

## ЛЕКСИКОГРАФІЧНА БАЗА ДАНИХ НА ОСНОВІ ЛЕКСИКО-СЕМАНТИЧНИХ ГРУП ДІЄСЛІВ УКРАЇНСЬКОЇ МОВИ

© Бігдай М. О., 2015

**Розглянуто основні переваги комп'ютерної лексикографії над традиційною. Обґрунтовано доцільність представлення семантичних характеристик дієслів у лексикографічній базі даних. За інформаційною теорією лексикографічних систем побудовано лексикографічну базу даних лексико-семантичних груп дієслів української мови.**

**Ключові слова:** формальна модель, лексикографічна база даних, інформаційна теорія лексикографічних систем.

**The article deals with the main advantages of computer lexicography comparing to the traditional one. The expediency of presenting the semantic characteristics of verbs in lexicographic database is proven. On the basis of the information theory of lexicographic systems the lexicographic database of lexical-semantic groups of the Ukrainian verbs is built.**

**Key words:** formal model, lexicographic database, information theory of lexicographical systems.

### Вступ. Загальна постановка проблеми

В останні десятиліття актуальними для лінгвістики є питання моделювання мовних явищ. Одним з аспектів такого моделювання є моделювання лексичної семантики дієслів. Для української мови вже давно актуальною є потреба у створенні семантичного словника, що передбачає насамперед розроблення його формальної моделі. Для розв'язання будь-якої лінгвістичної задачі за допомогою комп'ютерних технологій важливо вибрати спосіб представлення мовного матеріалу. В українській лексикографії єдина теорія, яка якнайповніше розв'язує поставлену мету, є інформаційна теорія лексикографічних систем, розвинута в працях В. А. Широкова [9, 10], тому саме ця теорія стала в дослідженні вихідною точкою. Таку формальну модель [3] ми побудували на основі лексико-семантичних груп дієслів української мови. Наступним етапом нашого дослідження є створення на її основі лексикографічної бази даних.

Розуміння мови як інформаційної системи, що є ключовим в інформаційній теорії лексикографічних систем, зосереджує увагу лінгвістики на вивченні структури парадигматичних та синтагматичних відношень між знаками мовної системи. Осмислення ролі інформаційних процесів мовних явищ як складників мовно-розумової діяльності людини в інформаційній теорії лексикографічних систем спирається на те, що мовні процеси не є замкненими, а відповідають різним видам діяльності людини. І оскільки в інформаційному апараті людини словникові підсистеми є дуже інтегрованими з образними, логічними та когнітивними підсистемами, "...прикладні лінгвістичні студії, орієнтовані на моделювання процесів розуміння... повинні застосовувати комплексні, нелінійні структури для представлення загальноновживаних знань, зокрема ті, що використовуються у когнітивних науках і серед них такі загальновідомі, як фрейми, сценарії (скрипти), продукції тощо" [9, с. 13].

### Аналіз останніх досліджень та публікацій

В Українському мовно-інформаційному фонді в межах дослідження лексичної та граматичної семантики дієслів було побудовано лексикографічну базу даних "Дієслово". Для отримання оптимальної репрезентації закономірностей лексико-граматичної системи дієслів сучасної

української мови, її цілісності, безперервності, ієрархічності було використано подану структуру елементарної лексикографічної системи одинадцятитомного “Словника української мови” (СУМа), а лінгвістичну інформацію представлено у вигляді лексикографічної бази даних, яку надалі у дослідженні використовували як інструмент для моделювання граматичної та лексичної семантики дієслів української мови [6, с. 35].

Побудова лексикографічної бази даних є новим прикладним аспектом лінгвістики. Якщо традиційна лексикографія орієнтувалася насамперед на опис окремого слова як самостійного об'єкта дослідження, то створення лексикографічної бази даних на перший план висуває дослідження класів слів, об'єднаних за одним чи декількома параметрами [4, с. 96–106]. У лінгвістиці неодноразово наголошувались переваги лексикографічної бази даних над традиційними словниками, оскільки саме в лексикографічній базі даних може практично реалізуватися ідея інтегрального опису мови, базованого на узгодженні зв'язку між словником та граматику [1, с. 57–70].

Очевидно, що впорядковано-класифікований лінгвістичний матеріал із багатопараметричним урахуванням і форми, і змісту мовних одиниць не може бути адекватно представленим у традиційних (некомп'ютерних) засобах подання інформації. Отже, на цьому рівні неможливе й забезпечення активного доступу користувачів до цієї інформації. Лише спеціалізовані комп'ютерні лексикографічні бази даних дають змогу досліджувати різноманітні системні, а не лише кількісні характеристики лексики, об'єктивно визначати системну значущість необхідних ознак, параметрів, категорій лексики, що є основою виділення тих чи інших угруповань, пояснити їх онтологічну сутність, системну зумовленість, а потім – прогнозувати й експериментально перевіряти кореляцію у співвідношенні форми і змісту – граматичної і лексичної семантики – досліджуваних одиниць. Книжковому варіантові словника, як правило, відповідає одиничний кількісний параметр – кількість словникових статей, а єдиним пошуковим апаратом є алфавіт, за яким впорядковано реєстрові одиниці словника. Комп'ютерна форма словника дає змогу широко оперувати будь-якими системно значущими характеристиками лексики, закладеними розроблювачами [6, с. 36].

Комп'ютерна лексикографія нараховує велику кількість різноманітних словників, що свідчить про її успішний розвиток внаслідок дедалі більших можливостей комп'ютера в лексикографії, серед яких слід згадати [5]:

1. Автоматична правка передрукованих текстів, що базується на машиночитаному орфографічному словнику, який є відкритою системою, постійно поповнюваною новими одиницями, що раніше не зустрічалися в текстах.

2. Зіставлення словників, наявних у пам'яті комп'ютера чи мережі комп'ютерів, формування зведеного реєстру зіставлюваних словників.

3. Збирання матеріалу, створення електронної картотеки на базі попередньо розмічених текстів, формування машинних фондів різного призначення, а, отже, й різної структури.

4. Створення системи автоматичного конвертування паперового словника відповідно до вимог комп'ютерної лексикографії.

5. Поповнення словника в автоматичному режимі опрацювання інформації.

6. Укладання словопокажчиків, конкордансів, частотних словників без використання чи з використанням програм лематизації.

7. Використання мультимедійних засобів представлення інформації: текст, колір, зображення, звуковий супровід.

8. Укладання перекладних словників на основі машинних корпусів паралельних текстів.

9. Укладання автоматичних словників.

10. Створення інтегрованих лексикографічних комп'ютерних систем.

Електронні словники перевершують паперові аналоги за своєю функціональністю, маючи при цьому ряд переваг [2, 8]:

1. Багатофункціональність — різноманітність додаткових функцій, які спрощують звернення до словника. Наприклад, можна вказати частини мови, походження, словотвір.

2. Використання засобів мультимедіа – озвучування заголовних слів, введення ілюстративного матеріалу з фотографіями, анімацією, відеофрагментами, а також використання різноманітних графічних засобів.

3. Актуальність і динамічність – можливість постійного оновлення інформації, а також вилучення застарілих даних. Це одна з важливих переваг перед “паперовими” словниками, оскільки вони неминуче застарівають на момент їх випуску.

4. Великий обсяг словникової бази. У більшості електронних словників термінологічна база перевищує базу паперових словників і надає зручніший доступ до інформації завдяки використанню гіперпосилань.

5. Варіативність у використанні – можливість використання словників у локальній і глобальній мережах, а саме використання офлайнової і онлайнової версії.

6. Універсальність – переважно програми дають змогу працювати відразу з декількома мовами і напрямками перекладу. Можливе використання будь-якого із введених до словника мов як вхідне.

7. Зручний пошук – можливість використання ефективної системи пошуку (повнотекстовий пошук, одночасний пошук у декількох словниках, висока швидкість пошуку). Також зникає необхідність точно пам'ятати слово – програма сама запропонує варіанти за першими літерами. В електронних словниках для доступу до вмісту використовуються різноманітні лінгвістичні технології, такі як морфологічний і синтаксичний аналіз, повнотекстовий пошук, розпізнавання і синтез мови.

Лексикографічна база даних забезпечить комплексний підхід із застосуванням інформаційно-лексикографічного моделювання для встановлення формальних (граматична семантика) та змістових (лексична семантика) характеристик слів. Запропонована в УМІФ модель лексикографічної бази даних лексико-граматичної структури дієслів української мови компактно та зручно репрезентує цю структуру порівняно з текстом словникових статей традиційних тлумачних словників, у яких форма зберігання та спосіб запису лінгвістичної інформації задають фіксований і визначений специфікою самої друкованої книжки доступ до цих даних. При формуванні лексикографічної бази даних її структуротвірні елементи відіграють роль елементів структури бази даних, ідентифікаторів різноманітних відношень у цій базі та її пошукових параметрів [6, с. 47].

Із використанням інформаційної теорії лексикографічних систем В. А. Широкова у дослідженні [7, с. 15] розроблено модель лексикографічної бази даних “Дієслово”, яка представляє декомпозицію дієслівних словникових статей за основними компонентами.

Основні положення інформаційної теорії лексикографічних систем можна сформулювати так [9]:

1. Структурна модель лексикографічних систем ґрунтується на багатоаспектному представленні знакової природи лексичних одиниць як центральних в системі будь-якої природної мови із врахуванням дихотомічної структури у відображенні співвідношення їх форми та змісту.

2. Дихотомічність структури слова в лексикографічній практиці (традиційних словниках) виявляється у:

- розмежуванні в словниковій статті  $V(x)$ , що подає опис заголовкового слова  $x$ , лівої  $L(x)$  і правої  $P(x)$  частин, які співвідносяться, відповідно, з формою і змістом мовного знака;
- констатуванні наявності спеціального оператора  $H$ , що забезпечує єдність словникової статті, й змістово визначає відображення її реєстрової та інтерпретаційної частин:  $H : L(x) \rightarrow P(x)$  (цей оператор у традиційних словниках зазвичай є імпліцитним і задається декларативно), а також множини автоморфізмів  $A$  – міжстатейних відображень або зв'язків між окремими словниковими статтями та їх структурними елементами;

3. Відображення  $H$  та  $A$  породжують макроструктуру лексикографічної системи; окрім цього, кожна лексикографічна система має свою мікроструктуру, в якій відбивається в неявному вигляді семантика предметної галузі, що є об'єктом цієї конкретної лексикографічної системи (мікроструктура стосується будови об'єктів  $V(x)$ ,  $L(x)$  і  $P(x)$ );

4. Структура, визначена твердженнями 1–3, рекурсивно продовжується на об'єкти  $V(x)$ ,  $L(x)$ ,  $P(x)$ ,  $H$ ,  $A$  та будь-які елементи їхньої структури (принцип рекурсивної редукції лексикографічних систем).

Відповідно до засад інформаційної теорії лексикографічних систем будь-який тлумачний словник як лексикографічна система з класом елементарних інформаційних одиниць “СЛОВО” підтримує структуру рекурсивної редукції другого порядку і набуває вигляду [9]:

$$\begin{array}{c}
 V(I^W(U)) = (L_0(I^W(U)); P_0(I^W(U)); H_0(I^W(U))) \\
 \begin{array}{ccc}
 & \longleftarrow & \\
 & L_0(I^W(U)) & H(I_0^W(U)) & P_0(I^W(U)) \\
 & \longrightarrow & \\
 \begin{array}{ccc}
 \longleftarrow & & \longrightarrow \\
 L^L_{01} & H^L_{01} & P^L_{01} \\
 \longrightarrow & & \longleftarrow \\
 L^P_{01} & H^P_{01} & P^P_{01}
 \end{array}
 \end{array}
 \end{array}
 \quad (1)$$

де символом  $I^W(U)$  позначено клас слів мови  $U$ , в нашому випадку – української; структурний елемент  $L_0(I^W(U))$  інтерпретується як носій граматичної, а  $P_0(I^W(U))$  – лексичної семантики. Їхній зв'язок і, отже, поєднання лінгвістичного об'єкта в єдине ціле забезпечує оператор  $H_0(I^W(U))$ . Своєю чергою, структурний елемент  $L_0(I^W(U))$ , яким репрезентується граматична семантика, інтерпретується як лексикографічна система, що підтримує відношення словозміни, словотвору, а також певні інші морфолого-дериваційні відношення на рівні лексичної системи. У конструкції репрезентанта лексичної семантики  $P_0(I^W(U))$  представлено цілу низку семантичних відношень, явно виділено ієрархію лексичних значень (для кожної лексеми  $x$  вони зосереджені в структурному елементі  $L^P_{01}(x)$  і представлені тлумаченнями) та відповідних прикладів слововживань, репрезентованих елементами  $P^P_{01}(x)$ . В елементах  $L^P_{01}(x)$  зосереджується й певна частина ідентифікаторів відношень синонімії, антонімії, меронімії, гіпо- та гіперонімії тощо, повним носієм яких є  $L^P_{01}(I^W(U))$  на всій множині описаних у словнику лексем.

Під час формування лексикографічної бази даних “Дієслово” було проведено такі етапи роботи [6]:

ЕТАП I. Скануванням та розпізнаванням отримано електронний варіант тексту тлумачного словника української мови в 11-ти томах.

ЕТАП II. Електронний текст СУМа наведено у вигляді масиву окремих словникових статей.

ЕТАП III. На основі формальних ознак, застосованих у СУМі, з позначеного у попередньому пункті масиву автоматично виділено дієслівні словникові статті.

Формальною ознакою при цьому стала видова ремарка – *док.* чи *недок.* (та їх різновиди), яка послідовно і регулярно представлена при лексикографічному описі дієслів. Саме видовий параметр є ідентифікаційним для розпізнавання дієслівних словникових статей: категорія виду характеризує кожен дієслівну форму. Отже, видова ремарка як обов'язкова позначається лише при дієсловах і не вживається при словах інших частин мови.

На цьому етапі автоматично було виділено 24550 повноцінних (невідсилкового типу) дієслівних словникових статей. До цієї вибірки не потрапили відсилкові дієслівні словникові статті, ліва частина яких містить лише реєстрове дієслово, а права за допомогою тлумачення відсилкового типу з елементом “Див.” відсилає до тієї словникової статті, в якій лексикографічно опрацьовано це реєстрове дієслово як один із компонентів реєстрового ряду. Ліва частина таких дієслів не містить видової ремарки, а права з тлумаченням відсилкового типу з елементом “Див.”, широко використовуючись при лексикографуванні не лише дієслів, не може бути ідентифікатором словникової статті як дієслівної.

Застосовуючи формалізм інформаційної теорії лексикографічних систем, через  $V(x)$  було позначено лексикографічний опис елементарної інформаційної одиниці (заголовкового слова словникової статті)  $x$  із множини реєстрових одиниць аналізованого словника. У такий спосіб до лексикографічної бази даних “Дієслово” занесено лексикографічні описи усієї дієслівної лексики СУМа. Цей опис представлено у вигляді лінійного тексту словникових статей невідсилкового типу (тобто в лексикографічній базі даних на цьому етапі не заносяться окремі записи, які містять відсилковий ідентифікатор “Див.”). У традиційних словниках лексикографічна інтерпретація

лексики через “Див.” є, окрім одного із способів представлення зв’язків у системі мови, також важливим прийомом економії місця, що непринципово для комп’ютерних лексикографічних баз даних.

ЕТАП IV. Автоматично декомпоновано масив дієслівних словникових статей, отриманий у результаті проведення ЕТАПу III, на ліву та праву частини.

ЕТАП V. Автоматично у правих частинах словникових статей дезактуалізовано фразеологічну частину та ілюстративний матеріал.

ЕТАП VI. Із використанням інформаційної теорії лексикографічних систем декомпоновано дієслівні словникові статті за такими основними компонентами:

- реєстрове дієслово (поле Word);
- ліва частина словникової статті, у якій отримує лексикографічне опрацювання реєстрове дієслово (поле Lpart);
- номер семантичного стану (поле Param) (окремі значення нумеруються арабськими цифрами – 1., 2., ... 30, а їх відтінки – після скісних рисок у межах значень – 1. // 1., 1 // 2., ... 28. // 1). (Семантичний стан – це конкретне значення реєстрового слова, що формально фіксується у словнику рубрикою основного значення чи його відтінку);
- власне тлумачення (поле Interpr).

Отже, сформована лексикографічна база даних, акцентуючи на виокремленні кожного ЛСВ (поле Param та поле Interpr), зафіксованого у тексті дієслівних словникових статей, дублює деяку інформацію (поле Word та поле Lpart). Таке дублювання уможливить надалі встановлення зв’язку лексичної семантики, що міститься у полі Interpr, з граматичними параметрами, поданими у полі Lpart.

Лексикографічна система СУМа у вигляді сукупності словникових статей із формально виокремленими структурно-функціональними компонентами (реєстрове дієслово, ліва частина словникової статті, номер семантичного стану, власне тлумачення) являє собою первинну лексикографічну базу даних “Дієслово”. Наступні лексикографічні бази даних, отримані за допомогою декомпозиції даних цієї бази, будуть вторинними лексикографічними базами даних, які, своєю чергою, будуть первинними для створення наступних лексикографічних баз даних. Лексикографічна база даних “Дієслово” являє собою автоматизовану інформаційну систему підтримки діяльності лінгвіста. Якщо застосувати інформаційну теорію лексикографічних систем до інтерпретації правих частин  $P(x)$  дієслівних словникових статей СУМа, то цілком формально – за спеціальними розділовими позначками, позиційним та шрифтовим виділенням – встановлюється їх структура. Отже, застосування інформаційної теорії лексикографічних систем до інтерпретації правих частин  $P(x)$  дієслівних словникових статей СУМа веде до встановлення формальних репрезентантів їхньої структури. Ними є [6]:

- факультативне уточнення параметрів граматичної семантики (ПАР, ВИД, ПЕРЕХ, КЕРУВ, СПОЛ, СТИЛЬ);
- тлумачення основного значення ( $S$ );
- тлумачення відтінку значення ( $V$ );
- ілюстрація до основного значення ( $I^S$ );
- ілюстрація до відтінку значення ( $I^V$ ).

У такий спосіб в УМІФ встановили формальну структуру правих частин усіх дієслівних словникових статей, що надало можливість для автоматичного виділення їхніх структурних елементів. Після розроблення необхідного алгоритмічно-програмного інструментарію із первинної лексикографічної бази даних “Дієслово” автоматично було одержано нову лексикографічну базу даних – “DiesLPart”, що має таку структуру [7]:

- **Word** – реєстрове дієслово (це поле відповідає полю **Compon** лексикографічної бази даних “DiesLPart”);
- **MWord** – дієслово-семантична тема, що автоматизовано встановлювалася шляхом індексування тлумачень;

- **Lpart** – ліва частина словникової статті реєстрового дієслова;
- **Param** – номер семантичного стану;
- **Interpr** – тлумачення.

Як бачимо, за допомогою інформаційної теорії лексикографічних систем в УМІФ було побудовано лексикографічні бази даних “Дієслово” та “DieslPPart”. Сам принцип рекурсивної редукції лексикографічних систем, який притаманний цій теорії, дає змогу доповнювати згадані лексикографічні бази даних. Крім того, аналогічно можна будувати і розширювати лексикографічні бази даних для інших частин мови.

### Формулювання мети статті

Мета статті — побудувати лексикографічну базу даних “VerbSemClass” на основі лексико-семантичної класифікації дієслів української мови розширенням лексикографічної бази даних “DieslPPart”, розробленої в УМІФ. Об’єктом дослідження є лексико-семантичні групи дієслів української мови, виділені нами. Предметом дослідження є лексикографічна база даних на основі лексико-семантичних груп дієслів української мови.

### Виклад основного матеріалу

Декомпозицією поля “MWord” з лексикографічної бази даних “DieslPPart”, розробленої в УМІФ [7], було одержано лексикографічну базу даних “VerbSemClass”, що має таку структуру:

- **Word\_ID** – ідентифікатор дієслова-семантичної теми;
- **Word** – дієслово-семантична тема, що відповідає полю “MWord” з лексикографічної бази даних “DieslPPart”;
- **LexSemGroup** – лексико-семантична група;
- **LexSemSubGroup** – лексико-семантична підгрупа.

Таблиця 1

Word_ID	Word	LexSemGroup	LexSemSubGroup
1	Поговорити	Дієслова дії та діяльності	Мовлення
2	Думати	Дієслова дії та діяльності	Мислення
3	Пізнавати	Дієслова дії та діяльності	Пізнання

Лексикографічна база даних “VerbSemClass” складається з трьох таблиць:

1. Головна таблиця “SemClass” має таку будову:
  - **Word\_ID** – ідентифікатор дієслово-семантичної теми;
  - **Word** – дієслово-семантична тема, що відповідає полю “MWord” з лексикографічної бази даних “DieslPPart”;
  - **LexSemGroup\_ID** – ідентифікатор лексико-семантичної групи;
  - **LexSemSubGroup\_ID** – ідентифікатор лексико-семантичної підгрупи.
  - Таблиця лексико-семантичних груп “LexSemGroup” складається з таких полів:
    - **LexSemGroup\_ID** – ідентифікатор лексико-семантичної групи;
    - **LexSemGroup\_Name** – назва лексико-семантичної групи;
    - **Comment** – нотатка.
2. Таблиця лексико-семантичних підгруп “LexSemSubGroup” має такі поля:
  - **LexSemSubGroup\_ID** – ідентифікатор лексико-семантичної підгрупи;
  - **LexSemSubGroup\_Name** – назва лексико-семантичної підгрупи;
  - **LexSemGroup\_ID** – ідентифікатор лексико-семантичної групи;
  - **Comment** – нотатка.

Для відображення структури лексикографічної бази даних “VerbSemClass” за допомогою програмних засобів було обрано систему керування базами даних з реляційним способом організації бази даних – Microsoft Access 2010.

У нашій лексикографічній базі даних було виділено такі зв'язки між таблицями, як: “один до багатьох” і “багато до багатьох” (рис. 1).

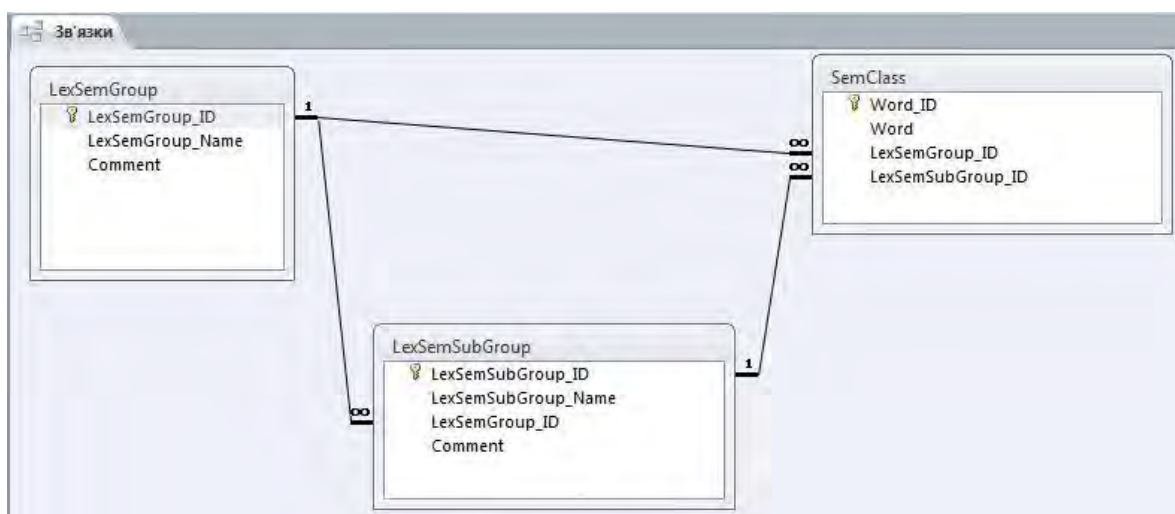


Рис. 1. Зв'язки між таблицями в лексикографічній базі даних “VerbSemClass”

Для отримання різноманітних вибірок з нашої лексикографічної бази даних ми створили набір найнеобхідніших запитів.

Серед створених запитів є такі:

1. **Count\_SemClass\_By\_LSG** – порахувати кількість дієслів, які входять до певної лексико-семантичної групи, наприклад “Дієслова дії та діяльності”:

```

SELECT Count(*) FROM (SemClass INNER JOIN LexSemSubGroup ON
[SemClass].LexSemSubGroup_ID=LexSemSubGroup.LexSemSubGroup_ID) INNER
JOIN LexSemGroup ON
[LexSemSubGroup].LexSemGroup_ID=LexSemGroup.LexSemGroup_ID WHERE
LexSemGroup_Name='Дієслова дії та діяльності';
    
```

Цей запит рахує кількість дієслів, які належать до лексико-семантичної групи з назвою “Дієслова дії та діяльності” з об’єднаних таблиць SemClass, LexSemSubGroup та LexSemGroup за спільними полями LexSemSubGroup\_ID та LexSemGroup\_ID.

Як результат запит видасть число 1755.

2. **Count\_SemClass\_By\_LSSG** – порахувати кількість дієслів, які входять до певної лексико-семантичної підгрупи, наприклад, “Дії”:

```

SELECT Count(*) FROM (SemClass INNER JOIN LexSemSubGroup ON
[SemClass].LexSemSubGroup_ID=LexSemSubGroup.LexSemSubGroup_ID) INNER
JOIN LexSemGroup ON
[LexSemSubGroup].LexSemGroup_ID=LexSemGroup.LexSemGroup_ID WHERE
LexSemSubGroup_Name='Дії';
    
```

Цей запит рахує кількість дієслів, які належать до лексико-семантичної підгрупи з назвою “Дії” з об’єднаних таблиць SemClass, LexSemSubGroup та LexSemGroup за спільними полями LexSemSubGroup\_ID та LexSemGroup\_ID.

Як результат запит видасть число 1471.

3. **View\_All\_SemClass** – показати досліджувані дієслова і відповідні їм лексико-семантичні групи і лексико-семантичні підгрупи:

```

SELECT Word_ID, Word, LexSemSubGroup_Name, LexSemGroup_Name
FROM (SemClass INNER JOIN LexSemSubGroup ON
    
```

```
[SemClass].LexSemSubGroup_ID=LexSemSubGroup.LexSemSubGroup_ID) INNER
JOIN LexSemGroup ON
[LexSemSubGroup].LexSemGroup_ID=LexSemGroup.LexSemGroup_ID ORDER BY
Word_ID;
```

Цей запит видає всі досліджувані дієслова і назви відповідних їм лексико-семантичних груп і лексико-семантичних підгруп з об'єднаних таблиць SemClass, LexSemSubGroup та LexSemGroup за спільними полями LexSemSubGroup\_ID та LexSemGroup\_ID.

Результат цього запиту продемонстровано на рис. 2.

Word_ID	Word	LexSemSubGroup_Name	LexSemGroup_Name
1	АНАЛІЗУВАТИ	Мислення	Дієслова дії та діяльності
2	АПЛОДУВАТИ	Дії	Дієслова дії та діяльності
3	АТАКУВАТИ	Дії	Дієслова дії та діяльності
4	БАЇТИСЯ	Діяльності	Дієслова дії та діяльності
5	БАЖАТИ	Дії	Дієслова дії та діяльності
6	БАЗІКАТИ	Мовлення	Дієслова дії та діяльності
7	БАЛАКАТИ	Мовлення	Дієслова дії та діяльності
8	БАРІТИСЯ	Дії	Дієслова дії та діяльності
9	БАЧИТИ	Сприйняття	Дієслова дії та діяльності
10	БАЧИТИСЯ	Діяльності	Дієслова дії та діяльності
11	БЕЗЧІСТИТИ	Дії	Дієслова дії та діяльності
12	БЕНКЕТУВАТИ	Діяльності	Дієслова дії та діяльності
13	БЕНТІЖИТИ	Психоемоційного стану	Дієслова стану і процесу
14	БЕРЕГТИ	Дії	Дієслова дії та діяльності
15	БЕРЕГТИСЯ	Дії	Дієслова дії та діяльності
16	БЕШКЕТУВАТИ	Дії	Дієслова дії та діяльності
17	БИТИ	Дії	Дієслова дії та діяльності
18	БИТИСЯ	Дії	Дієслова дії та діяльності
19	БІГАТИ	Руху та переміщення	Дієслова руху та переміщення
20	БІГТИ	Руху та переміщення	Дієслова руху та переміщення
21	БІЛИТИ	Дії	Дієслова дії та діяльності

Рис. 2. Результат запиту View\_All\_SemClass

4. **View\_LSG\_LSSG\_Relations** – показати відношення між кожною лексико-семантичною групою і відповідною їй лексико-семантичною підгрупою:

```
SELECT LexSemSubGroup_ID, LexSemSubGroup_Name, LexSemGroup_Name
FROM LexSemSubGroup INNER JOIN LexSemGroup ON
LexSemSubGroup.[LexSemGroup_ID]=LexSemGroup.[LexSemGroup_ID];
```

Цей запит видає ідентифікатор лексико-семантичної підгрупи та назви лексико-семантичних підгруп і лексико-семантичних груп з об'єднаних таблиць LexSemSubGroup та LexSemGroup за спільним полем LexSemGroup\_ID.

Результат цього запиту продемонстровано на рис. 3.

5. **View\_SemClass\_By\_LSG** – показати дієслова, які належать до певної лексико-семантичної групи, наприклад “Дієслова дії та діяльності”:

```
SELECT Word_ID, Word, LexSemSubGroup_Name, LexSemGroup_Name
FROM (SemClass INNER JOIN LexSemSubGroup ON
[SemClass].LexSemSubGroup_ID=LexSemSubGroup.LexSemSubGroup_ID) INNER
JOIN LexSemGroup ON
[LexSemSubGroup].LexSemGroup_ID=LexSemGroup.LexSemGroup_ID
WHERE LexSemGroup_Name='Дієслова дії та діяльності' ORDER BY
Word_ID;
```



LexSemSubGroup_ID	LexSemSubGroup_Name	LexSemGroup_Name
1	Початок буття	Дієслова буття
2	Власне буття	Дієслова буття
3	Кінець буття	Дієслова буття
4	Буттєво-просторові дієслова	Дієслова буття
5	Початок	Фазові дієслова
6	Кінець	Фазові дієслова
7	Власне відношення	Дієслова відношення
8	Залежності	Дієслова відношення
9	Приналежності	Дієслова відношення
10	Фізичного стану	Дієслова стану і процесу
11	Психоемоційного стану	Дієслова стану і процесу
12	Фізіологічного стану	Дієслова стану і процесу
13	Перехід від одного стану до іншого	Дієслова стану і процесу
14	Звучання та звуконаслідування	Дієслова стану і процесу
15	Руху та переміщення	Дієслова руху та переміщення
16	Дії	Дієслова дії та діяльності
17	Діяльності	Дієслова дії та діяльності
18	Мовлення	Дієслова дії та діяльності
19	Мислення	Дієслова дії та діяльності
20	Пізнання	Дієслова дії та діяльності
21	Сприйняття	Дієслова дії та діяльності
22	Можливості	Модальні дієслова
23	Примусу	Модальні дієслова
24	Дозволу	Модальні дієслова

Рис. 3. Результат запити *View\_LSG\_LSSG\_Relations*

Цей запит видає ідентифікатор дієслова-семантичної теми, саме дієслово-семантичну тему та назви відповідних лексико-семантичних підгруп і лексико-семантичних груп для лексико-семантичної групи за назвою “Дії та діяльності” з об’єднаних таблиць *SemClass*, *LexSemSubGroup* та *LexSemGroup* за спільними полями *LexSemSubGroup\_ID* та *LexSemGroup\_ID*. Дані посортовані за ідентифікатором дієслова-семантичної теми у порядку зростання.

Результат цього запити продемонстровано на рис. 4.

Word_ID	Word	LexSemSubGroup_Name	LexSemGroup_Name
114	ВІКИДАТИ	Дії	Дієслова дії та діяльності
115	ВИКЛАДАТИ	Дії	Дієслова дії та діяльності
116	ВИКЛИКАТИ	Мовлення	Дієслова дії та діяльності
117	ВИКЛЮЧУВАТИ	Дії	Дієслова дії та діяльності
118	ВИКЛЮЧАТИ	Дії	Дієслова дії та діяльності
119	ВИКОНУВАТИ	Дії	Дієслова дії та діяльності
120	ВИКОНУВАТИСЯ	Дії	Дієслова дії та діяльності
121	ВИКОРИСТОВУВАТИ	Дії	Дієслова дії та діяльності
122	ВИКОРИСТОВУВАТИСЯ	Дії	Дієслова дії та діяльності
123	ВИКОРИНЮВАТИ	Дії	Дієслова дії та діяльності
124	ВИКРАДАТИ	Дії	Дієслова дії та діяльності
125	ВИКРИВАТИ	Дії	Дієслова дії та діяльності
126	ВИКРИВЛЯТИ	Дії	Дієслова дії та діяльності
127	ВИКРИВЛЯТИСЯ	Дії	Дієслова дії та діяльності
128	ВИКРИКУВАТИ	Мовлення	Дієслова дії та діяльності
129	ВИКРІЧУВАТИ	Дії	Дієслова дії та діяльності
130	ВИКРІЧУВАТИСЯ	Дії	Дієслова дії та діяльності
133	ВИЛАМУВАТИ	Дії	Дієслова дії та діяльності
134	ВИЛАЯТИ	Мовлення	Дієслова дії та діяльності
135	ВИЛИВАТИ	Дії	Дієслова дії та діяльності
136	ВИЛИВАТИСЯ	Дії	Дієслова дії та діяльності
137	ВИЛІКОВУВАТИ	Діяльності	Дієслова дії та діяльності
138	ВИЛІКОВУВАТИСЯ	Діяльності	Дієслова дії та діяльності
141	ВИЛЮВЛЮВАТИ	Дії	Дієслова дії та діяльності
142	ВИЛУЧАТИ	Дії	Дієслова дії та діяльності
145	ВИМИВАТИ	Дії	Дієслова дії та діяльності
146	ВИМОВЛЯТИ	Мовлення	Дієслова дії та діяльності

Рис. 4. Результат запити *View\_SemClass\_By\_LSG*

6. **View\_SemClass\_By\_LSSG** – показати дієслова, які належать до певної лексико-семантичної підгрупи, наприклад, “Дії”:

```
SELECT Word_ID, Word, LexSemSubGroup_Name, LexSemGroup_Name
FROM (SemClass INNER JOIN LexSemSubGroup ON
[SemClass].LexSemSubGroup_ID=LexSemSubGroup.LexSemSubGroup_ID) INNER
JOIN LexSemGroup ON
[LexSemSubGroup].LexSemGroup_ID=LexSemGroup.LexSemGroup_ID WHERE
LexSemSubGroup_Name='Дії' ORDER BY Word_ID;
```

Цей запит видає ідентифікатор дієслово-семантичної теми, саме дієслово-семантичну тему та назви відповідних лексико-семантичних підгруп і лексико-семантичних груп для лексико-семантичної підгрупи з назвою “Дії” з об’єднаних таблиць SemClass, LexSemSubGroup та LexSemGroup за спільними полями LexSemSubGroup\_ID та LexSemGroup\_ID. Дані посортовано за ідентифікатором дієслово-семантичної теми за зростанням.

Результат цього запиту продемонстровано на рис. 5.

Word_ID	Word	LexSemSubGroup_Name	LexSemGroup_Name
919	КЛЯ́Кнути	Дії	Дієслова дії та діяльності
928	КОЛО́ти	Дії	Дієслова дії та діяльності
929	КОНОПА́тити	Дії	Дієслова дії та діяльності
930	КОПА́ти	Дії	Дієслова дії та діяльності
931	КОПА́ти	Дії	Дієслова дії та діяльності
932	КОПЮ́вати	Дії	Дієслова дії та діяльності
933	КОПІ́ти	Дії	Дієслова дії та діяльності
934	КОРИ́стуватися	Дії	Дієслова дії та діяльності
936	КО́рчити	Дії	Дієслова дії та діяльності
937	КО́рчитися	Дії	Дієслова дії та діяльності
940	КО́штувати	Дії	Дієслова дії та діяльності
941	КРА́сти	Дії	Дієслова дії та діяльності
942	КРАСУ́ватися	Дії	Дієслова дії та діяльності
943	КРЕ́слити	Дії	Дієслова дії та діяльності
944	КРИ́вдити	Дії	Дієслова дії та діяльності
947	КРИШІ́тися	Дії	Дієслова дії та діяльності
948	КРО́їти	Дії	Дієслова дії та діяльності
949	КРОПІ́ти	Дії	Дієслова дії та діяльності
953	КУВА́ти	Дії	Дієслова дії та діяльності

Рис. 5. Результат запиту *View\_SemClass\_By\_LSSG*

### Висновки та перспективи подальших наукових розвідок

Окреслено переваги комп’ютерної лексикографії над традиційною, зокрема багатофункціональність, актуальність і динамічність, використання засобів мультимедіа, великий обсяг словникової бази, варіативність у використанні, універсальність та зручний пошук. Розглянуто основні положення інформаційної теорії лексикографічних систем В. А. Широкова, яку покладено в основу дослідження.

Обґрунтовано доцільність подання семантичних характеристик дієслів у лексикографічній базі даних. На основі інформаційної теорії лексикографічних систем побудовано лексикографічну базу даних лексико-семантичних груп дієслів української мови “VerbSemClass”, яка є доповненням до лексикографічної бази даних “DieslPPart”, розробленої в Українському мовно-інформаційному фонді НАНУ України.

1. Апресян Ю. Д. *Интегральное описание языка и толковый словарь* / Ю. Д. Апресян // *Вопросы языкознания*. – 1986. – № 2. – С. 57–70.
2. Кашеварова И. С. *Электронный словарь как новый этап в развитии лексикографии* // *Молодой ученый*. – 2010. – № 10 (21). – С. 145–147.
3. Кульчицький І. М., Лукач М. О. *Формальна модель відображення семантичних характеристик у словниках* / І. М. Кульчицький, М. О. Лукач // *Інформаційні системи та мережі: Вісник Нац. ун-ту*

“Львівська політехніка”. – 2014. – № 805. – С. 298–305. 4. Кустова Г. И. Словарь как лексическая база данных / Г. И. Кустова, Е. В. Падучева // Вопросы языкознания, 1994. – № 4. – С. 96–106. 5. Перебийніс В. І., Сорокін В. М. Традиційна та комп'ютерна лексикографія: навч. посібник. – К.: Вид. центр КНЛУ, 2009. – 218 с. 6. Рабулець О. Г. Дієслово в лексикографічній системі / О. Г. Рабулець, Н. М. Сухарина, В. А. Широков, К. М. Якименко; НАН України; Укр. мов.-інформ. фонд. – К.: Довіра, 2004. – 259 с. 7. Сухарина Н. М. Граматична та лексична семантика українського дієслова в лексикографічній системі: автореф. дис. ... канд. філол. наук: 10.02.01 / Н. М. Сухарина; НАН України. Ін-т мовознав. ім. О. О. Потебні. – К., 2003. – 19 с. 8. Чепик Е. Ю. Политическое слово в структуре электронного словаря // Культура народов Причерноморья. 2005. – № 69. – С. 205–209. 9. Широков В. А. Інформаційна теорія лексикографічних систем / В. А. Широков. – К.: Довіра, 1998. – 331 с. 10. Широков В. А. Технологічні основи сучасної тлумачної лексикографії / В. А. Широков, О. Г. Рабулець, О. М. Костишин, І. В. Шевченко, К. М. Якименко // Мовознавство, 2002. – № 6. – С.49–86.

УДК 004.9

В. В. Литвин, В. А. Висоцька, Д. Г. Досин, М. Г. Гірняк  
Національний університет “Львівська політехніка”,  
кафедра інформаційних систем та мереж

## РОЗРОБЛЕННЯ МЕТОДІВ ТА ЗАСОБІВ ПОБУДОВИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ОПРАЦЮВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ З ВИКОРИСТАННЯМ ОНТОЛОГІЧНОГО ПІДХОДУ

© Литвин В. В., Висоцька В. А., Досин Д. Г., Гірняк М. Г., 2015

Розглянуто питання розроблення методів та програмних засобів опрацювання інформаційних ресурсів. Розроблено формальний опис процесу опрацювання текстового контенту. Сформульовано новий підхід до застосування та впровадження бізнес-процесів.

**Ключові слова:** контент, онтологія, інформаційний ресурс, бізнес-процес, система управління контентом, життєвий цикл контенту.

The article discusses the development of unified methods and software tools for processing information resources. The models of information resource processing are proposed. A new approach to application and implementation of business processes is formulated.

**Key words:** content, ontology, information resource, business-process, content management system, content lifecycle.

### Вступ. Загальна постановка проблеми

Зміст дослідження полягає у вирішенні важливої науково-прикладної проблеми побудови інтелектуальних систем опрацювання інформаційних ресурсів (ядром баз знань яких є онтології) для аналізу текстових масивів даних при формуванні спільних тематик досліджень у міжвузівських консорціумах. Використання онтологій у складі баз знань інтелектуальних систем опрацювання інформаційних ресурсів допомагає вирішити низку методологічних та технологічних проблем, які виникають під час розроблення таких систем. Зокрема, для України це відсутність концептуальної цілісності й узгодженості окремих прийомів та методів інженерії знань; нестача кваліфікованих фахівців у цій галузі; жорсткість розроблених програмних засобів та їх низька адаптивна здатність;