

ТЕРМІНОЛОГІЯ ГУМАНІТАРНИХ ЗНАНЬ

УДК 81.340.113

Владислав Шкурко, Світлана Ющенко
Український мовно-інформаційний фонд НАН України

ПІДХОДИ ДО ФОРМУВАННЯ ТА ВПОРЯДКУВАННЯ ТЕРМІНОЛОГІЇ ЗАКОНОДАВСТВА ЗАСОБАМИ КОМП'ЮТЕРНИХ ЛІНГВІСТИЧНИХ ТЕХНОЛОГІЙ*

© Шкурко В., Ющенко С., 2007

У статті викладено методику формування та впорядкування термінології законодавства засобами комп'ютерних лінгвістичних технологій та показано її дієвість на прикладі правової галузі «Автоматизація законотворчого процесу».

The procedure of the formation and adjustment of the legislative terminology by use of computer linguistics technologies is under consideration and its efficiency is shown in the field «The automation of the legislative process».

Суттєвий елемент змісту правових документів складають терміни – відтворювані спеціалізовані одиниці мови, що характеризуються певною функціональною** стійкістю та відносно чіткою семантичною окресленістю і вербально оформлюють певне поняття, що входить до системи понять предметної галузі. У правовій теорії поширена й інша думка: «Займаючи, в принципі, незначний обсяг нормативного тексту, юридична термінологія становить його базу, основний смисловий фундамент» [1].

Терміни, що є неподільними в смисловому плані одиницями, носіями «елементарного правового смислу», відіграють роль елементарних інформаційних одиниць законодавства, своєрідного його семантичного каркасу; решта мовних об'єктів виконує роль інформаційного середовища, «натягнутого» на цей каркас.

Попри існування національних стандартів, у яких закріплено систему вимог до терміносистеми вітчизняних нормативних документів [2; 3; 4], стихійність термінології є однією з наріжних проблем вітчизняного законодавства.

Зазначене явище, частково зумовлене перетином сфер правового регулювання та «вмонтованістю» кожного документа до окремого законотворчого процесу, полягає в існуванні в термінології системно-мовних відношень, які лінгвістична теорія кваліфікує як прояви лінгвістичної неоднозначності – синонімії, омонімії та полісемії.

* Автори щиро вдячні директорові Українського мовно-інформаційного фонду НАН України В. А. Широкову та старшому науковому співробітнику Міжнародного науково-навчального центру інформаційних технологій та систем НАН та Міністерства освіти і науки Т. Г. Дрогаль за надану допомогу.

** функційною – *ред.*

Звідси походить нагальність завдання унормувати правничу термінологію – комплексний захід, метою якого є визначити та ухвалити як стандартні варіанти термінів, найадекватніші нормам української мови, семантичним законам та специфіці предметної галузі. Нагадаємо, що етапи унормування термінології такі: формування експериментального корпусу термінів, систематизація, аналіз, уніфікація та оптимізація [5].

У цій праці обмежуємося етапами добору та систематизації елементів терміносистеми, а запропонована методика експлікації чинників, що вносять неоднозначність у розуміння правового змісту, стосується лише синонімії.

Мета цієї праці, яку виконано в межах державної науково-дослідної роботи «Розробка концептуальної та логічної моделі законотворчого процесу Верховної Ради України з урахуванням динаміки змін технологічних процесів законотворчої діяльності в різних скликаннях» (реєстраційний номер 0106U011851) – викласти методику формування та впорядкування термінології законодавства в максимально автоматизованому режимі.

1. Формування термінологічного масиву законодавства

З часу проголошення незалежності вітчизняні лексикографи провели колосальну роботу із забезпечення законотворчої та правової діяльності словниками – і традиційними, «паперовими», і електронними. Відзначимо такі видання, як [6, 7, 8] та багато-багато інших.

Словник є не лише продуктом лексикографічної техніки – його можна розглядати як певну модель конкретних феноменів реальності [9]. Проте будь-яка модель апіорі не може бути цілком адекватною до оригіналу, й особливо це стосується моделей мови, зокрема лексикографічних, зафіксованих в словниках юридичної тематики. Досвід переконує, що наявні словники дуже приблизно відбивають реалії правового поля навіть теоретично. Звідси випливає, що джерелами формування термінологічного масиву законодавства мають радше бути юридичні тексти (законодавчі та нормативно-правові акти, регламенти, стандарти, постанови, накази тощо), ніж юридичні словники.

Через неточність формальних границь багатьох законодавчих термінів унаслідок особливості правової комунікації, формування термінологічного масиву не можна сьогодні виконати інакше, ніж у ручному режимі, хоча певні операції все-таки вже можна автоматизувати.

Формальною ознакою терміна є вмотивована смислова єдність та граматична узгодженість слів, з яких він складається. Щодо питання мінімальної довжини терміна в термінознавстві панує визначеність – одне слово; питання максимальної довжини навряд чи має однозначну відповідь. Навіть належність певного словосполучення* до множини елементів терміносистеми часто є дискусійною. Істотний вплив людського чинника ілюструє такий факт: вважають, що термінологія Конституції України [10] налічує 917 одиниць, а дві групи науковців УМІФу, що діяли незалежно одна від одної, встановили 1127 та 1432 термінологічні одиниці відповідно. Отже, для забезпечення максимальної репрезентативності та коректності формування терміносистеми, термінологічний масив, на нашу думку, треба формувати в експертному режимі, узагальнивши думки щонайменше трьох експертних груп (експертів) [11]¹.

2. Упорядкування загального термінологічного масиву законодавства

2.1. Встановлення повного збігу (дублювання) термінів

Природно, що множини термінів, які сформовано кожною з експертних груп, переважно будуть збігатися, адже експерти використовують однакові джерела.

Методику встановлення повного збігу термінів ми ґрунтуємо на положеннях структурної теорії лексикографічних систем, згідно з якою конкретна природна мова має скінченну множину

* певної словосполуки – *ред.*

¹ Проблеми організації експертних досліджень деяких мовних феноменів правового поля ми розглянули в окремій статті, яку з незалежних від нас причин досі не опубліковано.

символів, які використовують при графічному зображенні її об'єктів [6]².

Узагальнений алфавіт української мови має таку структуру:

$$A(U) = AL(U) \cup F(U) \cup P(U) \cup S(U) \cup p,$$

де $AL(U)$ – фонетичний алфавіт української мови; $F(U)$ – множина символів іншомовних алфавітів, які вживають у текстах і не містяться у множині $AL(U)$, $P(U)$ – множина знаків пунктуації, $S(U)$ – множина спеціальних символів (арабських та римських цифр, математичних, музичних символів тощо), p – пробіл.

Своєю чергою, до кожного зі складників структури (окрім, звісно, p) входить певна скінченна множина елементів. Підструктура $S(U)$, на відміну від підструктур $AL(U)$, $F(U)$ та $P(U)$, вирізняється більшою відкритістю.

Таким чином, термін має вигляд послідовності символів в узагальненому алфавіті $AL(U)$:

$$T = \{v_1 v_2 v_3 \mathbf{K} v_n\}, v \in AL(U).$$

Визначення.

Терміни T_i і T_j є такими, що повністю збігаються, якщо:

– будь-якому $n \in T_i$ відповідає єдиний $n \in T_j$ та навпаки;

– порядок слідування n є строго визначеним³.

Оскільки для вирішення цієї поточної задачі, ми розглядаємо терміни суто як комбінацію символів, наведене визначення є справедливим для термінів будь-якої довжини.

Отже, для встановлення множини термінів, що відповідають цьому визначенню, достатньо розмістити терміни за алфавітом, користуючись вбудованим сервісом впорядкування табличних записів.

Випадком дублювання також є терміни з прямим та інверсним порядком слів: <первинний документ – документ первинний>, <нормативний акт – акт нормативний>, <правові дефініції – дефініції правові> та ін. Такі форми трапляються тому що у словниках, якими послуговуються експерти, класифікувальні ознаки зазвичай подаються після реєстрової одиниці (наприклад: *ДОКАЗИ* – речові, непрямі, початкові, особисті, письмові, похідні, прямі).

Це спричиняє певну проблему у встановленні такого дублювання, адже відстань між записами перебуває в прямій залежності від розміру термінологічного масиву.

Зазначену проблему вирішуємо у такий спосіб.

Попередні зауваження.

Щоб уникнути занурення у філософію лінгвістики, під *словом* розуміємо сукупність символів поміж двома пробілами.

$$S = \{p v_1 v_2 v_3 \mathbf{K} v_n p\}.$$

Отже, термін довжини 2 представляється як $T^2 = \{pS_1 pS_2 p\}$, довжини 3 – відповідно $T^3 = \{pS_1 pS_2 pS_3 p\}$ і так далі.

Визначення.

Терміни $T_i^n = \{pS_i^1 \mathbf{K} pS_i^n p\}$ та $T_j^n = \{pS_j^1 \mathbf{K} pS_j^n p\}$ є такими, що повністю збігаються (дублюються), якщо будь-якому $S \in T_i$ відповідає єдине $S \in T_j$ та навпаки.

На основі цього визначення розроблено комп'ютерну програму, яка встановлює подібне дублювання.

² Звісно, представлення мови як системи знаків є лише одним зі способів її моделювання: «знаковість – це лише одна з функцій мови. Вважати мову лише системою знаків – означає спрощувати та загрублювати сутність справи» [12].

³ Це визначення відображає погляд на мову як на семіотичну систему. Згідно з ним, випадками повного збігу також є лексичні омоніми. Для скорочення викладу вважаємо омонімію знятою, а самі терміни такими, що не містять орфографічних помилок.

2.2. Нормалізація термінів

Оскільки багато термінів в законодавчих текстах трапляються у різних формах та відмінках, для усунення дубльованих записів виникає необхідність їх нормалізації. Нормалізованим вважають термін, форма якого збігається із вихідною⁴.

Множину вихідних форм для всіх однослівних термінів нескладно отримати засобами будь-якого електронного граматичного словника⁵. Щодо термінів більшої довжини, зазначимо, що через відсутність надійних програмних засобів аналізу синтаксичних структур, у теперішній час їх нормалізують у ручному режимі.

Для забезпечення роботи експертів в УМІФі розроблено окремий програмний комплекс «АРМ юриста-експерта»⁶. У проектуванні зазначеного комплексу, який орієнтовано на обробку великих текстових масивів, враховано як останні тенденції в галузі мовно-інформаційних технологій, так і побажання фахівців юридичної галузі. Наприклад, для ідентифікації лексичної конструкції як терміна, в усьому масиві документів, занесених у базу даних комплексу, автоматично маркуються усі такі конструкції. Іншою практично корисною функцією є пошук термінів із заданням відстані між словами (наприклад: конструкція «касаційні та апеляційні суди» містить два терміни – касаційні суди та апеляційні суди. Відстань між словами терміна касаційні суди – 2).

Після нормалізації повторюють процедуру, описану в п. 2.1.

2.3. Експлікація синонімії та поліморфізму термінів

Явище синонімії полягає у називанні того самого поняття (того самого змісту) різними термінами. Воно досить поширене в законодавстві внаслідок природного бажання нормопроектувальників урізноманітнити виклад, уникнути повторів тощо. Проте наявність синонімів у мові права є певною проблемою через їх властивість мати різні відтінки значення. Наприклад, терміни затриманий та заарештований, подані у Конституції (ст. 29, 126) як синоніми, фактично позначають різні правові стани суб'єкта правовідносин. Подібна ситуація із термінами призначення та обрання (ст. 85.15), що вживаються як синоніми, хоча передбачають різні правові процедури і, отже, мають різну семантику.

Розрізняють синоніми лексичні та синтаксичні; останні – це конструкції, що збігаються за значенням за відмінності формальної будови та виражаються одиницями однакового синтаксичного рівня [14]. Смысловая близькість синтаксичних синонімів, так само, як і лексичних, не означає тотожності їх значень (закони України – українські закони).

Поліморфізмом називаємо паралельне вживання різних форм запису того самого терміна. Попри те, що поліморфізм є порушенням логічного принципу однаковості «подібне позначати подібним»⁷ та одноманітності викладу, це доволі поширене явище в законодавстві. Наприклад, існує декілька варіантів написання кон'юнктивно-диз'юнктивного сполучника: **або/та; та/або; та (або); або (та)**; числівники (числа, дати, кількісні характеристики) подають то цифрою, то прописом, не кажучи вже про численні варіанти скорочень та лексичних конструкцій.

⁴ Для однослівних термінів вихідною формою є називний відмінок, однина, однослівних множинних – називний відмінок, множина. Щодо складених – встановлюючи вихідну форму, орієнтуємось на головне слово, яке має відповідати вищезазначеним положенням.

⁵ При здійсненні впорядкування масиву термінів правової галузі «Автоматизація законотворчого процесу» було використано засоби Інтегрованої лексикографічної системи (ІЛС) «Словники України» (розробник – Український мовно-інформаційний фонд НАН України).

⁶ Детально про комплекс «АРМ юриста-експерта» та організацію його інформаційного наповнення див. [13].

⁷ «Різні слова не можуть вживатися для виразу одного і того ж самого поняття. Одне й те ж саме слово не може вживатися для виразу різних понять» (з Регламенту Верховної Ради України, Відомості Верховної Ради, 1994 р., № 35).

Подекуди досить вільне поводження нормопроектувальників із мовним матеріалом негативно впливає на сприйняття правового змісту. Наприклад, терміни нерівномірне податкове навантаження та нерівномірний розподіл податкового навантаження є випадком поліморфізму, хоча постають як імена різних понять.

Експлікувати синонімічні відношення однослівних термінів рекомендуємо використовуючи інструментарій ІЛС «Словники України».

Експлікація синонімічних відношень та поліморфізму термінів більшої довжини потребувала розробки спеціальної методики, яка передбачає попередній аналіз галузей законодавства (у вузькому сенсі – окремих документів) для встановлення репрезентативних детермінантів їх семантичного поля. Зазначені детермінанти будуть виступати назвами сегментів поля; у кожному з них, своєю чергою, виділяються групи. Наприклад, детермінантами семантичного поля Конституції України є назви деяких розділів (рис. 1).

Після цього в автоматизованому режимі розподіляють терміни по групах. Зазначимо, що якість розподілу збільшується в міру розвитку уявлень про специфіку групи. Розмір масиву термінів, віднесених до однієї групи, знаходиться в межах можливостей мовно-розумового апарату людини сприймати та опрацьовувати текстову інформацію.

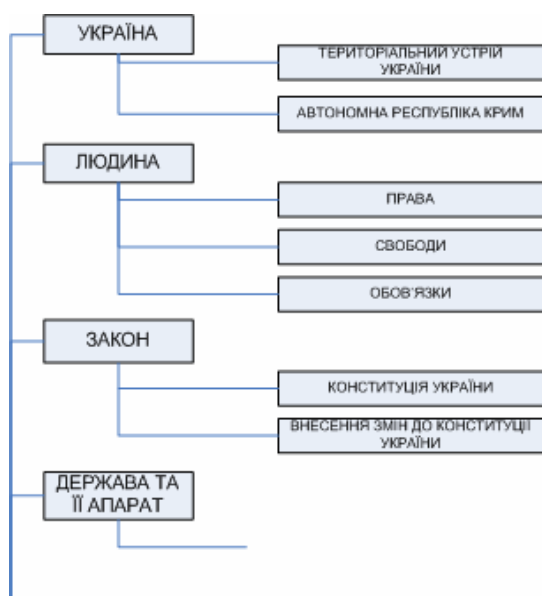


Рис. 1. Семантичне поле Конституції України (фрагмент)

Отже, експлікація синонімії та поліморфізму в фахівців-лінгвістів не викликає складнощів. Після її проведення вирішують, який саме варіант терміна найповніше відповідає стандартам українського правопису та стане нормативним.

3. Апробація запропонованої методики

Запропоновану методику було застосовано у формуванні та впорядкуванні термінологічного масиву правової галузі «Автоматизація законотворчого процесу».

З метою формування загального масиву термінів трьома експертними групами (відповідно E1, E2, E3) із документів, які є нормативно-правовою базою або стосуються зазначеної галузі, проведено відбір лексичних конструкцій, що, на погляд експертів, є термінами обраної предметної галузі. Окрім повнотекстових документів експерти використовували такі словники: тезаурус ЄВРОВОК, Термінологічний словник НАУ (ТСН), Словник Законодавства у сфері інформації та інформатизації в Україні (СЗІ) та науково-практичне видання «Законотворча діяльність. Словник термінів і понять» (СЗД).

Кількість проаналізованих документів становила 2392 одиниці.

Результати діяльності кожної з експертних груп зведено в одну таблицю (табл. 1), загальна кількість рядків якої становила 8193.

Таблиця 1 (фрагмент)

№ з/п	Термін	ОМ	ТС Н	СЗІ	СЗД	ЄВР ОВ ОК	ТГ	БД	Е1	Е2	Е3
1.	Автоматизована система	*	*	*			*			*	*
2.	Автоматизоване оброблення даних	*	*					*	*	*	*
3.	Автоматизоване робоче місце	*	*							*	*
4.	Автор	*	*			*	*		*		*
5.	Адаптація законодавства		*		*				*	*	
7.	Акт		*		*		*	*			
8.	База даних	*	*	*	*		*			*	*
	База знань	*		*	*		*			*	*
	Використання інформації	*		*			*		*	*	

Унаслідок виконання процедур нормалізації та видалення дублів скорочено вихідний масив (8193 " 5058)⁸.

Для кожного з 360-ти однослівних термінів в автоматичному режимі було визначено належність до певного синонімічного ряду; загальна кількість рядів становила 39 (табл. 2).

Таблиця 2 (фрагмент)

Термін	Синонімічний ряд
Акт	УКАЗ; ПОСТАНОВА; ДЕКРЕТ; УНІВЕРСАЛ; ЕДИКТ
Етап	ПЕРІОД; СТАДІЯ; СМУГА; ФАЗА; ФАЗИС <i>рідше</i>
Коаліція	ОБ'ЄДНАННЯ <i>кого, чого або яке</i> ; СОЮЗ; СПІВДРУЖНІСТЬ; СПІВТОВАРІСТВО; ФЕДЕРАЦІЯ; КОНФЕДЕРАЦІЯ; БЛОК; ЛІГА
Код	ШИФР
Наказ	РОЗПОРЯДЖЕННЯ
Пленарний	СПІЛЬНИЙ; КОЛЕКТИВНИЙ; СУКУПНИЙ; ГУРТОВИЙ; СУМІСНИЙ; ГРУПОВИЙ; КОЛЕГІАЛЬНИЙ; ОБ'ЄДНАНИЙ; ЗАГАЛЬНИЙ
Посвідчення	ДОВІДКА; ПОСВІДЧЕННЯ, ПОСВІДКА
Працівник	РОБІТНИК <i>розм.</i>
Резолюція	УХВАЛА; ПОСТАНОВА; РІШЕННЯ; ПРИГОВОР <i>розм.</i> ; ВОТУМ
Результат	НАСЛІДОК
Розгляд	ОБГОВОРЕННЯ
Термін	СТРОК
Повноваження	ПРАВА

У правовій галузі «Автоматизація законотворчого процесу» виділено 20 концептів, загальна кількість груп сягала 52.

Фрагмент структури семантичного поля галузі наведено на рис. 2.

⁸ Дані про джерела, враховуючи їх важливість, винесено в окрему таблицю.

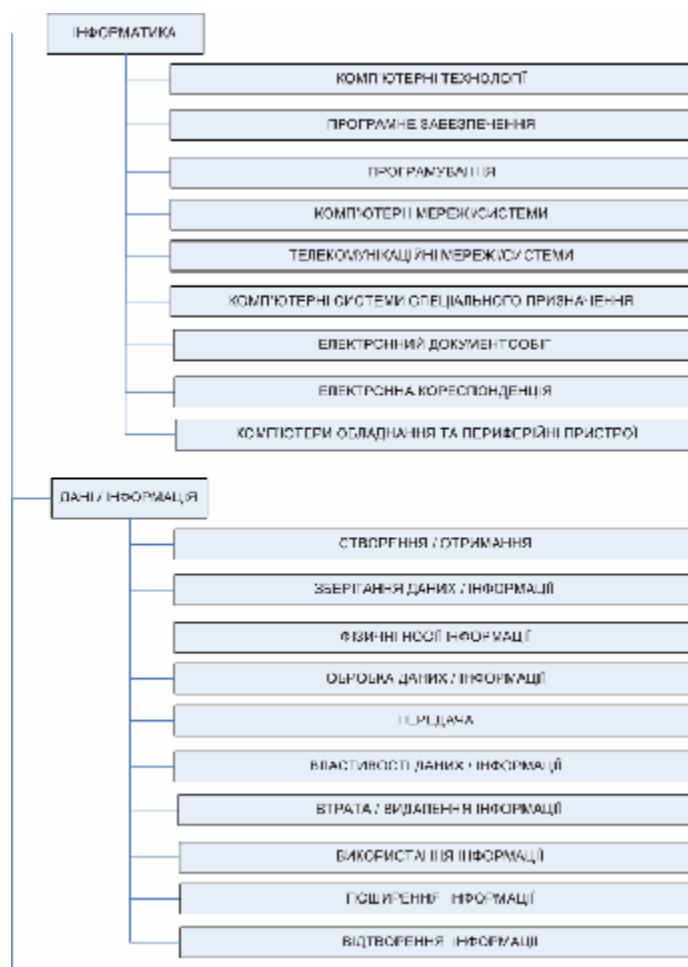


Рис. 2. Семантичне поле правової галузі «Автоматизація законотворчого процесу» (фрагмент)

У результаті встановлено 69 синонімічних груп (табл. 3) та 42 групи поліморфічних термінів (табл. 4).

Таблиця 3 (фрагмент)

Цифровий підпис
Електронний підпис
Кваліфікований працівник
Кваліфікований робітник
Посвідчення народного депутата України
Мандат депутатський
Засідання Верховної Ради України
Засідання Парламенту
Постачальник послуг мережі Інтернет
Провайдер мережі Інтернет
Державна програма
Всеукраїнська програма
Закони, повернені Президентом
Закони, відхилені Президентом
Користувач системи електронної пошти
Абонент системи електронної пошти
Депутатські фракції
Депутатські групи
Оприлюднення закону
Опублікування закону
Національний стандарт
Державний стандарт України

Таблиця 4 (фрагмент)

Аналітична нормативна документація
Аналітично-нормативна документація
Засоби технічного захисту інформації
Технічні засоби захисту
Комплекс технічного захисту інформації
Забезпечення тендерної пропозиції
Тендерне забезпечення
Об'єкт експертизи
Об'єкт експертної оцінки
Висновок державної експертизи
Висновок експерта
Висновок експертизи
Експертний висновок
Висновок експертний комплексний

Остаточний висновок щодо того, який варіант терміна стане нормативним для правової галузі, як вже зазначалось, – за фахівцями-лінгвістами.

4. Статистичні дослідження на термінологічному масиві

Залучення засобів інформаційно-лінгвістичних технологій дає змогу швидко отримати будь-які статистичні дані стосовно як масиву (загальна кількість термінів, розподіл за кількістю слів та джерелами), так й окремого терміна (кількість документів, до яких входить термін, кількість прямих входжень, а також з урахуванням лематизації та синонімії).

Наприклад, для галузі «Автоматизація законотворчого процесу» статистичні відомості за всіма термінами мають такий вигляд: кількість прямих входжень – 308788; кількість входжень з урахуванням лематизації: 1226399; синонімії – 2533537.

Розвиток лінгвістичних технологій триває і, сподіваємось, у недалекому майбутньому впорядкування багатотисячного термінологічного масиву за запропонованою методикою займатиме менше часу, ніж потрібно для прочитання цієї статті.

Ураховуючи значний обсяг опрацьованого матеріалу та інтенсивність законодавчих процесів, основним засобом розвитку терміносистеми законодавства та основним інструментом її дослідження має стати спеціалізований програмний комплекс, що ґрунтуватиметься на сучасних інформаційно-лінгвістичних технологіях.

Вважаємо, що запропонована методика формування та впорядкування термінології, що пройшла апробацію на одній із галузей законодавства, може бути корисною у формуванні бази даних зазначеного комплексу.

1. Алексеев С. С. *Теория государства и права*. – М.: БЕК, 200. – 486 с. 2. ДСТУ 3325-96 *Термінологія. Визначення основних понять: Вид. офіц.* – К.: Держстандарт України, 1996. 3. ДСТУ ISO 860-1999 *Термінологічна робота. Гармонізування понять та термінів: Вид. офіц.* – К.: Держстандарт України, 1999. 4. ДСТУ 3966-2000 *Термінологія. Засади і правила розроблення стандартів на терміни та визначення понять: Вид. офіц.* – К.: Держстандарт України, 2000. 5. Гринев С. В. *Введение в терминоведение*. – М.: Московский лицей, 1993. – 309 с. 6. *Юридична енциклопедія в 6-ти томах / Під ред. Ю. Шемшученка*. – К.: Українська енциклопедія, 1998. 7. *Словник законодавчих термінів*. – К.: Основа, 2000. – 604 с. 8. Гіжевський В. К., Головченко В. В. та ін. *Популярна юридична енциклопедія*. – К.: Юрінком, 2003. – 528 с. 9. Широков В. А. *Інформаційна теорія лексикографічних систем*. – К.: Довіра, 1998. – 330 с. 10. *Конституція незалежної України. Кн. II / Під ред. С. Головатого*. – К., 1997. – 162 с. 11. Литвак Б. Г. *Експертная информация. Методы получения и анализа*. – М.: Радио и связь, 1982. – 184 с. 12. Морковкин В. В. *Идеографические словари* // <www.rhymes.atlab.ru>. – 1970. 13. Шкурко В. В., Юценко С. С. *Концептуальні засади та системотехнічні підходи до проектування системи пошуку протиріч* // *Вісник НАУ*. – 2006. – № 4. – С. 212–215. 14. *Терминоведение* // <ru.wikipedia.org>. – 2007.