

УДК 004.774.6:34(051)

**Олександр Кузнєцов,**

наук. співроб. відділу довідково-консультаційної допомоги НЮБ НБУВ

**Андрій Кузнєцов,**

провід. бібліотекар НБУВ

## **ВИЯВЛЕННЯ ЛАТЕНТНИХ ЗАКОНОМІРНОСТЕЙ У НАУКОВИХ СТАТТЯХ ЮРИДИЧНОЇ ПРОБЛЕМАТИКИ (НА ПРИКЛАДІ НАУКОВОГО ЖУРНАЛУ «АДВОКАТ» ЗА 2009–2011 РР.)**

Надано алгоритм знаходження латентних закономірностей та зміни термінів у статтях юридичної проблематики за певний період одного напрямку за допомогою статистичних методів обробки тексту на фоні досліджуваного періоду.

*Ключові слова:* контент-аналіз, юридична література, журнал «Адвокат».

Обсяги наукових знань, які циркулюють в Інтернеті, цій головній комунікаційній артерії майбутнього суспільства знань, стрімко зростають [1]. За даними аналітичної компанії IDS обсяг інформації в Інтернет незабаром становитиме 500 екзобайтів (500 млрд Гб), що несе загрозу інформаційного калапсу. Ситуацію погіршує те, що у 2008 р. більшість компаній скоротили свої ІТ-бюджети через кризу.

Таким чином, утворився розрив між кількістю виробленої інформації та кількістю засобів, які забезпечують накопичення, обробку та використання даних. Для отримання корисної інформації потрібен різноманітний інструментарій і велика кількість шляхів доступу до неї. За умови прогресивного розвитку наукової інформації стають дедалі більш актуальними розробка та використання нових механізмів для її структурування, аналізу та оцінки [2].

Для автоматизованих систем найбільш зручним є застосування кількісних методів, одним з яких є контент-аналіз [3], що бере свій початок у Швеції 1640 р. Там, під час дискусії з офіційним лютеранством, теологи порівняли апокрифічну збірку релігійних гімнів «Пісні Сіону» з гімнами офіційної церкви та підраховали кількість основних релігійних ідей та їх висвітлення (позитивне, негативне чи реальне) [4].

Іntenсивний розвиток методів контент-аналізу почався з часів розвитку комп'ютерної техніки та Інтернету. Ось чому, згідно з [4] винахідниками

та лідерами контент-аналізу небезпідставно вважають американців, бо в Америці комп'ютерна індустрія була і є більш розвинутою за інші країни світу. І ще фактор мови: англійська мова – мова світу, мова комп'ютерних технологій. Більшість прикладів дослідження текстів за допомогою контент-аналізу або лінгвістичних методів, у наукових статтях вітчизняної періодиці та країн СНД були саме тексти англійською мовою. Це говорить на користь розвитку вітчизняного контент-аналізу, адже від мови багато залежить. Контент-аналіз текстів українською або російською мовами робити набагато складніше, ніж з англійською.

Контент-аналіз юридичних матеріалів виявив питання взаємозв'язку права й мови, які виникли з моменту розуміння їх соціальної природи: мова і право були і є засобами спілкування та взаємодії в суспільстві, і тому форми, через які суспільство створює еталон поведінки, можуть бути збережені і передані суспільству тільки, якщо будуть закріплені мовою.

Відомо, що першими з правників обґрунтували ідею взаємодії права і мови в процесі їх природного розвитку представники історичної школи права, засновниками якої були Ф.-К. Савінії (1779–1861) та Г.-Ф. Пухта (1798–1846). Останній, власне, був першим із вчених історичної школи права, хто окреслив проблему інтерпретації нормативно-правового закону, загалом, і юридично-лінгвістичного тлумачення, зокрема, називаючи його «граматичним тлумаченням» [5].

Проведення системних досліджень взаємодії та взаєморозвитку системи національного права та її терміносистеми є особливо важливим для формування розвитку державно-правового українського суспільства та структури правничої терміносистеми, динаміки змін, трансформацій у ній. Актуалізація системних досліджень зумовлена, по-перше, практичною необхідністю цілісного, системного і комплексного аналізу державно-правових явищ, а по-друге, через поглиблення, диференціацію загальної картини досліджуваного суб'єкта на окремі фрагменти, потребою їх аналізу як елементів цілісного явища [6].

Право нерозривно пов'язане з мовою в різних її проявах та має специфічне мовне забарвлення, зумовлене культурно-історичними традиціями народу, якими вони формуються.

Становлення й розвиток української правничої термінології безпосередньо пов'язані з етапами формування української національної державності та становлення національного права. Загалом, процес виникнення та розвою української правничої термінології відбувався за складних, суперечливих, нерідко навіть несприятливих, історико-політичних умов, що зумовлювалося специфікою характеру і змісту

боротьби за встановлення державності на українських землях у різні історичні періоди [7].

Проблеми аналізу юридичної інформації цікавлять суспільство і не стають менш актуальними. Складність і суперечливість соціально-економічного й політичного розвитку країни зумовлюють потребу в їх належному теоретичному осмисленні й розробці науково обґрунтованого інструментарію, перегляді науково виваженої термінології. Потрібен розвиток вітчизняного юридичного контент-аналізу [8].

Одне з прямих завдань контент-аналізу – виявлення тематичної близькості документів, визначається за принципом – наскільки часто терми, тобто слова, використовувались у документах (статтях) [9]. Загальні слова надають інформацію про тематичний напрям [10], але для виявлення динаміки змін цього недостатньо. Визначення появи нових слів – очевидний факт і не завжди є зв'язок минулого із сучасним і навіть життя нових слів, як правило, недовготривале. Пошук латентних закономірностей відкриває більш корисної інформації з текстів, тому і стає більш актуальним для сьогодення.

«Латентний», згідно з тлумаченням у словнику ABBYY Lingvo (Uk-Uk), означає – прихований; який не виявляється зовнішньо [11]. Слово «латентний» має широке застосування: медицина, юридична наука тощо.

Латентні (приховані) закономірності – це закономірності, які можуть бути виявлені інформаційним аналізом (контент-аналізом) у разі впливу багатьох чинників, які залежать, в основному, від часу. Латентні закономірності неможливо знайти звичайним пошуком, бо в явному виді невідомо наперед, що конкретно шукати.

Мета цієї статті – розробка методики виявлення латентних закономірностей на прикладі інформаційного аналізу текстів юридичної проблематики.

Для цього треба:

- провести загальний частотний аналіз;
- створити частотний словник юридичної тематики;
- розробити алгоритм пошуку латентних слів та словосполучень;
- перевірити наявність їх зв'язку з нормативними актами;
- виявити нормативні акти через латентні словосполучення.

Для прикладу візьмемо масив випусків журналу «Адвокат» за 2009: випуски з 1 по 11, 2010: випуски з 1 по 12 та 2011: випуски з 1 по 9 на сайті НБУВ, портал наукової періодики: [http://www.nbuv.gov.ua/portal/Soc\\_Gum/Advokat/texts.html](http://www.nbuv.gov.ua/portal/Soc_Gum/Advokat/texts.html). Дослідження проведемо для 2009 р. У цьому досить великому масиві юридичної інформації треба виявити

словосполучення, які відобразили б найбільш актуальні проблеми дослідного терміну – 2009 р. та як ці терміни пов’язані з нормативними актами України.

Визначення корпусу є досить складним процесом, і за відсутності юридичного корпусу візьмемо за нього випуски за 2010 та 2011 рр. (корпус). Таким чином, створений корпус суттєво заощаджує час роботи програми на обробку текстів. Також слова корпусу співпадають за тематикою та повністю відповідають визначенню тематичного динамічного корпусу [13].

Для пошуку латентних закономірностей визначимо загальні слова випуску 2009 р. у порівнянні з корпусом, таким чином, щоб частота появи слова у 2009 р. була більше, ніж у корпусі. Впорядкуємо слова у зворотній послідовності відношення частот появи слів ( $F1/F2$ ).  $F1$  – частота появи слова в дослідному тексті,  $F2$  – у корпусі. Декілька цих слів наведемо в табл. 1.

Таблиця 1

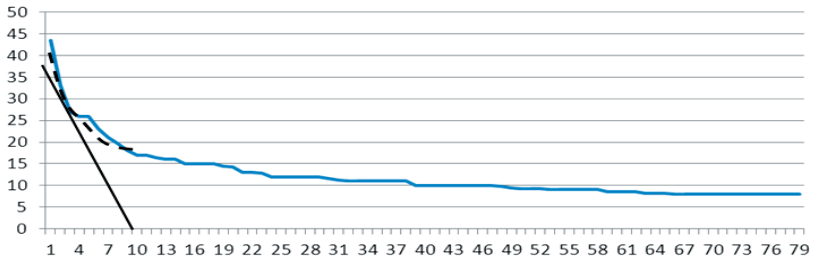
#### Частотні показники слів

| Слова           | F1 | F2 | F1/F2    |
|-----------------|----|----|----------|
| СУДИМОСТІ       | 87 | 2  | 43,5     |
| ВСЕУКРАЇНСЬКОГО | 33 | 1  | 33       |
| ДОБРОСОВІСНОСТІ | 27 | 1  | 27       |
| РИМСЬКОГО       | 52 | 2  | 26       |
| ПОГРОЗА         | 23 | 1  | 23       |
| РОЗБІЙ          | 21 | 1  | 21       |
| ПАРТІЇ          | 59 | 3  | 19,66667 |
| ІДЕМ            | 36 | 2  | 18       |

У результаті отримали слова, які пов'язані з певними змінами в законодавстві і які, імовірно, мають латентні якості.

Побудуємо графік зворотної послідовності відношення частот ( $F1/F2$ )

### $F1/F2$



Крива ( $F1/F2$ ) має експонентний характер з виразною крутою частиною. Ця частина кривої ( $F1/F2$ ) відповідає найбільш вживаним словам, які з великою імовірністю заслуговують на увагу. Для виявлення латентних закономірностей треба визначитись із кількістю слів для подальшого аналізу. Застосуємо графічний метод:

- визначимо на графіку круту частину (---);
- побудувати дотичну в точці приблизно на середині крутої частини графіку, до перетинання з віссю  $Ox$ .

Значення на осі  $Ox$  і буде відображати оптимальну кількість слів, гідних для подальшого аналізу. Щодо фактора часу опрацювання текстових масивів оптимальне зменшення кількості слів значно покращує цей показник – час роботи зменшується на певний порядок. Також зменшуються потреби в потужності комп'ютерної техніки.

Ці слова найчастіше використовуються в дослідному році випуску журналу «Адвокат» у порівнянні з корпусом, і можна з впевненістю казати, що знайдена кількість слів буде відображати латентні закономірності.

Через знайдені слова знайдемо словосполучення шляхом кількісного аналізу: словосполучення, які значно частіше зустрічались у тексті – стійкі словосполучення. Ці словосполучення мають відношення до відповідних законів. Також за їх допомогою можна виявити тематичний напрям або деякі метадані дослідного тексту чи зробити кластеризацію [13].

Отже для 2009 р. маємо (див. табл. 2):

Таблиця 2

## Частотні показники слів з пов'язаними словами та законами

| № | Слова           | F1 | F2 | F1/F2 | Пов'язані слова | Законо                                       |
|---|-----------------|----|----|-------|-----------------|--|
| 1 | СУДИМОСТІ       | 87 | 2  | 43,5  | Зняття          | Кримінальний кодекс України                  |
| 2 | ВСЕУКРАЇНСЬКОГО | 33 | 1  | 33    | референдум      | Кодекс адміністративного судочинства України |
| 3 | ДОБРОСОВІСНОСТІ | 27 | 1  | 27    | принцип         | Цивільне законодавство                       |
| 4 | РИМСЬКОГО       | 52 | 2  | 26    | Права           | -  |
| 5 | ДОРОЖНЬОГО      | 26 | 1  | 26    | Руху            | ДТП  |
| 6 | ПОГРОЗА         | 23 | 1  | 23    | розбійний напад | Кримінальний кодекс України                  |
| 7 | РОЗБІЙ          | 21 | 1  | 21    | Напад           | ст. 187 Кримінального кодексу України        |

| № | Слова  | F1 | F2 | F1/F2  | Пов'язані слова   | Закони  |
|---|--------|----|----|--------|---|---|
| 8 | ПАРТІЇ | 59 | 3  | 19,667 | політичні,<br>Керівник,<br>представник,<br>організація,<br>документ | КАС України, Закон України «Про політичні партії в Україні» |
| 9 | IDEM   | 36 | 2  | 18     | Принцип <i>pop<br/>bis in idem</i>                                  | Міжнародне право  |

Для виявлення кількісних співвідношень знайдених стійких словосполучень та законів перебудуємо попередню таблицю. У лівому стовпчику запишемо закони, у наступному – словосполучення. Як бачимо, деякі слова взаємно об'єднались у словосполучення (РОЗБІЙ, ПОГРОЗА) та словосполучення об'єднались за законами.

Розглянемо отримані стійкі словосполучення с позиції нормативних актів України (табл. 3).

Таблиця 3

## Рейтингові показники законів

| Закони   | Словосполучення                            | F1/F2  | Рейтинг |
|--|--|--------|---------|
| Кримінальний кодекс України  | Зняття судимості                           | 43,5   | 1       |
|  | Правила дорожнього руху – ДДП              | 26     | 3       |
|  | погроза, розбій, розбійний напад<br>ст.187 | 23, 21 | 5       |
| Кодекс адміністративного судочинства України   | Всеукраїнський референдум                  | 33     | 2       |
| Цивільне законодавство   | Принцип добросовісності,                   | 27     | 4       |
|  | Римське право<br>політичні партії          |        |         |
| Кодекс адміністративного судочинства України, Закон України «Про політичні партії в Україні» | керівник партії                            | 19,5   | 6       |
|  | представник партії                         |        |         |
|  | документ партії                            |        |         |
| Міжнародне право   | IDEM Принцип non bis in idem               | 18     | 7       |



Кримінальний кодекс України поєднав такі словосполучення: «Зняття судимості», Правила дорожнього руху – ДТП, «розбійний напад»; Кодекс адміністративного судочинства України – «Всеукраїнський референдум», Цивільне законодавство – «Принцип добросовісності», «Римське право».

Створимо зведену таблицю для визначення рейтингу законів, про які задували у випусках журналу «Адвокат» за 2009 р. Для цього в лівому крайньому стовпчику запишемо закони, далі суму відношень частот відповідних словосполучень, у наступному стовпчику – діаграму (див. табл. 4).

Таблиця 4

## Зведена таблиця законів та частотних співвідношень

| Закони   | F1/F2 |    | Діаграма |
|--|-------|----|----------|
|  | Абс.  | %  |          |
| Кримінальний кодекс України                    | 113,5 | 53 |          |
| Кодекс адміністративного судочинства України   | 33    | 16 |          |
| Цивільне законодавство                         | 27    | 13 |          |
| Закон України «Про політичні партії в Україні» | 19,5  | 9  |          |
| Міжнародне право                               | 18    | 9  |          |

Як бачимо, з великим відривом на першому місті Кримінальний кодекс України – 53 % з питаннями щодо зняття судимості, порушення правил дорожнього руху та розбій, напад, погрози; далі Кодекс адміністративного судочинства України – 16 % з питанням Всеукраїнського референдуму, потім Цивільне законодавство – 13 % з питаннями щодо принципу добро-

совісності, Закон України «Про політичні партії в Україні» та Міжнародне право – по 9 %.

Описаний метод виявлення латентних закономірностей на прикладі журналу «Адвокат» спрацьовує, латентні закономірності щодо зміни нормативних актів виявлено.

Отже вищеописаний метод складається з таких етапів:

- визначається тематичний корпус та дослідна частина;
- знаходимо загальні слова дослідної частини та корпусу;
- з масивів загальних слів відокремлюємо ті слова, які застосовуються частіше в частині, ніж у корпусі;
- будемо гістограму відношення частот слів частини та корпусу;
- графоаналітичним методом визначасмо оптимальну кількість слів для дослідження;
- знаходимо словосполучення статистичним методом: за максимальною частотою зустрічі сусідніх слів у дослідному тексті;
- серед знайдених словосполучень виявляємо латентні закономірності щодо нормативних актів.

Таким чином можна зробити висновок, що описаний метод дає змогу виявити латентні закономірності з мінімальним втручанням людини. За результатами дослідження видно, що цей метод некритичний до тематичного корпусу. Його динамічно створено з інших випусків журналу «Адвокат». Водночас слід зазначити, що краще використовувати спеціалізований юридичний корпус і результат буде точнішим.

### **Список використаних джерел**

1. *Симоненко Т. В.* Мережеве інформаційне забезпечення наукових досліджень : автореф. дис. ... канд. наук із соц. комунікацій / Т. В. Симоненко. – К., 2011. – 18 с.
2. *Редькина Н. С.* Библиометрия: история и современность / Н. С. Редькина // Молодые в библиотечном деле. – 2003. – № 2. – С. 76–86.
3. *Чернокозинский С. А.* Использование текстовых анализаторов для защиты информации в образовательной среде / С. А. Чернокозинский. – Информационное противодействие угрозам терроризма. – 2005. – № 4. – С. 239–242.
4. Вікіпедія: вільна енциклопедія [Електронний ресурс] : веб-сайт. – Електрон. дані та прогр. – Режим доступу: <http://uk.wikipedia.org/wiki/контент-аналіз>. – Назва з екрана.
5. *Пухта Г. Д.* Энциклопедия права / Г. Д. Пухта. – Ярославль, 1872.

6. *Карвацька С. Б.* Системний підхід у дослідженні взаємодії національної державно-правової історії терміносистеми українського права / С. Б. Карвацька // Держава і право : зб. наук. пр. Юридичні і політичні науки. Спецвипуск. – К. : Ін-т держави і права ім. М. В. Корецького НАН України, 2007. – С. 38–43.

7. *Козловський А. А.* Право як пізнання: Вступ до гносеології права / А. А. Козловський. – Чернівці : Рута, 1999. – С. 217.

8. *Мірошниченко М. І.* Відношення понять «системи права» та «правової системи» синергетичний аналіз / М. І. Мірошниченко // Науковий вісник Чернівецького університету : зб. наук. пр. Вип. 187: Правознавство. – Чернівці : Рута, 2003. – С. 33.

9. *Пескова О. В.* Методы автоматической классификации документов / О. В. Пескова // Информационные процессы и системы. – 2006. – № 3. – С. 13–20.

10. *Сбойчаков К. О.* Распределение ключевых слов по рубрикам ГРНТИ в базе данных Электронного каталога ГПНТБ России [Электронный ресурс] / К. О. Сбойчаков // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса : материалы конф. (Судак, А. Р. Крым, Украина 5–13 июня, 2004 г.). – Режим доступа: <http://www.nbuv.gov.ua/articles/crimea/2004/doc/292.pdf>. – Загл. с экрана.

11. *Lingvo Online* – безкоштовний онлайн-словник [Електронний ресурс] : веб-сайт. – Електрон. дані та прогн. – Режим доступу: <http://www.lingvo.ua/uk/Interpret/uk/лагентний>. – Назва з екрана.

12. *Аглямова Г. М.* Криминологическое исследование по проблемам латентной преступности / Г. М. Аглямова // Ученые записки Казанского государственного университета. – 2008. – Т. 150, кн. 5. – С. 185–189. – (Гуманитарные науки).

13. *Мамчич А. А.* Алгоритмы индексирования и поиска документов на основе динамических корпусов текстов / А. А. Мамчич // Информатика. – 2010. – № 1. – С. 82–90.