

УДК 004.44:02



Олександр Кузнєцов,

науковий співробітник відділу довідково-консультаційної допомоги
Національної юридичної бібліотеки

Дослідження динаміки змін термінів у бібліотечній справі

Наведено методику визначення змін термінів у наукових документах для відслідковування змін напрямів у бібліотечній справі.

Ключові слова: бібліотечна сфера, наукові документи, терміни, аналіз текстів, "стоп-слова", електронні видання.

Постановка проблеми. Сучасний етап розвитку бібліотечної справи в Україні характеризується динамічним оновленням теоретичних концепцій. З'являються нові терміни, розробляються інноваційні напрями та форми роботи книгозбірень, практичну діяльність яких визначають процеси інтеграції, розвитку прогресивних інформаційних технологій, соціокультурної діяльності. Вітчизняні бібліотеки, перебуваючи під помітним впливом світових тенденцій розвитку інформаційного суспільства, все активніше долучаються до процесів глобалізації та уніфікації інформаційно-бібліотечної сфери [1].

Розвиток сучасної бібліотечно-інформаційної термінології позначається потужним впливом електронного інформаційного середовища на українські бібліотечно-інформаційні поняття та терміни — це вже реальність, яка потребує уваги [2]. Впровадження комп'ютерних технологій сприяє запозиченню з англійської мови українською архівною, й особливо бібліотечно-інформаційною, галуззю групи понять на позначення цієї діяльності [3].

Відкритість сучасного світу, глобалізація, спрощення міжнародних контактів, розвиток телекомунікацій викликають масове проникнення запозичень в українську мову. На її словниковий склад впливає й неконтрольована демократизація, поширення жаргонної лексики та її експансія в літературну мову. Доцільно зазначити, що часто це відбувається за посередництва російської мови: спочатку термін іншомовного походження вливається в російську лексику, а згодом уже в перекладі з неї стрімко запозичується в українську. Іншомовні слова, що увійшли до української мови, починають функціонувати за законами її лексико-граматичної системи, утворюються за аналогією й існуючими правилами [4]. Отже, зміна термінів та напрямів має певну залежність, деякі пояснення нових слів у статтях носять несистемний характер і не відповідають викликам сучасності. Тому треба знайти методику визначення змін напрямів у бібліотечній справі через зміни термінів з використанням новітніх технологій.

Аналіз останніх досліджень. Поява нових термінів знайшла своє відображення у діяльності бібліотекарів: друкувалися не тільки статті з визначенням та правилами правопису нових слів, а й тлумачні словники з їхнім поясненням [5]. Але все це було у площині традиційних технологій видавництва — без автоматизації бібліотечних процесів. На початку розвитку інтернет-технологій АБІС були єдиним інструментом створення пошукових систем в інформаційному середовищі [6]. Технологія пошуку нерозривно пов'язувалась із аналізом термінів тексту. Однією з перших спроб автоматизації роботи з ними було створення

розробниками АБІС інструментарію опрацювання ключових слів — термінів — для автоматичного визначення рубрики [7], але робота загальмувалась. Проте у пошукових системах продовжували вдосконалювати методи дослідження термінів [9, с. 70—71; 10].

У принципах пошуку на перше місце вийшли семантичні алгоритми, а не організація логічного пошуку з індексацією баз даних [9, с. 60]. Компанія Google однією з перших розробила цю технологію, що забезпечило їй пріоритетне місце серед пошукових служб. Адаже не тільки бажання лідерства змусило її застосувати нові алгоритми: за оцінками зарубіжної преси, дев'ять із десяти великих американських компаній витрачають щорічно в середньому мільйон доларів на моніторинг діяльності конкурентів. Рішення, прийняті на підставі відібраної та проаналізованої інформації, дають змогу побудувати корпоративну стратегію таким чином, щоб окупили витрати й отримати прибуток від цих капіталовкладень [11].

Мета статті — запропонувати методику виявлення тенденцій змін основних напрямів діяльності у бібліотечній справі на основі різних тлумачень термінів у авторефератах різних років.

Викладення основного матеріалу. Для вирішення цієї проблеми була розроблена нова методика, що базується на корпусному підході. Її основні особливості:

- використання комп'ютерних методів опрацювання мовних даних;
- залучення як джерела вивчення репрезентативного корпусу текстів обраної проблемної сфери;
- створення словника термінів і його застосування для опрацювання корпусу текстів;
- суміщення контент-аналізу з методами опрацювання даних, що використовуються у галузі інформаційного пошуку. Надзвичайно важливо, що проект комп'ютерної системи тематичного дослідження зміни термінів певним чином вирішує проблему об'єктивізації його результату, оскільки використання динамічно утвореного корпусу не залежить від актуального стану когнітивної системи експерта [11].

Програма здійснює опрацювання документів за всіма словами, з яких потім створює частотні масиви для подальшого оброблення. Зрозуміло, фахівець може деяким чином впливати на цей процес, відбираючи ті слова, які становлять інтерес для конкретного етапу аналізу [12]. Для вивчення динаміки змін термінів було взято два масиви документів (авторефератів) з бібліотечної справи за 2000 та 2011 роки у кількості восьми та шести назв відповідно та досліджено за допомогою програми "Аналіз текстів" (рис. 1).

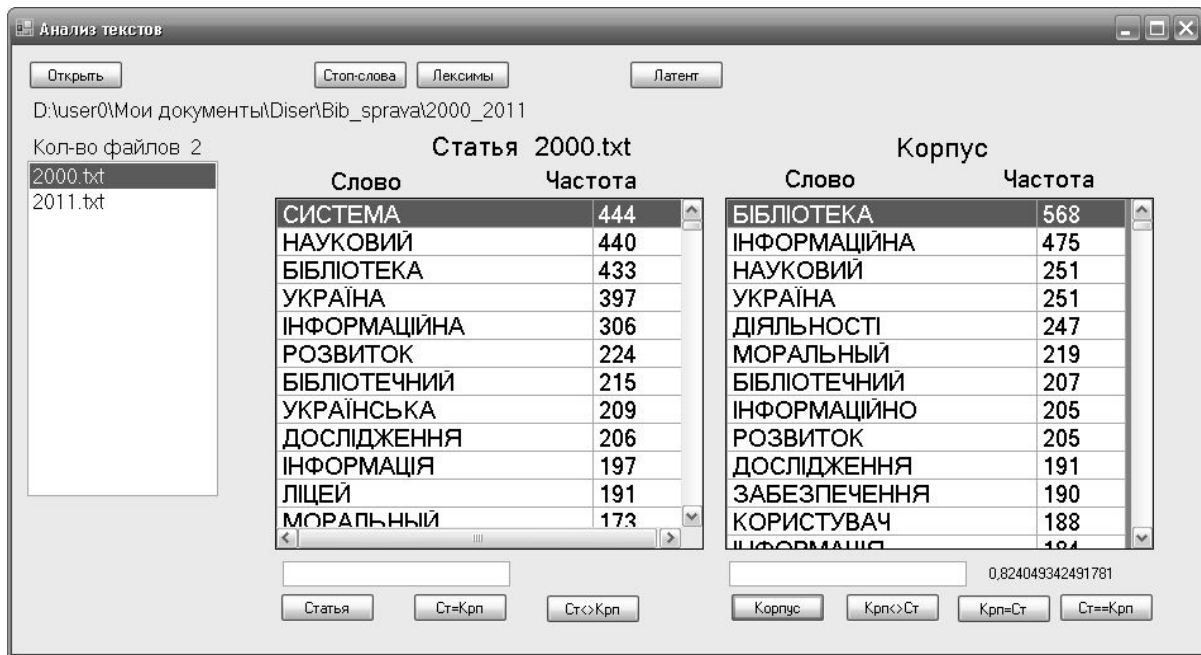


Рис. 1. Динаміка змін термінів

Всі тексти було об'єднано за роком випуску в два файли: 2000.txt та 2011.txt, і опрацьовано таким чином:

1. З тексту статті видалено усі службові символи.
2. Вилучено "стоп-слова": англійські, російські та українські. "Стор-слова" — це слова, які не несуть інформаційного навантаження щодо змісту статті, тобто на них можна не звертати увагу.

База даних "стоп-слів" дає змогу позбавитись не тільки беззмістовних слів, а й редакторських, граматичних авторських помилок, ненормативної лексики тощо. До цієї категорії належать сполучники, прийменники й інші слова-зв'язки, які не розпізнаються і не сприймаються пошуковими роботами. При цьому треба враховувати, що набір "стоп-слів" залежить від тематичного напрямку. У нашому випадку вони підібрані до "загальної тематики" та "бібліотечної справи" [13].

3. Нормалізовано ключові слова за словником синонімів: замінено всі слова-синоніми або словоформи одним словом також трьома мовами, яке вважається нормалізованим. Наприклад, для слова "бібліотека" є відповідний набір словоформ: "бібліотеки", "бібліотеку", "бібліотекою", "бібліотеці", "бібліотеках", "бібліотекам", "бібліотек", "бібліотеками". Програма замінює будь-яку словоформу або її синонім (якщо він є) на слово "бібліотека". Заміна синонімів та словоформ значно підвищує статистичну точність смислового змісту дослідних документів, тобто отримано чіткішу смислову статистику.

4. Створено такі частотні масиви слів: а) частотний масив слів 2000.txt та 2011.txt; б) частотний масив слів 2000.txt, які є і в 2011.txt; в) частотний масив слів 2011.txt, які є і в 2000.txt; г) частотний масив слів 2000.txt, яких нема в 2011.txt; д) частотний масив слів 2011.txt, яких нема 2000.txt.

У статті аналізується тільки змістовний аспект значення слів, а експресивно-емоційні оцінки, конотації¹ та відповідні їм поняття не розглядаються [14].

За результатом дослідження загальних слів двох масивів: б) і в) — слів та їх відповідних частот вживання, отри-

мали коефіцієнт схожості 0.82405, за яким можна зробити висновок про єдиний тематичний напрям. Цей показник є очікуваним, бо вивчалися тексти одного тематичного напрямку — бібліотечна справа. Масиви а) та б) — основні, з них отримано решту.

Аналіз частотних масивів слів г) і д) дає право зробити висновки про минуле та майбутнє бібліотечної справи. Так, слова "Ліцей", "Просвіта", "Рубрикатор", "АБІС", "Бібліографічні", які є в 2000 і яких нема в 2011 році, нагадують про популярність ліцеїв — платну альтернативу середнім школам; актуальність рубрикаторів, бо з появою нових електронних технологій виникла потреба в оновленні останніх; "АБІС" — розвиток автоматизації бібліотечних процесів і пов'язане з ним слово "Бібліографічні" — "бібліографічні записи", які згадувались скрізь.

Слова "Дистанційний", "Мультимедійний", "Університетська" є тільки в документах 2011 року, що свідчить про пріоритет у майбутньому дистанційних методів навчання на мультимедійній основі, починаючи з університетів. Ці матеріально потужні заклади першими оволодіють (якщо вже не володіють) новими електронними технологіями. І для цього є економічні підстави — платне навчання. Сайти деяких університетів вже мають відеолекції основних предметів, інтернет-варіанти оплати дистанційного навчання і навіть електронні дипломи.

Впровадження електронних видань пов'язане з підвищенням рівня освіти. Основне збільшення дає освітній контент: закупівля електронних видань університетськими бібліотеками сягає 40%. Більше того, вже сьогодні великі наукові книгозбірні, особливо академічних інститутів та університетів, відмовляються від традиційної передплати, навіть з фундаментальних наук, віддаючи перевагу електронним повнотекстовим виданням [15].

При аналізі майбутніх тенденцій розвитку в авторефератах, на жаль, не відображено технологічний аспект розвитку. Прикро, адже і у виробництві, і в науці відбувається революція, яка, безумовно, впливає на термінологію. Так, у публікаціях щодо бібліотечної справи зустрічаються технологічні терміни, наприклад, "мобільний". Це можна пояснити тим, що науковці досліджують лише професійні питання і не враховують ці новації. Хто знає, може з'явиться термін "мобільна бібліотека"?

¹ Додаткові семантичні та прагматичні особливості ("співзначення") лексичного значення та значень інших мовних рівнів, які нашаровуються на їхній предметно-понятійний аспект і можуть бути зумовлені як змістовою стороною, так і внутрішньою формою слова.

Зміни, характерні для бібліотечної сфери, можна спостерігати у загальних словах документів 2000 та 2011 років, послідовність яких впорядковується за зменшенням співвідношення частоти вживання слова 2000 року на відповідну частоту 2011. Отже, отримана послідовність слів, які у 2000 році згадувались частіше, ніж у 2011: "Реферування", "Класифікація", "Предметизація", "Систематизація" тощо, дає змогу чітко простежувати, що сучасні інформаційні технології знизили процеси "ручного" реферування, класифікації, предметизації. Тож реферативні бази даних у звичайному форматі нині втратили свою актуальність, що дасть змогу продуктивніше використовувати творчий потенціал редакторів.

Висновки

1. Запропонована методика виявлення динаміки змін основних напрямів діяльності у предметній галузі на основі відстеження хронологічних змін терміносистеми цієї галузі.

2. Ця методика апробована на текстових масивах авторефератів дисертацій за тематикою "Бібліотечна справа" за 2000 та 2011 роки.

3. Універсальність методики дає змогу застосовувати її в усіх галузях для дослідження якісних змін у явищах науки на базі кількісних характеристик вказаних масивів. Отримані дані за досить тривалий термін (від 10 років) уможливають дослідження тенденцій розвитку сучасної науки.

Список використаної літератури

1. *Проект КОНЦЕПЦІЇ "Української бібліотечної енциклопедії"*. — Режим доступу: www.nbuv.gov.ua/portal/Soc_Gum/Bdil/2011_2/2.pdf. — Назва з екрана.
2. *Стрішенець Н. В.* Метадані у сучасному бібліотекознавстві. Метадані — нове чи старе поняття? / Н. В. Стрішенець // *Бібліотекознавство. Документознавство. Інформологія*. — 2010. — № 2. — С. 4—11. — Режим доступу: www.nbuv.gov.ua/portal/Soc_Gum/bdil/2010_2/1.pdf. — Назва з екрана.
3. *Стрішенець Н. В.* "Хайдеггер чи Гайдеггер", або Поняття авторитетного контролю / Н. В. Стрішенець // *Студії з архівної справи та документознавства*. — 2010. — Т. 18. — С. 175—190. — Режим доступу: http://www.nbuv.gov.ua/portal/soc_gum/Studii/2010_18/27.pdf. — Назва з екрана.
4. *Добко Т. В.* До питання розвитку термінології бібліотечно-інформаційної сфери / Тетяна Добко // *Вісник Книжкової палати*. — 2011. — № 8. — С. 13—17. — Режим доступу: http://www.nbuv.gov.ua/portal/soc_gum/vkp/2011_8/st_5.pdf. — Назва з екрана.
5. *Англо-український тлумачний словник з бібліотечної справи та інформатики* / підгот.: Г. Воскресенська [та ін.]. — К. : [б. в.], 1999. — 57 с. — Режим доступу: <http://nbuv.gov.ua/books/19/glossary.pdf>. — Назва з екрана.
6. *Соловяненко Д. В.* Академічні бібліотеки у новому соціотехнічному вимірі. Ч. 3 : Роль академічних бібліотек у конструванні інтегрованих інформаційних е-середовищ / Д. В. Соловяненко // *Бібліотечний вісник*. — 2010. — № 6. — С. 3—17.
7. *Сбойчаков К. О.* Распределение ключевых слов по рубрикам ГРНТИ в базе данных Электронного каталога ГПНТБ России / К. О. Сбойчаков // *Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : материалы конф. (Судак, АР Крым, Украина 5—13 июня, 2004 г.)*. — Режим доступа: <http://www.nbuv.gov.ua/articles/crimea/2004/doc/292.pdf>. — Загл. с экрана.
8. *Маришкова И. В.* Система цитирования научной литературы как средство слежения за развитием науки. — М. : Наука, 1988. — 288 с.
9. *Ландэ Д. В.* Поиск знаний в INTERNET. Профессиональная работа : пер. с англ. / Д. В. Ландэ. — М. : Вильямс, 2005. — 272 с.
10. *Интернетика: Навигация в сложных сетях: Модели и алгоритмы* / Д. В. Ландэ, А. А. Снарский, И. В. Безсуднов. — Москва : URSS, 2009. — 258 с.
11. *Опарин А.* Системы мониторинга и анализа СМИ / А. Опарин // *PCWEEK (Москва)*. — Режим доступа: http://www.newart.ru/orarin/smi_oparin.htm. — Загл. с экрана.
12. *Баранов А. Н.* П50 Политический дискурс: методы анализа тематической структуры и метафоры / Баранов А. Н. [и др.]. — М. : [Фонд ИНДЕМ], 2004. — 94 с. — (Труды фонда ИНДЕМ / Регион. обществ. фонд Информатика для демократии).
13. *Коротич А.* Семантический анализ текста — основа качества статьи / А. Коротич. — Режим доступа: <http://fb.ru/article/25046/semanticheskiy-analiz-teksta-osnova-kachestva-statii>. — Загл. с экрана.
14. *Гак В. Г.* Лексическое значение слова / В. Г. Гак // *Языкознание : большой энцикл. слов. / гл. ред. Ярцева В. Н.* — М., 1998. — С. 261—263.
15. *Шрайберг Я. Л.* Интеграция библиотек в развивающееся информационное общество: что нас ждет впереди? : ежегодный доклад Конференции "Крым" / Я. Л. Шрайберг // *Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. (Судак, АР Крым, Украина 4—12 июня, 2012 г.)*. — С. 27.

Представлена методика определения изменений терминов в научных документах для отслеживания изменений направлений в библиотечном деле.

The method for determine changes in terms of scientific documents to monitor changes in directions of librarianship.

Надійшла в редакцію 14 вересня 2012 року