В результаті аналізу 25 статей фахових видань з галузі УДК 004, встановлено, що найчастіше інтертекст трапляється у вигляді цитат. Цитату можна розглядати як дослівне відтворення фрагменту претексту, вона передає чужий смисл за допомогою чужих слів і має максимально виражену автономність в новому текстовому оточенні [Чернявская 2009, с. 205]. Спеціальні експліцитні маркери вираження цитат: [цифра], [цифра:цифра], [цифра:с.цифра], [цифра:р.цифра], [цифра - цифра], [цифра, цифра]. Для програми АКТ важливо задати точні маркери визначення цитат, оскільки квадратні дужки використовуються з різною метою, наприклад, для позначення формул [max, min] та ін.

Інтертекст визначається як текст у тексті, у науковій статті є окрема структурна частина, яка відповідає таким критеріям - опис бібліографії, який можна розглядати як інтертекст у тому сенсі, що ключові поняття, які є в переліку списку літератури, відсилають до більших за обсягом матеріалів. Аналіз статей з галузі УДК 004 виявив, що ключові слова з усього тексту належать до тієї ж категорії, що й слова з опису бібліографії (з заголовків книг, підручників, монографій та ін.). Науковці часто користуються іноземними джерелами, таким чином, бібліографічний опис часто містить позиції різними мовами, зокрема у цьому дослідженні – англійською та російською. У такому випадку тезаурус, приєднаний до програми АКТ повинен містити ключові терміни як мінімум трьома мовами, а саме українською, англійською та російською. Ще одним аспектом, який ускладнює використання бібліографії для АКТ – інтернет посилання, оскільки сама електронна адреса не містить термінів, на основі яких можна віднести текст до певної категорії чи підкатегорії.

Другим етапом дослідження був ручний відбір інтертексту з досліджуваних статей і його автоматизований контент-аналіз програмою Concordance 3.2.0.0 Аналіз довів ефективність використання інтертексту як вхідних даних для АКТ. Ефективність стосується швидкості (табл. 1) і точності. Швидкість аналізу інтертексту в 4 рази більша ніж всього тексту, що дуже важливо для програмного продукту, оскільки контент-аналіз – лише один з етапів встановлення класу тексту.

Що стосується точності, то на прикладі однієї статті [<http://www.nbuv.gov.ua/portal/>] класу УДК 004. 21 (алгоритми) можна показати ефективність застосування АКТ для наукових текстів. За допомогою програми

Кількість лексичних одиниць у всьому тексті (середнє значення)	Швидкість контент-аналізу всього тексту (сек) (середнє значення)	Кількість лексичних одиниць в інтертексті (середнє значення)		Швидкість контент-аналізу інтертексту (сек) (середнє значення)
		число	%	
2220	12,5	314	19	3

Табл. 1. Ефективність контент – аналізу тексту/інтертексту (швидкість)

Concordance 3.2.0.0 аналізуємо статтю, визначаємо кількість слововживань кожної лексичної одиниць; зі всього переліку вибираємо тільки ті, які представляють певну галузь. Ті самі дії виконуємо над інтертекстом і порівнюємо результати (табл.2)

Значимі лексичних одиниць у тексті	Відсоток значимих лексичних одиниць у тексті (%)	Відсоток значимих лексичних одиниць у інтертексті (%)
Координація	9,1	3,6
Багаторівневий	5,6	2,7
Елемент	4,6	1,7
Ієрархічний	2,8	1,4
Керувати	2,3	3,3
Алгоритм	2,29	1,8

Табл.2. Ефективність контент – аналізу тексту/інтертексту (точність)

Аналіз виявив, що найчастотніше слово у всьому тексті є найчастотнішим і у інтертексті. Така схема діє для кожної проаналізованої статті з невеликою похибкою, яка не впливає суттєво на результат.

Ручний та автоматизований аналіз інтертекстів наукових статей доводить, що ефективність АКТ зростає, якщо програмний продукт аналізуватиме заголовок, анотацію та інтертекст для встановлення класу всього тексту.

Література

Пьеге-Гро Н. Введение в теорию интертекстуальности; [Пер. с фр.]/ Натали Пьеге-Гро / Общ. ред. Г.К.Косикова. – М.: Издательство ЛКИ, 2008. – 240 с. Чернявская В.Е. Лингвистика текста: Поликодовость, интертекстуальность, интердискурсивность. Учеб. пособие.— М.: Либроком, 2009. — 284 с. Bliss H. E. The Organization of Knowledge in Libraries: and the Subject-Approach to Books, 2nd ed. — New York: H. W. Wilson, — 1939. Svenonius E. The intellectual foundations of information organization. — Cambridge: MIT Press, — 2000. <http://www.nbuv.gov.ua/portal/>

САВЧУК В. І.

(Київський національний лінгвістичний університет)

ТИПОЛОГІЯ ЗВЕРНЕНЬ ПРЕЗИДЕНТІВ США

У статті розглядається типологія президентських звернень і визначаються найбільш вагомі промови у плані їхньої впливовості на сучасний американський соціум. Крім того, виділяються типи аудиторії в залежності від характеру президентського звернення.

Ключові слова: президентське звернення, політичний дискурс, інтернальний і екстернальний типи комунікації.

Савчук В. І. Типология обращений президентов США. В статье рассматривается типология президентских обращений и определяются наиболее существенные речи в плане их воздействия на современный американский социум. Кроме того, выделяются типы аудитории в зависимости от характера президентского обращения.

Ключевые слова: президентское обращение, политический дискурс, интернальный и экстернальный типы коммуникации.

Savchuck V. I. Types of U.S. presidents' addresses. The article investigates the types of presidential addresses and defines the most significant speeches in the sense of their impact on the contemporary American society. Also it differentiates the audience according to the type of presidential address.

Key words: presidential address, political discourse, internal and external types of communication.

Президентство є чимось відмінним від інших явищ у сфері політичної діяльності. Так, наприклад, під час зіткнення нації з кризою, президент має можливість говорити з національної трибуни, збираючи з'їзд або звернутися до масового слухача, чого не може зробити жоден інший політичний діяч. Звертаючись до нації, президент має можливість не лише заспокоїти настрої суспільства, але й безпосередньо позиціонувати себе та свою діяльність певним чином, що відображається в доборі специфічних комунікативних стратегій і тактик у його вербальній поведінці, що, у свою чергу, визначає специфіку використання мовних одиниць усіх рівнів. Звідси цілком очевидно є доцільність проведення