

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
Національна академія наук України
Національна бібліотека України імені В. І. Вернадського

Кваліфікаційна наукова
праця на правах рукопису

МАЦЮК ГАЛИНА РОСТИСЛАВІВНА

УДК 025.4.06:303.436.2:004/001.4

ДИСЕРТАЦІЯ
**ТЕЗАУРУС ЯК ІНСТРУМЕНТ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ
МІЖДИСЦИПЛІНАРНИХ НАУКОВИХ ДОСЛІДЖЕНЬ**

27.00.03 – Книгознавство, бібліотекознавство, бібліографознавство
(27 – соціальні комунікації)

Подається на здобуття наукового ступеня кандидата наук із соціальних
комунікацій

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

Г. Р. Мацюк

Науковий керівник:
Кунанець Наталія Едуардівна, доктор наук
із соціальних комунікацій, професор, с.н.с.

Київ – 2020

АНОТАЦІЯ

Мацюк Г. Р. Тезаурус як інструмент інформаційного забезпечення міждисциплінарних наукових досліджень. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата наук із соціальних комунікацій за спеціальністю 27.00.03 – книгознавство, бібліотекознавство, бібліографознавство (27 – соціальні комунікації). – Тернопільський національний технічний університет імені Івана Пулюя, Національна бібліотека України імені В. І. Вернадського. Київ, 2020.

У дисертаційному дослідженні подано результати комплексного дослідження й узагальнення теоретичних і прикладних аспектів формування інформаційно-пошукових тезаурусів та представлено розроблений автором тезаурус предметної області «Розумне місто», який використовується як ефективний інструмент інформаційного забезпечення міждисциплінарних наукових досліджень.

Здійснено огляд основних видів та змістовного наповнення інформаційно-пошукових тезаурусів, які використовуються у зарубіжних та вітчизняних бібліотеках, як інформаційний ресурс структури певної предметної області та інструмент формування пошукових образів документів і пошукових запитів у бібліотеках.

Проаналізовано моделі подання знань для вибору ефективного способу представлення знань предметної області «Розумне місто». Основу будь-якої предметної області складає сукупність знань та даних про об'єкти й процеси, які підлягають проектуванню й зберіганню в базах знань та базах даних. Концепти слугують інформаційними одиницями, що мають зв'язки між собою. Форма подання знань вагомо впливає на якість формування пошукових образів документів, що зберігаються у фондах бібліотек. При формуванні метаданих доцільно використовувати структуровані, приведені до однозначного формалізованого виду концепти, що характеризують зміст документа.

Застосовано метод аналізу ієрархій для вибору моделі представлення знань предметної області «Розумне місто». Це дозволило обрати на науково-обґрунтованих підставах найзручніший і найінформативніший спосіб подання концептів даної предметної області для формування пошукових образів у автоматизованих бібліотечних інформаційних системах та представити його у формі тезауруса з використанням онтологічного підходу.

Розглянуто еволюцію тезаурусів від трактування їх у працях мислителів стародавнього світу, відзначено, що поняття пов'язане з фундаментальними проблемами пізнання, відображеннями картини світу та закономірностями сприйняття світу. У роботах: «Ономастикон» Юлія Поллукса, «Амаракоша» Амара, «Книга про скарб» Брунетто Латіні, «Тезаурус англійських слів і виразів» П. М. Роже та ін. окреслено засадничі моменти використання тезаурусів. Термін «тезаурус» у сучасному трактуванні розпочав використовуватися у сімдесятих роках ХХ ст. У цей період створювалися інформаційно-пошукові тезауруси для індексування документів та проведення їх як у ручному, так і автоматизованому режимі.

Уточнено визначення поняття «інформаційно-пошуковий тезаурус предметної області» як сукупність структурованих знань, поданих у вигляді набору дескрипторів та аскрипторів, які пов'язані між собою певними відношеннями, що відображають семантичні зв'язки між поняттями певної предметної області. Тезаурус дозволяє представити зміст концепту не тільки з допомогою його визначення, а й за допомогою співвіднесення його з іншими концептами або групами концептів. Таке співвіднесення, як правило, задається шляхом встановлення між концептами парадигматичних зв'язків. Наявність таких зв'язків створює передумови для використання тезаурусів у задачах індексування документів та інформаційного пошуку в бібліотечних системах.

Проаналізовано роль галузевих інформаційно-пошукових тезаурусів у процесах систематизації термінології предметних областей та можливість їхнього застосування як засобу підвищення ефективності пошуку в сучасних

автоматизованих бібліотечних інформаційних системах. Завдяки тезаурусу можна здійснювати змістовне опрацювання документів, формуючи уніфікований пошуковий образ документа й пошуковий образ запиту користувачів, формувати авторитетні файли, предметні рубрики або визначати ключові слова, індексувати документи й запити засобами дескрипторної мови, замінюючи ключові слова відповідними концептами, уніфікувати та удосконалювати пошукові алгоритми електронних каталогів різних бібліотек.

Сформовано алгоритм розроблення двомовного інформаційно-пошукового тезауруса на основі консолідації принципів та підходів до побудови тезаурусів, який складається з чотирьох етапів: 1) визначення меж предметної області, пошук джерел та накопичення галузевої інформації; 2) формування реєстрової частини тезауруса; 3) побудова двомовних словникових статей для кожного концепту; 4) уточнення та редагування отриманого тезауруса. Алгоритм апробовано під час формування інформаційно-пошукового тезауруса предметної області «Розумне місто», яка формується як міждисциплінарна. Запропонований підхід сприяв вирішенню актуальних завдань у контексті інформаційного забезпечення користувачів, що проводять наукові дослідження на міждисциплінарних засадах: автоматизації процесів розроблення двомовних пошукових образів; створення алгоритмів формування пошукового запиту у вузькоспеціалізованій предметній області.

Розроблено двомовний інформаційно-пошуковий тезаурус «розумного міста», який складається з трьох частин: вступ, алфавітний лексико-семантичний показник термінів та пермутаційний показник. Алфавітний лексико-семантичний показник термінів, у якому подано всі дескриптори та аскриптори зі словниковими статтями, використовується для пошуку дескриптора та його парадигматичних зв'язків і забезпечує можливість якісного формулювання запитів. Пермутаційний показник використовується

для пошуку дескрипторів чи аскрипторів у відповідному словниковому гнізді за окремими словами, що входять до складу термінів-словосполучень.

Подано систему відношень між термінами тезауруса, яка містить два типи зв'язків: традиційні – ієрархічні й еквівалентні та проблемно-специфічні (неієрархічні або асоціативні). Традиційні ієрархічні відношення показують логічне підпорядкування дескрипторів і встановлюються, якщо відношення є істинним незалежно від контексту. Відношення еквівалентності встановлюється між дескриптором і аскриптором, тобто між термінами-синонімами, якщо заміна одного з них на інший не призводить до зміни змісту тексту, істотного для пошуку інформації. Використання асоціативних зв'язків дозволяє сформувати додаткові дескриптори, корисні при індексуванні чи пошуку.

Двомовний інформаційно-пошуковий тезаурус «розумного міста» використовується для формування уніфікованого пошукового образу документа й пошукового образу запиту користувачів, індексування документів, здійснення релевантного пошуку в електронних каталогах, базах даних бібліотек.

Наукова новизна одержаних результатів полягає в тому, що під час розв'язання поставлених дослідницьких завдань було:

у перше: застосовано метод аналізу ієрархії для вибору способу представлення знань предметної області «Розумне місто» при формуванні інформаційних ресурсів бібліотек; на основі аналізу наукових публікацій консолідовано підходи до формування двомовного тезауруса, який сприяє ефективному формуванню метаданих інформаційних ресурсів бібліотек, та сформовано алгоритм його розроблення; розроблено двомовний інформаційно-пошуковий тезаурус «розумного міста» на основі формалізованого представлення знань предметної області, що забезпечує інформаційний пошук релевантних документів у бібліотеках, сприяє підвищенню ефективності інформаційного забезпечення наукових

досліджень, слугує самостійним термінологічним довідником, який допомагає орієнтуватися в даній галузі;

удосконалено: визначення поняття «інформаційно-пошуковий тезаурус предметної області» для використання в бібліотечних інформаційних системах;

доповнено: класифікацію типів семантичних відношень з урахуванням міждисциплінарних тенденцій сучасного наукового знання (виділено ряд проблемно-специфічних відношень характерних для предметної області «Розумне місто»), що сприяло підвищенню інформаційного пошуку у бібліотеках;

набули подальшого розвитку: технологія поповнення бібліотечними працівниками тезауруса «розумного міста»: наповнення новими лексичними одиницями для двомовного відображення найновіших термінів галузі, відслідковування міжпредметних зв'язків; методи розроблення електронної версії двомовного тезауруса предметної області «Розумне місто» для бібліотечних інформаційних систем.

Ключові слова: інформаційно-пошуковий тезаурус, дескриптор, аскриптор, термін, онтологія, предметна область, інформаційно-пошукова система, інформаційний пошук, пошуковий образ, автоматизована бібліотечна інформаційна система, бібліотеки.

SUMMARY

Matsiuk H. R. Thesaurus as the instrument for information support of interdisciplinary scientific research. – Qualification scientific work on the rights of manuscripts.

Thesis for a Candidate Degree in Social Communications (PhD) by specialty 27.00.03 – book science, library science, bibliography. – Ternopil Ivan Puluj National Technical University, V. I. Vernadsky National Library of Ukraine. Kyiv, 2020.

The results of the comprehensive research and generalization of theoretical and applied aspects of the information retrieval thesauruses formation are given and developed by the author the “Smart City” domain-specific thesaurus which is used as an effective tool for informational support of interdisciplinary scientific researches is presented.

The overview of the main types and content of information retrieval thesauruses used in foreign and national libraries, as the information resource of the particular domain-specific structure and the tool for search document images and library retrieval requests formation is carried out.

Models of knowledge representation for choosing the effective way of the “Smart City” domain knowledge presentation are analyzed. The basis of any domain is the body of knowledge and data about objects and processes subjected to designing and storing in knowledge bases and databases. The concepts serve as interrelated informational units. The knowledge presentation form significantly influences the quality of the formation of search documents images stored in the libraries foundations. While forming metadata, it is reasonable to use structured, uniquely formalized concepts characterizing the document content.

The hierarchy analysis method is used in order to select the “Smart City” domain knowledge presentation model. This makes it possible to select the most convenient and informative way of presenting the concepts of the given domain for search images formation in automated information library systems on the scientifically proved basis, and presenting it in the form of thesaurus using ontological approach.

The thesauruses evolution from their interpretation in the ancient papers is considered, it is noted that the concept is connected with the fundamental problems of cognition, world picture reflections and world perception patterns. In papers: “Onomasticon” by Julius Pollux, “Amarokosha” by Amar, “The Book of Treasure” by Brunetto Latini, “Thesaurus of English Words and Phrases” by P. M. Roget and others the cornerstone principles of thesauruses use were outlined. The term “thesaurus” in modern interpretation came into common use in the seventies of the

XX century. During this period, information retrieval thesauruses were created for documents indexing in both manual and automated mode.

The definition “domain information retrieval thesaurus” is clarified as the set of structured knowledge, presented in the form of descriptors set interrelated by certain relationships reflecting semantic connections between the particular domain concepts. The thesaurus makes it possible to present the concept meaning not only by defining it, but by correlating it to other concepts or groups of concepts as well. This correlation is usually determined by establishing paradigmatic connections between concepts. The availability of such links creates the preconditions for the thesauruses use in document indexing and information retrieval in library systems.

The role of specific purposes information retrieval thesauruses in the processes of domain terminology systematization and the possibility of their use as the means of increasing search efficiency in modern automated information library systems are analyzed. Due to thesaurus it is possible to carry out the document meaningful processing, generating the unified document search image and the user search image request, to form the authority files, subject headings or to define keywords, to index documents and requests by descriptive language, replacing keywords with appropriate concepts, unifying and improving the search algorithms of e-catalogs of various libraries.

The algorithm for bilingual information retrieval thesaurus construction based on the consolidation of principles and approaches to thesauri construction is developed. It consists of four stages: 1) determination of the domain boundaries, search for sources and accumulating specialized information; 2) formation of the thesaurus register part; 3) construction of bilingual dictionary articles for each concept; 4) clarification and editing of the received thesaurus. The algorithm is tested during the “Smart City” domain information retrieval thesaurus construction which is being developed as the interdisciplinary one. The proposed approach provides the relevant problems solution in the context of information support for users carrying out the scientific research on interdisciplinary basis: processes

automation for the bilingual search images development; creation of algorithms for search request formation in the specialized domain.

The “Smart City” bilingual information retrieval thesaurus consisting of three parts: introduction, alphabetic lexical-semantic index of terms and permuted index is developed. The alphabetic lexical semantic index of terms listing all descriptors and non-descriptors with lexical entry is used in order to search the descriptor and its paradigmatic relationships and provide high-quality request formulation. The permutation term index containing descriptors and non-descriptors without lexical entry is used to search single keywords for the descriptors in appropriate group of words with the same root presented in one dictionary entry.

The system of relations between the thesaurus terms containing two types of relations: traditional - hierarchical and equivalent and problem-specific (non-hierarchical or associative) is developed. Traditional hierarchical relationships show the logical descriptors subordination and are established if the relationship is true regardless of context. The equivalence relation is established between descriptor and non-descriptor, i. e., between the terms-synonymous, if replacing one of them with another does not change the content of the text relevant to the information search. Associative links use makes it possible to generate additional descriptors useful for indexing or searching.

The “Smart City” bilingual information retrieval thesaurus is used to form the unified document search image and users request search image; document indexing; performing the relevant search in electronic catalogs and libraries databases.

The scientific novelty of the obtained results is as follows: while solving the given research assignments:

- for the first time: the hierarchy analysis method was applied in order to choose the way of the “Smart City” domain knowledge presentation while forming the libraries information resources; approaches to the bilingual thesaurus development contributing to the efficient formation of the libraries information

resources metadata were consolidated and the algorithm for its creation was developed on the basis of scientific publications analysis; the “Smart City” bilingual information retrieval thesaurus was developed on the basis of a formalized domain knowledge, which provides information retrieval of relevant documents in libraries; contributing to the effectiveness of scientific research information support; besides it also serves as a separate terminological guide assisting in the given field navigation;

- definition of the term “domain information retrieval thesaurus” for use in library information systems are improved;
- classification of types of semantic relations taking into account interdisciplinary tendencies of modern scientific knowledge that promoted increase of information search in libraries was supplemented;
- further developments: technologies for the “Smart City” thesaurus updating by the library workers: filling new lexical items for bilingual reflection of the latest domain terms, tracking the interdisciplinary relations; methods for the development of the electronic version of the “Smart City” domain bilingual Thesaurus for library information systems.

Key words: information retrieval thesaurus, descriptor, non-descriptor, term, ontology, domain, information retrieval system, information request, information search, search image, libraries, automated library information system.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті в наукових фахових виданнях України:

1. Мацюк Г. Представлення знань у вузькоспеціальних предметних областях за допомогою тезауруса / Г. Мацюк, Н. Кунанець // Бібліотечний вісник. – 2019. – № 1. – С. 19–23.

Розкрито роль галузевого інформаційно-пошукового тезауруса в процесах систематизації термінології предметної області.

2. Мацюк Г. Алгоритм формування двомовного тезауруса інформаційного масиву предметної області «Розумне місто» / Н. Кунанець, Г. Мацюк // Бібліотекознавство. Документознавство. Інформологія. – 2019. – № 1. – С. 110–117.

Проаналізовано технологічні засади створення двомовного тезауруса предметної області «Розумне місто».

3. Мацюк Г. Інформаційно-пошукові тезауруси: світовий та вітчизняний досвід формування / Г. Мацюк // Бібліотекознавство. Документознавство. Інформологія. – 2019. – № 2. – С. 106–115.

4. Мацюк Г. Р. Удосконалення технологій інформаційного пошуку на основі тезауруса предметної області / Г. Р. Мацюк // Бібліотечний вісник. – 2019. – № 5. – С. 20–25.

Статті в наукових періодичних виданнях інших держав:

5. Мацюк Г. Р. Підготовка масиву термінів для побудови двомовного тезауруса предметної області «Розумне місто» / Г. Р. Мацюк // Social and Human Sciences. PolishUkrainian scientific journal. – 2018. – № 3 (19). – URL : https://sp-sciences.io.ua/s2634980/matsiuk_halyna_2018_.stages_of_preparation_of_terms_array_for_creation_of_bilingual_thesaurus_of_the_smart_city_data_domain._social_and_human_sciences._polish-ukrainian_scientific_journal_03_19.

6. Matsiuk H. Selective Dissemination of Information – Technology of Information Support of Scientific Research / A. Rzhenskyi, H. Matsiuk,

N. Veretennikova, R. Vaskiv // Advances in Intelligent Systems and Computing III. – 2018. – Vol. 871. – P. 235–245. *(Індексовано в Scopus)*.

Подано технологію вибіркового поширення інформації.

7. Matsiuk H. Formation of Search Queries Based on Thesaurus of Narrowly Specialized Subject Areas / H. Matsiuk, A. Rzheuskyi, N. Kunanets, N. Veretennikova // Econtechmod. An international quarterly journal. – 2019. – Vol. 08. – No. 3. – P. 11 – 14.

Проаналізовано роль галузевого інформаційно-пошукового тезауруса для підвищення ефективності інформаційного забезпечення наукових досліджень, що здійснюється з використанням технологій вибіркового розповсюдження інформації.

8. Halyna Matsiuk Open online training courses for engineering purpose / Ihor Baran, Nataliya Kunanets, Halyna Matsiuk, Mykola Mytnyk, Khristina Shunevich, Yuriy Skorenky and Volodymyr Jaskilka // Workshop Proceedings of the VIII International Scientific Conference «Mathematics. Information Technology. Education», MoMLeT&DS'2019 (Ukraine, Shatsk, June 2-4, 2019). – Shatsk, 2019. – Vol. 2386. – Pp. 331-339. – ISSN 1613-0073. *(Індексовано в Scopus)*.

Запропоновано використовувати тезаурус при створенні онлайн-курсів дистанційного навчання інженерного спрямування.

9. Matsiuk H. Use of the Smart City Ontology for Relevant Information Retrieval / Nataliya Kunanets, Halyna Matsiuk // Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019) (Ukraine, Kharkiv, April 18-19, 2019). – Kharkiv, 2019. – Pp. 322–333. – ISSN 1613-0073. *(Індексовано в Scopus)*.

Подано алгоритм побудови онтології предметної області «Розумне місто».

Наукові праці, які засвідчують апробацію матеріалів дисертації:

10. Мацюк Г. Р. Онтологічне дослідження предметної області «Розумне місто» / Н. Е. Кунанець, В. В. Пасічник, Г. Р. Мацюк, О. В. Мацюк,

О. П. Левченко // Математика. Інформаційні технології. Освіта : VI Міжнарод. наук.-практ. конф. (Україна, Луцьк-Світязь, 5-7 червня 2017 р.). – Луцьк-Світязь, 2017. – С. 48–52.

Подано онтологічне дослідження предметної області «Розумне місто як основи для побудови бази знань про «інтелектуальне» місто, запропоновано класифікаційне дерево онтології «Розумне місто» на етапі представлення першого рівня.

11. Мацюк Г. Р. Двомовний тезаурус як соціокомунікаційна складова проекту «Розумне місто» / Г. Р. Мацюк, Н. Е. Кунанець // Современные информационные технологии в экономике и управлении предприятиями, программами и проектами: материалы XV Международной научно-практической конференции (Украина, Харьков, 29 сентября 2017 г.). – Харьков : Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт», 2017. – С. 85–87.

Обґрунтовано актуальність створення комплексного двомовного тезауруса як соціокомунікаційної складової проекту «Розумне місто».

12. Мацюк Г. Р. Вербальне моделювання предметної області засобами побудови тезауруса / Н. Е. Кунанець, Г. Р. Мацюк // Матеріали V науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 1-2 лютого 2018 р.). – Тернопіль, 2018. – С. 36.

Подано особливості вербального моделювання предметної області засобами побудови тезауруса, консолідовано підходи до технологій формування тезауруса.

13. Мацюк Г. Метадані інформаційних ресурсів у контексті інтеграції різнотипних даних у соціокомунікаційних інформаційних системах / Г. Липак, О. Липак, Н. Кунанець, Г. Мацюк // Матеріали V Міжнародної науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 1–2 лютого 2018 р.). – Тернопіль, 2018. – С. 38.

Здійснено порівняльний аналіз стандартів опису метаданих інформаційних ресурсів різних за профілем установ.

14. Мацюк Г. Р. Застосування методу аналізу ієрархій для вибору моделі представлення предметної області «Розумне місто» / Н. Е. Кунанець, А. В. Ржеуський, Г. Р. Мацюк // Математика. Інформаційні технології. Освіта : тези доп. VII Міжнар. наук.-практ. конф. (Україна, Луцьк – Світязь, 3-5 червня 2018 р.). – Луцьк – Світязь, 2018. – С. 68 – 70.

Розглянуто вибір моделі представлення знань предметної області «Розумне місто» у рамках створення інформаційно-пошукового тезауруса.

15. Matsiuk Halyna Application of Saaty Method while Choosing Thesaurus View Model of the “Smart city” Subject Domain for the Improvement of Information Retrieval Efficiency / Kunanets Nataliia, Matsiuk Halyna // 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Ukraine, Lviv, September 11-14, 2018). – Lviv, 2018. – P. 21–25. (Індексовано в Scopus, Web of Science).

Подано методику вибору моделі представлення концептів для побудови тезауруса предметної області «Розумне місто» методом Сааті.

16. Мацюк Г. Р. Формування тезауруса для виконання інформаційного пошуку у вузькоспеціальних предметних областях / Н. Е. Кунанець, Г. Р. Мацюк // Бібліотека. Наука. Комунікація: 100-річчя Національної бібліотеки України імені В. І. Вернадського: матеріали міжнар. наук. конф. (Україна, Київ, 6-8 листопада 2018 р.) / Нац. б-ка України ім. В. І. Вернадського. – Київ, 2018. – С. 366–369.

Розкрито поняття тезауруса як інструмента інформаційного пошуку ресурсів у автоматизованих інформаційно-бібліотечних системах для створення пошукових образів для вузькоспеціальних предметних областей.

17. Мацюк Г. Стандарти, що регламентують формат подання інформаційно-пошукового тезауруса / Г. Мацюк, Н. Кунанець // Інформаційні моделі, системи та технології: матеріали VI науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 12-13 грудня 2018 р.). – Тернопіль, 2018. – С. 39.

Проаналізовано міжнародні та вітчизняні стандарти, що регламентують побудову інформаційно-пошукових тезаурусів.

18. Мацюк Г. Побудова онтології як інтероперабельної моделі подання знань у процесі наукового дослідження предметної області «Розумне місто» / Г. Мацюк, Н. Кунанець // Інформаційні моделі, системи та технології: матеріали VI науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 12-13 грудня 2018 р.). – Тернопіль, 2018. – С. 40.

Обґрунтовано обрання онтологічного підходу для методологічного забезпечення застосування системного підходу до аналізу та вивчення предметної області.

19. Halyna Matsiuk Application of the “Smart City” data domain thesaurus as the tool for representing knowledge while improving the problem-oriented Web search effectiveness / Halyna Matsiuk, Nataliya Kunanets // 2019 IEEE 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Ukraine, September 17-20, 2019). – Lviv, 2019. – P. 31–34. (Індексовано в Scopus, Web of Science).

Розкрито використання тезауруса предметної області «Розумне місто» як інструмента термінологічного контролю в процесі аналізу, опрацювання документів та пошуку інформації.

20. Мацюк Г. Р. Тезаурус як засіб семантизації понять предметної області / Г. Р. Мацюк, Н. Е. Кунанець // Бібліотека. Наука. Комунікація: актуальні тенденції у цифрову епоху: матеріали міжнар. наук. конф. (Україна, Київ, 8-10 жовтня 2019 р.) / Нац. б-ка України ім. В. І. Вернадського. – Київ, 2019. – С. 94–98.

Розкрито поняття тезауруса як засобу семантизації понять предметної області.

21. Мацюк Г. Р. Лексикографічна систематизація галузевої термінології за допомогою тезауруса / Н. Е. Кунанець, Г. Р. Мацюк // Х Львівський міжнародний бібліотечний форум «Нова бібліотека – відповідальність кожного» : збірник матеріалів. – Київ : УБА, 2019. – С. 64–66.

Розкрито роль галузевого інформаційно-пошукового тезауруса як лексикографічного систематизатора термінології певної галузі.

22. Мацюк Г. Р. Лінгвістичний аспект спілкування з експертом при формуванні пошукового тезауруса / Г. Р. Мацюк, Н. Е. Кунанець // Інформаційні моделі, системи та технології: матеріали VII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 11-12 грудня 2019 р.). – Тернопіль, 2019. – С. 70.

Розглянуто особливості лінгвістичного аспект спілкування з експертом під час отримання інформації для побудови інформаційно-пошукового тезауруса для використання в бібліотеці.

Публікації, що додатково відображають наукові результати дисертації:

23. Мацюк Г. Р. Тезаурус предметної області «Розумне місто» / Н. Е. Кунанець, Г. Р. Мацюк // Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі». – 2017. – № 872. – С. 179–189.

Окреслено особливості формування тезауруса предметної області «Розумне місто», подано приклад побудови фрагменту онтології з предметної області «Розумне місто».

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ.....	19
ВСТУП	20
РОЗДІЛ 1 ІНФОРМАЦІЙНО-ПОШУКОВИЙ ТЕЗАУРУС ЯК ОБ'ЄКТ БІБЛІОТЕКОЗНАВЧИХ ДОСЛІДЖЕНЬ.....	30
1.1 Аналіз стану дослідження інформаційно-пошукових тезаурусів.....	30
1.2 Теоретичні засади міждисциплінарності та інформаційного забезпечення наукових досліджень	49
1.3 Методологічні засади дослідження.....	58
Висновки до Розділу 1.....	73
РОЗДІЛ 2 ВИБІР МОДЕЛІ ПРЕДСТАВЛЕННЯ ЗНАНЬ ПРЕДМЕТНОЇ ОБЛАСТІ «РОЗУМНЕ МІСТО» ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В БІБЛІОТЕКАХ.....	77
2.1 Аналіз моделей представлення знань.....	77
2.2 Вибір способу представлення знань предметної області «Розумне місто» методом аналізу ієрархій.....	84
2.3 Онтології як системи зберігання знань.....	94
Висновки до Розділу 2	104
РОЗДІЛ 3 ТЕЗАУРУС ЯК ЗАСІБ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СУЧАСНИХ ІНФОРМАЦІЙНО- ПОШУКОВИХ СИСТЕМ БІБЛІОТЕК.....	106
3.1 Понятійно-термінологічний аналіз тезауруса.....	106
3.2 Стандарти, що регламентують формат представлення тезауруса.....	119
3.3 Структура тезауруса як засобу систематизації наукових понять.....	122
3.4 Практичні рекомендації стосовно визначення термінів тезауруса....	130
3.5 Технологічні фази формування тезауруса предметної області.....	133
3.6 Отримання інформації для побудови тезауруса.....	142
3.7 Тезаурус як засіб лінгвістичного забезпечення автоматизованих бібліотечних інформаційних систем.....	151

Висновки до Розділу 3.....	161
РОЗДІЛ 4 ДВОМОВНИЙ ТЕЗАУРУС ПРЕДМЕТНОЇ ОБЛАСТІ	
«РОЗУМНЕ МІСТО» ЯК ІНСТРУМЕНТ ВИРІШЕННЯ	165
ІНФОРМАЦІЙНО-ПОШУКОВИХ ЗАДАЧ У БІБЛІОТЕКАХ.....	
4.1 Підготовка масиву термінів для побудови двомовного інформаційно-пошукового тезауруса «розумного міста».....	165
4.2 Визначення відношень між термінами інформаційно-пошукового тезауруса «розумного міста».....	188
4.3 Структура інформаційно-пошукового тезауруса «розумного міста»	202
4.4 Функції двомовного інформаційно-пошукового тезауруса «розумного міста».....	218
4.5 Здійснення інформаційного пошуку релевантної інформації з допомогою тезауруса «розумного міста».....	222
Висновки до Розділу 4.....	230
ВИСНОВКИ	233
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	237
ДОДАТКИ	264

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

АБІС – автоматизована бібліотечна інформаційна система

ВІУ – випадковий індекс узгодженості

ВУ – відношення узгодженості

ЄС – Європейський Союз

ІНІСН РАН – Інститут наукової інформації з питань суспільних наук

Російської академії наук

ІУ – індекс узгодженості

МАІ – метод аналізу ієрархій

ХДУХТ – Харківський державний університет харчування та торгівлі

ЮНЕСКО – Організація Об'єднаних Націй з питань освіти, науки і культури

ВСТУП

Обґрунтування вибору теми дослідження. Обмін інформацією, отримання релевантних даних є важливими аспектами функціонування наукових комунікацій. Потреба якісного пошуку документів та швидкого інформаційного обміну ресурсами генерується в роботі бібліотек. Тому сучасним напрямом діяльності бібліотечних установ є організація ефективних процесів опрацювання та пошуку інформації, що сприяє покращенню процедур виконання інформаційних запитів користувачів. З огляду на це виникає необхідність уніфікувати пошуковий образ документа й пошуковий образ запиту користувачів з метою їхнього подальшого пошуку в автоматизованих бібліотечних інформаційних системах (АБІС). Цьому слугують штучні інформаційно-пошукові мови, призначені для передачі змісту та формальних ознак документів, а також формування інформаційних запитів. Вагомими серед них є інформаційно-пошукові тезауруси, використання яких сприяє швидкому та результативному пошуку релевантних документів.

Активне впровадження в роботу бібліотек автоматизованих технологій ставить високі вимоги до лінгвістичного забезпечення АБІС, тому воно потребує постійного вдосконалення.

Важливим у розробленні лінгвістичного забезпечення для створення уніфікованого пошукового образу та пошукового запиту документа є спеціально створені тезауруси, які базуються на штучній мові, призначеній для відображення змісту документів і запитів користувачів з метою їхнього подальшого пошуку в АБІС.

Сьогодні існує значна кількість тезаурусів, які використовуються для оброблення та пошуку інформації в базах даних. Серед таких ресурсів загальномовні – WordNet, GermaNet, DanNet, MultiWordNet тощо; спеціалізовані (представляють терміносистему певної галузі науки) – SNOMED CT, MeSH, EUROVOC, AGROVOC, INICH РАН тощо. Українські

інформаційні ресурси представлені інформаційно-пошуковим тезаурусом Національної бібліотеки України імені Ярослава Мудрого, галузевим тезаурусом лінгвістичного забезпечення електронної бібліотеки ХДУХТ, тезаурусами з комп'ютерної лексикографії та лінгвістичних термінів Інституту філології Київського національного університету імені Тараса Шевченка, а також вузькоспеціалізованими галузевими тезаурусами з нанотехнології, Державної прикордонної служби України, будівельної термінології, авіації, надтвердих матеріалів тощо.

Дослідженню проблем тезауруса присвячені роботи таких вітчизняних і зарубіжних науковців як Т. Вилегжаніна, Н. Гендіна, А. Гладун, В. Дубичинський, О. Збанацька, Ю. Караулов, І. Кудашев, Є. Купріянов, К. Лобузін, Н. Лукашевич, Л. Пшенична, Р. Робертс (R. Roberts), Ю. Рогушина, Н. Соколова, В. Шульгіна та ін.

Але, незважаючи на це, слід констатувати, що комплексні, системні дослідження проблеми розроблення інформаційно-пошукового тезауруса предметної області «Розумне місто» не проводилися.

Важливим у становленні понятійного апарату будь-якої галузі є питання нормалізації термінологічної лексики. Таке завдання також постало при формуванні предметної області «Розумне місто». «Розумне місто» – це ефективна інтеграція фізичних, цифрових і людських ресурсів у штучному середовищі для забезпечення сталого, благополучного майбутнього своїх мешканців.

З огляду на той факт, що наукові дослідження в інноваційній міждисциплінарній галузі «Розумне місто» (оскільки до досліджень залучені фахівці з інформаційних технологій, соціальних комунікацій та телекомунікацій, економіки та електронного урядування) набирають інтенсиву, для інформаційного забезпечення її дослідників виникає потреба формування єдиної понятійної бази та відповідної спеціальної термінології, щоб створювати пошукові образи документів, що висвітлюють результати наукових розвідок. Зазначене завдання лежить у площині лінгвістичного

забезпечення бібліотечних інформаційно-пошукових систем, узгодження українських термінів з іншомовними при формуванні пошукових образів та пошукових запитів. Такий стан речей генерує потребу розроблення двомовного інформаційно-пошукового тезауруса «розумного міста», який відповідав би вимогам сучасної бібліотечної спільноти щодо формування систематизованої терміносистеми як інструмента створення пошукових образів документів та пошукових запитів користувачів, дозволяв би при формуванні інформаційних ресурсів у бібліотеках окреслити межі термінологічного поля предметної області «Розумне місто» на рівні двох мов. Це буде важливим кроком до впорядкування, стандартизації термінологічного підґрунтя досліджуваної галузі знань, оскільки двомовний тезаурус, як результат лексикографічного опису даної терміносистеми, виступатиме нормативним документом кодифікованих термінів, які орієнтуватимуть бібліотекарів та користувачів на правильне їхнє застосування, та забезпечить пошук релевантної інформації в АБІС.

Саме в цьому контексті визначається актуальність дисертаційної роботи, у якій представлено технологію побудови інформаційно-пошукового тезауруса як інструмента інформаційного забезпечення міждисциплінарних наукових досліджень на основі формалізованого подання термінології предметної області «Розумне місто». Інформаційно-пошуковий тезаурус при використанні в бібліотечних та інформаційних установах, електронних бібліотеках сприяє підвищенню ефективності технологій формування пошукових образів документів та пошукових образів запитів користувачів. У дослідженні запропоновано теоретико-методологічні засади формування тезауруса «розумного міста» та представлено його функціональність.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана в Тернопільському національному технічному університеті імені Івана Пулюя в рамках науково-дослідних тем: «Класи соціокомунікаційних технологій в проєктах «Розумне місто» (2017-2019 рр., державний реєстраційний № 0117U002240), «Класи інформаційних

технологій в проєктах «Розумне місто» (2017-2019 рр., державний реєстраційний № 0117U002241) та науково-дослідної теми Національного університету «Львівська політехніка» «Науково-освітнє соціокомунікаційне середовище великого міста: моделювання, прототипування, інформаційні технології» (2016-2019 рр., № 0116U006723).

Мета та завдання дослідження. Мета – розроблення концептуальних засад формування двомовного інформаційно-пошукового тезауруса як формалізованого представлення знань предметної області «Розумне місто» та інструмента ефективного пошуку релевантної інформації в автоматизованих бібліотечних інформаційних системах для інформаційного забезпечення наукових досліджень, що проводяться на міждисциплінарних засадах.

Для досягнення поставленої мети передбачено вирішення таких **завдань**:

- виявити та проаналізувати джерельну базу дослідження щодо формування пошукових образів та запитів на документи, що містять результати міждисциплінарних досліджень;
- обґрунтувати вибір моделі представлення знань предметної області «Розумне місто» з метою підвищення ефективності інформаційного пошуку в автоматизованих бібліотечних інформаційних системах для забезпечення наукових досліджень;
- простежити історію розвитку технологій формування тезаурусів та проаналізувати роль тематичного інформаційно-пошукового тезауруса в процесі систематизації термінології певної галузі та можливості його застосування, як засобу підвищення ефективності пошуку в автоматизованих бібліотечних інформаційно-пошукових системах;
- удосконалити в сучасних бібліотеках технології побудови інформаційно-пошукового тезауруса, що описує категоріально-понятійний апарат певної предметної області та взаємозв'язки між концептами;
- розробити двомовний інформаційно-пошуковий тезаурус «розумного міста» для формування уніфікованих пошукових образів

документів та пошукових запитів користувачів з метою їхнього подальшого пошуку в автоматизованих бібліотечно-інформаційних системах.

Об'єкт дослідження – процеси формування пошукових образів документів та пошукових запитів користувачів у бібліотеках, створених на основі двомовного подання термінів предметної області «Розумне місто».

Предмет дослідження – структуроване подання знань предметної області «Розумне місто» у формі двомовного інформаційно-пошукового тезауруса для проведення ефективної процедури пошуку в бібліотеках.

Методи дослідження. Для досягнення мети та розв'язання завдань, поставлених у дисертаційній роботі, використано методи загальнонаукові, інформатики, математичні, соціокомунікаційний та історико-хронологічний.

У процесі дослідження були поєднані загальнонаукові методи та прийоми дослідження: джерелознавчого пошуку й аналізу профільних публікацій за темою дослідження для визначення стану розробленості наукової проблеми; дедуктивний та індуктивний методи для побудови тезауруса на основі переходу від ширших термінів до вузких та навпаки; метод зіставлення для виявлення відмінних рис мовної форми вираження та змістового наповнення термінів, їхніх структурних компонентів та елементів в українській та англійській мовах; термінологічний аналіз для уточнення поняттєвого апарату дослідження та термінів галузі «розумне місто» з метою їх унормування та упорядкування.

Методи інформатики (метод інформаційного підходу, консолідації інформації, метод онтології, тезаурусний підхід) було використано для виявлення особливостей та розроблення інформаційно-пошукового тезауруса; математичний інструмент системного підходу до складних проблем прийняття рішень – метод аналізу ієрархій – для вибору способу представлення знань предметної області «Розумне місто».

Застосований соціокомунікаційний метод було використано для обґрунтування сутності тезаурусів, які реалізуються в цілеспрямованих комунікаційних зв'язках між суб'єктами та об'єктами системи, суб'єктами

системи та зовнішнім середовищем. Історично-хронологічний метод дозволив дослідити виникнення та розвиток тезаурусів у хронологічній послідовності.

Наукова новизна дослідження полягає в тому, що дисертація є одним з першим досліджень, в якому проведена робота над створенням формалізованого представлення знань інноваційної предметної області «Розумне місто» у формі двомовного інформаційно-пошукового тезауруса, що відображено у таких положеннях:

уперше:

- *застосовано* метод аналізу ієрархії для вибору способу представлення знань предметної області «Розумне місто» при формуванні інформаційних ресурсів бібліотек;

- на основі аналізу наукових публікацій *консолідовано* підходи до формування двомовного тезауруса, який сприяє ефективному формуванню метаданих інформаційних ресурсів бібліотек, та сформовано алгоритм його розроблення;

- *розроблено* двомовний інформаційно-пошуковий тезаурус «розумного міста» на основі формалізованого представлення знань предметної області, що забезпечує інформаційний пошук релевантних документів у бібліотеках, сприяє підвищенню ефективності інформаційного забезпечення наукових досліджень, слугує самостійним термінологічним довідником, який допомагає орієнтуватися в даній галузі;

удосконалено:

- визначення поняття «інформаційно-пошуковий тезаурус предметної області» для використання в бібліотечних інформаційних системах;

додовнено:

- класифікацію типів семантичних відношень з урахуванням міждисциплінарних тенденцій сучасного наукового знання (виділено ряд проблемно-специфічних відношень характерних для предметної області

«Розумне місто»), що сприяло підвищенню інформаційного пошуку у бібліотеках;

набули подальшого розвитку:

- технологія поповнення бібліотечними працівниками тезауруса «розумного міста»: наповнення новими лексичними одиницями для двомовного відображення найновіших термінів галузі, відслідковування міжпредметних зв'язків;

- методи розроблення електронної версії двомовного тезауруса предметної області «Розумне місто» для бібліотечних інформаційних систем.

Практичне значення дослідження. Практичне значення дисертаційної роботи полягає в розробленні двомовного інформаційно-пошукового тезауруса предметної області «Розумне місто» як інструмента пошуку релевантних документів для інформаційного забезпечення міждисциплінарних наукових досліджень.

Сьогодні для опису будь-якої предметної області та створення інтелектуальних систем у бібліотеках чимале значення має розроблення нових алгоритмів та методик формування тезаурусів на основі онтологій. Для адекватного відображення моделей подання знань, здатних до взаємодії, у роботі запропонована методика, яка використовує проблемно-специфічні відношення для реалізації пошукових запитів та дозволяє формувати якісні знання про досліджувану галузь, що сприяє підвищенню ефективності інформаційного забезпечення.

Результати дослідження впроваджені та використовуються в роботі науково-технічної бібліотеки Тернопільського національного технічного університету імені Івана Пулюя, Державного архіву Тернопільської області, а також у навчальному процесі кафедри комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя. Водночас результати дослідження можна використовувати в дослідницькій, навчальній та практично-фаховій діяльності в різних суміжних галузях наук. Вони можуть слугувати науково-методологічним базисом для розширення та

поглиблення навчальних курсів із інформаційного забезпечення наукових досліджень та інформаційного менеджменту. Їхнє активне впровадження сприятиме розвитку бібліотечних та інформаційних установ, де аналіз документів здійснюється при їхньому опрацюванні в бібліотечних установах з використанням штучної мови. У бібліотечних та архівних установах тезаурус можна використовувати як інструмент формування уніфікованих пошукових образів документів та пошукових запитів користувачів, а також для пошуку релевантних документів в інформаційних ресурсах глобальної мережі «Інтернет». Використовуючи тезаурус, інформаційні установи, у тому числі й бібліотеки, можуть формувати предметні рубрики. Якісне індексування документів забезпечує адекватність пошуку в автоматизованих інформаційно-бібліотечних системах.

Впровадження матеріалів дослідження підтверджено відповідними актами.

Особистий внесок здобувача. Дисертаційне дослідження виконано самостійно, викладені наукові результати, висновки й рекомендації належать дисертанту. Внесок здобувача в працях, які виконані в співавторстві, конкретизовано в списку публікацій.

Апробація результатів дисертації. Основні теоретичні положення, висновки й результати дисертаційного дослідження було викладено в опублікованих матеріалах та збірниках праць, обговорено та апробовано на міжнародних наукових та науково-практичних конференціях, серед яких: VI Міжнародна науково-практична конференція «Математика. Інформаційні технології. Освіта» (Луцьк-Світязь, 5-7 червня 2017 р.); XV Международная научно-практическая конференция «Современные информационные технологии в экономике и управлении предприятиями, программами и проектами» (Харьков, 29 сентября 2017 г.); V Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 1-2 лютого 2018 р.); VII Міжнародна науково-практична конференція «Математика. Інформаційні технології. Освіта» (Луцьк-Світязь, 3-5 червня 2018 р.); 2018

IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Lviv, September 11-14, 2018); Міжнародна наукова конференція «Бібліотека. Наука. Комунікація: 100-річчя Національної бібліотеки України імені В. І. Вернадського» (Київ, 6-8 листопада 2018 р.); VI Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 12-13 грудня 2018 р.); The 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019) (Kharkiv, April 18-19, 2019); The VII International Scientific Conference «Mathematics. Information Technology. Education», MoMLeT&DS'2019 (Shatsk, June 2-4, 2019); 2019 IEEE 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Lviv, September 17-20, 2019); X Львівський міжнародний бібліотечний форум «Нова бібліотека – відповідальність кожного» (Львів, 18-21 вересня 2019 р.); Міжнародна наукова конференція «Бібліотека. Наука. Комунікація: актуальні тенденції у цифрову епоху» (Київ, 8-10 жовтня 2019 р.); VII Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 12-13 грудня 2019 р.).

Публікації. Основні положення та результати дисертаційного дослідження викладено у 23 наукових публікаціях, серед яких 4 статті у вітчизняних наукових фахових журналах, 5 статей у наукових періодичних виданнях інших держав; 13 публікацій у збірниках матеріалів міжнародних конференцій; 1 стаття, що додатково відображає наукові результати дослідження. П'ять із вказаних наукових праць опубліковано в журналах та збірниках, що індексуються в міжнародних наукометричних базах: Scopus, Web of Science.

Структура та обсяг дисертації. Дисертація складається з титульного аркуша, анотації українською та англійською мовами, списку публікацій здобувача за темою дисертації, змісту, переліку умовних скорочень, вступу, 4 розділів, висновків у кінці розділів та загальних висновків, списку використаних джерел та додатків. Загальний обсяг дисертації – 292 сторінки:

основна частина – 212 сторінок (8,9 авт. арк.) у тому числі 22 рисунки та 6 таблиць, список використаних джерел – 253 позиції на 27 сторінках, додатки – 26 сторінок.

РОЗДІЛ 1

ІНФОРМАЦІЙНО-ПОШУКОВИЙ ТЕЗАУРУС ЯК ОБ'ЄКТ БІБЛІОТЕКОЗНАВЧИХ ДОСЛІДЖЕНЬ

1.1 Аналіз стану дослідження інформаційно-пошукових тезаурусів

В умовах розвитку сучасного інформаційного суспільства користувачам для того, щоб орієнтуватися у накопичених великих фондах і потужному потоці інформації, знаходити дані, що відповідають їхнім запитам, необхідні допоміжні засоби, якими є спеціальні методи опрацювання даних, організації швидкого та ефективного пошуку. Для вирішення цієї проблеми призначені інформаційно-пошукові тезауруси [1].

Розроблення інформаційних тезаурусів для ефективного формування пошукових образів документів та пошукових запитів користувачів через постійне зростання інформації набуває все більшої актуальності. Одним з найвідоміших інформаційних ресурсів для реалізації цієї мети є універсальний комп'ютерний тезаурус WordNet [2], створений командою Принстонського університету, США під керівництвом психолінгвіста Джорджа Міллера. Це велика база даних англomовних лексичних одиниць. Онлайн-версію WordNet 3.1. містить 155 327 слів, організованих у 175 979 синсетів та 207 016 пар лексем. Тезаурус складається з 4 файлів, які містять головні частини мови (іменники, дієслова, прикметники та прислівники). Особливість цього тезаурусу в тому, що базовою одиницею є не окреме слово, а синонімічний ряд – синсет (synset), який об'єднує слова зі схожим значенням, які і є вузлами мережі. Для зручного використання тезаурусу кожен синсет доповнений дефініцією й прикладами вживання слів у контексті. Синсети взаємопов'язані за допомогою концептуально-семантичних і лексичних відношень: гіперонім – hyperonymy (breakfast – meal), гіпонім – hyponymy (meal – lunch), має ієрархію підпорядкування – has-member (faculty – professor), учасник – member-of (pilot – crew), меронім

– мерониму (*table – leg*) та ін. (див. Додаток Б). Найбільш вживаним між іменниками є відношення підпорядкування (гіперонім, гіпонім чи is-a), яке пов'язує загальні синсети з конкретними (*bed – bunkbed*). Усі ієрархії іменників піднімаються вгору до кореневого вузла. Відношення гіпонімії є перехідним. Для іменників є характерний і зв'язок меронімії (*chair – back, leg*). Частини не успадковуються вгору, оскільки вони можуть бути лише частиною конкретного предмета, а не цілого класу. Синсети дієслова також розташовані ієрархічно, дієслова на нижчих рівнях виражають все більш конкретні характеристики дії (*communicate – talk – whisper*). Прикметники пов'язані з позиції антонімії (*wet – dry, young – old*), кожен прикметник з пари має відношення з низкою семантично подібних прикметників (*dry* пов'язаний з *parched, arid, dessicated, bone-dry, wet to soggy, waterlogged*). У WordNet-і є лише декілька прислівників, так як більшість англійських прислівників мають такі ж форми як і прикметники. Перехресні зв'язки між різними частинами мови будуються між семантично подібними словами: *observe* (дієслово), *observant* (прикметник), *observation, observatory* (іменники).

Тезаурус WordNet знаходиться у вільному доступі.

Концепція WordNet-у лягла в основу створення ряду тезаурусів аналогічної тематики іншими мовами. Серед них тезаурус EuroWordNet [3], який об'єднав терміни європейськими мовами, проте робота над його створенням була завершена влітку 1999 року. Дизайн бази даних, визначені відношення, топ-онтологія та інтерактивний індекс зараз заморожені. Разом з тим, завдяки створенню цього тезауруса вперше була реалізована ідея об'єднання окремих термінологічних мереж у загальну систему.

Відповідно до засадничих принципів WordNet-у розроблений тезаурус термінів німецькою мовою GermaNet [4], який розробляється і підтримується на кафедрі загальної та обчислювальної лінгвістики в Тюбінгенському університеті (University of Tübingen) з 1997 року. Версія GermaNet 14.0 містить 144113 синсетів, 185000 лексичних одиниць, 157921 концептуальних відношень. На відміну від WordNet-у, в якому застосований принцип

роздільного опису різних частин мови та відсутності відношень між різними частинами мови, що мають однакове значення, у GermaNet семантично пов'язані німецькі іменники, прикметники, дієслова шляхом групування лексичних одиниць, які виражають однакові чи близькі поняття в синсети, за допомогою визначень семантичних зв'язків між цими синсетами. Таким чином, усі лексичні одиниці, які згруповані в одному синсеті, є синонімами.

За такими ж принципами було створено DanNet [5], що містить 65 тис. синсетів датською мовою, MultiWordNet [6] для об'єднання термінів італійською мовою, а також низку національних тезаурисів, що містять концепти шведською, норвезькою, грецькою, португальською, баскською, каталонською, румунською, литовською, російською, болгарською, словенською мовами. Тезауруси структуровані на тих самих засадах, що і американський WordNet для англomовних термінів (Princeton WordNet). Проте кожен з них є унікальною мовною системою.

У 2001 році була створена Всесвітня Асоціація WordNet (Global WordNet Association), завдання якої полягає в об'єднанні інформаційних ресурсів тезаурусного типу, розробленні загальних стандартів, що сприяють реалізації моделі WordNet для різних мов.

Використовуючи загальний принцип побудови WordNet-у, було побудовано російськомовний тезаурус RussNet [7], укладений дослідницькою групою Санкт-Петербурзького державного університету під керівництвом І. Азарової. Цей ресурс складається з чотирьох взаємозв'язаних частин, сформованих з іменників, дієслів, прикметників та прислівників, які пов'язані між собою парадигматичними та синтагматичними відношеннями. Базовою одиницею є синонімічний ряд – синсет, що об'єднує слова з подібними значенням. Тезаурус включає приблизно 15 тисяч лексико-семантичних одиниць, які організовані в 5500 синсетів. Тезаурус RussNet на сьогодні є найбільш великим проєктом подібного роду та є закритим.

До того ж класу, що і WordNet, належить російський тезаурус РуТез [8], розроблений на базі суспільно-політичної тематики. Тезаурус

РyТез є лінгвістичним ресурсом концептуального типу, тобто ієрархічною мережею понять, до яких під'єднані текстові вирази. При цьому, на відміну від WordNet-y, який створювався як модель системи понять, присутніх у людській пам'яті (роздільний опис частин мови, спеціальні типи відношень та ін.), тезаурус РyТез розроблений саме як ресурс для автоматизованого опрацювання текстів. Даний ресурс містить термінологію економічної, політичної, військової, фінансової, законодавчої, соціальної, культурної та інших галузей діяльності. За принципами розроблення тезаурус РyТез об'єднує три способи розроблення комп'ютерних лінгвістичних ресурсів: традиційних інформаційно-пошукових тезаурусів, лінгвістичних ресурсів типу WordNet та формальних онтологій [9]. Нова версія тезауруса РyТез 2.0 містить 31.5 тис. понять, 111.5 тис. різних текстових входів (слів та виразів російської мови) і є інформаційним ресурсом закритого типу.

У Національному університеті «Львівська політехніка» було зроблено спробу розроблення WordNet-подібного тезауруса, що містить концепти українською мовою. У результаті дослідження було розроблено перші фрагменти української версії WordNet-y, в якому реалізовано 194 синсети, з яких 183 пов'язані гіпо-гіперонімічними, 14 антонімічними зв'язками, а також 150 зв'язками меронімії / гіпонімії та 6 also see [10], що дає підстави стверджувати важливість продовження роботи над розробленням цього інформаційного ресурсу.

На сьогодні є понад 40 тезаурусів, розроблених різними мовами на основних засадах побудови WordNet-y, що демонструє ефективність інформаційних ресурсів такої структури.

Міжнародною Організацією Об'єднаних Націй з питань освіти, науки і культури, ЮНЕСКО (United Nations Educational, Scientific and Cultural Organization, UNESCO) створено багатомовний тезаурус ЮНЕСКО [11], укладений англійською, іспанською, французькою та російською мовами. Він містить 7 тис. термінів з галузей освіти, культури, природничих, соціальних та гуманітарних наук, комунікації та інформації, за ієрархічною

класифікацією. Тезаурус складається із семи тематичних областей, розбитих на мікро масиви, що дозволяють швидко ознайомитися з темами: освіта, природничі науки, культура, соціальні та гуманітарні науки, інформація та комунікація, політика, право і економіка, країни і групи країн. Кожен термін вищого рівня супроводжується посиланням UF (Used For) та спадковою ієрархією описів, перед кожним з яких знаходиться символ NT (Narrower Term – «вузький термін»). Для розкриття кожного терміна використовуються такі позначення: SN (Scope Note – «зміст терміна»), MT (Microthesaurus – «мікротезаурус»), UF (Used For – «синонім»), BT (Broader Term – «широкий термін»), NT (Narrower Term – «вузький термін»), RT (Related Term – «споріднений термін») (див. Додатки В, Г). Тезаурус постійно оновлюється та поповнюється.

Видавничим бюро Європейського Союзу (ЄС) підтримується тезаурус EUROVOC [12]. Багатомовний політематичний інформаційно-пошуковий тезаурус EUROVOC відіграє роль міжнародного термінологічного стандарту. Він був спеціально розроблений для роботи з документною інформацією, яка зберігається в інституціях Європейського Союзу. EUROVOC з 1984 року використовується для індексування та пошуку даних в інформаційно-пошукових системах Європейського парламенту, Бюро офіційних публікацій ЄС, парламентів європейських країн.

Бібліотеки інститутів Європейського Союзу, служби баз даних і їхні користувачі використовують EUROVOC як довідкову термінологічну базу даних. Інституції Європейського Союзу використовують EUROVOC для створення пошукових образів для документації, що зберігається в створених ними базах даних, так Європейський Парламент у документальній базі даних EPOQUE, яка сформована з матеріалів парламентських документів і електронного каталогу з метаданими на них; Відділ офіційних публікацій ЄС – у каталогах системи CATEL, Офіційний журнал ЄС – в електронному архіві випусків.

Тезаурус EUROVOC містить терміни на 23 офіційних мовах ЄС, що поєднані зв'язками, характерними для кожної з мов. Однак між недескрипторами не визначено еквівалентності в різних мовах. Інформаційно-пошуковий тезаурус EUROVOC побудовано за ієрархічним принципом від загального до часткового. Верхній рівень складається з 21 домену та 127 мікротезаурусів. Кожному домену присвоєно двозначний код та назву на природній мові, наприклад, 36 Science (Наука), 48 Transport (Транспорт), 64 Production, technology and research (Виробництво, технологія та дослідження) та ін. Їх поділено на тематичні мікротезауруси. Кожен тезаурус ідентифікується чотиризначним кодом (перші дві цифри співпадають з кодом розділу, до складу якого входить мікротезаурус) та назвою на природній мові, наприклад, 3606 Natural and applied sciences (3606 Природничі та прикладні науки). З огляду на можливі зміни в EUROVOC, розробники пронумерували розділи та мікротезауруси не суцільною нумерацією, а залишили «лакуни» в номерах: 04, 08, 10, ..., 76 (див. Додаток Д). Числові коди розділів та мікротезаурусів однакові для всіх мовних версій. Пошук здійснюється з допомогою першої букви дескриптора, цілого слова чи ключових слів.

Тезаурус має такі семантичні співвідношення: «*SN*» (*Scope Note*) – примітка, що пояснює значення дескриптора; «*MT*» (*Microthesaurus*) – посилання на мікротезаурус; відношення еквівалентності: «*UF*» (*Used For*) між дескриптором та аскриптором, «*USE*» між аскриптором та дескриптором; «*BT*» (*Broader Term*) – ширший термін; «*NT*» (*Narrower Term*) – вужчий термін; «*RT*» (*Related Term*) – асоціативні зв'язки між дескрипторами.

Основні розділи тезауруса EUROVOC пов'язані з різними видами діяльності органів, установ та інститутів ЄС: політика, міжнародні відносини, законодавство, право, економіка, торгівля, фінанси, соціальні питання, культура, освіта, комунікація, наука, бізнес, транспорт, навколишнє середовище, промисловість, географія тощо.

У грудні 2019 року було опубліковано EuroVoc 4.10 [13]. Новий випуск доступний для завантаження у форматі XML, SKOS-Core, SKOS-AP-EU (RDF), Marc-XML та Excel.

У 2001 році видавництво Державної Думи Росії видало російську версію тезауруса EUROVOC – «Информационно-поисковый тезаурус Парламентской библиотеки», який використовується для індексування та пошуку інформації в електронному каталозі Парламентської бібліотеки.

Науково-дослідним центром правової інформатики НАПрН України (під керівництвом члена-кореспондента НАПрН України М. Я. Швеця) у співпраці з Інститутом законодавства Верховної Ради України та Управлінням комп'ютеризованих систем ВР України в період з жовтня 2003 року по червень 2009 року велася робота над розробленням та впровадженням електронної версії інформаційно-пошукового тезауруса EUROVOC 4.2, що містить дескриптори українською мовою. Тезаурус включає 21 розділ; 127 мікротезаурусів; 6439 дескрипторів (з них 511 верхнього рівня). Ця версія тезауруса відображала 6448 ієрархічних зв'язків, 3501 асоціативних зв'язків. На перших етапах робота над українською версією інформаційно-пошукового тезауруса EUROVOC 4.2 спрямовувалася на досягнення відповідності української версії офіційним версіям країн ЄС, проте надалі для подолання обмежень щодо використання тезауруса при індексації термінів національної законодавчої бази планувалося використання дескрипторів та приміток і це передбачало формування версій тезауруса EUROVOC 4.3 та 4.4. За результатами виконання робіт, окрім інформаційно-пошукової системи, українська версія тезауруса EUROVOC була також видана в абетково-пермутаційному, тематичному та багатомовному представленнях у складі електронної «Системи (бібліотеки) баз даних і знань у галузі держави і права» [14]. Українська версія тезауруса EUROVOC, яка включає 8670 термінів розміщена на сайті Верховної Ради України у розділі «Законодавство України» (<https://zakon.rada.gov.ua/rada/main/eurovoc/sp:java-:max25#Langs>).

Розроблення тезаурусів відбувається і в медичній галузі. Так, Національною службою охорони здоров'я Великобританії та Колегією американських патологів (College of American Pathologists) розроблено багатомовний тезаурус SNOMED CT [15], що містить систематизовану медичну номенклатуру. До складу SNOMED CT входять медичні терміни (terms), коди термінів (codes) і визначники кодів (definitions). У 2011 році номенклатура SNOMED CT включала 311 000 концептів. Кожен з концептів має унікальний числовий ідентифікатор. Так, терміну «інфаркт міокарда» відповідає «22298006», «застида» – «82272006» та ін.

Функціонує SNOMED CT для американського варіанту англійської мови, британського варіанту англійської мови, іспанської, датської, шведської, французької і нідерландської мов. Застосовується в медичній документації і звітах для підвищення ефективності роботи з клінічними даними.

Національною бібліотекою медицини Національного інституту здоров'я (США) (National Library of Medicine, National Institutes of Health (U.S.)) сформований тезаурус Medical Subject Headings або MeSH (Медичні предметні рубрики) [16], який використовується для індексування, каталогізації та пошуку інформації в галузі біології, медицини, охорони здоров'я та суміжних науках. Мовою оригіналу тезауруса є англійська, проте існують національні версії MeSH іншими мовами, україномовної версії тезауруса MeSH не існує.

Тезаурус складається з алфавітного переліку термінів та ієрархічної структури. Алфавітний перелік термінів містить перехресні посилання: дескриптори або предметні рубрики, що характеризують предмет (назви органів, хвороб, препаратів, професій та інше); модифікатори, тобто підрубрики (аспекти дослідження), які використовуються при індексуванні способом приєднання до предметних рубрик; синоніми, які приєднані перехресними посиланнями до дескрипторів. Ієрархічна структура має 11 рівнів, які включають різні категорії.

Станом на 2019 рік база даних MeSH містить 29351 дескриптор. Тезаурус MeSH доступний он-лайн для перегляду та безкоштовного завантаження (див. Додаток Е).

SNOMED-CT та MeSH входять у масштабний уніфікований термінологічний та онтологічний ресурс, розроблений для інформаційних систем аналізу медичних текстів UMLS (Universal Medicine Language System).

Центральною науковою медичною бібліотекою Московської медичної академії ім. І. М. Сеченова Міністерства охорони здоров'я РФ спільно з Національною медичною бібліотекою США сформовано російсько-англійську версію тезауруса MeSH [17]. Тезаурус складається з двох частин:

1. Алфавітний перелік термінів з перехресними посиланнями (дескриптори (рубрики), модифікатори (підрубрики) та синоніми (перехресні посилання));
2. Ієрархічна структура, що має 11 рівнів. Перший рівень ієрархії включає 16 основних категорій. Основними лексичними одиницями MeSH є дескриптори і модифікатори. Кожен дескриптор має один або декілька унікальних номерів (кодів), що позначають його положення в ієрархічному дереві, яке може змінюватися по мірі оновлення MeSH (див. Додаток Є).

Серед закордонних розробок також потрібно назвати комплекс тезаурусів Фонду Гетті (Getty Vocabularies) (США, Каліфорнія) (див. Додаток Ж), призначених для уніфікації описів об'єктів історико-культурної спадщини в найширшому сенсі цього слова, які широко застосовуються різними музеями, бібліотеками та архівами для опису власних зібрань. Для перегляду вони доступні в онлайн-режимі [18], проте повноцінне використання вимагає ліцензування.

На сторінці словників Getty Vocabularies представлені:

Тезаурус з мистецтва та архітектури –The Art and Architecture Thesaurus (AAT) – структуроване зібрання понять, що включає терміни з галузей мистецтва, архітектури, декоративного мистецтва, матеріальної культури та ін. Тезаурус містить близько 35 тисяч дескрипторів і понад 130 тисяч

англомовних термінів. Дескриптори тезауруса поділяються на 7 фасетів: асоційовані поняття, фізичні властивості, стилі й періоди, агенти (люди й організації), діяльність, матеріали, об'єкти, а кожен фасет поділяється на ієрархії. Всього налічується 33 ієрархічні рівні. Дескриптори тезауруса забезпечені стандартними для тезаурусів відношеннями вищий–нижчий рівень і асоціативними зв'язками [19].

Тезаурус географічних назв – The Getty Thesaurus of Geographic Names (TGN) [20], містить 1115000 дескрипторів, що включають назви держав, міст, археологічних об'єктів та їхніх фізичних особливостей, важливих для дослідження мистецтва й архітектури.

Тезаурус імен авторів художніх творів – The Union List of Artist Names (ULAN) [21], що включає 375000 імен, біографічну й бібліографічну інформацію про художників і архітекторів, містить варіанти імен, псевдонімів та варіанти написання імен на різних мовах.

Тезаурус найменувань культурних об'єктів – Cultural Objects Name Authority (CONA) – офіційних назв об'єктів історико-культурної спадщини, створений у 2011 році [22]. У тезаурусі зібрані офіційні назви творів мистецтва, архітектури, архівної документації, фондів рідкісної книги, необхідні для здійснення їхнього уніфікованого опису. До складу словника включені авторитетні іконографічні поняття, складені спеціальною робочою групою Дослідницького інституту Гетті. Ця частина тезауруса включає власні імена персонажів (вигаданих, легендарних і реальних), а також найменування міфологічних, релігійних і літературних сюжетів, подій і місць, взятих з найбільш відомих творів літератури, живопису та декоративно-прикладного мистецтва.

Тезауруси Дослідницького інституту Гетті узгоджені з міжнародними стандартами й можуть бути використані для підвищення ефективності пошуку в окремих базах даних та в інтернеті. Вони можуть використовуватися при індексуванні архівних та бібліотечних ресурсів.

Міжнародною Продовольчою й сільськогосподарською організацією ООН – ФАО (Food and Agriculture Organization, FAO) розроблено тезаурус AGROVOC [23] сільськогосподарської тематики. Він охоплює термінологію сільсько-, лісо-, водно-, рибогосподарського спрямування, проблем механізації й будівництва, охорони природних ресурсів, запобігання забруднення навколишнього середовища та ін. Тезаурус містить понад 40 тис. концептів 29 мовами (зокрема й українською). У ньому можна знайти, наприклад, назву будь-якої рослини мовою, що вас цікавить, а також семантичні зв'язки, що існують між сировинним товаром і сільськогосподарською культурою, з якої він виготовлений (див. Додаток 3).

Сьогодні AGROVOC використовується науковцями, бібліотекарями та інформаційними менеджерами для індексування, пошуку та організації даних у сільськогосподарських інформаційних системах і на веб-сторінках.

Найбільшим у Росії центром наукової інформації й багатoproфільним науково-дослідним інститутом у області соціальних і гуманітарних наук Інститутом наукової інформації з питань суспільних наук Російської академії наук (ІНІСН РАН) (Институтом научной информации по общественным наукам (ИНИОН РАН)) розроблено комплекс галузевих тезаурусів з правознавства, економіки та демографії, політології, релігієзнавства, мовознавства, гендерних досліджень, соціології, які знаходяться на сайті інституту <http://inion.ru/resources/bazy-dannykh-inion-ran/> [24] (див. Додаток И).

Тезауруси ІНІСН РАН – це перш за все допоміжний інструмент для роботи з бібліографічними базами даних, що дозволяє уточнити пошуковий запит шляхом підбору додаткових термінів, логічно пов'язаних з вихідним набором ключових слів. Більшість тезаурусів доступні в двох версіях: 1) Тезауруси, інтегровані безпосередньо в систему бібліографічних баз даних. Їхньою сильною стороною є широкі можливості для пошуку, але в змістовному відношенні ці тезауруси вже досить сильно застарілі. 2) PDF-версії тезаурусів, що містять велику кількість термінів, але технічно кожен

такий тезаурус є звичайним текстовим документом з можливістю швидкого переміщення до потрібних розділів за посиланнями в змісті.

У 2018 році інститутом розроблено Інформаційно-пошуковий тезаурус ІНІСН з соціології [25], який включає 4760 словникових статей (з них 3126 дескрипторів і 1634 аскриптори), що відображають різні аспекти соціологічної науки, і призначений для індексування вхідного потоку документів і запитів, забезпечуючи інтелектуальний пошук інформації в бібліографічному банку даних, зокрема – пошук у режимі віддаленого доступу. Включає 3 показники: алфавітний лексико-семантичний, пермутаційний і систематичний.

Основною частиною тезауруса є алфавітний лексико-семантичний показник, де представлені всі дескриптори й аскриптори (синоніми) зі словниковими статтями. У тезаурусі використані недиференційовані ієрархічні відношення, асоціативні відношення (посилання на семантичні (неієрархічні) зв'язки з іншими дескрипторами) та відношення синонімії заголовного дескриптора з аскриптором.

Пермутаційний показник є допоміжним до алфавітного лексико-семантичного. Містить дескриптори й аскриптори без словникових статей. Специфічною особливістю показника є формування словникових гнізд за ключовими словами, що входять до складу термінів-словосполучень.

Систематичний показник містить тільки дескриптори. Терміни розподілені в алфавітному порядку за рубриками ієрархічної класифікації рубрикатора ІНІСН. При цьому окремий дескриптор може бути включений у кілька рубрик.

У тому ж році інститутом розроблено Інформаційно-пошуковий тезаурус ІНІСН РАН з етнології та антропології [26]. Містить 6698 термінів, що відображають різні аспекти етнології та антропології з дотриманням принципів історизму; включає два показники: алфавітний лексико-семантичний і пермутаційний.

У 2019 році ІНІСН РАН було випущено PDF-версії інформаційно-пошукових тезаурусів: «Економіка. Демографія» [27], «Філософія» [28] та «Релігієзнавство» [29]. Тезаурус з економіки містить близько 6 600 термінів, що відображають різні аспекти економіки й демографії з дотриманням принципу історизму; включає два показники: алфавітний лексико-семантичний і пермутаційний. Тезаурус з філософії включає більше 6500 термінів, що відбивають галузеву термінологію; термінологію інших наук, коли відповідні проблеми розглядаються в аспекті філософії; загальну лексику. Тезаурус містить 6 544 словникові статті, з них 4 325 дескрипторів і 2 219 аскрипторів; включає два показники: алфавітний лексико-семантичний і пермутаційний. Тезаурус з релігієзнавства містить близько 8300 термінів, що відображають різні аспекти релігієзнавства з дотриманням принципу історизму; включає два показники: алфавітний лексико-семантичний і пермутаційний.

Галузеві тезауруси ІНІСН РАН входять до складу Великого інформаційного словника з суспільних наук (Большого информационного словаря по общественным наукам), забезпечуючи як галузевий, так і міжгалузевий пошук інформації. Призначені для використання в наукових бібліотеках незалежно від відомчої належності та юридичного статусу (самостійна юридична особа, структурний підрозділ юридичної особи), в центрах науково-технічної інформації, в інших установах, що надають бібліотечно-інформаційні послуги в галузі соціальних і гуманітарних наук.

Ще одним відомим ресурсом є Тезаурус NASA [30], який містить авторизовані тематичні терміни Національного управління з аеронавтики й дослідження космічного простору (НАСА) (National Aeronautics and Space Administration, NASA), що використовуються для індексації та отримання матеріалів у Сервері технічних звітів НАСА. Сфера цього контрольованого словника включає не тільки аерокосмічну інженерію, але й усі галузі техніки та фізики, природничі космічні науки (астрономія, астрофізика, планетарна наука), науки про Землю і біологічні науки. Тезаурус NASA містить понад

18400 тематичних термінів, 4300 визначень та більше 4500 перехресних посилань.

Науково педагогічною бібліотекою імені К. Д. Ушинського (Москва, Росія) розроблений «Тезаурус по педагогике и образованию» [31], який ведеться з 1993 року та нараховує понад п'ять тисяч термінів. Тезаурус розроблений на основі Тезауруса ЮНЕСКО. Застосовується як інформаційно-пошукова мова вербального типу в АБІС. Тезаурус трансформується в робочу програму шляхом переміщення даних через буфер обміну операційної системи WINDOWS. Він постійно поповнюється російськомовною педагогічною термінологією з англійськими еквівалентами.

Терміни (дескриптори) розташовані у восьми лексико-семантичних полях, які визначаються як: 1. Контекст. 2. Управління та дослідження. 3. Навчання (викладання). 4. Люди. 5. Розвиток і навчання. 6. Зміст. 7. Речі. 8. Визначники місця і часу. Поля тезауруса поділяються на лексико-граматичні групи (фасети). Фасета складаються з термінів (дескрипторів), що охоплюють пов'язані між собою поняття. Структура тезауруса представлена двома списками: алфавітним (дескриптори розташовані в алфавітному порядку) і деревоподібним (дескриптори розташовані в ієрархічному порядку). Система передбачає можливість пошуку дескриптора по його фрагменту (пермутації).

Тематична область тезауруса відображає педагогічну науку й практику, а також суміжні галузі знань.

У бібліотеці Російської академії наук (Санкт-Петербург) було розроблено двомовний російсько-англійський тезаурус «Інформаційно-пошуковий тезаурус зі збереження документів» («Информационно-поисковый тезаурус по сохранности документов») [32]. Тезаурус налічує 5166 термінів. Оскільки основною мовою вибрано англійську, то дескрипторами виступають англійські терміни, а російські терміни наводяться як синоніми.

Завдання цього інформаційно-пошукового тезауруса – індексування документів і опрацювання запитів щодо забезпечення збереження документів у традиційних і нетрадиційних інформаційних системах. Він може використовуватися як спеціальний двомовний словник при перекладі з російської на англійську чи з англійської на російську мову, а також як термінологічний словник-довідник.

До складу тезауруса входять: лексико-семантичний показник, ієрархічний показник та хронологічний ідентифікатор. У лексико-семантичному показнику тезауруса розташовані дескрипторні й аскрипторні статті.

У дескрипторній статті терміни розташовуються за такими індексами: «с» – синонім, «в» – вищий дескриптор, «н» – нижчий дескриптор. У дескрипторних статтях тезауруса відсутні лексичні примітки та дескриптори, які пов'язані іншими видами відношень (рис. 1.1).

ACCESS CONTROL
с Visitors control
Контроль и управление доступом
в ORGANIZATION OF SECURITY SYSTEMS
н ACCESS CONTROL SYSTEMS

DIGITAL LIBRARIES
с Electronic libraries
Электронные библиотеки
в DIGITIZING
н DIGITAL RECORDINGS

Библиотечные фонды
см LIBRARY STOCK

Документация
см DOCUMENTATION

Рис. 1.1 Статті інформаційно-пошукового тезауруса
зі збереження документів

Як показує аналіз джерел, в Україні є спроби розроблення інформаційно-пошукових тезаурусів, які можуть безпосередньо входити до складу лінгвістичного забезпечення автоматизованих інформаційних систем,

зокрема й автоматизованих бібліотечних інформаційних систем. При цьому доцільно відмітити, що робота по розробленні комп'ютерних термінологічних тезаурів різних типів, як загальномовних, так і вузькогалузових, проводиться співробітниками лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка в рамках наукових студій з формалізації мовних досліджень. Так науковцями розроблено тезаурус з комп'ютерної лексикографії [33] (онлайн-версія розміщена на сторінках Лінгвістичного порталу MOVA.info у розділі «Словники» (<http://www.mova.info/Page.aspx?l1=61>), що представляє терміносистему комп'ютерної ідеографії, яка налічує 75 термінів.

Макроструктура тезауруса репрезентована родо-видовим деревом термінів. Нульовий рівень тезауруса представлено терміном *комп'ютерна лексикографія*, який є родовим для концепту першого рівня *комп'ютерна ідеографія*. Другий рівень має чотири концепти: одиниці комп'ютерного тезауруса, відношення між одиницями тезауруса, комп'ютерний тезаурус та укладання тезауруса, які містять відповідно 5, 8, 10 і 6 термінів третього рівня. Максимальна глибина ієрархізації тезауруса з комп'ютерної лексикографії складає шість інтервалів. Терміни розташовані в тематично-алфавітному порядку й об'єднуються видовими, родовими та синонімічними відношеннями. Словникова стаття тезауруса складається з заголовної одиниці-терміна та дефініції (див. Додаток І).

Співробітниками лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка також розроблено тезаурус лінгвістичних термінів [34], який знаходиться у вільному доступі на сторінках Лінгвістичного порталу MOVA.info у розділі «Словники» (<http://www.mova.info/Page.aspx?l1=61>).

Тезаурус містить вживані загальнолінгвістичні терміни; терміни окремих прикладних напрямів лінгвістики; терміни з комп'ютерної лінгвістики, пов'язані з автоматизацією лінгвістичних процесів. Кількість

термінів у тезаурусі – 3394 одиниці, які складають 9265 семантичних відношень. Основними відношеннями є гіпонімія, парціація (частина – ціле), синонімія, кореляція, асоціація, локалізація об'єкта, його призначення, функція, відношення, імплікація тощо. Дескриптори мають перекладні аналоги англійською та російською мовами, тоді як дефініція – українською.

Відповідь на запит користувач отримує у формі тезаурусного графа, зображеного у вигляді семантичної мережі – ієрархічно організованої структури даних (термінів-вузлів та дуг), які представляють різні типи відношень та подаються у тезаурусі в текстовому вигляді (див. Додаток І).

Активне впровадження в практичну роботу бібліотек автоматизованих технологій вимагає високого лінгвістичного забезпечення АБІС. Уніфікації пошукових термінів та веденню діалогу між людиною та комп'ютером з використанням інформаційно-пошукової мови сприяють саме тезауруси. Цьому допомагає розроблення фахівцями відділу лінгвістичного забезпечення інформаційно-пошукових систем Національної бібліотеки України імені Ярослава Мудрого (Національної парламентської бібліотеки України) «Інформаційно-пошукового тезауруса» Національної бібліотеки України імені Ярослава Мудрого (див. Додаток Й), який призначений для предметизації і координатного індексування документів. Тезаурус представлено в електронному та друкованому варіантах [35, 36]. Це перший україномовний інформаційно-пошуковий тезаурус.

Інформаційно-пошуковий тезаурус одночасно використовується для двох видів індексування: предметизації та координатного індексування; є авторитетним джерелом для складання предметних рубрик на документ; ним користуються для формування авторитетних файлів предметних рубрик Національної бібліотеки України імені Ярослава Мудрого [37].

Тезаурус Національної бібліотеки України імені Ярослава Мудрого використовується в бібліотечних та інформаційних установах, де аналітико-синтетичне опрацювання документів здійснюється за допомогою уніфікованої пошукової мови, а також як термінологічний словник зі

структурними зв'язками, що відображають місце терміна в системі понять. Будь-яка бібліотека або інша інформаційна установа, за допомогою цього тезауруса може формувати власні авторитетні файли предметних рубрик або визначати ключові слова [37].

Тезаурус містить 36730 термін, зокрема більше 15 тис. дескрипторів, 5 тис. аскрипторів (станом на 2020-08-13). У ньому наявні три класи основних відношень: еквівалентні («*B*» означає «вживається» – дескриптор, «*HB*» означає «не вживається» – аскриптор), ієрархічні («*III*» – «ширший термін» – родовий термін, «*H*» означає «вужчий термін» – видовий термін), асоціативні («*A*» – «асоціативний термін» – термін якимось чином пов'язаний з даним поняттям, але не є його синонімом, родовим або видовим поняттям).

До тезауруса включені універсальні терміни, що означають: дії й процеси; поняття наук; назви організацій і установ; види, типи й назви матеріалів, біологічних організмів, хімікатів, сільськогосподарських культур, порід тварин тощо; назви народів, мов (природних, штучних, комп'ютерних), небесних тіл; види мистецтва, літератури й архітектури; соціальні й природні явища (війни, революції, битви, катастрофи та ін.); назви священних книг, міфічних персонажів; континенти й регіони; країни, їхній адміністративно-територіальний поділ (головним чином України); назви етнічних земель, елементів земної поверхні (океанів, морів, гір тощо), парків, заповідних зон.

Інформаційно-пошуковий тезаурус складається з 8 розділів: Лексико-семантичне зібрання термінів (загальний розділ); Показчик географічних назв; Показчик персоналій; Показчик установ і організацій; Показчик формальних підзаголовків; Показчик російських відповідників; Показчик термінів латинською мовою; Показчик термінологічних джерел.

Проте до дескрипторів у тезаурусі подаються лише російські відповідники та відповідники латинською мовою на переважну більшість термінів з біології та медицини, а словникова стаття – тільки українською мовою.

Актуалізація тезауруса здійснюється саме власником цього інформаційного продукту, Національною бібліотекою України імені Ярослава Мудрого, централізовано, за визначеною періодичністю.

Науковцями Харківського державного університету харчування та торгівлі (ХДУХТ) створено двомовний (українсько-російський) галузевий тезаурус лінгвістичного забезпечення електронної бібліотеки ХДУХТ [38]. Тезаурус складається з алфавітного і систематичного покажчиків. Алфавітний покажчик містить алфавітний перелік дескрипторних статей. Дескрипторна стаття має заголовний дескриптор; ключові слова з класу еквівалентності; дескриптори, що підпорядковуються заголовному; дескриптори, асоційовані з заголовним. Систематичний покажчик служить для розкриття, обліку та контролю парадигматичних відношень між дескрипторами. Текстовий файл тезауруса становить контрольовану систему рубрик, пов'язаних ієрархічними й асоціативними зв'язками. Галузевий тезаурус ХДУХТ містить терміни з розділів «Плоди та овочі» і «Рослини та прянощі». Він постійно доповнюється новими даними (див. Додаток К).

Проведений аналіз літератури показує, що маємо спроби розроблення вузькоспеціалізованих галузевих тезаурусів. Так, науковці Таврійського національного університету працюють над створенням тезауруса предметної області «Нанотехнології» [39], Національної академії Державної прикордонної служби України ім. Б. Хмельницького – тезауруса управління підрозділами, органами, управліннями Державної прикордонної служби України [40], Одеського національного університету імені І. І. Мечникова – тезауруса будівельної термінології [41], Житомирського державного університету імені Івана Франка над побудовою двомовного (українсько-англійського) тезауруса авіації [42], Інституту надтвердих матеріалів ім. В. М. Бакуля Національної Академії Наук України над створенням автоматизованої онтолого-тезаурусної системи управління знаннями в предметній області «Надтверді матеріали» [43], Національної наукової сільськогосподарської бібліотеки Національної академії аграрних наук

України (ННСГБ НААН) над розробленням Галузевого тезауруса з сільського господарства [44] та ін.

1.2 Теоретичні засади міждисциплінарності та інформаційного забезпечення наукових досліджень

У сучасному суспільстві міждисциплінарність стає основною рисою сучасної науки, оскільки міжнауковий синтез, інтеграція та взаємопроникнення наук набувають масштабного характеру. У час створення «розумних міст», що передбачає покращення та спрощення управління містом, благоустрій міського середовища, забезпечення безпеки та підвищення якості життя жителів міста, зростає роль наукових знань, роль теоретичної, фундаментальної науки. Розумне місто (англ. Smart City) – це ефективна інтеграції фізичних, цифрових і людських ресурсів у штучному середовищі заради сталого майбутнього своїх громадян. Таким чином маємо формування нової галузі науки, яке неможливе без збагачення методології досліджень, наповнення її інструментарієм міждисциплінарного характеру.

Міждисциплінарний підхід розглядається у науковій літературі як підхід, що спрямований на створення практичних досліджень, які використовують декілька дисциплін, і який веде до об'єднання цих дисциплін для налагодження діяльності однієї предметної галузі.

Для комплексної характеристики категорії міждисциплінарності, на думку професора А. Колота, необхідно розглядати її з декількох позицій, серед яких щодо впровадження міждисциплінарного підходу до використання тезауруса як інструмента інформаційного забезпечення міждисциплінарних досліджень доцільно виділити такі:

- міждисциплінарність – це можливість побачити, розпізнати, сприйняти те, що було прихованим у окремо взятій науці за умови використання методів та інструментарію інших наук;

- міждисциплінарність – це запозичення методів, інструментарію, результатів дослідження науками, що перебувають у взаємному зв'язку, а також використання їхніх теоретичних схем, моделей, категорій, понять;
- міждисциплінарність – це не лише запозичення методів, інструментарію з інших наук, а й інтеграція останніх з метою конструювання міждисциплінарних об'єктів, предметів, опрацювання яких дозволяє отримати нове наукове знання;
- міждисциплінарність – це науково-педагогічна новація, яка дає можливість побачити, розпізнати, сприйняти те, що є недоступним у межах однієї науки з її специфічним, вузькоорієнтованим об'єктом, предметом і методами дослідження;
- міждисциплінарність – у широкому, функціональному розумінні – це синергія різних наук, що передбачає розвиток інтеграційних процесів, зростаючу взаємодію методів, інструментарію задля отримання нового наукового знання [45]. Науковець акцентує на тому, що йдеться не про механічне запозичення, а про об'єднання, побудову нових парадигм, утворення нових міждисциплінарних об'єктів та предметів дослідження.

На практиці міждисциплінарний підхід, на думку А. Колота, може реалізовуватися двома підходами. За першим підходом, міждисциплінарність «образно кажучи наводить мости між різними науками, неформально об'єднує їх, не порушуючи їхньої самостійності, унікальності, своєрідності» [45]. За другим підходом міждисциплінарність «постає як реальний інструмент об'єднання наук, появи інтегрованих продуктів, проєктів, міждисциплінарних об'єктів дослідження, подальше опанування яких є принципово важливим для науки» [45].

Суть міждисциплінарного підходу у сфері наукових досліджень – це збагачення споріднених наук на основі запозичення методологічного інструментарію, об'єднання зусиль для пояснення природи нових явищ і процесів та з'ясування тенденцій суспільного розвитку. Саме міждисциплінарність сприяє залученню методологічного інструментарію

суміжних галузей науки і формування наукового знання. Міждисциплінарність наукових досліджень ґрунтується на використанні можливостей наявних концепцій, теорій, учень, сформованих науковцями різних галузей знань. Аналіз понять різних галузей сприяє формуванню нового теоретичного знання, нової галузі.

Будь-яке міждисциплінарне дослідження зводиться до виявлення нових відношень між поняттями вихідних дисциплін, установлення нової системи законів, що їх пов'язують, та синтезі прагматики вирішення нових завдань. Знання вихідних дисциплін можуть подаватися двома формами:

- найпростіший лінійний випадок міждисциплінарної взаємодії – це коли знання залишаються незмінними або можуть бути включеними повністю / вибірково в нову ієрархічну структуру;
- нелінійна міждисциплінарна взаємодія – знання можуть зазнавати модифікацій, розвиватися завдяки процесам обміну парадигмальними настановами, поняттями й методами різних наук, що є відображенням сутності міждисциплінарної проблеми [46].

Міждисциплінарність є актуальною для усіх галузей, які є дотичними до формування предметної області «Розумне місто».

Розвиток «розумних міст» супроводжується чималою кількістю проблем, розв'язання яких (в окремому місті) повинно отримати загальне визнання, чітке формулювання й приводити до консолідації наукових, організаційно-технічних, матеріальних і фінансових зусиль суспільства для вирішення цих проблем у вигляді науково-технічних досліджень і програм відповідного рівня.

Загальносвітовою тенденцією сучасного розвитку суспільства є удосконалення інформаційних та комунікаційних технологій, розроблення алгоритмів та підходів до опрацювання інформаційних ресурсів. Накопичення та оновлення наукової інформації потребує нових підходів до організації процесу інформаційного забезпечення міждисциплінарних наукових досліджень, а одночасно й роботи бібліотек загалом. Відповідного

інформаційного забезпечення потребують наукові дослідження, спрямовані на отримання та опрацювання інформації, що балансує на межі різних галузей знань. Використання сучасних інформаційних та комунікаційних технологій забезпечує інформаційно-комунікаційні зв'язки між творчими колективами вчених та бібліотечно-інформаційний супровід наукових досліджень. Своєчасне отримання інформації, її використання для аналізу отриманих у певній предметній області результатів необхідні для проведення успішної роботи науковців [47]. Безперечно, на якість міждисциплінарних наукових досліджень впливає ефективне інформаційне забезпечення науковців, оскільки саме пошук інформації є основним етапом будь-якого дослідження, тобто своєчасне й повноцінне виконання їхніх інформаційних потреб бібліотечними працівниками.

Інформаційне забезпечення – це комплексна ресурсна система, що включає інформаційні ресурси, інформаційні технології, технічні засоби та програмне забезпечення, дія яких спрямована на задоволення інформаційних потреб для вирішення соціальних, економічних і управлінських завдань [48]. Інформаційне забезпечення наукових досліджень охоплює систему пошуку, опрацювання, накопичення, збереження, перетворення інформації в дані, які використовуються науковцями в процесі наукових досліджень. Завдання системи полягає в інформуванні науковців про стан об'єктів, які досліджуються. Одночасно виступає засобом комунікації наукових колективів.

Інформаційне забезпечення як техніко-технологічний процес ґрунтується на системі інформаційних ресурсів, спільне використання їх спрямоване на формування коректних інформаційних потоків і баз даних, дотримання порядку збирання, опрацювання, передачі й перетворення інформації, що досягається застосуванням сучасних інформаційних та телекомунікаційних технологій [48].

Термін «інформація» в перекладі з латинської мови «*informatio*» означає «пояснення», «повідомлення» й спочатку використовувався лише до

засобів зв'язку, а потім у ширшому значенні. Інформація – це знання, дані, які отримуються й нагромаджуються в процесі розвитку науки та практичної діяльності людства. У вузькому розумінні інформація – це дані, що є об'єктом опрацювання, зберігання та передавання. Відповідно до Закону України «Про інформацію» інформація – це будь-які відомості та / або дані, які можуть бути збережені на матеріальних носіях або відображені в електронному вигляді [49].

Розглядаючи інформаційне забезпечення наукових досліджень, потрібно звернути увагу на поняття науково-технічної інформації. Згідно з чинним законодавством України [49] науково-технічна інформація – це будь-які відомості та / або дані про вітчизняні та зарубіжні досягнення науки, техніки й виробництва, отримані під час науково-дослідної, дослідно-конструкторської, проектно-технологічної, виробничої та громадської діяльності, які можуть бути збережені на матеріальних носіях або відображені в електронному вигляді.

Роль інформації полягає в тому, що будь-яке наукове дослідження не може бути проведено без неї. Проте слід пам'ятати, що інформація швидко старіє, тому потрібне постійне оновлення даних. Швидкість старіння інформації, за даними зарубіжних джерел, складає для газет понад 10% на день, для журналів 10% на місяць, для книг та монографій 10% на рік. Результатом праці дослідників виступає також інформація, але уже якісно нова, яку вони продукують, опрацьовуючи релевантну інформацію, на пошуки якої витрачають приблизно 50% свого часу. Тому вміння швидко знаходити й опрацьовувати необхідну інформацію є важливим етапом наукового дослідження.

Інформація, для того, щоб бути придатною для використання науковцями, повинна бути:

- правдивою – реально відбивати факти, події і ситуацію в певних галузях;
- актуальною – новою, відповідати сьогоденню;

- достатньою – повно, глибоко й всебічно з'ясовувати проблеми, що досліджуються;
- автентичною – подаватися в чіткій зрозумілій формі, лексично відповідати моделям сучасної державної мови.

Сучасні інформаційні та комунікаційні технології широко використовуються в процесі міждисциплінарних наукових досліджень. Інтернет-простір – складний об'єкт і потужний засіб наукових досліджень, ефективність якого зумовлена розвитком інформаційно-пошукових систем. Завдання таких систем – пошук релевантних документів у інформаційних масивах та базах даних.

Інформаційне забезпечення наукових досліджень це сукупність інформаційних процесів, що забезпечують підготовку й надання наукової інформації користувачу для вирішення наукових завдань відповідно до етапів їхнього виконання.

У ході підготовки й проведення наукових досліджень виділяють п'ять основних етапів:

1. Загальне ознайомлення з проблематикою дослідження та встановлення рівня її розробленості: етап накопичення наукової інформації, пошук наукової бібліографічної інформації, вивчення документів, відбір основних джерел із теми дослідження, формування списку літератури, вибір плану дослідження.

2. Формулювання теми, мети й цілей дослідження, визначення проблеми, обґрунтування об'єкта й предмета, розробка гіпотези дослідження, постановка головних завдань дослідження.

3. Проведення теоретичного дослідження – обґрунтування напрямів, вибір загальної методики, методів, розроблення концепції, параметрів, формулювання висновків дослідження.

4. Організація та проведення експерименту – розроблення програми, методики, одержання та аналіз даних, формулювання висновків і результатів дослідження; узагальнення та синтез експериментальних даних.

5. Оформлення результатів наукового дослідження, висновків, рекомендацій, уточнення наукової новизни та практичної цінності [50].

Результати наукових досліджень сфокусовані в різних видах публікацій (книгах, статтях тощо). Отриманню нових наукових даних науковцями передуює проведення громіздкої роботи зі систематизації накопичених знань у предметній області дослідження. Опрацювання наукової інформації потребує від науковців певних навиків та вмінь роботи з інформаційними джерелами. Інформаційне забезпечення цього етапу дослідження передбачає наявність у бібліотекаря вміння знаходити, опрацьовувати, зберігати й використовувати інформаційні повідомлення, що допомагає швидко отримувати актуальну інформацію та ефективно переробляти її.

Для підвищення результативності інформаційного забезпечення бібліотечний працівник повинен бути обізнаним із національною та зарубіжними системами науково-технічної інформації, зокрема науково-інформаційним, програмно-технологічним і організаційно-методичним забезпеченням системи формування інформаційних ресурсів в Україні та за кордоном, а також особливостями реферування вітчизняної і зарубіжної наукової літератури. Сьогодні в Україні існує система інформаційно-аналітичного обслуговування користувачів, яка використовує національні та світові комп'ютерні мережі, формування баз даних. Інформаційні послуги дослідникам надають бібліотеки та наукові установи. Загальносвітові тенденції стрімкого оновлення інформації зумовлюють зростання кількості електронних бібліотек, які дозволяють працювати як з бібліографічними базами даних, так і з повнотекстовими електронними виданнями наукової літератури [51].

Організація та методика проведення інформаційного забезпечення ускладнюються постійним зростанням кількості наукової інформації, що потребує систематизації та впорядкування. Науковий матеріал потрібно збирати системно й володіти навиками класифікації інформації. З огляду на завдання інформаційного забезпечення та його ролі в науково-дослідному

процесі основою його класифікації є професійна інформаційна комунікація [50]. Опублікована інформація допомагає детальніше зрозуміти практичний та науковий матеріал інших дослідників, з'ясувати рівень досліджуваної теми. Необхідну наукову інформацію дослідник отримує в бібліотечних установах, де йому допомагають бібліотекарі, використовуючи інформаційно-пошуковий апарат установи, основою якого є традиційні (на паперових носіях) та електронні каталоги – упорядковані бібліографічні записи, що допомагають швидко зорієнтуватися в швидкоплинному науково-інформаційному потоці.

У процеси інформаційного забезпечення активно задіюються інформаційно-пошукові системи (ІПС), призначені для пошуку, опрацювання, збереження і видачі інформації, що цікавить користувача, тобто релевантної, що складає відповідність між бажаною інформацією та отриманою інформацією. Релевантність трактується як міра близькості між реально одержаними документами й тою інформацією, яку хотілося б отримати із системи. Виникає дві задачі, які варто вирішити: представлення інформації в системі й формулювання інформаційних потреб користувача. Абонент звертається до інформаційно-пошукової системи з інформаційним запитом, що відображає його інформаційну потребу, для цього створюється пошуковий образ документа – текст інформаційно-пошуковою мовою, що виражає основний зміст документа [52].

Досконала інформаційно-пошукова система повинна видавати документи, релевантні запиту користувача. Проте на практиці спостерігається відсутність повної відповіді, тобто всіх релевантних документів, а разом з тим і видача зайвих документів (шуму). Інакше кажучи, користувач як результат запиту отримує масив документів, серед яких релевантні та нерелевантні документи. Цілий ряд факторів має вплив на отримані результати: як внутрішні властивості пошукової системи (обсяг і характеристика інформаційного масиву, інформаційно-пошукова мова), так і зовнішні умови (ступінь специфічності інформаційних запитів, здатність

користувача правильно сформулювати свої інформаційні потреби природною мовою, правильність побудови конкретного запиту). Таким чином, саме через помилки чи неточності, які з'являються на кожному етапі роботи як користувача, так і системи, отримані результати можуть значно відрізнятись від тих, які хотів би одержати користувач, використовуючи інформаційно-пошукові системи.

Серед поширених пошукових систем наукової інформації є Open Ukrainian Citation Index (OUCI), Google Scholar, Meta, BASE: Bielefeld Academic Search Engine, WorldWideScience.org, EconBiz (Virtual Library for Economics and Business Studies) (економіка), FreeFullPDF (медицина), ScienceResearch.com, «Наукова періодика України» (загальнодержавна технологічна платформа на базі Open Journal Systems) та ін., завдання яких полягає в спрощенні пошуку наукових публікацій українських та зарубіжних науковців.

Наукова діяльність сьогодення потребує відповідного інформаційного підґрунтя, надійного та оперативного інформаційного забезпечення. Тому необхідно знайти серед функціонуючих механізм або створити такий, що:

- інтегрує знання всіх учасників процесу;
- забезпечує комунікацію та рольове порозуміння;
- дозволяє усвідомити інноваційний характер дій;
- здійснює м'яке управління процесом взаємодії;
- використовує та відтворює інфраструктуру продуктивної інноваційної діяльності [53].

Подібним механізмом може стати механізм створення й підтримки тезаурусів. Усвідомлення особливостей інформаційно-інтелектуального забезпечення наукових досліджень епохи інформатизації потребує не тільки використання потенціалу тезаурусного підходу до організації діяльності, а й орієнтує на розуміння його перспектив, які пов'язані з:

- формою знання, що існують у інформаційних мережах;
- логістизацією інноваційних потокових ресурсів;

– особливостями становлення тезаурусних структур [53].

Швидко відшуковувати, узагальнювати, ефективно використовувати релевантну інформацію для виконання відповідних завдань науковому досліднику дозволяє саме чітке знання термінів і їхньої сутності, а також галузей інформації.

Складання запитів для пошукових машин – складний етап, що пов’язаний з опрацюванням великого масиву інформації в неструктурованих чи слабоструктурованих масивах (в мережі «Інтернет»). Формуванню якісного запиту до вибраних пошукових серверів для отримання релевантної інформації сприяє використання тезауруса, що використовує семантичні відношення між термінами та передбачає опрацювання синонімів, омонімів і морфологічних варіацій ключових слів. Він може виступати як основа для незалежної програми-асистента, завдання якої – формування запитів до пошукових машин.

1.3 Методологічні засади дослідження

Чіткої парадигми розроблення тезаурусів вузькоспеціалізованих предметних областей для пошуку інформації релевантної інформаційним потребам користувача на сьогодні ще не вироблено. Беручи до уваги багатоаспектність та міждисциплінарний характер дисертаційного дослідження, методологічні засади роботи опираються на широкий спектр принципів та засад, що представлені на різних рівнях методологічної науки.

Існують різні тлумачення змісту методології. Найчастіше під цим поняттям подають сукупність підходів, способів, методів і прийомів, які використовуються в пізнавальному процесі й практичній діяльності та допомагають досягненню пізнавальної чи практичної мети. Так, Р. Лукич під методологією розуміє сукупність методів, що використовуються в науці чи іншій сфері пізнання [54]. За М. Вебером методологія – це засоби, що виправдали себе на практиці. Методи наук формувалися в ході вирішення

конкретних проблем [55]. М. Кельман подає методологію як учення про структуру, логічну організацію, методи, засоби та форми роботи дослідника в процесі пізнання явищ, які він досліджує [56]. Вітчизняні науковці методологію розглядають як учення про науковий метод пізнання або як систему наукових принципів, на основі яких базується дослідження й здійснюється вибір сукупності пізнавальних засобів, методів, прийомів дослідження [57].

Методологія дослідження – процес, спрямований на одержання нових наукових результатів, досягнення поставленої мети та виконання запланованих завдань, що базується на загальнонаукових принципах та методах і передбачає застосування певних прийомів, зумовлених особливостями роботи. Будь-який метод, система принципів та інших засобів діяльності не можуть забезпечити найбільш успішного вирішення теоретичних і практичних проблем. Кожне наукове дослідження потребує вибору раціональної внутрішньої організації та регулювання процесу пізнання.

Вирішальну роль у досягненні результатів нашого дослідження відіграла оптимізація вибору методів, що склали основу вибраного інструментарію, визначення можливих сфер застосування та відбору найефективнішого методу в певному конкретному випадку. Вибір та використання методу залежали від поставленої мети та часу, відведеного на проведення аналізу й дослідження окремих завдань. Методологія нашого дослідження включає, зокрема, загальнонаукові методи, методи інформатики, математичні, соціокомунікаційний та історико-хронологічний методи.

У процесі дослідження були поєднані загальнонаукові методи та прийоми дослідження: джерелознавчого пошуку й аналізу профільних публікацій з теми дослідження – для визначення стану розробленості наукової проблеми; синтез, дедуктивний метод (побудова тезауруса на основі переходу від ширших термінів до вузких); індуктивний метод (тезаурус

формується на основі переходу від вузьких термінів до ширших); зіставлення (для виявлення відмінних рис мовної форми вираження та змістового наповнення термінів, їхніх структурних компонентів та елементів в українській та англійській мовах); термінологічний аналіз (для уточнення термінів з метою їх унормування та упорядкування; для формування критеріїв вдалого терміна тощо); узагальнення (процес встановлення загальних властивостей і ознак предметів); а також аналогія, моделювання, формалізація та інші методи, які носили не лише методологічний, а й прикладний характер, оскільки сприяли критичному осмисленню наявного матеріалу і його змістовій інтерпретації у вирішенні поставлених завдань.

Для кожного теоретичного дослідження потрібен опис, аналіз та уточнення понятійного апарату, тобто термінів і понять досліджуваної галузі знань. Вивчення термінів і понять певної сфери, формулювання та уточнення змісту й обсягу понять, визначення їхнього місця в понятійному апараті, на основі якого базується дослідження – усе це вирішується за допомогою методу термінологічного аналізу.

Дефініції понять потрібно формулювати, використовуючи тлумачні та професійні словники. Обсяг та зміст поняття дається через родову ознаку й найближчу видову відмінність. Спочатку називається родове поняття, яке поєднує інші поняття, що визначаються. Потім виділяється та ознака поняття, яка відрізняє його від усіх подібних, важливо, щоб ця ознака була найважливішою і найсуттєвішою. Ю. Тарелкін та В. Цикін у своїй роботі [58] подають певні правила визначення понять: 1) правило розмірності вимагає, щоб обсяг поняття, що визначається, відповідав обсягу поняття, яке визначає, тобто ці поняття мають бути тотожними; 2) нове поняття не повинно бути тавтологічним; 3) поняття має бути чітким і однозначним. Якщо при визначенні поняття важко зазначити одну ознаку, називають декілька ознак, достатніх для розкриття специфіки його обсягу й змісту.

Загальнонаукові методи та прийоми використовуються в більшості наук, так як будь-які наукові знання, крім предметного змісту, мають і

методологічний зміст, спричиняють критичний перегляд сформованого та прийнятого понятійного апарату, чинників, передумов і підходів до інтерпретації матеріалу, що вивчається.

Серед загальнонаукових методів вагоме місце займає системний підхід, застосування якого потребує будь-який об'єкт наукового дослідження, зокрема й розроблення інформаційно-пошукових тезаурисів. В основі системного підходу лежить аналіз об'єктів як системи. Він дозволяє здійснювати комплексне дослідження великих та складних об'єктів-систем як єдиного цілого при узгодженому функціонуванні всіх елементів і складових.

Важливим аспектом системного підходу є вироблення нового принципу його використання – створення оптимального способу пізнання, для застосування до будь-якого пізнавального об'єкта, з метою отримання повного й цілісного уявлення про нього [59].

Швидке зростання кількості інформації в «розумних містах» є однією з передумов, що визначила роль системного підходу в дослідженні предметної області «Розумне місто». Застосування системного аналізу до дослідження терміносистеми «розумного міста» та формування тезауруса для здійснення пошукових запитів забезпечує підвищення ступеня обґрунтованості із врахуванням таких методичних принципів системи, як: динамізм, структурність, органічна єдність суб'єктивного та об'єктивного, міждисциплінарний характер системних досліджень, органічна єдність формального та неформального підходів до системного аналізу.

Методологія системного аналізу базується на відомих законах взаємозв'язку та взаємозумовленості явищ [59], що в контексті суспільних процесів потребує аналізу будь-якої предметної області не тільки як самостійної системи, а як підсистеми значно більшої та складнішої системи – соціокомунікаційної.

Використання системного підходу, тобто представлення об'єкта дослідження у формі системи, дає можливість отримати найбільш

ефективний опис об'єкта. Об'єкт розглядається тільки в тому відношенні, у якому він виступає як система.

Відповідно до системного підходу, система – це «цілісність, яка становить єдність закономірно розташованих і взаємопов'язаних частин» [58].

Предметну область «Розумне місто» ми розглядаємо як певну систему, якій притаманна множина взаємозв'язаних елементів, компонентів, підсистем, визначених функцій, цілей та структура. Ця галузь подається як багатофункціональна матеріальна та концептуальна, відкрита, складна, динамічна, детермінована, цілеспрямована, регульована система.

Виходячи із загальних характеристик будь-якої системи, науковці виділяють відповідні характеристики системи, які забезпечують системну спрямованість практичного дослідження об'єкта та його наукового пізнання [60].

Так системний підхід у дослідженні термінології предметної області «Розумне місто» та її уніфікації сприяв виділенню та вивченню таких характеристик системи як:

- цілісність, що дозволило розглядати елементи системи одночасно і як єдине ціле, і як підсистему рівнів, що знаходяться вище;
- структуризація дозволила аналізувати елементи системи з їхніми взаємозв'язками в рамках конкретної організаційної структури, як правило, функціонування системи обумовлене не стільки властивостями її окремих елементів, скільки властивостями самої структури;
- ієрархічність дозволила аналізувати систему як множину елементів, які функціонують на основі підпорядкування елементів нижчого рівня елементам вищого рівня;
- цілеспрямованість дозволила досягнути мети – розроблення тезауруса «розумного міста»;
- взаємозв'язок із зовнішнім середовищем – кожна система змінюється та вдосконалюється відповідно до змін зовнішнього середовища.

Дотримання перелічених характеристик системи сприяло глибшому розумінню системи знань предметної області «Розумне місто» в її організаційній єдності, цілісності, взаємозалежності, зв'язку із зовнішнім середовищем.

Потреба у використанні системного підходу виникає при наявності таких ситуацій [61]:

- необхідність розв'язання нової проблеми, сформульованої за допомогою методів та підходів системного аналізу;
- вирішення проблеми передбачає координацію цілей з множиною засобів їхнього досягнення;
- досліджувана проблема має розгалужені зв'язки, які спричиняють наслідки в різних галузях, і прийняття рішення в таких випадках потребує врахування факторів сукупної ефективності;
- існують варіанти розв'язання проблеми або досягнення взаємопов'язаного комплексу цілей, які важко порівняти;
- створюються нові складні системи;
- здійснюється вдосконалення, реконструювання виробництва, що обумовлює необхідність реінженерії технологічних процесів;
- при створенні інформаційних систем та комп'ютеризованих систем керування;
- коли важливі рішення повинні прийматися за наявності невизначеності, ризику та мають стратегічний характер.

Застосування системного підходу до дослідження особливостей розроблення інформаційно-пошукового тезауруса базувалося на використанні таких аспектів [61]:

- системно-комплексний аспект сприяв дослідженню окремих складових досліджуваної предметної області «Розумне місто» як системи;
- системно-структурний аспект було використано для з'ясування взаємозв'язків та взаємозалежностей складових елементів предметної

області, що дало можливість чіткого представлення її внутрішньої організації;

- системно-функціональний аспект покликаний виявляти, досліджувати та вдосконалювати функції, що становлять суть функціонування галузі як системи;
- системно-цільовий аспект передбачав визначення цілей і їхніх взаємозв'язків;
- системно-ресурсний аспект полягав у комплексному аналізі джерел забезпечення успішного функціонування системи;
- системно-інтеграційний аспект полягав у визначенні сукупності якісних властивостей системи, що забезпечують її цілісність;
- системно-комунікаційний аспект допоміг виявленню зовнішніх зв'язків цієї системи з іншими, тобто, її зв'язків з навколишнім середовищем;
- системно-історичний аспект дозволив з'ясувати виникнення тезаурусів та етапи розвитку.

Застосування системного підходу забезпечило змістове розуміння терміносистеми предметної області «Розумне місто» як системного об'єкта, що характеризується власною структурою, стійкими взаємозв'язками між його елементами та підсистемами, специфічними функціями складових компонентів та системи в цілому.

Системний характер мовних явищ виявляється у взаємодії кожної одиниці мови з іншими зіставними одиницями, що функціонують у межах однієї системи. У таких підсистемах мови, якими є терміносистеми певних галузей, ця взаємодія простежується особливо чітко. Усі вони складаються з компонентів (окремих понять), між якими виникає певна взаємодія.

Наступним методом, використаним у дисертаційному дослідженні, є системно-діяльнісний підхід, що отримав широке застосування в сучасних наукових дослідженнях. Через систему компонентів діяльнісного підходу (потреба – суб'єкт – об'єкт – процеси – умови – результат) було комплексно досліджено фази побудови тезаурусів та сформовано алгоритм розроблення

тезауруса «розумного міста», певні етапи якого носили як аналітико-теоретичний, так і практичний характер. Системно-діяльнісний підхід дозволив комплексно вивчити можливість пошуку релевантної інформації за допомогою інформаційно-пошукового тезауруса, що відповідає сучасним запитам інформаційного суспільства. Завдяки цьому підходу розглядаємо тезаурус як систему, що пов'язана з діяльністю людини, яка має потреби інформаційного характеру, а взаємодія користувачів з цією системою, у процесі якої досягаються мета та завдання, провокує певні зміни в суспільному бутті та в той самий час є реакцією на деякі з цих змін.

Важливі результати досягнення поставленої мети дослідження дало застосування структурно-функціонального підходу до розроблення тезауруса «розумного міста». Структуризація дозволяє розглядати тезаурус як структурно-функціональну цілісну систему, кожна підсистема якої має певне функціональне призначення, що узгоджується із загальними цілями системи в цілому. Цілісність системи залежить від рівня відповідності її структури й функцій головній меті системи. Завдяки структурно-функціональному підходу було розроблено функціональну модель тезауруса, у якій визначено такі складові: вхідні дані моделі – метадані інформаційного ресурсу; вихідні дані – релевантна інформація, що отримується як відповідь на запит користувача; керуючими даними виступають правила формування пошукових запитів. Усі елементи системи розглядаються як окремі складові, проте вони є пов'язані взаємозв'язками в єдину системно організовану цілісність.

У роботі було застосовано дедуктивний та індуктивний методи при формуванні тезауруса. Використаний нами дедуктивний метод передбачав видобування термінів з документів на підготовчій стадії розроблення тезауруса, проте ні укладання контрольованого словника, ні визначення відношень між термінами не здійснювалося, поки не було зібрано достатньої кількості лексичних одиниць. Усі терміни потім переглядалися. Спочатку ідентифікувалися терміни, що представляють найширші категорії. Усі інші

терміни на основі їхніх логічних відношень поділялися за цими категоріями. Таким чином, категорії в тезаурусі будувалися на основі переходу від ширших понять до вужчих.

Використання індуктивного методу передбачає, що нові терміни вводяться у тезаурус одразу ж після того, як вони були отримані з документів. Контролювання тезауруса ведеться з самого початку. Кожний новий термін, який вводиться в тезаурус, позначається як член однієї чи кількох сформованих ширших категорій, які побудовані на основі списку на попередній стадії. У цьому випадку тезаурус формується на основі переходу від вужчих термінів до ширших.

Фактично на практиці на різних стадіях побудови тезауруса застосовувалися як дедуктивний, так й індуктивний методи. Категорії термінів, які були визначені індуктивно, пізніше переглядалися з дедуктивного погляду. Оскільки дедуктивний та індуктивний методи є емпіричними, природно, що деякі рішення, прийняті на ранніх стадіях формування тезауруса, вимагали їхнього перегляду з отриманням подальшого досвіду, оскільки поповнення структури тезауруса новими термінами робило зрозумілими попередньо прийняті рішення відносно зафіксованих термінів та їхніх взаємовідношень з нововведеними одиницями. Тому доводилося часто перевіряти терміни та ієрархії, щоб переконатися, що взаємовідношення між термінами будуть відповідати загальним, послідовно здійснюваним принципам. Доводилося звертатися до експертів за порадою, коли виникали проблеми вибору між варіантними формами термінів.

Вагоме місце в методології дослідження формування пошукових образів займають методи інформатики. Одним із таких методів є метод онтології.

Онтологія в інформатиці – це формалізація досліджуваної галузі знань за допомогою концептуальної схеми. Онтологія предметної області – формальний опис предметної області. Використовуються онтології предметних областей для уточнення понять певної предметної області та

визначення загальної термінологічної бази цієї галузі. Онтологія забезпечує адекватну специфікацію понять досліджуваної галузі та забезпечує словник для представлення й обміну знаннями про цю галузь та систему відношень між термінами в даному словнику. У найпростішій формі побудова онтології зводиться до виділення базових понять (концептів) досліджуваної галузі, побудови зв'язків між концептами (див. Розділ 2). Онтології дозволяють подати основні поняття так, що вони стають основними для формального інформаційного пошуку. Формальна мова опису онтології сприяє підвищенню ефективності опрацювання даних.

Для ефективного пошуку релевантних даних використовуються онтології як методологічний підхід до опису та класифікації інформації, оскільки вони є провідною парадигмою структурування інформаційних потоків як ієрархічних концептуальних структур, що формуються на основі вивчення та структурування потоків інформації [62].

Іншим методом інформатики, застосованим у дослідженні, є тезаурусний підхід (див. Розділ 3). В основі тезаурусного підходу лежить однойменний метод, який широко застосовують у теорії та практиці інформаційного пошуку. У ході структурування й моделювання знань, систематизації та впорядкування наукової термінології, формування та використання термінологічних банків даних чимале значення відіграють тезауруси. Основний акцент цього методу робиться на змістовий аналіз призначених для відбору термінів із текстів, які є важливими для запиту користувача.

Тезаурус «розумного міста» репрезентує систему термінів досліджуваної галузі, сприяє формуванню термінологічної компетенції фахівців, обумовлює оновлення термінології даної галузі знань та забезпечує галузь ефективним пошуковим інструментарієм для інформаційного забезпечення на основі наукових джерел інформації. Технологічно було опрацьовано архітектуру й сформовано прототип відкритої інформаційної системи, яку можна доповнювати.

Онтологічний і тезаурусний опис контенту інформаційних ресурсів належать до методів формалізації знань суб'єкта про предметну область із використанням формалізованих знакових систем.

Особливе місце серед запозичених методів інформатики в дослідженні посідає метод консолідації інформації. Автори підручника «Вступ до спеціальності: "Консолідована інформація"» Наталія Кунанець та Володимир Пасічник під консолідованою інформацією розглядають різнотипні інформаційні ресурси, «одержані з декількох джерел та системно інтегровані ... які в сукупності наділені ознаками повноти, цілісності, несуперечності та складають адекватну інформаційну модель предметної області з метою її аналізу, опрацювання та ефективного використання в процесах підтримки прийняття рішень» [63]. Беручи до уваги це означення ми використали переваги методу консолідації інформації з двох позицій: як групування опрацьованих джерел та як засіб досягнення мети дослідження. Перша позиція – позиція науковця-дослідника, що спочатку відбирає дані в процесі наукових розвідок, а потім консолідує опрацьовані дані для отримання цілісного уявлення про предмет дослідження. Друга позиція – позиція практика, що в процесі моделювання об'єднаного інформаційного ресурсу добирає методи та прийоми, які забезпечують консолідацію різних за походженням та поданням інформаційних ресурсів.

Консолідація інформації передбачає консолідування отриманих відомостей у базу даних та знань, яка дозволяє здійснювати управління знаннями. Таким чином, база знань, сформована для консолідації інформації є сукупністю даних галузі, організована для забезпечення зручного їхнього представлення. Створювати консолідовану інформацію означає перетворювати дані в інформацію – знання – інформаційний продукт. Процес консолідації інформації складається з таких етапів: виявлення вихідних вимог та даних; формування інформаційних ресурсів; витяг та придбання знань; надання інформації споживачеві. Кінцевий продукт повинен містити

відомості, що базуються на фактах, поданих як підстави для прийняття правильних рішень [59].

Вагоме місце в методології розроблення тезаурусів займає інформаційний підхід, суть якого полягає в тому, що при вивченні терміносистеми певної галузі знань перш за все виявляються найхарактерніші для неї інформаційні аспекти. Пізнавальні можливості інформаційного підкоду полягають у тому, що предмет дослідження вивчається в контексті інформації, її численних форм подання. Він передбачає використання пізнавальних можливостей інформаційної теорії, методів, засобів, організаційних форм і технологій, вироблених інформатикою, для визначення специфічних рис предмета дослідження. Основний дослідницький актив інформаційного підкоду полягає в тому, що всі об'єкти, процеси та явища є інформаційними, так як пов'язані зі створенням, накопиченням, обміном або використанням інформації (відомостей, знання) задля здійснення соціальної комунікації [59].

У більш вузькому значенні інформаційний підхід означає ефективне використання пізнавального потенціалу інформаційної діяльності, що розглядаються як сукупність процесів одержання, збирання, аналітико-синтетичної переробки, зберігання, пошуку та розповсюдження інформації, що використовується комунікаційними посередниками та його споживачами [59]. Інформаційна діяльність для науковців – це один із елементів наукового дослідження, які є важливими та обов'язковими, один із основних засобів досягнення мети і завдань дослідження, забезпечення достовірності його наукових положень, висновків і рекомендацій.

У рамках проведеного дослідження було використано й математичний інструмент системного підходу до складних проблем прийняття рішень – метод аналізу ієрархій, який використовується для прийняття рішень у різноманітних ситуаціях. Використовуючи цей метод, особа, що приймає рішення, може в інтерактивному режимі віднайти такий варіант (альтернативу) вирішення проблеми, який буде найкращим чином

узгоджуватися з його розумінням суті проблеми та вимогами до її вирішення. За допомогою методу аналізу ієрархій зрозумілим і раціональним чином структурується складна проблема прийняття рішень у вигляді ієрархії, порівнюється й виконується кількісна оцінка альтернативних варіантів рішення.

Нами було застосовано метод аналізу ієрархій для вибору способу представлення знань предметної області «Розумне місто». Способом декомпозиції проблеми на більш прості складові та в подальшій обробці послідовності думок експертів щодо попарних порівнянь критеріїв і альтернатив було вибрано модель представлення знань досліджуваної галузі. Так, серед запропонованих експертам моделей: логічні, фреймові, продукційні моделі, семантичні мережі та онтології, на основі множини сформульованих критеріїв (див. Розділ 2) і проведених обчислень пріоритетів було вибрано онтологічний підхід як найкращий спосіб для представлення знань досліджуваної області «Розумне місто» та формування пошукових образів у інформаційних системах різної складності.

Зважаючи на новизну та важливе соціальне значення для сучасного соціально-комунікаційного простору ідеї формування пошукових запитів на основі тезаурисів вузькоспеціалізованих предметних областей, у нашому випадку тезауруса предметної області «Розумне місто», виникла необхідність у використанні специфічних методологічних підходів, крім вище названих, до якісного вивчення цієї проблеми. Чільне місце серед них займає соціокомунікаційний підхід.

Передусім звернімося до визначення соціальних комунікацій: під соціальними комунікаціями розуміється така система суспільної взаємодії, яка «включає визначені шляхи, способи, засоби, принципи встановлення і підтримання контактів на основі професійно-технологічної діяльності, що спрямована на розроблення, провадження, організацію, удосконалення, модернізацію відносин у суспільстві, які складаються між різними соціальними інститутами, де, з одного боку, у ролі ініціаторів спілкування

найчастіше виступають соціальнокомунікаційні інститути, служби, а з іншого – організовані спільноти (соціум, соціальні групи) як повноправні учасники соціальної взаємодії. Ці комунікації є соціально маркованими, бо передбачають взаємодію із соціально визначеними групами людей. Соціальні комунікації утворюються за законами спілкування, але, як і будь-які технологічні речі, передбачають використання наукових знань про спілкування та про все, що використовується для організації суспільно комунікаційної справи» [64]. Відповідно до цього визначення соціальних комунікацій ми розглядали предметну область «Розумне місто» як соціальний інститут, тезаурус «розумного міста» як комунікаційний канал.

Професор О. Холод соціальні комунікації визначає як комплекс організованих дій (збір інформації, її опрацювання, трансляція і верифікація впливу інформації на споживача), що спрямовані на обмін соціально важливою інформацією та регулювання соціальних дій, взаємодій і відносин між соціальними суб'єктами й об'єктами в суспільстві. Соціальними комунікаціями науковець називає системні процеси формування соціальних практик за допомогою комунікаційних технологій під час створення інформації, її пошуку, упорядкування, трансляції, обміну, збереження й вимірювання впливу на реципієнта [65].

Професор В. Різун соціальну комунікацію подає як «спілкування, дане в умовах штучно створеної технологічної системи духовного єднання людей» [64]. Під соціально-комунікаційною технологією він розглядає найвищу форму прояву культури народу, культуру мислення, праці, слова, дії; науково виважений, вироблений спосіб думання [64]. Щоб досягнути цієї найвищої форми прояву культури мислення, слова, особливо при дослідженні та формуванні інноваційних галузей дослідження, якою і є галузь «Розумне місто», постає проблема формування єдиної термінологічної системи для розширення контактів і глобалізації процесів професійної комунікації [66]. Мовною моделлю даної предметної області, систематизатором термінології цієї галузі в нашому випадку виступає інформаційно-пошуковий тезаурус,

що є нормативним документом кодифікованих термінів, які орієнтують користувачів на правильне їхнє застосування. Таким чином використання соціокомунікаційного підходу як методологічного інструментарію дослідження дозволило обґрунтувати сутність тезаурусів, які реалізуються в цілеспрямованих комунікаційних зв'язках між суб'єктами та об'єктами системи, суб'єктами системи та зовнішнім середовищем.

Виходячи із позицій соціокомунікаційного підходу комунікаційна природа взаємодії суб'єктів системи ґрунтується на цілеспрямованому обміні соціально значущою інформацією, необхідною для прийняття рішень. Соціальне спілкування передбачає суб'єктно-суб'єктний зв'язок між комунікаторами. На основі соціального спілкування утворюється професійно-виробнича діяльність, метою якої стає саме соціальне спілкування як продукт діяльності, розробляються різні види соціального спілкування, що й отримує назву соціальні комунікації.

Соціокомунікаційні процеси досліджуваної галузі спрямовані на налагодження взаємозв'язків між учасниками комунікації, тобто мають комунікативний характер. Науковець у процесі спілкування чи пошуку необхідної інформації змушений сформулювати свою інформаційну потребу. Це стає важливим чинником релевантності його інформаційного пошуку. У цьому плані важливим є розуміння ним самої інформаційної потреби, форми представлення свого інформаційного запиту, щоб інші суб'єкти системи могли зрозуміти цю інформацію. Результатом такої комунікації стає підвищення культурного та комунікативного рівня суб'єктів системи (науковців, керівників, розробників та ін.), що створює відповідні умови для реалізації та розвитку їхніх творчих здібностей.

Концептуальне дослідження теми дисертаційної роботи можливе за умови цілісного підходу до об'єкта вивчення, розгляду його у виникненні та розвитку, тобто використанні історичного підходу до його вивчення. Перед тим як вивчати сучасний стан об'єкта дослідження – розроблення тезауруса предметної області «Розумне місто» – необхідно вивчити генезис та розвиток

тезауруса в загальному. У зв'язку з цим особливого значення набуває вивчення історії виникнення та розвитку самого терміна «тезаурус» та тезаурусів як одної з форм формування пошукових образів та запитів. Застосований у цьому дослідженні історико-хронологічний метод дозволив дослідити виникнення та розвиток тезаурусів у хронологічній послідовності. З'ясувати, що перші тезауруси були пов'язані з фундаментальними проблемами пізнання, відображення засобами природної мови сутності навколишнього світу та закономірності його сприйняття. Вони не враховували особливостей інформаційної діяльності.

У межах дослідження застосовувався компаративний метод, який дозволяв виявляти схожість і відмінність між явищами, що вивчалися, загальне й специфічне в їхньому розвитку. Компаративний метод дозволив шляхом порівняння зіставити різні визначення терміна «тезаурус» та сформулювати наше розуміння інформаційно-пошукового тезауруса предметної області як системи знань, поданої у вигляді набору дескрипторів та аскрипторів, пов'язаних між собою певними семантичними відношеннями.

Таким чином, методологія нашого дослідження розглядається як процес, що передбачає використання цілої системи прийомів, підходів, дій, спрямованих на досягнення поставленої мети та виконання визначених завдань, отримання нових наукових досліджень, і включає ряд принципів та методів.

Висновки до Розділу 1

У розділі проаналізовано особливості інформаційно-пошукових тезаурусів, які використовуються у зарубіжних та вітчизняних бібліотеках, як інструментів ефективного формування метаданих, створення пошукових образів та пошукових запитів. Досліджено основні види, змістовне наповнення та кількісні показники цього різновиду інформаційних ресурсів бібліотек.

Теоретичну базу дослідження становили інформаційні ресурси:

– велика база даних англомовних лексичних одиниць – комп’ютерний тезаурус WordNet (створений у Принстонському університеті, США), а також ряд тезаурусів аналогічної тематики іншими мовами (GermaNet, DanNet, MultiWordNet, RussNet та PyТез тощо), в основу розроблення яких лягла концепція WordNet-у, що свідчить про ефективність інформаційних ресурсів такої структури;

– комп’ютерні тезауруси, що використовуються для індексування, каталогізації документів і пошуку інформації в галузевих інформаційних ресурсах бібліотек з проблем біології, медицини, охорони здоров’я та суміжних наук – SNOMED CT, MeSH, Web-ИРБИС; освіти, культури, природничих, соціальних та гуманітарних наук, комунікації та інформації – багатомовний тезаурус ЮНЕСКО; галузей діяльності інститутів Європейського Союзу – тезаурус EUROVOC; комплекс інформаційно-пошукових тезаурусів ІНІСН РАН; сільського господарства – тезаурус AGROVOC та інші.

Певний внесок у цю галузь зроблено й в Україні. Розроблено тезаурус з комп’ютерної лексикографії, тезаурус лінгвістичних термінів (Інституту філології Київського національного університету імені Тараса Шевченка), інформаційно-пошуковий тезаурус Національної бібліотеки України імені Ярослава Мудрого, галузевий тезаурус лінгвістичного забезпечення електронної бібліотеки ХДУХТ та ін.

Унаслідок опрацювання наукових робіт з питання створення галузевих комп’ютерних тезаурусів було виявлено вагомі напрацювання із розроблення вузькоспеціалізованих галузевих тезаурусів з проблем нанотехнології, будівельної сфери, авіації, для прикордонної служби України та ін. Дослідження інформаційно-пошукових тезаурусів, принципів їхньої побудови й функціонування в різних галузях дає підстави зробити висновок, що специфіка кожної предметної області знаходить відображення в структурі її тезауруса.

Виявлено, що удосконалення підходів до організації процесу інформаційного забезпечення міждисциплінарних наукових досліджень, включаючи і роботу бібліотек, потребує відповідних технологій – інформаційно-пошукових систем, що забезпечують пошук релевантної інформації, яка є основою наукових досліджень. Відповідним механізмом може слугувати тезаурусний підхід до організації наукової діяльності, адже використання науковцями саме релевантної інформації дає їм чіткі знання термінів досліджуваної галузі.

Таким чином, розроблення інформаційних тезаурусів для ефективного формування пошукових образів документів у бібліотеках через постійне зростання обсягів різногалузевої інформації набуває все більшої актуальності. Формуванню якісних запитів для пошукових машин сприяють тезауруси.

Методологічну основу дисертації склали загальнонаукові методи, методи інформатики, математики, соціокомунікаційний та історико-хронологічний методи.

Основу дослідження склав системний підхід, що дало змогу поглибити концептуальне філософське розуміння предметної області як цілісної системи, яка характеризується власною структурою, стійкими взаємозв'язками між окремими елементами та підсистемами, специфічними функціями складових компонентів і системи в цілому, та аналізу предметної області не тільки як самостійної системи, а як підсистеми значно більшої та складнішої системи – соціокомунікаційної.

Для підвищення ефективного пошуку релевантної інформації застосовувалися онтологічний та тезаурусний підходи, які дозволяють представити поняття «розумного міста» так, що вони стають базовими для формального інформаційного пошуку.

Застосування методів консолідованої інформації забезпечило ефективний пошук джерел, які найповніше відповідали завданням наукового дослідження, сприяло глибшому аналізу отриманої інформації, дозволило

здійснити перетворення отриманої інформації з різних джерел у цілісні системні знання.

Інформаційний підхід дозволив здійснити різностороннє дослідження інформаційної функції тезаурусів. З одного боку, це функція здійснення соціальної комунікації, пов'язана із створенням, зберіганням, опрацюванням та використанням інформації в процесі комунікування між джерелом інформації та її споживачем. З другого боку, це функція трансляції знань через ті ж процеси створення, накопичення, зберігання, опрацювання, пошуку, розповсюдження знань.

За допомогою соціокомунікаційного підходу як методологічного інструментарію дослідження обґрунтовано сутність тезаурусів, які реалізуються в цілеспрямованих комунікаційних зв'язках між суб'єктами та об'єктами системи, суб'єктами системи та зовнішнім середовищем; доведено, що інформаційно-комунікаційне середовище інформаційного суспільства суттєво впливає на формування інноваційних галузей дослідження, у нашому випадку «розумного міста».

Встановлено, що міждисциплінарність – це синергія різних наук, що передбачає розвиток інтеграційних процесів, зростаючу взаємодію методів, інструментарію задля отримання нового наукового знання (йдеться про утворення нових міждисциплінарних об'єктів, що формують інноваційну галузь «Розумне місто»).

У першому розділі дисертації подано основні результати наукових праць, включених до Списку публікацій здобувача за темою дисертації (пп. 3, 11) та представлених у Списку використаних джерел: [1], [66].

РОЗДІЛ 2

ВИБІР МОДЕЛІ ПРЕДСТАВЛЕННЯ ЗНАНЬ ПРЕДМЕТНОЇ ОБЛАСТІ «РОЗУМНЕ МІСТО» ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В БІБЛІОТЕКАХ

2.1 Аналіз моделей представлення знань

Базою будь-якої інтелектуальної системи є база знань і закладений у систему механізм виведення рішень. Названі компоненти окреслюють дві основні інтелектуальні характеристики системи: здатність зберігати знання про щось і вміння оперувати цими знаннями [67].

Знання – результат теоретичної та практичної діяльності людини, що відображає накопичення попереднього досвіду й відрізняється високим ступенем структурованості [68].

Під знаннями розуміють спеціальну форму подання інформації, яка дозволяє мозку зберігати, відтворювати й розуміти її. Проте не вся інформація є знання. Знання – це особлива інформація, зафіксована й виражена в мові. Типи відношень, що визначають зв'язок знань з позамовним світом, один з одним і з системою людських дій підпорядковуються особливим закономірностям – семантиці, синтаксису і прагматиці [69].

Знання – сукупність даних, які мають розвинену й складну структуру. Знання в інформаційних системах (базі знань) подаються у вигляді певних інформаційних одиниць знань, які мають зв'язки між собою. Форма представлення знань впливає на властивості та характеристики інформаційних систем, тому це є однією з найважливіших проблем, які характерні для систем, орієнтованих на знання. Знання не можуть бути подані безпосередньо у звичайному вигляді, а саме в тій формі, в якій їх використовують користувачі системи, так як дії над знаннями та логічні висновки в інтелектуальних інформаційних системах виконуються не користувачем, а програмою [70].

Знання – це принципи, зв'язки, закони, які отримані в результаті практичної діяльності та професійного досвіду, що дають можливість спеціалістам ставити й вирішувати задачі в певній галузі, тобто закономірності предметної області.

Знання, як особливий ресурс, можуть постійно нарощувати їхні обсяги та якісні характеристики у відповідності до змін у науці, суспільстві або культурі. Аналогічно до будь-якого іншого ресурсу, знання піддаються швидкому старінню. Тому для роботи зі знаннями потрібні особливі методи їхнього опрацювання, зберігання та використання [71]. Дослідження в галузі штучного інтелекту показали, що знання експертів є необхідними для вирішення проблеми отримання знань і здійснення їхнього виведення. Проблема подання знань полягає в тому, що знання часто мають хоча і якісний, але суперечливий характер, оскільки вони є продуктом експерта в певній предметній області. Знання повинні бути зведені до однозначного формалізованого виду для подальшого їхнього опрацювання та використання.

Вибір форми представлення знань є основним завданням побудови систем, які засновані на знаннях. Подання знань – це спосіб формального вираження знань про предметну область у комп'ютерно-інтерпретованій формі, який передбачає визначення деяких початкових об'єктів, правил формування на їхній основі нових об'єктів і, як результат, отримання опису знань. Вхідними неподільними об'єктами є значення даних. Відношення між даними встановлюють правила утворення нових об'єктів. Виконуючи окремі процедури над відношеннями між даними, структурують дані й формують знання. Відповідні формалізми, які забезпечують зазначене представлення, називаються моделями представлення знань [72].

Моделі подання знань є предметом дослідження й розробок вузьких спеціалістів, як-от програмістів і математиків, та спеціалістів різних предметних областей, які використовують інтелектуальні системи у своїй діяльності. В інтелектуальних системах можна визначити кілька напрямів

досліджень, які стосуються подання знань. Наприклад, дослідники зі штучного інтелекту когнітивного напрямку у своїх дослідженнях спиралися на принципи організації людської пам'яті й моделювання виводу на їхній основі. Знання подаються за допомогою моделей, які базуються на використанні продукцій, семантичних мереж, фреймів. Дослідники логічного напрямку працювали в напрямку вивчення теоретичних методів подання знань, у результаті чого було створено модель подання знань на основі логіки предикатів першого порядку.

Кожній предметній області притаманні свої поняття й зв'язки між ними, своя термінологія, закони, які пов'язують між собою об'єкти даної предметної області, свої процеси й події. Разом з тим, кожна предметна область має певні методи розв'язання задач. Вирішуючи такі завдання необхідно використовувати такі моделі, які містять основні характеристики предметних областей.

Існують десятки моделей представлення знань для різноманітних предметних областей [73, 74]. Більшість із них можуть бути зведені до таких класів: логічні моделі, продукційні моделі, фреймові моделі, семантичні мережі, онтології.

Мова, що використовується для розроблення систем, спроектованих на основі цих моделей, називається мовою представлення знань.

Вищеподані моделі подання знань можна об'єднати у дві великі групи (рис. 2.1) [75]:

- модульні (модульні мови оперують не пов'язаними (окремими) елементами знань, чи то правилами, чи аксіоми предметної області),
- мережеві (мережеві мови надають можливість пов'язувати елементи або фрагменти знань через відношення в семантичні мережі або мережі фреймів).

При проєктуванні моделі подання знань предметної області «Розумне місто» необхідно враховувати однорідність подання й простоту розуміння. Однорідне подання спрощує механізм керування логічним висновком і

управління знаннями. Подання знань має характеризуватися простотою розуміння, тобто бути зрозумілим експертам і користувачам системи, так як у іншому випадку ускладнюється отримання знань та їхня оцінка.

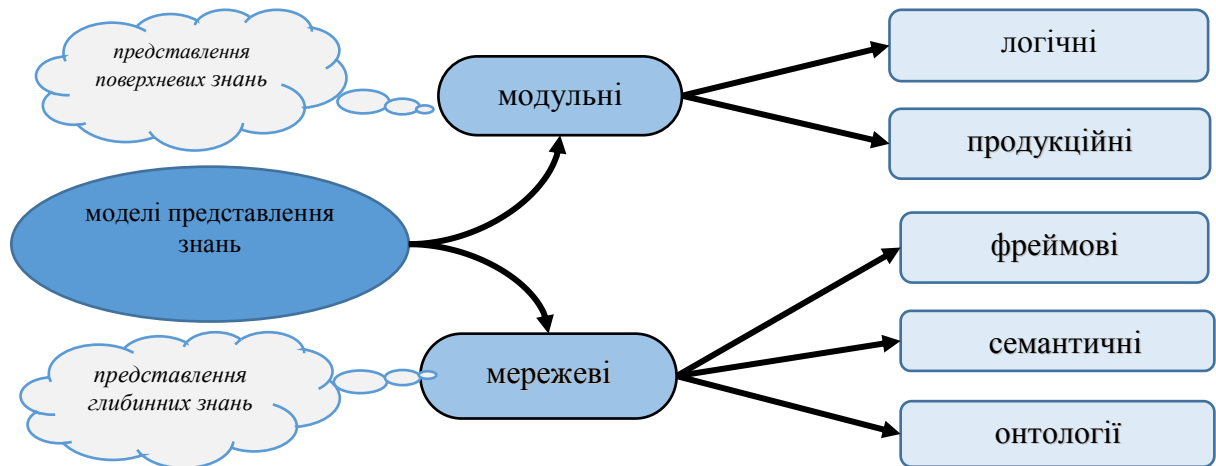


Рис. 2.1 Класифікація моделей представлення знань

Для того щоб вибрати модель представлення знань предметної області «Розумне місто», коротко подамо основні характеристики кожної із зазначених моделей.

Логічні моделі здійснюються засобами логіки предикатів. Знання про предметну область подаються у формі сукупності логічних формул. Вони дуже формалізовані, і в цьому полягає один з основних їхніх недоліків. Логічні моделі включають 1) множину основних елементів різної природи; 2) множину синтаксичних правил, за допомогою яких з базових елементів можна побудувати синтаксичні коректні сукупності (синтаксично правильні фрази, речення будуються зі слів обмеженого словника); 3) у безлічі синтаксично правильних сукупностей виділяється деяка підмножина, елементи якої називаються аксіомами, тобто правильно побудовані формули теорії, що вважаються істинними; 4) множину правил виведення, які дозволяють виводити з аксіом нові твердження (теореми) [76]. Головним недоліком логічних моделей є відсутність чітких принципів організації фактів у базі знань.

Продукційні моделі (моделі правил) – це моделі, засновані на правилах, в яких знання представлені у вигляді пропозицій типу «якщо (умова), то (дія)». Під «умовою» розуміється деякий зразок, який використовують для здійснення пошуку в базі знань, а під «дією» – дії, що виконуються при вдалому результаті пошуку (вони можуть бути проміжними, які виступають далі як умови, й термінальними або цільовими, завершальними роботу системи).

Основу моделі складають системи продукцій. Кожна продукція в найбільш загальному вигляді містить 1) ім'я продукції, за допомогою якого вона виділяється серед усієї множини продукцій; 2) сферу застосування продукції, предметна область, до якої вона відноситься; 3) умови застосування продукції; 4) ядро продукції (інтерпретація залежить від конкретної ситуації, часто ядро інтерпретується як «якщо маєш А, зроби В»); 5) постумову продукції, описує дії й процеси, які треба виконати після реалізації ядра [77].

Переваги продукційних моделей: простота створення і розуміння окремих правил, модифікацій і розширення бази знань, механізму логічного висновку; природна модульність організації баз знань. Недоліками є неясність взаємних відношень правил, складність оцінки цілісного образу знань, дуже низька ефективність опрацювання, відмінність від людської структури знань, відсутність гнучкості в логічному висновку [78].

Фреймові системи – моделі подання знань, засновані на теорії М. Мінського, як один з підходів до опису знань, придатний для розуміння просторових сцен і мови. У фреймовій системі одиницею подання інформації є об'єкт, що називається фреймом (англ. *frame* – *рамка, каркас*). Він є формою подання деякої ситуації, що описується сукупністю понять і даних та дозволяє подати всі знання про певний об'єкт у одній структурі даних, а не розпорошувати їх між більш дрібних структур. Таким чином, фрейм є системно-структурним описом предметної області [79], основною структурною одиницею якого є слот.

Слот (англ. *slot* – *щілина, пропіз*) – це незаповнені підструктури фрейму, які перетворюють його в носія конкретної інформації про предметну область після заповнення їх певними даними. Слот може містити тільки ім'я слота й значення імені або бути фреймом нижнього рівня, що забезпечує побудову мережі фреймів.

У фреймових моделях закріплюється жорстка структура інформаційних одиниць, на відміну від моделей інших типів. Така структура називається протофреймом (фрейм-прототип). У загальному вигляді структура протофрейму може бути представлена як список властивостей:

Ім'я фрейму:

(ім'я 1-го слота: значення 1-го слота),

(ім'я 2-го слота: значення 2-го слота),

...

(ім'я N-го слота: значення N-го слота).

Таким чином, фрейм представляє собою деяку закінчену структуру, яка при заповненні слотів значеннями описує цілком конкретний факт, подію, явище чи процес. Фреймова система зручна для опису структури й характеристик однотипних об'єктів (процесів, подій).

Головним недоліком фреймового подання знань і одночасно його особливістю є відсутність механізму висновку, а саме розробник системи несе відповідальність за правильне опрацювання знань. Оскільки він представляє предметну область, то повинен не тільки детально описати кожен фрейм і слот, а й правильно задати процедуру їхнього опрацювання.

Семантичні мережі – це моделі предметної області, подані у вигляді графа, вершини якого відображають поняття, а дуги – відношення між ними. Вершини й дуги семантичної мережі позначаються іменами відповідних об'єктів і відношень предметної області. За допомогою семантичної мережі відображається семантика предметної області у вигляді понять і відношень між ними.

Семантичні мережі включають множину інформаційних одиниць; множину типів зв'язків між інформаційними одиницями; відображення, яке задає зв'язки між інформаційними одиницями.

Будуючи семантичні мережі для представлення знань предметних областей, прийнято використовувати стандартизовані типи відношень. До зв'язків, які найбільш широко використовуються, належить відношення типу «is-a», яке означає, що окремий об'єкт «є екземпляром» певного класу. Іншим, типом зв'язку, є «part-of», що показує відношення «ціле – частина». Це відношення задається між екземплярами класу.

Перевагами семантичних мереж, як засобів подання знань, є наочність представлення знань, близькість структури мережі до структури предметної області. З їхньою допомогою зручно зображати причинно-наслідкові зв'язки між елементами (підсистемами) та структуру складних систем. Недоліком семантичних мереж є складність виведення пошуку підграфа, що відповідає запиту.

На сучасному етапі фрейми та семантичні мережі розглядаються, як правило, у спільному контексті, оскільки, з одного боку, вузли семантичної мережі можуть розглядатися як фрейми з власною внутрішньою структурою, з другого боку, вводячи різні зв'язки між слотами фреймів, надаємо фрейму рис семантичної мережі.

Зі збільшенням потоку інформації, з'являється потреба пошуку нових способів її зберігання, подання, формалізації й систематизації, а також автоматичного опрацювання. Таким чином зростає інтерес до інформаційних ресурсів, одним із типів яких є онтології.

Онтологічний підхід використовується «для підвищення ефективності від застосування моделі і одночасного спрощення комп'ютерного опрацювання моделей і компонентів у якості інструментарію для формальної специфікації понять і відношень, що характеризують певну галузь знань експертів» [70].

Під онтологією (в інженерії знань) розуміється широкий опис предметної області, що використовується для формального та декларативного визначення концептуалізації цієї області [80]. Головне завдання онтології полягає в спрощенні пошуку інформації шляхом

систематизації знань, створення єдиної ієрархії понять, стандартизації термінів та правил інтерпретації.

2.2 Вибір способу представлення знань предметної області

«Розумне місто» методом аналізу ієрархій

Предметна область – «сукупність об'єктів, понять, зв'язків, відношень і способів перетворення та взаємодії цих об'єктів під час розв'язання задач, що відносяться до певної сфери людської діяльності» [81]. Це частина реального світу, яка є об'єктом деякої діяльності або ділянкою дослідження. У нашій роботі предметною областю, яка досліджується є «розумне місто».

«Розумне місто» (концепція соціополіусів на основі смарт-рішень) – це концепція, в основі якої лежить місто, що використовує різноманітні інформаційні технології для підвищення ефективності функціонування основних сфер життєдіяльності міст. Розвиток інформаційно-комунікаційних технологій сприяє розробленню високоякісних програмних продуктів, призначених для виконання інформаційних потреб користувачів. За допомогою техніки, що використовує штучний інтелект, підвищується рівень комфорту й безпеки, вона стає характерною рисою повсякденного побуту в соціополіусах. Набувають розвитку інтелектуальні інформаційні системи – автоматизовані, засновані на знаннях, інформаційні системи для здійснення підтримки діяльності людини та пошуку інформації в режимі забезпечення діалогу між обчислювальною технікою та користувачем.

Використання інформаційних технологій сприяє підвищенню ефективності отримання релевантної інформації. Вагома роль у представленні знань предметної області «Розумне місто» відводиться формуванню терміносистеми, яка спирається на ідеї, принципи, концепції, теорії, напрацьовані різними науковими галузями. Понятійний апарат цих галузей шляхом адаптації застосовується для побудови понятійної бази

предметної області «Розумне місто». Таким чином формується термінологічна система, яка потребує уніфікації та стандартизації.

Формування знаннєвої системи предметної області, її наукової лексики, понятійного апарату не може бути хаотичним процесом. Наукове обґрунтування створення терміносистеми «розумного міста» має стати окремим напрямом наукових досліджень. Тому необхідно вибрати спосіб представлення пошукових концептів предметної області «Розумне місто». На наш погляд, серед напрацьованих підходів визначення оптимального рішення в умовах багатокритеріальності особливе місце посідає метод аналізу ієрархій (MAI), оскільки він дає можливість повноцінно врахувати всі критерії, висунені до вибору моделі представлення знань.

Розроблений відомим американським математиком Томасом Сааті метод аналізу ієрархій є одним із основних методів науково-технічного прогнозування, який використовується для вирішення складно структурованих багатокритеріальних завдань з ієрархічною структурою і ґрунтується на припущенні [82]. MAI дозволяє «структурувати задачу та формалізувати зв'язки між її елементами, моделювати процедури критеріального оцінювання й вподобань особи, що приймає рішення, виконувати синтез правила, що є розв'язуючим для певного типу задач, та встановлювати вподобання на множині альтернатив з урахуванням цього правила» [83].

Метод аналізу ієрархій представляє математичну процедуру для ієрархічного зображення елементів з метою визначення суті будь-якої проблеми. Він полягає в декомпозиції проблеми на більш прості складові й подальшій обробці послідовності думок експертів щодо парних порівнянь критеріїв і альтернатив. У результаті виникає ієрархічна структура. На найвищому рівні ієрархії знаходиться мета, проміжними рівнями є критерії, за якими оцінюють нижчі рівні. Найнижчий рівень ієрархії – це альтернативи, які складають перелік різних варіантів рішень, серед котрих

необхідно вибрати найкращі. У результаті одержують відносний ступінь впливу компонентів нижнього рівня на компоненти верхнього рівня.

Системи з ієрархічною структурою є найбільш пристосованими для розв'язання складних завдань, так як у таких структурах використовується принцип декомпозиції – розбиття проблеми на простіші складові, підпорядкованість елементів нижчих рівнів елементам вищих та агрегування інформаційних потоків у напрямку «знизу – догори». Крім того, зазвичай вважається, що елементи кожного рівня є незалежними, хоча й можуть обмінюватися інформацією між собою [84].

МАІ характеризується такими ознаками як системність, економність щодо часу застосування, залучення незначної кількості експертів, отримання нових знань стосовно предмета оцінки, перевірка обґрунтованості та погодженості отриманих експертних суджень.

МАІ використовується для прийняття рішень у різноманітних ситуаціях: управління, вирішення галузевих і приватних проблем у бізнесі, промисловості [85, 86, 87]; охорона здоров'я [88, 89]; освіта [84, 90, 91]; дистанційне бібліотечне обслуговування осіб із особливими потребами [92] та ін. Ці роботи відзначаються переважно розвитком принципів побудови й галузей застосування методу аналізу ієрархій.

МАІ складається з декількох етапів (рис. 2.2):

перший етап – окреслення проблеми дослідження – за допомогою методу аналізу ієрархій знайти рішення поставленої задачі;

другий етап – структуризація проблеми у вигляді ієрархії – розкладання проблеми на елементарні складові: мету, альтернативні варіанти досягнення мети й критерії для оцінки якості альтернатив;

третій етап – побудова матриці – визначення пріоритетів усіх елементів ієрархії (альтернатив) з використанням методу парних порівнянь;

четвертий етап – оцінка локальних пріоритетів елементів, що порівнюються;

п'ятий етап – узгодженість локальних пріоритетів;

шостий етап – ухвалення рішення на основі отриманих результатів.

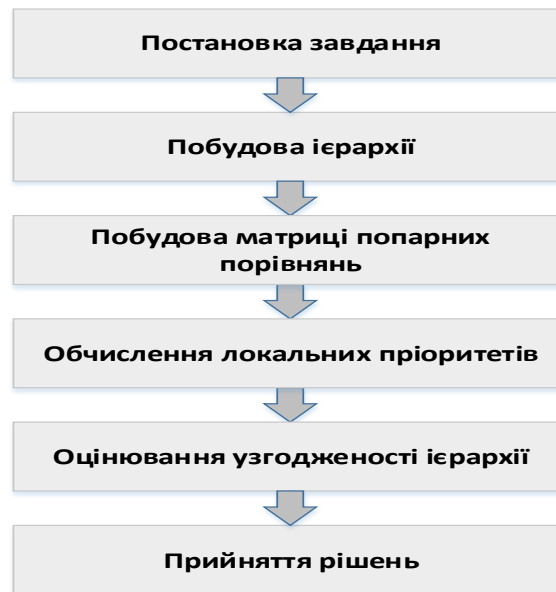


Рис. 2.2 Етапи методу аналізу ієрархій

Першим кроком МАІ є здійснення декомпозиції поставленої задачі – вибору моделі представлення знань предметної області «Розумне місто» – та подання її у вигляді ієрархії. Для цього побудовано відповідну ієрархію за трирівневою структурою. Ієрархічна модель задачі подана на рис. 2.3.

Запропонована ієрархічна трирівнева структура вибору моделі представлення знань предметної області «Розумне місто» дозволить виділити модель, яка займає провідні позиції за кожним з досліджуваних показників.

У відповідності до методу аналізу ієрархій на верхньому рівні ієрархії знаходиться мета поставленої задачі – вибір способу представлення знань предметної області «Розумне місто», – щодо якої приймаються рішення. Другий рівень вказаної ієрархії формує множина критеріїв, за якими здійснюється вибір альтернативних платформ для представлення знань предметної області «Розумне місто». Вони формуються на основі аналізу властивостей та можливостей використання обраних платформ та уточнюють мету. Для того, щоб ієрархія вважалася повною, потрібно, щоб кожен

елемент заданого рівня функціонував як критерій для всіх елементів рівня, що знаходяться нижче. В іншому випадку ієрархія є неповною.

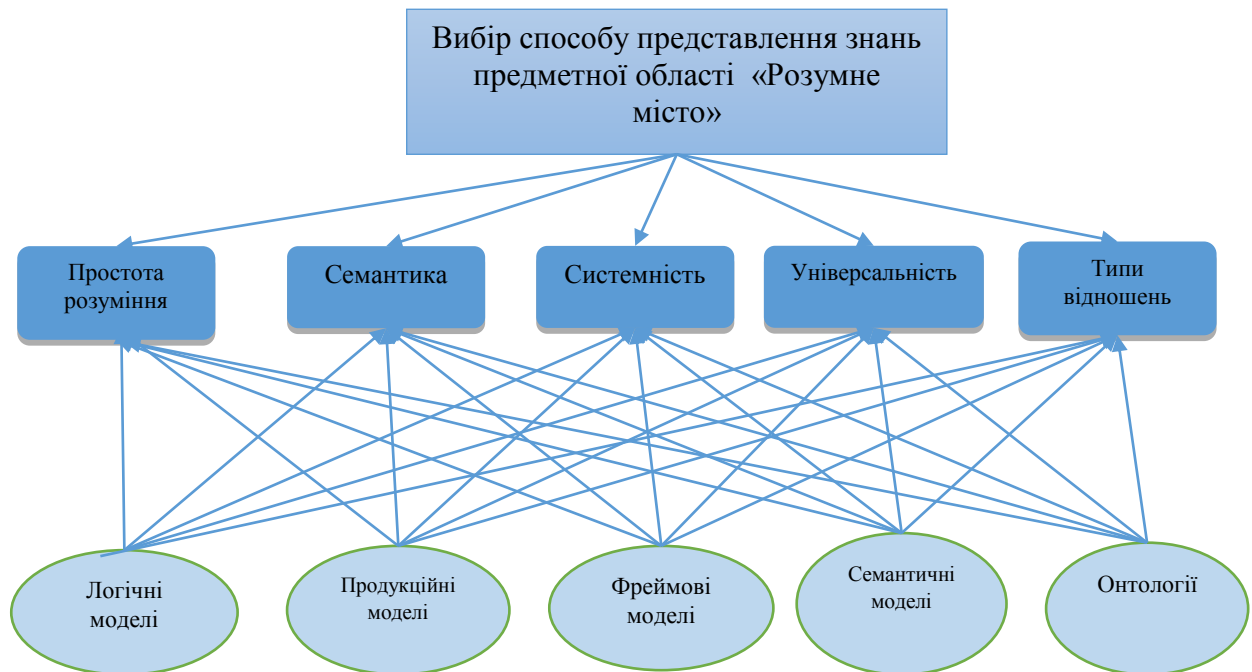


Рис. 2.3 Ієрархічна модель задачі вибору моделі представлення знань предметної області «Розумне місто»

Для попарного порівняння альтернатив, у нашому випадку моделей представлення знань, кожна з яких має свої переваги та недоліки, нами було використано такі критерії оцінювання моделей і методів подання знань:

- простота розуміння – рівень складності (абстрактності) елемента знань;
- відображення семантики предметної області;
- системність – формування цілісних уявлень про об'єкт чи процес;
- універсальність – можливість описання знань з різних предметних галузей;
- типи відношень.

На останньому (третьому) рівні ієрархії знаходяться такі п'ять альтернатив: «Логічні моделі», «Продукційні моделі», «Фреймові моделі»,

«Семантичні мережі» та «Онтології», які оцінюються за критеріями другого рівня.

Реалізація поставленої мети полягає у виборі однієї з наведених альтернатив на основі множини сформульованих критеріїв. Вибір альтернативи виконується обчисленням елементів вектора пріоритетів альтернатив. Прийнятим рішенням поставленої проблеми є альтернатива з найбільшим значенням елемента [93].

Для встановлення відносної важливості елементів ієрархії експертами використовувалася стандартна шкала відносної важливості об'єктів по відношенню до загальної мети [82] (див. Додаток Л), яка дозволяла експертам оцінити моделі представлення знань у відповідності зі ступенями важливості одного порівнюваного об'єкта перед іншим.

Наступним кроком є укладання квадратних матриць шляхом проведення попарних експертних порівнянь з використанням шкали оцінки відносної важливості та встановлення локальних пріоритетів показників n .

При проведенні попарних порівнянь моделей представлення знань предметної області «Розумне місто» перед особою, що приймає рішення, ставиться питання, яка із двох порівнюваних моделей важливіша або дієвіша. Міра переваги моделі A_i над моделлю A_j виражається елементом матриці a_{ij} . Матриця парних порівнянь $[A]$ має властивість зворотної симетрії і виконується умова:

$$a_{ij} = \frac{1}{a_{ji}}.$$

У таблиці 2.1 приведена матриця парних порівнянь для критерію «типи відношень» десятковими значеннями елементів.

Таблицю потрібно читати таким чином: зазначена в рядку модель подання знань порівнюється з моделлю подання знань, зазначеною в стовпці. Якщо зазначена в рядку модель краща, то відповідне число від 1 до 9 записується в комірку на перетині рядка й стовпця. Якщо ж краща модель, зазначена в стовпці, то 1 ділиться на відповідне число від 1 до 9, і результат

записується в комірку на перетині рядка і стовпця. При порівнянні моделі з самою собою відношення дорівнює одиниці й у всі діагональні клітинки заноситься значення 1.

Таблиця 2.1

Матриця попарних порівнянь альтернатив
за критерієм «типи відношень»

<i>Моделі представлення знань</i>	<i>Логічні моделі</i>	<i>Продукційні моделі</i>	<i>Фреймові моделі</i>	<i>Семантичні мережі</i>	<i>Онтології</i>
Логічні моделі	1	2	0,5	0,2	0,25
Продукційні моделі	0,5	1	3	2	0,2
Фреймові моделі	2	0,33	1	0,33	0,17
Семантичні мережі	5	0,5	3	1	0,5
Онтології	4	3	6	2	1

Маючи побудовану ієрархію, складену матрицю і висловлені суб'єктивні парні судження, потрібно з'ясувати, яким чином поєднуються ієрархічна декомпозиція і шкала відносної важливості для отримання обдуманих підходів до багатокритеріальної проблеми вибору моделі представлення знань предметної області «Розумне місто».

Після виконання всіх попарних порівнянь матрицю необхідно нормалізувати. Для цього з груп попарних порівнянь формуються набори локальних пріоритетів. Процес формування локальних пріоритетів ґрунтується на обчисленні власних векторів для кожної матриці й нормалізації результату до одиниці.

Значення пріоритетів обчислюється за допомогою спеціальної процедури. У методі аналізу ієрархій рекомендовані такі чотири способи опрацювання даних для виконання обчислення пріоритетів:

1) обчислити суму елементів у кожному рядку й нормалізувати діленням кожної суми на суму всіх елементів;

2) обчислити суму елементів кожного стовпчика й отримати обернені елементи цих сум, нормалізувати їх діленням кожної оберненої величини на суму всіх обернених величин;

3) розділити елементи кожного стовпчика на суму елементів цього стовпчика, після цього обчислити суму елементів кожного отриманого рядка і розділити суму на число елементів у рядку;

4) помножити n елементів кожного рядка і обчислити з добутку корінь n -го степеню, нормалізувати отримані числа [82].

На основі матриць попарних порівнянь визначаються вектори локальних пріоритетів для кожного з критеріїв. Це виконується відповідно до способу 4 [82], шляхом множення n елементів кожного рядка й обчислення з добутку кореня n -го степеню та нормування отриманих чисел.

Розрахунки вектора локальних пріоритетів та нормування значень елементів цього вектора подано у Додатку М.

Процес обчислення можна здійснювати як вручну, так і за допомогою різних програмних засобів, зокрема Microsoft Excel за допомогою функції GEOMEAN (табл. 2.2)

Таблиця 2.2

Розрахунки вектора локальних пріоритетів
за критерієм «типи відношень»

№	Моделі представлення знань	Номер моделі представлення знань					Вектор локальних пріоритетів
		1	2	3	4	5	
1	Логічні моделі	1	2	0,5	0,2	0,25	0,549
2	Продукційні моделі	0,5	1	3	2	0,2	0,903
3	Фреймові моделі	2	0,33	1	0,33	0,17	0,517
4	Семантичні мережі	5	0,5	3	1	0,5	1,303
5	Онтології	4	3	6	2	1	2,702

Після того як компоненти локального вектора отримані для всіх n рядків, стає можливим їхнє використання для подальших обчислень – нормалізації (табл. 2.3) (див. Додаток М).

Нормована матриця попарних порівнянь
за критерієм «типи відношень»

№	Моделі представлення знань	Номер моделі представлення знань					Вектор локальних пріоритетів	Нормативний вектор цих пріоритетів
		1	2	3	4	5		
1	Логічні моделі	1	2	0,5	0,2	0,25	0,549	0,092
2	Продукційні моделі	0,5	1	3	2	0,2	0,903	0,151
3	Фреймові моделі	2	0,33	1	0,33	0,17	0,517	0,087
4	Семантичні мережі	5	0,5	3	1	0,5	1,303	0,218
5	Онтології	4	3	6	2	1	2,702	0,452
Сума		12,5	6,83	13,5	5,53	2,12	5,974	1

Таким чином отримали нормовану матрицю попарних порівнянь за критерієм «типи відношень».

Дані з таблиці 2.4 можна представити графічно, подавши процентне співвідношення, таким чином (рис. 2.4):

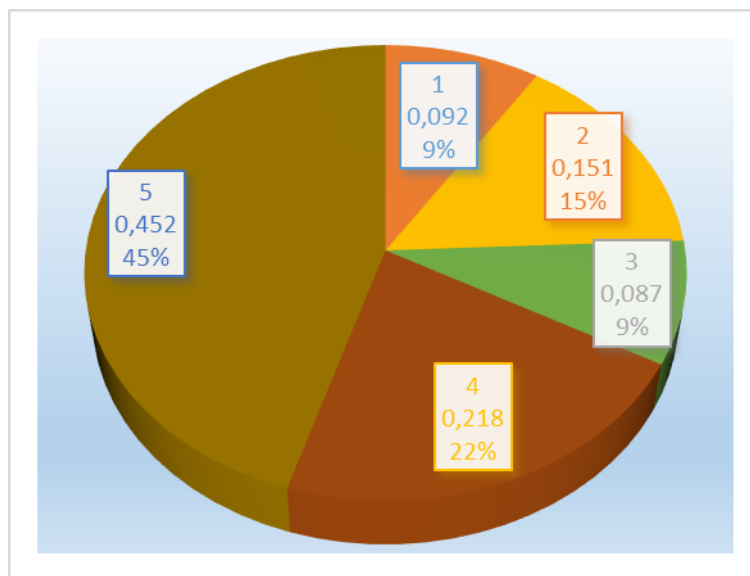


Рис. 2.4 Нормований вектор локальних пріоритетів
за критерієм «типи відношень»

Отже, як бачимо з діаграми, найнижчий нормований вектор пріоритетів за критерієм «типи відношень» мають альтернативи: логічні моделі – 0,092 (9%) та фреймові моделі – 0,087 (9%). Найкращою альтернативою є

онтології, оскільки вони мають найбільше значення вектора – 0,452 (45%) [94].

Оцінки відносної важливості альтернатив, що порівнюються, повинні бути узгоджені, тому визначається узгодженість локальних пріоритетів: максимальне власне значення матриці, індекс узгодженості та відношення узгодженості (див. Додаток М). Оскільки відношення узгодженості є в допустимих межах (менше 10%, у деяких випадках менше 20%) (у нашому випадку – це 0,12), то матриця попарних порівнянь за критерієм «типи відношень» вважається узгодженою.

Аналогічно будуються матриці попарних порівнянь альтернатив за критеріями «простота розуміння», «відображення семантики предметної області», «системність», «універсальність» (див. Додаток Н).

Після визначення всіх нормативних векторів локальних пріоритетів будується зведена таблиця B і для неї будується вектор N за аналогією з вектором локальних пріоритетів. Для визначення глобальних пріоритетів будується вектор U , який отримується множенням матриць $B \times N$, використовуючи операцію MMULT (масив 1, масив 2) (табл.2.4).

Таблиця 2.4

Зведена таблиця нормативних векторів локальних пріоритетів

B						N	U
1. Логічні моделі	0,092	0,087	0,074	0,068	0,055	0,074	0,064
2. Продукційні моделі	0,151	0,080	0,033	0,056	0,029	0,058	0,048
3. Фреймові моделі	0,087	0,081	0,174	0,079	0,137	0,106	0,115
4. Семантичні мережі	0,218	0,299	0,272	0,318	0,250	0,269	0,265
5. Онтології	0,452	0,453	0,446	0,479	0,529	0,471	0,485

В останньому стовпчику маємо значення вектора U , який містить значення глобальних пріоритетів моделі представлення знань предметної області «Розумне місто», що візуалізовані на рис. 2.5.

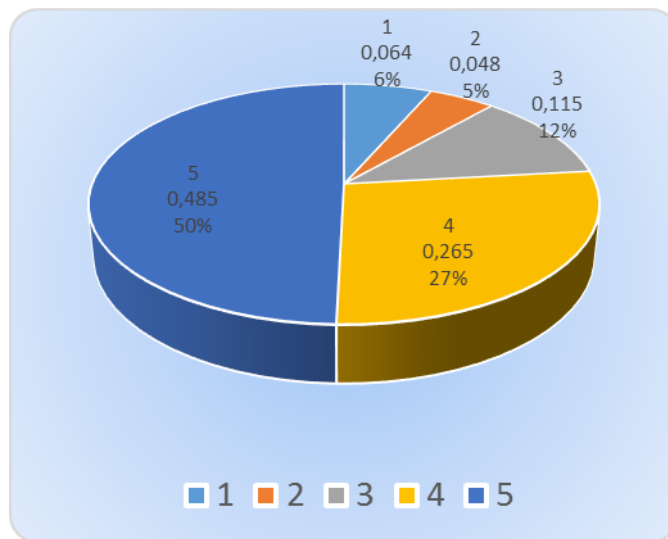


Рис. 2.5 Нормований вектор глобальних пріоритетів

Проаналізувавши діаграму, можемо обґрунтовано стверджувати, що онтологічний підхід є найкращим способом для представлення знань предметної області «Розумне місто» та формування пошукових образів у інформаційних системах різної складності, включаючи і автоматизовані бібліотечні системи.

Алгоритм розв'язання розглянутої задачі реалізується за допомогою інструментарію програмного додатку MS Excel.

2.3 Онтології як системи зберігання знань

На сьогодні в найрізноманітніших галузях людської діяльності накопичено велику кількість інформації, яка використовується далеко не в повному обсязі. Ця інформація характеризується різними форматами й способами представлення, тому для того, щоб накопичені дані набули широкої практичної цінності для розвитку науки й виробництва, необхідно звести різноманітно представлену інформацію до такого виду, що забезпечить можливість її спільного використання різними системами. Вирішення цієї проблеми пов'язано із завданням вилучення та представлення знань, що знаходиться в центрі уваги багатьох учених. На сьогодні в області

штучного інтелекту розроблено ряд засобів представлення знань і до найбільш ефективним з них відноситься онтологія.

Останніми роками розроблення онтологій переходить зі світу лабораторій із штучного інтелекту на робочі місця експертів з предметних галузей. У багатьох дисциплінах використовуються стандартні онтології для полегшення обміну знаннями та повторного використання цих знань. Онтологія предметної області застосовується для того, щоб визначити та уточнити її загальну термінологічну базу. Вона є угодою «про спільне використання понять (термінів), яка містить засоби подання предметних знань і домовленості про методи логічного виведення. Онтологія може розглядатися як формалізований опис погляду на світ у конкретній сфері інтересів, який складається з набору термінів і правил використання цих термінів, які обмежують їхнє значення в рамках конкретної предметної області» [95].

Існує чимало досліджень, у яких розглядається розв'язання задач розроблення та розвитку онтологій. Так, роботи Т. Бернерса-Лі, Дж. Хандлер, О. Лассіла (Т. Berners-Lee, J. Handler, O. Lassila) [96], Т. Грубера (Т. Gruber) [97] – присвячені концепції онтологій, Н. Ной (N. Noy) та Д. МакГіннесс (D. McGuinness) [98] розробили методологію створення онтологій. В. Соловйов, Б. Добров, В. Іванов, Н. Лукашевич [99] займаються навчальною роботою з метою формування базових знань про онтології та практичних навичок з проєктування й застосування онтологій. С. Дубров та Ю. Нечаєв [100], Н. Соколова, М. Голубєва [101] у своїх роботах розглядають створення електронних освітніх середовищ, що реалізують онтологічні моделі.

Також маємо приклади розроблення онтологій у медицині [102, 103, 104], сейсмології [105], природокористуванні [106] та в інших галузях. Питаннями організації та управління знаннями з використанням онтологій у галузі інформаційного пошуку та бібліотечних системах займалися А. Рідкокаша, Б. Шевченко [107], Д. Фенсел [108], М. Кагаловський,

С. Парінов [109]. Популярність онтологій полягає в тому, що вони є інструментом узагальнення й загального розуміння певної предметної галузі та забезпечують взаємодію між людьми через прикладні інформаційні системи [110].

Термін онтологія вперше вживався у філософії і був пов'язаний з Арістотелем, який намагався класифікувати предмети для побудови системи знань. Існує чимало визначень цього поняття. Найпоширенішим у філософії є таке визначення: онтологія (з латинської *ontologia* походить від давньогрецького *onto* – *сущє, те, що існує*, і грецького *logos* – *учення, наука*) – це вчення про буття, розділ філософії, у якому з'ясовуються фундаментальні проблеми існування, розвитку сутнісного, найважливішого. Іншими словами, онтологія – це наука про буття, наука про природу речей і взаємозв'язки між ними [111].

Онтологія в інформаційних технологіях та комп'ютерних науках – формалізоване представлення знань про певну предметну область, придатне для автоматизованого опрацювання. Онтологію неодмінно супроводжує деяка концепція цієї галузі інтересів, яка найчастіше виражається за допомогою визначення базових об'єктів (атрибутів, процесів) і відношень між ними, що називається концептуалізацією [111].

Незважаючи на те, що термін «онтологія» походить з філософії, в інформаційних технологіях він набув самостійного значення. Основні відмінності зводяться до того, що онтологія в інформатиці має формалізовану мову представлення, яку комп'ютер може обробляти без участі людини, та створюється для вирішення певної задачі, і тому вона оцінюється більше з погляду застосування, ніж повноти.

Одне з найвідоміших визначень онтології є сформульоване Т. Грубером: онтологія – це формальна специфікація узгодженої концептуалізації [112]. Іншим важливим визначенням онтології є формулювання, запропоноване Н. Гуаріно, сформоване на основі аналізу семи різних визначень: онтологія – це формальна теорія, що обмежує

можливі концептуалізації [113]. Під «концептуалізацією» він розумів абстрактний, спрощений погляд на світ, який використовується людьми для здійснення певної мети [112]. Завдання концептуалізації полягає в тому, що для обміну знаннями між програмними системами (в контексті концепції штучного інтелекту) необхідно відкрито специфікувати їхню концептуалізацію, тобто побудувати опис цих знань, до того ж достатньою мірою формальний, щоб його «розуміли» інші системи. Результат такої специфікації Т. Грубер назвав онтологією.

Таким чином, поняття «інженерна онтологія» можна представити як специфікацію (формальний опис) якоїсь концептуалізації (подання предметної області досліджуваного завдання, як це необхідно для певної задачі). Т. Грубер вважав, що концептуалізація проводиться в термінах класів і атрибутів. Середовище досліджуваного завдання подається у вигляді понять, які описуються класами, разом з їхніми властивостями (атрибутами) і конкретними об'єктами – екземплярами класів [114]. Отже, концептуалізація – опис системи понять, об'єктів та відношень, що зв'язують їх один з одним. Вона ділить будь-яку галузь знань, яка існує в цілісному вигляді, виділяє з неї окремі об'єкти, а потім формулює відношення, властиві для даної галузі. Основна частина формально представлених знань базується на концептуалізації.

Представлення онтології Т. Грубером та Н. Гуаріно лягло в основу визначення онтології предметної області іншими дослідниками відповідно до їхньої конкретної галузі професійних інтересів. Як робоче визначення онтології Едвард Хові подає таку дефініцію: онтологія – це структура даних із заданими в ній символами, що дозволяють представляти концептуалізації для опрацювання комп'ютерними програмами [115].

Д. Ланде пояснює онтологію як специфікація концептуалізації предметної області, або спрощено, документ чи файл, що формально задає зв'язки між поняттями. Це свого роду словник понять предметної області й сукупність явно визначених припущень щодо змісту цих понять. Найчастіше

онтологія представляється як ієрархія понять, пов'язаних відношеннями певних видів [116].

Часто онтологією називають базу знань спеціального виду, яку можна розділяти, відчужувати й самостійно використовувати в рамках розглянутої предметної області. Онтологія – це точна специфікація певної галузі, що містить словник термінів цієї галузі та множину зв'язків, що описують, як ці терміни співвідносяться між собою. Саме онтології є адекватними та ефективними засобами для моделювання уявлень щодо різноманітних предметних областей, які дозволяють формально подавати їх семантику [117].

Отже, комп'ютерна онтологія – це формалізоване представлення знань певної предметної області, яке можна використовувати для автоматизованого опрацювання. А комп'ютерна онтологія предметної області «Розумне місто» – це:

- ієрархічна структура скінченної множини понять «розумного міста», що описують задану предметну область;
- структура онтографа, вершинами якого є поняття «розумного міста», а дугами – семантичні відношення між ними;
- поняття «розумного міста» і відношення між ними, подаються відповідно до загальноприйнятих функцій, які взяті з електронних «джерел знань»;
- визначення понять і відношень виконуються на основі аксіом і обмежень (правил) їхньої області дії;
- засіб існування формального опису онтографа.

Проаналізувавши визначення онтології в інформатиці, можна зробити висновок, що серед вчених існує єдиний підхід до представлення онтології як засобу детальної та різносторонньої формалізації знань про дані за допомогою концептуальної схеми. Така концептуальна схема включає опис структури даних, що містить формулювання всіх релевантних класів об'єктів, відношення та правила, які встановлені в певній галузі.

Основна роль онтології – інтеграція інформації. Вона пов'язує два важливі аспекти: по-перше, визначає формальну семантику інформації, дозволяючи опрацювання цієї інформації комп'ютером, і, по-друге, визначає семантику реального світу, дозволяючи на основі загальної термінології зв'язувати інформацію, представлену у вигляді, необхідному для комп'ютерного опрацювання, з інформацією, представленою в зручній формі для сприйняття людиною [118].

Онтології є новими інтелектуальними засобами для пошуку ресурсів у мережі «Інтернет», новими методами подання та опрацювання знань і запитів. Вони здатні точно та ефективно описувати семантику даних для певної предметної області та вирішувати проблему несумісності й суперечливості понять. Онтології володіють власними засобами опрацювання інформації (логічного виводу), відповідними задачами семантичного опрацювання інформації. Так, завдяки онтологіям, при зверненні до пошукової системи користувач має можливість отримувати у відповідь ресурси, семантично релевантні запиту [119]. Онтологія «розумного міста» може використовуватися як посередник між користувачем та інформаційною системою, вона дозволить формалізувати домовленості про термінологію.

Онтологія використовується як джерело даних для багатьох комп'ютерних програм: для інформаційного пошуку, аналізу текстів, вилучення знань і в інших інформаційних технологіях, дозволяючи більш ефективно обробляти складну й різноманітну інформацію.

У загальному вигляді структурою онтології є набір елементів таких категорій: поняття (класи, концепти); екземпляри; аксіоми; відношення (зв'язки); атрибути (властивості).

Поняття в онтології розглядаються як концептуалізації класу всіх представників якоїсь сутності або явища. Вони є загальними категоріями, які можуть бути впорядковані ієрархічно [120]. Поняття можуть бути абстрактними й конкретними, елементарними й складовими, реально

існуючими й уявними. Іншими словами, поняттям може бути будь-яка сутність або явище, про які може бути дана будь-яка інформація. Вони можуть містити в собі екземпляри, інші класи, або ж класи та екземпляри одночасно. Отже, поняттям може бути будь-яка сутність, про яку маємо будь-яку інформація.

Поняття пов'язуються між собою відношеннями різного роду. Найпоширенішим типом відношень, який характерний для всіх онтологій, є категоризація (subsumption) (відношення до певної категорії), синонімами якої виступають такі поняття: таксономічне відношення, відношення is-a, клас-підклас, відношення a-kind-of.

Аксіома – це твердження, що вводиться в онтологію в готовому вигляді, і з якого можуть впливати інші твердження. За допомогою аксіом можемо висловити ту інформацію, що не може бути відображена в онтології за допомогою побудови ієрархії понять і установки різних відношень між поняттями. Як приклад аксіоми можна навести таке твердження: «Якщо біоенергія є відновлювальна енергія, то вона виробляється з відновлювальних джерел енергії». Аксіоми дозволяють надалі здійснювати умовиводи в рамках онтології. Кількість і ступінь деталізації аксіом зазвичай залежать від типу онтології.

Останнім елементом онтології виступають окремі екземпляри. Це окремий представник класу, тобто конкретний елемент якоїсь категорії. Наприклад, екземпляром класу «Електромобіль» буде «Tesla».

Одиниці онтології (поняття та екземпляри) можуть мати властивості – атрибути. Кожен атрибут зазвичай має ім'я і значення та використовується для зберігання інформації, яка специфічна для цієї одиниці.

Складові онтології утворюють своєрідну ієрархію. На нижньому рівні цієї ієрархії знаходяться екземпляри, конкретні індивіди, вище йдуть поняття, тобто категорії. На рівень вище розташовуються відношення між цими поняттями, а узагальнюючою і сполучною є ступінь правил або аксіом.

Онтологія предметної області, як правило, – структура даних, що забезпечує можливість зберігання знань предметної області за допомогою тлумачення термінів предметної області та інтерпретації таких типів парадигматичних відношень як «частина-ціле», «клас-підклас» та деяких видів асоціативних зв'язків [121].

Побудова онтології зазвичай розпочинається зі складання словника термінів, що використовуються для опису певної галузі знань, створення системи точного визначення цих термінів, а також документування основних логічних зв'язків між термінами. Таким чином, онтологія містить сукупність термінів та правил, відповідно яким ці терміни можуть бути скомбіновані для представлення знань певної галузі.

Згідно з методологією стандарту IDEF5 (Integrated Definition for Ontology Description Capture Method) [122], який призначений для онтологічного аналізу предметної області, аналізу основних термінів і понять, що використовуються для опису об'єктів і процесів певної предметної області, меж їхнього використання, а також взаємозв'язків між ними, процес побудови онтології складається з п'яти основних етапів:

- 1) вивчення і систематизація початкових умов (Organize and Define Project) – цей етап передбачає визначення мети та контексту проєкту розроблення онтології, а також розподіл ролей між членами проєкту;
- 2) збір та накопичення даних (Collect Data) – на цьому етапі відбувається збір і накопичення необхідних початкових даних для побудови онтології;
- 3) аналіз даних (Analyze Data) – завдання цього етапу полягає в аналізі та групуванні зібраних даних, призначених для побудови термінології;
- 4) початкове розроблення онтології (Develop Initial Ontology) – на цьому етапі формується попередня онтологія, на основі відібраних даних;

5) уточнення і затвердження онтології (Refine and Validate Ontology) – ця діяльність включає уточнення та перевірку онтології, завершує процес розроблення.

Відповідно до стандарту IDEF5 використовуються основні концепції дослідження онтології предметної області:

- при побудові концептуальної моделі використовуються предметні знання у вигляді набору понять і відношень, що з'єднують їх, кожне поняття має ім'я і може мати атрибути, кожен атрибут може мати значення з урахуванням специфіки предметної області;

- модель може бути побудована з використанням словника термінів, що використовується при описуванні характеристик об'єктів і процесів відповідної предметної області, точних і однозначних визначень всіх термінів цього словника і класифікації логічних зв'язків між цими термінами;

- основою проектування концептуальної моделі інформаційної системи є предметна область з відпрацьованою та вивіреною термінологічною системою, правильно побудованою ієрархією визначень предметно-орієнтованих понять [122].

Основними засадами при системному аналізі предметної області є зокрема принципи кінцевої мети, єдності, зв'язності, ієрархії, розвитку (інтеграція з іншими фрагментами онтології предметної області або «родинними» онтологіями) [123].

Онтологія предметної області «Розумне місто» – це модель, що містить систему понять, пов'язаних між собою, які використовуються для опису цієї галузі. Це концептуальна модель реального світу, тому її поняття повинні відображати цю реальність.

Основними задачами, які повинні вирішувати онтології, є: спільне використання людьми та програмними агентами загального розуміння структури; накопичення і повторне використання знань у предметній області; можливість робити явними припущення в предметній області; відокремлення

знань предметної області від операційних знань; аналіз знань предметної області [98].

Онтологія як інтероперабельна модель подання знань предметної області «Розумне місто» є забезпеченням відповідного формалізованого опису об'єктів цієї галузі. У ній визначаються терміни предметної області, подається їхнє тлумачення, містяться твердження, які обмежують зміст цих термінів. Сукупність термінів предметної області утворює її термінологічну систему, яка представляється тезаурусом «розумного міста», що відображає онтологію предметної області «Розумне місто». Це повний систематизований набір даних галузі «розумне місто», що дає змогу людині чи комп'ютеру в ній орієнтуватися.

Використаний нами підхід для представлення понять, властивих досліджуваній галузі, містить базовий набір термінів, який використовується для опису «розумного міста». Тезаурус дозволяє представити основні поняття таким чином, що вони стають базовими для формального інформаційного пошуку.

Когнітивно-інформаційна складова в постановці та вирішенні задач предметної області «Розумне місто» базується на розробленні інтелектуальних систем аналізу текстів природної мови. Основним елементом системи семантичного аналізу є ефективна словникова база. Одним із завдань створення інтелектуальних систем онтології та тезауруса предметної області є сукупність понять та зв'язків між ними [123, 124].

Представлення знань предметної області «Розумне місто» у формі придатній для інформаційного пошуку слугували такі причини:

- аналіз знань предметної області «Розумне місто»;
- можливості повторного використання знань цієї предметної області;
- спільне використання людьми або програмними агентами загального розуміння структури інформації предметної області «Розумне місто».

Таким чином, для представлення знань предметної області «Розумне місто», нами було розроблено тезаурус як вид онтологічного ресурсу, який сприяє продуктивній реалізації механізмів інформаційного пошуку в бібліотечних системах.

Висновки до Розділу 2

У розділі акцентовано увагу на тому, що основу будь-якої предметної області складає сукупність знань та даних про об'єкти, процеси та явища, які підлягають проєктуванню й зберіганню в базах знань та базах даних. Знання подаються як інформаційні одиниці знань, що мають зв'язки між собою. Форма подання знань вагомо впливає на характеристики інтелектуальних інформаційних систем. Так як дії над знаннями в інформаційних системах відбуваються за допомогою програм, то знання повинні бути структуровані, приведені до однозначного формалізованого виду. Формальний опис предметної області є моделлю представлення знань.

Подано коротку характеристику моделей представлення знань (логічних, продукційних, фреймових, семантичних мереж, онтологій), серед яких було вибрано найкращий варіант для представлення знань предметної області «Розумне місто» та формування пошукових образів документів. Усі із зазначених класів моделей подання знань мають переваги та недоліки в залежності від галузі застосування. Оскільки кожна предметна область має свої поняття та зв'язки між ними, термінологію, закони, процеси, задачі та методи їхнього розв'язання, то для представлення знань цих областей потрібно використовувати саме такі моделі, які містять основні характеристики даних предметних областей.

Використано метод аналізу ієрархій для вибору оптимальної моделі подання знань предметної області та формування пошукових образів у інформаційних системах різної складності, як один із методів теорії прийняття рішень, що використовується для розв'язання

багатокритеріальних задач та базується на ієрархічному представленні елементів. Коректність відбору кращого способу представлення знань забезпечено шляхом побудови ієрархічної моделі вибору способу подання знань досліджуваної галузі, побудови матриць попарних порівнянь альтернатив та проведення розрахунків на основі множини вагових коефіцієнтів, що дало підстави для отримання найкращого способу подання концептів.

Обґрунтовано визначено онтологічний підхід як найкращий спосіб формування пошукових образів у інформаційних системах різної складності, включаючи й бібліотечні, на основі побудованих матриць попарних порівнянь альтернатив та їх візуалізації у формі діаграм, та подання його у формі тезауруса, як окремого випадку онтологічного представлення знань.

Розкрито поняття «комп'ютерна онтологія», на основі аналізу наукових робіт як українських, так і закордонних учених, як форма подання формалізованого представлення знань предметної області, що придатна для автоматизованого опрацювання. Онтологія є дійовим засобом формального подання семантики галузі дослідження. Оскільки основне завдання онтологічного підходу полягає в тому, щоб полегшити користувачеві пошук інформації у великому наборі ресурсів шляхом систематизації знань, створення єдиної ієрархії понять, уніфікації термінів та правил інтерпретації, то він є придатним для використання в бібліотечних системах. Видом онтологічного ресурсу, який сприяє продуктивній реалізації механізмів інформаційного пошуку в бібліотечних системах є інформаційно-пошуковий тезаурус.

У другому розділі дисертації подано основні результати наукових праць, включених до Списку публікацій здобувача за темою дисертації (пп. 7, 9, 10, 11, 14, 15, 23) та представлених у Списку використаних джерел: [123], [75], [119], [124], [66], [94], [93].

РОЗДІЛ 3

ТЕЗАУРУС ЯК ЗАСІБ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СУЧАСНИХ ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМ БІБЛІОТЕК

3.1 Понятійно-термінологічний аналіз тезауруса

Основні напрямки розвитку інформаційних технологій пов'язані зі створенням інформаційних систем, заснованих на знаннях відповідних предметних областей. Знання користувача про конкретну предметну область, яка його цікавить, необхідно подавати у формі, придатній для автоматичного опрацювання. Тезауруси предметних областей поряд з онтологіями предметних областей є саме такою формою подання знань [125].

Історія виникнення тезаурусів сягає коренями праць великих мислителів стародавніх часів. У II ст. н. е. з'являється робота лексикографа й софіста Юлія Поллукса (Полідевка) «Ономастикон» (J. Pollux. Julii Pollucis Onomasticon. Cum praefatione Simonis. Grynaeri. Basilea) грецькою мовою. Це словник, що складається з 10 книг, в яких подаються короткі пояснення слів, що відносяться до певних тематичних груп. Наприклад, у першій книзі говориться про віру, богів, релігійні звичаї; у другій – про людину, її фізіологічну будову; у четвертій – про освіту, у шостій – про купівлю та продаж; у дев'ятій – про адміністративний поділ. Слова в цьому словнику супроводжуються короткими тлумаченнями.

Між II і III ст. н. е. з'являється санскритський словник «Амаракоша», автором якого є давньоіндійський поет, граматик та лексикограф Амара Сіна (за іншими джерелами Амара Дева) [126]. Словник містить близько 10 тис. слів, які об'єднані в три книги, кожна з яких ділиться на глави, глави – на секції. Так, перша книга присвячена небу, богам і всьому тому, що з ними пов'язано. У ній є секції: «Небесне склепіння, атмосфера, планети, зірки», «Слово, мова, твір», «Потойбічний світ, душі покійних, нещастя, страждання», «Моря, вода, острови, річки, риба, водойми, водорості» та ін.

Друга книга містить слова, що відносяться до землі, рослинного й тваринного світів та людини. Третя книга є власне мовною, що впливає з назв шести її глав: «Відмінювані омоніми; невідмінювані омоніми», «Визначення роду: імена жіночого роду, імена чоловічого роду...» та ін.

Слово «тезаурус» як науковий термін було вперше використано у XIII столітті флорентійським ученим Брунетто Латіні (італ. Brunetto Latini) в заголовку його праці – систематизованої енциклопедії «Книга про скарб» («*Livres dou trésor*»), виданій Ф. Шабейлем (Chabeille, François, Париж, 1863) [127], що повністю збігалось зі семантикою слова «thesauros», тобто «скарб», «багатство», «запас». Це енциклопедія, у якій подано різні відомості про навколишній світ. Так, у першій частині розповідається про створення світу, у другій – про пророків та чесноти, у третій – про риторику та мистецтво управління державою.

Сучасний етап історії ідеографічних словників відкривається роботою П. Роже «Тезаурус англійських слів і виразів» (Peter Mark Roget “*Thesaurus of English Words and Phrases*”) (1852 р.) [128]. П. Роже намагався охопити весь лексичний фонд англійської мови. Усі елементи лексики англійської мови він розбив на такі понятійні кластери: абстрактні відношення, простір, фізика, матерія, почуття, інтелект, воля, психологічні стани. Далі ці категорії поділяються на 24 класи, класи – на підкласи, підкласи розпадаються на категорії, категорії – на секції, секції – на групи. Усього в тезаурусі Пітера Роже більше тисячі понятійних груп, у кожен з яких він зібрав слова, близькі за значенням.

Тезаурус Роже – це «своєрідний концептуальний каталог здогадно відомих користувачам слів» [129], завдання якого – полегшити пошук взаємопов’язаних лексичних одиниць. У тезаурусі не подано дефініцій, немає розмежування значення слів, похідних слів та словосполучень, авторських прикладів, також не уточнюються галузі вживання слів. Основна увага приділяється семантичній єдності слів.

Тезаурус Роже вважається першою науково обґрунтованою спробою логічного впорядкування лексики. Він був важливим кроком у розвитку світової лексикографії, про що свідчить той факт, що тезаурус неодноразово перевидавався та його схема класифікації була адаптована для інших мов.

Через 10 років після «Тезауруса англійської мови» П. Роже у Франції виходить у світ «Аналогічний словник французької мови» П. Буассьєра (P. Boissiere «Dictionnaire analogique de la langue française») [130]. Французький лексикограф відмовився від створення логіко-філософської схеми класифікації понятійного змісту лексики та скористався алфавітно-змістовим розташуванням слів, що, на його думку, створювало оптимальні умови для користування словником. Таким чином, словниковий склад мови він поділив на активний і пасивний. П. Буассьєр розбив активну лексику французької мови на 2 тисячі тематичних груп, що складаються із центрального слова та групи слів, що співвідносяться з ним за змістом. Утворені лексико-семантичні групи розташовуються в словнику за алфавітом слів-центрів. Паралельно з цими групами автор подає повний алфавітний список усіх слів словника з відповідними посиланнями на слова-центри.

Найбільш значущим ідеографічним словником німецької мови є словник Ф. Дорнзайфа «Німецький словниковий склад за предметними групами» (Берлін, 1959) (F. Dornseiff *Der deutsche Wortschatz nach Sachgruppen*) [131]. Весь лексичний масив розбито на 20 класів. Поділ здійснюється в такому порядку: від аргі (1-4 класи) до слів, що виражають поняття, пов'язані зі сферою соціальних відносин і культури (13-20 класи), через групи слів, що позначають поняття, що стосуються природи (5-8 класи) і людини (9-12 класи). Кожен з двадцяти класів складається з певного числа понятійних груп. У словнику є алфавітний покажчик усіх слів із зазначенням їхнього місця в ідеографічній класифікації.

Іншим вагомим ідеографічним словником є словник іспанського мовознавця Х. Касареса «Ідеологічний словник іспанської мови» (J. Casares. *Diccionario ideologico de la lengua espanola*) [132]. Словник складається з

трьох частин: синоптичної, аналогічної й алфавітної. Синоптична частина складається з 38 основних понятійних класів, які потім розпадаються на 2000 понятійних груп. В аналогічній частині кожна з двох тисяч понятійних груп наповнюється словами. Алфавітна частина є звичайним тлумачним словником сучасної іспанської мови, в якому слова мають вербальне визначення та забезпечені вказівками на їхнє місце в ідеографічній класифікації [133].

Однак, варто зазначити, що перші тезауруси не враховували особливостей інформаційної діяльності, вони були пов'язані з фундаментальними проблемами пізнання, відображенням засобами природної мови сутності навколишнього світу та закономірностей його сприйняття. Словники такого типу, подаючи семантичну системність, не могли забезпечити повне відображення лексичної системи з усіма зв'язками, що існують між її елементами [134].

Використання тезаурусів, що мали відношення до обчислювальних машин, було пов'язано з ім'ям Анетти Мастерман, яка в 1954 році звернула увагу на те, що тезаурус – досить загальний засіб для опису семантичних структур природної мови й запропонувала використати цей засіб у галузі машинного перекладу. У сімдесятих роках розпочався «тезаурусний бум», пов'язаний з розробленням інформаційно-пошукових систем. Відтоді термін «тезаурус» міцно узвичаївся як у лінгвістів, так і у фахівців з інформаційних технологій [135].

Термін «*тезаурус*», що в перекладі з грецької мови «*thesauros*» означає «скарбниця», не є новим у науковій літературі, адже продуктивно функціонує в лінгвістиці, педагогіці, лінгвокультурології, психолінгвістиці, кібернетиці, семіотиці та в інших науках протягом тривалого періоду, що дає підстави говорити про нього як про міждисциплінарний термін.

Тезауруси активно використовуються в штучному інтелекті та інформатиці. Все більше інтелектуальних завдань, пов'язаних з опрацюванням знань, наприклад, таких як індексування документів,

інформаційний пошук і автоматичне опрацювання текстів, вирішуються з допомогою тезаурусів [136].

Тезаурус – багатозначне слово, що має декілька значень. Можна виділити такі подання тезауруса, наявні в сучасній науці:

1. Як абстрактно існуючий об'єкт – сукупність знань, накопичених людиною або деяким колективом; упорядкований і відбитий у свідомості людини «лексикон», «світ» окремої особистості, «світ знань та інтересів», наприклад, «світ знань та інтересів дитини – тезаурус дитини», «світ знань та інтересів художника – тезаурус художника», «світ знань та інтересів бізнесмена – тезаурус бізнесмена» або, інакше кажучи, тезаурус предметної області «Бізнес» і т. п. [137]; загальний словниковий запас людини, що активно використовується нею в певній галузі професійної діяльності [138]. Практично тезаурус людини – це певною мірою вербалізована сукупність її уявлень про світ, яка базується на пізнавальних, оціночних і діяльнісних установках.

2. Як матеріально існуючий об'єкт – словник, що можна розглядати у двох аспектах – лінгвістичному та інформаційно-пошуковому.

У лінгвістичному розумінні тезаурус:

1) словник, в якому якнайповніше подано всі слова мови з переліком прикладів їхнього використання в текстах (у повному обсязі це може бути реалізовано лише для мертвих мов); у лінгвістиці термін «тезаурус» використав Л. Щерба [139] в його первинному значенні «сховище», «скарб», щодо словника, в якому зафіксовано по-можливості всі контексти, в яких вживається конкретне слово;

2) ідеографічний словник, в якому відображаються семантичні відношення (синонімія, антонімія, гіперо-гіпонімія тощо) між лексичними одиницями.

Тезаурус як ідеографічний словник у ідеалі передбачає подання ієрархічної системи понять, що забезпечує пошук від змісту до лексичних одиниць, тобто пошук слів, виходячи з понять [140]. Тезаурус, за

визначенням Ю. Караулова, – це семантичний словник упорядкованих зібрань термінів, які відображають сукупність знань людини в певній сфері й зберігають перелік понять разом з їхньою інтерпретацією та відношеннями [140].

Тезаурус – семантичний об'єкт, який встановлює відношення між термінами. Він відображає такі логіко-сміслові параметри: синонімію, ієрархічні й асоціативні зв'язки. Зазвичай, у тезаурусах не встановлюється ієрархія термінів, але його зміст дозволяє їх виявити [141].

В інформаційному аспекті активне використання тезаурусів розпочалося разом з розвитком автоматизованих пошукових систем. Саме тоді тезауруси стали розглядатися як засіб для формування пошукових образів документів і пошукових запитів користувачів. У 70-их роках ХХ століття з'явився новий вид тезаурусів – інформаційно-пошукові, склад яких обмежувався термінами певної предметної області.

Одним з перших науковців, хто переніс поняття «тезаурус» з лексикографії в інформатику був Браян Вікері (Brian Campbell Vickery), англійський учений. Саме в галузі інформатики тезауруси знайшли широке практичне застосування. На міжнародній конференції із загальних принципів побудови тезаурусів (Варшава, 23-27 березня 1970 р.) було дано таке розуміння інформаційно-пошукового тезауруса. Тезаурус – лексичний інструмент інформаційно-пошукової системи, який складається з контрольованого, але змінного словника термінів зі змістовими зв'язками між ними. Такий словник розкриває будь-яку галузь знань і є переліком дескрипторів (основних термінів) і недескрипторів (допоміжних термінів), упорядкованих за систематичним і алфавітним принципами й містить відношення як ієрархічного (родо-видового), так і неієрархічного типу [142].

Сьогодні поняття «тезаурус» тлумачиться значно ширше, ніж воно мало значення у лінгвістиці. Під тезаурусом розуміється найприйнятніша форма опису знань певної предметної галузі, як повна систематизована форма інформації (знань) та установок у певній сфері життєдіяльності, що

дають змогу орієнтуватися в ній [143]. У комп'ютерних дисциплінах під тезаурусом розуміють «знання, подані у вигляді понять і відношень між ними, тобто певним чином структуровані знання» [144].

Тезаурус представляє собою словник понять та класифікаційних зв'язків між ними. Основним завданням тезауруса є переклад на інформаційно-пошукову мову смислового змісту документів і запитів, що ставляться до різних галузей науки й техніки, які охоплені класифікаційною схемою тезауруса. Тезаурус повинен забезпечити єдине уніфіковане й формалізоване подання інформації в інформаційних системах [138].

Тезаурус – це словник, в якому слова й словосполучення з близькими значеннями згруповані в одиниці, що називаються поняттями, концептами або дескрипторами, і в якому у вигляді відношень, ієрархії вказуються семантичні зв'язки між цими поняттями (концептами, дескрипторами) [136].

У роботі [145] тезаурус подається як словник, що подає семантичні відношення між лексичними одиницями дескрипторної інформаційно-пошукової мови (дескрипторами) і призначений для пошуку слів за їхнім смисловим змістом; це контрольований словник термінів предметної області, що розробляється для покращення якості інформаційного пошуку в певній предметній області.

Тезаурус насамперед представляє термінологію певної галузі знань. Зважаючи на специфіку семантики терміна, завдання упорядкування термінів співвідноситься із завданням систематизації наукових знань. Це дає підстави розуміти статус тезауруса як способу представлення наукових знань [146].

Стандарт ISO 2788-1986 подає тезаурус як набір термінів, пов'язаних між собою відповідними зв'язками (відношеннями) [147].

Український стандарт ДСТУ 4032-2001 «Інформація та документація. Одномовний тезаурус» подає дефініцію тезауруса як словник контрольованої мови індексування (контрольованої множини вибраних з природної мови термінів, що використовуються для представлення в стислій формі предметів документів), формально організований таким чином, що апріорні відношення

між поняттями (наприклад, «ширший» і «вужчий») мають явне вираження [148].

У роботі [149] тезаурус розглядається як одиниця системології, складові елементи якої перебувають у певних відношеннях між собою. Таке визначення було сформульоване в результаті аналізу подання тезауруса в лінгвістиці та комп'ютерних дисциплінах. Авторка зазначає, що у визначеннях тезауруса в цих галузях спостерігається як спільне, так і відмінне. Спільне полягає в наявності в тезаурусі структурованих, систематизованих відношення між його компонентами, а відмінне – у назві цих компонентів. Якщо з позицій лінгвістів певні відношення є саме між лексичними одиницями, то у визначенні комп'ютерників йдеться про структурованість знань. На думку авторки, «жодну з названих одиниць використовувати у визначенні тезауруса не доцільно. По-перше, тому, що не лише лексичні одиниці, здатні вступати в певні відношення з іншими, а й такі мовні одиниці, як словосполучення, речення та ін. По-друге, знання є характерними лише для живої істоти, і тому структурування знань у комп'ютерних електронних тезаурусах втратить сенс». Тому подання тезауруса як «лексикографічної одиниці системології, побудованої на семантичних взаємовідношеннях між інформативно значущими складовими елементами, у ролі яких виступають вербалізовані концепти певної концептосфери», на думку авторки, мало б задовольнити представників різних наук.

Виходячи з такого подання тезауруса, актуальності набуває проблема дослідження як самих компонентів тезауруса, так і відношень, які виникають між ними.

З-поміж багатьох визначень поняття «тезаурус» найоптимальнішим для використання в галузевій лексикографії можна вважати тлумачення тезауруса як лексичного інструмента інформаційно-пошукових систем, що складається з контрольованого, але змінного словника термінів, між якими зазначено смислові зв'язки. Такий словник, вичерпно покриваючи певну

специфічну галузь знань, є переліком дескрипторів і недескрипторів, який упорядкований систематичним і алфавітним способами й містить вказівку на смислові відношення між ними ієрархічного та неієрархічного типів [133].

Автори робіт [93, 150, 151] тезаурус розглядають як окремий випадок онтології, який простіше формувати та обробляти, та представляють тезаурус як повний систематизований набір даних про будь-яку галузь знань, які дають можливість людині чи комп'ютеру орієнтуватися в цій сфері.

Найпоширенішим різновидом електронних тезаурусів вважається інформаційно-пошуковий тезаурус, під яким розуміється призначений для автоматичного оброблення й пошуку інформації одномовний контрольований і змінюваний «дескрипторний словник з вказаними в ньому парадигматичними відношеннями» [152].

У роботі [153] інформаційно-пошуковий тезаурус розглядається як частковий випадок моделювання знань, який «з одного боку, є способом формалізованого представлення термінології, тому що досить строго представляє семантичні відношення між термінами, а з іншого – вважається важливим джерелом постійного вдосконалення систем знань конкретних наук». Актуальність розроблення тезаурусів зводиться до необхідності створення термінологічних банків даних для автоматичного семантичного аналізу науково-технічних текстів і термінологічних банків знань для накопичення значного обсягу термінологічної інформації.

У рамках дослідження термін «інформаційно-пошуковий тезаурус предметної області» розглядається як сукупність структурованих знань, поданих у вигляді набору дескрипторів та аскрипторів, які пов'язані між собою певними відношеннями, що відображають семантичні зв'язки між поняттями певної предметної області. Наявність зв'язків між термінами створює передумови для використання тезаурусів у бібліотеках у процесах індексування документів та інформаційного пошуку.

Тезаурус дозволяє представити зміст терміна не тільки з допомогою його визначення, а й за допомогою співвіднесення його з іншими термінами

або групами термінів. Таке співвіднесення, як правило, задається шляхом встановлення між термінами різних відношень. Наявність зв'язків у термінів створює передумови для використання тезаурусів у задачах індексування та інформаційного пошуку. Основним призначенням тезауруса є підвищення ефективності пошуку необхідної інформації: на основі семантичних відношень тезауруса відбувається розширення запиту, навігація по зв'язках допомагає чіткіше сформулювати сам запит [154].

За твердженням Л. Пшеничної, одна з функцій тезауруса може бути зведена до зняття багатозначності між планом вираження й планом змісту в природній мові [155].

А. Купцова і А. Стерлигова вважають, що тезаурус повинен стати ще одним продуктом роботи по систематизації термінології, а збір термінографічних даних у формі тезауруса є специфічним засобом для контролю над спеціалізованою мовою науки й техніки, що застосовується для індексації предметного змісту об'єктної галузі, вираження відношень між концептами [156].

Таким чином, тезаурус – це спосіб систематизації знань певної предметної області та інструмент ефективного інформаційного пошуку.

Проблематики тезаурусного моделювання лексики подається у працях В. Морковкіна [133], Ю. Караулова [140], І. Воронцової [157] та ін. Вони зробили спроби класифікувати тезауруси на основі типів смислових зв'язків між елементами словника, обсягу словника, способів розроблення значень лексем, кількості мов, граматико-стилістичної класифікації лексем. Виходячи з цих критеріїв, тезауруси було класифіковано на власне ідеографічні, тематичні, систематичні, тематико-систематичні.

У лінгвістичній літературі з'являється поняття системної лексикографії. Український лексикограф В. Широков [158] сформулював принцип консолідованості й інтегрованості елементів лексикографічних систем, в основі – ідея інтегрованості лінгвістичного моделювання мови і її подання в лексикографічній системі.

Тезауруси можна класифікувати за різними ознаками. Так, за мовою розроблення тезауруси поділяються на:

- 1) одномовні тезауруси, структура яких подана однією мовою, до такого типу належать Тезаурус Роже, WordNet, GermaNet, RussNet та ін.;
- 2) багатомовні тезауруси, структура яких подана одночасно декількома мовами, наприклад, міжнародні проєкти типу EuroWordNet, EUROVOC, тезаурус ЮНЕСКО, тезаурус SNOMED та ін.

Однією з головних характеристик тезауруса, що визначає як його структуру, так і обсяг, є його тематична спрямованість або підмова опису.

За тематичною спрямованістю тезауруси поділяються на:

- 1) загальномовні тезауруси, котрі представляють лексичний склад усієї мови й здебільшого є дуже об'ємними, наприклад, Тезаурус Роже, WordNet, EuroWordNet, PyТез, тезаурус ЮНЕСКО та ін.;
- 2) спеціалізовані тезауруси, які представляють певну галузь науки або просто об'єднують слова за конкретною темою, наприклад, Тезаурус NASA з аеронавтики, медичні тезауруси SNOMED та MeSH, тезаурус сільськогосподарської тематики AGROVOC та ін.

Загальномовні тезауруси відрізняється від спеціалізованих певними ознаками:

- 1) в основі спеціалізованого тезауруса покладена домінуюча наукова концепція, тоді як синоптична схема загальномовного тезауруса будується під впливом ідеологічних і світоглядних факторів;
- 2) особливості лексичного матеріалу, що подається в тезаурусах цих типів, принципово відмінний – відображення лексичної системності в спеціалізованому тезаурусі наперед визначене позамовними обставинами, терміносистемою галузі, що описується, а загальномовний тезаурус здебільшого сам моделює лексико-семантичну систему;
- 3) у спеціалізованій тезаурус переважно вносяться тільки іменники, так як саме ця частина мови переважно є термінами, а в загальномовному

представлені слова практично всіх частин мови й стійкі сполучення слів (фразеологізми і прислів'я) [159].

Більш детальну класифікацію інформаційно-пошукових тезаурусів подано у роботі Н. Гендіної [137]. Науковець виділяє типи тезаурусів за такими параметрами: широта тематичного охоплення, призначення в автоматизованих інформаційних системах, системність побудови, особливості складу лексики та внутрішньоструктурної побудови, кількість природних мов, форма подання і знакова природа інформації.

Тезауруси можна використовувати в різних галузях. Однією з головних сфер їхнього застосування є інформаційно-пошукова робота, при якій здійснюється заміна лексичних одиниць тексту на дескриптори при індексуванні документів та використанні ієрархічних і асоціативних зв'язків. На основі відношень тезауруса відбувається розширення пошукового запиту: розширення слів запиту синонімічними, більш загальними за сенсом термінами. У пошуковий образ кожного документа включаються поняття, які подаються термінами в тезаурусі. Користувач, здійснюючи пошук, використовуючи тезаурус, може знайти всі поняття, які цікавлять його в певній предметній області, та всі терміни, що характеризують їх.

Лексичні одиниці тезауруса пов'язуються між собою за допомогою ієрархічних та неієрархічних зв'язків. У такий спосіб будується система організації знань, які в цілому, на думку К. Лобузіної, призначені для організації пошуку інформації та обслуговування колекцій електронних документів, забезпечують можливість пошуку необхідних користувачеві інформаційних джерел навіть без попереднього знання про їхнє існування [110].

На думку А. Гладуна, Ю. Рогушиної [160], тезаурус може стати ефективним інструментом формування запитів та помітно підвищити якість інформаційного пошуку в спеціалізованій тематичній галузі при дотриманні таких умов:

- представляє термінологію досить вузької предметної області;

- використовується набір сильно диференційованих семантичних відношень;
- незалежний від машини пошуку.

К. Лобузін у монографії «Бібліотечні класифікації в сучасних інформаційних сервісах» [161] вважає, що проєктом майбутнього може стати база даних, яка включає багатомовний словник слів та словосполучень, зв'язаних з авторитетними записами для різних класифікацій і предметних рубрик (тезаурусів). Саме використання такого масиву даних могло б забезпечити пошук у міжнародних базах даних. Бібліотеки не будуть залежати від розходжень у використанні пошукових мов і типу класифікації власних електронних каталогів або придбаної бази. Такий проєкт може сприяти організації тематичного пошуку через інтернет.

Тезаурус застосовується як інструмент термінологічного контролю в процесі аналізу та індексування документів та інформаційних запитів, а також у процесі автоматизованого пошуку інформації. Функціональна роль тезауруса в інформаційно-пошукових системах висуває високі вимоги до якості підготовки тезауруса, від ступеня досконалості якого переважно залежить ефективність пошуку. Створення тезауруса є одним із засобів моделювання предметної області у формі множини термінів та множини відношень між цими термінами [162].

Сучасні підходи до проведення інформаційного пошуку ґрунтуються на вивченні інформаційних потреб та формуванні зручних, ефективних інформаційних запитів. Інформаційний запит відображає інформаційну потребу, сформульовану природною мовою. Перекодування інформаційного запиту інформаційно-пошуковою мовою дозволяє сформувати пошуковий образ запиту. Удосконалені пошукові технології на основі пошукового образу запиту забезпечують вивчення документів, що містять релевантну наукову інформацію. В умовах постійного зростання кількості документів ефективнішим є підхід, що ґрунтується на поєднанні двох методів: 1) аналізу контенту документів на основі пошукового образу запиту;

2) співставлення пошукового образу запиту з пошуковим образом документа. Взаємодія обох методів пошуку забезпечує зростання ефективності пошукових процесів, проте потребує формування коректних пошукових запитів.

Процедуру інформаційного пошуку можна подати у вигляді наступного алгоритму:

Крок 1. Формулювання інформаційної потреби і запиту.

Крок 2. Окреслення інформаційних масивів, у яких здійснюватиметься пошук.

Крок 3. Співставлення пошуковому образу запиту з пошуковими образами документа та вмістом документа.

Крок 4. Видобування інформації, що відповідає пошуковому образу запиту.

Крок 5. Формування відповіді користувачу.

Такий підхід працює, якщо при формуванні пошукового образу й пошукового запиту використовується інформаційно-пошукова мова, яка ґрунтується на використанні спільних термінологічних засад. Безперечно, синтаксис і семантика інформаційно-пошукової мови визначається структурою й змістом документів. Завдання з формування пошукового образу документа й пошукового образу запиту спрощується, при чіткому представленні знань про конкретну предметну область, з використанням системи, придатної для використання інформаційних технологій. Тезауруси предметних областей є саме такою системою подання знань [163].

3.2 Стандарти, що регламентують формат представлення тезауруса

Існує низка міжнародних та вітчизняних стандартів, що регламентують побудову інформаційно-пошукових тезаурусів.

Основними документами, що регламентують формат представлення тезаурусів, є міжнародні стандарти ISO, стандарти ГОСТ та ДСТУ [164].

Міжнародні стандарти ISO (International Organization for Standardization) розроблені міжнародною організацією зі стандартизації. Існують такі стандарти ISO для представлення тезауруса: ISO 2788:1986. Documentation – Guidelines for the establishment and development of monolingual thesauri (ISO 2788:1986 «Вказівки щодо створення і розвитку одномовного тезауруса») [147] та ISO 5964:1985. Documentation – Guidelines for the establishment and development of multilingual thesauri (ISO 5964:1985 «Вказівки щодо створення і розвитку багатомовного тезауруса») [165].

Стандарт ISO 2788:1986 для одномовних тезаурусів був розширений та уточнений американським стандартом ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (Periodic Review) (ANSI/NISO Z39.19-2005 «Вказівки щодо побудови, форматування та управління одномовними контрольованими словниками (періодичний огляд)») [166]. Стандарт ISO 2788:1986 залишався актуальним протягом 25 років до появи стандарту ISO 25964-1:2011 Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval (ISO 25964-1:2011 «Інформація та документація – Тезаурус та зв'язок з іншими словниками – Частина 1: Тезаурус для інформаційного пошуку») [167] в 2011 році та його продовження ISO 25964-2:2013 Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies (ISO 25964-2:2013 «Інформація та документація – Тезаурус та зв'язок з іншими словниками – Частина 2: Зв'язок з іншими словниками») [168].

Структура багатомовних тезаурусів регламентується стандартом ISO 5964-1985. У ньому, крім зв'язків між термінами, є також зв'язки між еквівалентними термінами на різних мовах, серед яких існують такі типи зв'язків: – повна еквівалентність; – неповна еквівалентність (значення

термінів не співпадають, але перетинаються); – часткова еквівалентність (значення одного терміна ширше, ніж значення іншого); – еквівалентність «один до багатьох» (значення одного терміну відповідає сукупності значень декількох термінів).

Стандарти серії ISO 13250 [169, 170, 171] визначають тематичні карти, інформація подається у форматі xml з поясненнями до визначень xmt. У стандартах є ключові елементи та поняття.

Стандарт ГОСТ 7.25-2001 [172] для одномовного тезауруса та стандарт ГОСТ 7.24-2007 [173] для багатомовного тезауруса. Ці стандарти були створені на базі вищеописаних стандартів ISO та ANSI і, фактично, мало чим відрізняються від них. Вони встановлюють правила розроблення, структуру, склад і форму подання інформаційних тезаурусів, орієнтованих на використання лексики російської мови.

Стандарт ГОСТ Р 7.0.91-2015 [174] на базі ISO 25964-1:2011 регламентує рекомендації, що стосуються розвитку та формування інформаційно-пошукових тезаурусів.

Національні стандарти, що переважно базуються на міжнародних стандартах (ISO) цієї серії: ДСТУ 4031-2001 (ISO 5964:1985) «Інформація та документація. Багатомовний тезаурус. Методика розроблення» [175], що регламентує загальні вимоги до укладання багатомовного тезауруса, управлінські рішення в цій галузі, інші мовні проблеми тощо; ДСТУ 4032-2001 (ISO 2788:1986) «Інформація і документація. Одномовний тезаурус. Методика розроблення» [176]. В останньому стандарті наведено рекомендовані процедури керування складом словника, встановлено терміни індексування та складені терміни, визначено базові відношення в тезаурусі, розглянуто засоби встановлення та відтворення певних різновидів відношень між термінами індексування.

На рис. 3.1 представлено еволюцію вищезазначених стандартів.

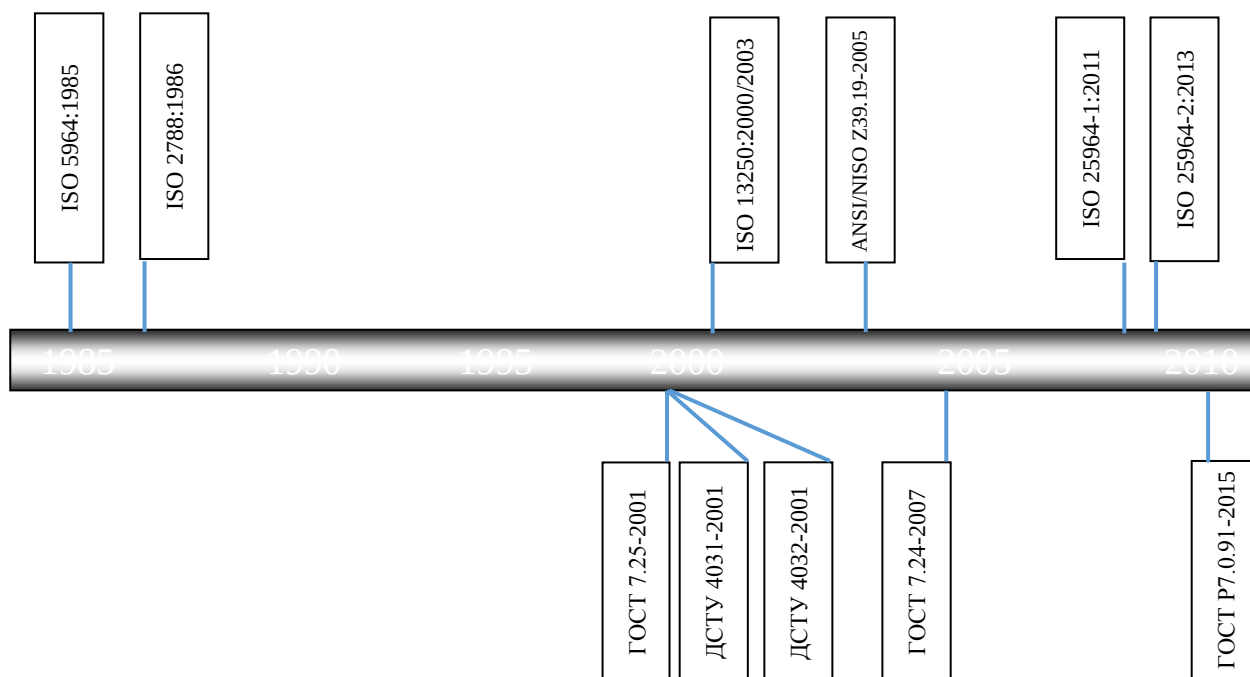


Рис. 3.1. Еволюція стандартів представлення тезауруса

3.3 Структура тезауруса як засобу систематизації наукових понять

Створення тезауруса є одним із засобів моделювання предметної області. Структура тезауруса складається з множини термінів та множини відношень між цими термінами, які є скінченними [177].

Оскільки основною одиницею тезауруса є термін, тому розглянемо його представлення як об'єкта дослідження.

Термін (лат. «*terminus*» – «рубіж», «межа») – це слово, словосполучення, що відповідає поняттю певної галузі знань (науки, техніки) і перебуває у відношенні з іншими словами та словесними конструкціями й утворює разом з ними систему, що відрізняється високою інформативністю, однозначністю, точністю й експресивною нейтральністю [178]. Множина термінів тезауруса відповідає множині концептів онтології.

Науковий термін – «слово, стале словосполучення або скорочення, яке виражає та певною мірою класифікує в даній системі наукової термінології конкретне наукове поняття, відображаючи у своїй смисловій структурі

характерні ознаки об'єкта термінування та взаємозв'язки цього об'єкта з іншими з достатньою для взаємного спілкування точністю» [179].

Для терміна характерні такі основні ознаки:

- 1) чітке визначення, зафіксоване в словнику;
- 2) однозначність у межах одної термінологічної системи або має тенденцію до однозначності;
- 3) точність та самостійність (не залежить від контексту);
- 4) стилістична нейтральність;
- 5) системність;
- 6) коротке вираження;
- 7) відсутність синонімів у межах одної термінологічної системи [180].

Термінологія – це сукупність термінів, що виражають поняття певної галузі знань. Галузеві термінології називають термінологічними системами чи терміносистемами. Говорячи про терміносистему предметної області «Розумне місто», доцільно відзначити динаміку її виникнення та інтенсивність поповнення новими лексемами.

Кожна терміносистема характеризується спільними рисами:

- 1) сукупність термінів, що виконує комунікативно-тематичне чи комунікативно-ситуативне завдання;
- 2) елементи структури безпосередньо чи опосередковано пов'язані між собою;
- 3) кожен термін належить до певної терміносистеми або одночасно до кількох;
- 4) чіткі межі між терміносистемами;
- 5) терміносистеми можуть перетинатися, утворюючи певну групу термінів, спільних для кількох терміносистем;
- 6) певна терміносистема може використовуватися як окремою галуззю, так і кількома (міжгалузева термінологія) [181].

Терміносистема предметної області «Розумне місто» є основою для побудови тезауруса «розумного міста», що описує її семантику та дає можливість здійснювати інформаційний пошук у ній.

Більшість стандартів по інформаційно-пошукових тезаурусах указують на зв'язок термінів з поняттями певної предметної області. В американському стандарті ISO 25964-1:2011 [167] зазначається, що терміном є одне або декілька слів, що позначають поняття. При цьому поняття розглядається як одиниця думки, яка формується для відображення всіх або деяких властивостей конкретного або абстрактного, реально існуючого або уявного об'єкта. Поняття існують як абстрактні сутності, незалежно від термінів, які їх висловлюють.

Стандарт ГОСТ 7.74-96 [182], подаючи визначення тезауруса, не оперує поняттям «термін». Одиницею тезауруса є лексичні одиниці, взяті з природної мови, що позначають окремі поняття. Лексичними одиницями можуть бути слова, стійкі словосполучення, аббревіатури, символи, дати, загальноприйняті скорочення та ін.

Як зазначає Н. Лукашевич, що не всі розробники тезаурусів чітко розділяють поняття і терміни. Так, розробники тезауруса AGROVOC характеризують його як терміно-орієнтований (term-oriented), що знаходить свій вияв у тому, що до терміна неможливо додати синоніми [136]. Ця особливість тезауруса розглядається автором як недолік, який необхідно виправити.

Ми пропонуємо таке визначення терміна: слово, словосполучення або аббревіатура, які означають чітко окреслені поняття, що відносяться до певної предметної області, входять у систему понять цієї області (хоча не обмежуються тільки цією сферою використання) і, як правило, регламентуються дефініцією.

Під термінологією «розумного міста» ми розуміємо сукупність мовних (лексичних) одиниць – термінів, що позначають поняття предметної області

«Розумне місто» та виступають засобом отримання, зберігання, оброблення, передавання знань у цій галузі.

Оскільки одиницями тезауруса виступають терміни предметної області, то, таким чином, тезаурус можна розглядати як модель термінологічної системи, більш високого рівня організації термінів.

Під «термінологічною системою» розуміється сукупність термінів певної галузі діяльності, які створюють специфічний шар лексики, що характеризується сталою системою внутрішніх відношень: матеріальних і семантичних, парадигматичних і синтагматичних. Семантичні відношення є родо-видовими відношеннями між частиною й цілим, вони з'єднують ті чи інші терміни в термінологічній системі й мають яскраво виражений логічний характер [183]. Термінологічна система – це складна динамічна стійка система, елементами якої є відібрані за певними правилами лексичні одиниці будь-якої природної мови, структура якої ізоморфна структурі логічних зв'язків між поняттями спеціальної галузі знань і діяльності, а функція полягає в тому, щоб служити знаковою (мовною) моделлю цієї галузі знань і діяльності [138].

Моделювання термінологічної системи останнім часом частіше здійснюється за допомогою тезауруса, що дозволяє правильно відобразити складність семантичних взаємозв'язків між термінами.

Для досягнення певної системи термінів потрібно оволодіти вміннями описувати предметну сферу відповідної галузі знань, тобто подавати сукупність об'єктів, предметів, процесів, явищ, що виокремлюються з погляду певної діяльності; відношення між виділеними об'єктами, предметами, процесами та явищами чи їхніми частинами; результатами певних відношень, взаємодій між виділеними предметами. Така сукупність дає можливість людині зрозуміти поняттєву картину предметної сфери певної галузі знань, що може бути побудована за допомогою тезаурусного моделювання термінолексики [184].

Словник – це абстрактний мовно-інформаційний об’єкт, визначальною рисою якого передусім є членоване розміщення матеріалу; основною композиційною та комунікативною одиницею слугує відносно самостійний відрізок тексту, який називають словниковою статтею [185]. Якщо словники розкривають значення слова в словниковій статті з наведеними прикладами його вживання, то тезауруси створюються таким способом, щоб пошукове слово виражало певне поняття на основі взаємозв’язків, якими воно оперує. Тезауруси створюються за смисловою близькістю і є ієрархічною структурою понять, що відповідають значенням певних слів або текстових виразів. Вони є невід’ємною частиною пошукової системи.

Головна відмінність інформаційно-пошукового тезауруса від традиційного словника полягає в тому, що він подає не лише компіляцію статей, а й систему термінів-дескрипторів.

Дескриптор – уніфікований мовний вираз, який подається засобами природної мови, відображає предмет документа й використовується для отримання контрольованого доступу до його змісту. Дескриптор відповідає групі ключових слів природної мови, відібраних з тексту досліджуваної предметної області. Кожне поняття виражається лише одним дескриптором, і кожен дескриптор виражає лише одне поняття або групу споріднених понять. Якщо слово, виражене природною мовою, багатозначне, то йому відповідає декілька дескрипторів, тоді як декільком синонімічним словам і виразам – один дескриптор. Дескрипторами можуть бути терміни, що позначають окремі поняття предметної області та відповідають принципам загальноновживаності, поширеності, стислості й термінологічної точності. Дескриптором можуть виступати окремі слова або словосполучення залежно граматичних категорій, що використовуються.

Ключові слова, що близькі за значенням, і які використовуються для опрацювання та пошуку інформації, об’єднуються в клас умовної еквівалентності, кожен з яких є словниковою одиницею (дескриптором), яка

подається у формі окремого слова, словосполучення або коду. Класифіковані за змістом дескриптори утворюють тезаурус.

У роботі [150] відмінність тезауруса від інших словників автори вбачають у тому, що тезаурус представляє когнітивну модель певної галузі знань за допомогою декількох входів: 1) від «концепту до концепту», тобто від одного поняття до іншого з поданням вертикальних і горизонтальних зв'язків; 2) від «концепту до знака» через ідеографічну частину тезауруса з поданням поняттєво-лексичних зв'язків; 3) від «знака до концепту» за допомогою алфавітного покажчика слів у понятійних полях дескрипторів з поданням лексико-понятійних зв'язків; 4) «від знака до знака» за допомогою пермутаційного покажчика, що використовується для пошуку словосполучень по одному зі складових компонентів. Таким чином, структура тезауруса задає лексичну систему підмови в чотирьох вимірах: «концепт-знак», «концепт-концепт», «знак-концепт», «знак-знак».

Характерною особливістю тезауруса є те, що статті в ньому побудовані таким чином, що відображають семантичні відтінки близьких за значенням слів. Якщо традиційний тлумачний словник дає тільки визначення слова та приклади його використання в реченнях, то статті тезауруса розкривають значення слова і через визначення, і через зіставлення з іншими словами й понятійними групами.

В. Шульгіна [149] статтю тезауруса гіпотетично ототожнює з моделлю хімічної речовини, де речовина, що має свої фізичні й хімічні властивості, складається з окремих молекул різнорідних хімічних речовин, що так само мають власні фізичні й хімічні ознаки завдяки своїй внутрішній структурі – атомів та їхніх складових частин. Усі компоненти складної речовини при цьому перебувають у складних відношеннях не лише з сусідніми компонентами, а й з найбільш віддаленими. Окрім цього порівняння вона наводить аналогію й зі структурою Всесвіту, де його компоненти планети та інші космічні тіла, перебувають у постійних специфічних зв'язках між

собою. Тому основні принципи побудови тезауруса є природними, універсальними.

Тезаурус інформаційно-пошукової системи – це словник, що автоматично формується або заздалегідь підготовлений словниковий масив, забезпечений рядом додаткових атрибутів. Терміном «тезаурус», або «deskрипторний словник», позначається упорядкований список лексичних одиниць, що представляє лексику певної предметної області зі зазначенням зв'язків між лексичними одиницями. Тезаурус відображає онтологію предметної області. Системи опису документів і запитів за допомогою deskрипторів та інформаційно-пошукових тезаурусів називають інформаційно-пошуковими мовами deskрипторного типу або deskрипторними інформаційно-пошуковими мовами.

Тезаурус – це ієрархія понять певної предметної області. У загальному структура тезауруса подібна на перевернуте дерево. Пошук може проводитися як від загального до конкретного так і навпаки. Запит, сформований користувачем, перед тим як потрапити в пошукову систему, спочатку надходить у електронний тезаурус, який визначає до якої категорії понять відноситься слово, уточнює запит та відправляє його в пошукову систему [186].

Інформаційно-пошуковий тезаурус робить можливим вхід через будь-яку характеристику терміна, який не обмежений тільки своєю словниковою статтею, як маємо це у паперовому словнику. Електронне подання тезауруса дозволяє через систему гіперпосилань встановлювати логічні та асоціативні зв'язки між термінами.

Електронні тезауруси, що базуються на великих можливостях сучасної комп'ютерної техніки для зберігання інформації, можуть містити значно більше інформації, ніж окремий паперовий словник. Це дає можливість реалізації критеріїв комплексності й вичерпності в поданні та зберіганні інформації, що має відношення до реальних понять.

К. Лобузіна у роботі [187], розглядаючи системи організації знань, які забезпечують пошук інформації й обслуговування колекції електронних документів, подає типи систем організації знань за ступенем семантичного узагальнення: лексикографічні, інформаційно-пошукові, концептуальні. Одним із підтипів інформаційно-пошукових систем організації знань є тезауруси, які формуються з ідентифікованих та стандартизованих термінів з певними відношеннями між ними.

Інформаційно-пошукові тезауруси застосовуються для індексування документів та використання ієрархічних, синонімічних та асоціативних зв'язків при опрацюванні пошукових запитів користувачів. На рисунку 3.2 подано приклад використання тезауруса в інформаційно-пошуковій системі.

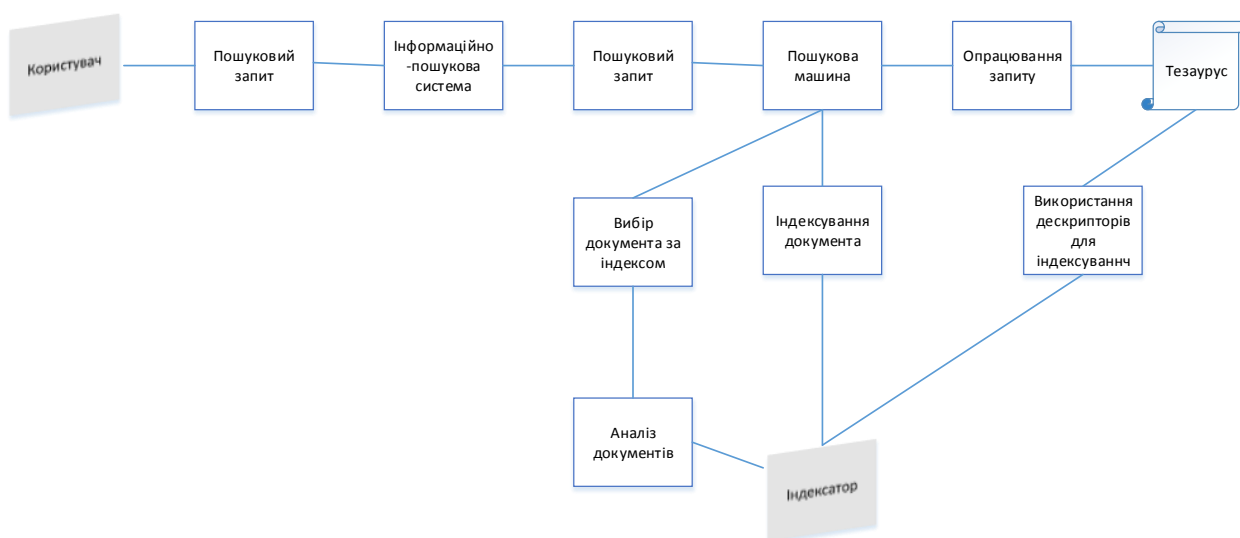


Рис.3.2 Приклад використання тезауруса в інформаційно-пошукових системах

Головна функція тезауруса – сприяти підвищенню пошуку та відбору необхідної інформації, проте структура й принципи організації, що забезпечують достатнє наповнення термінами й глибину розкриття семантичних взаємозв'язків, дозволяють тезаурусу моделювати термінологічну систему предметної області найбільш повно й системно. Практична цінність тезауруса предметної області обумовлюється

можливістю використовувати його одночасно для аналізу й конструювання лексичної системи, класифікації й зберігання термінологічних даних, опрацювання інформації в пошукових системах для підвищення ефективності проблемно орієнтованого пошуку в Web [134].

3.4 Практичні рекомендації стосовно визначення термінів тезауруса

При побудові тезауруса предметної області ми маємо справу не із загальноновживаною лексикою, а з термінами. Властивості термінологічних одиниць – системність, стійкість та регулярність зв'язків, відсутність експресії, точність, спрямованість на об'єктивність опису – дають можливість зробити адекватний опис галузі, що досліджується, за допомогою тезауруса.

Питання терміносистеми й відбору термінів при укладанні словників досліджували В. Широков (запропоновано формальні моделі структури словників для формування інтегрованих лексикографічних систем, розроблено комп'ютерну технологію побудови сучасних тлумачних словників [185, 188, 189, 190], О. Кучеренко (термінологія цивільного захисту) [191], Л. Симоненко (біологічна термінологія) [192], Б. Шуневич (способи відбору термінів для укладання словників) [193] та інші науковці.

Визначення термінів тезауруса можуть створюватися шляхом аналізу інформаційних джерел, наприклад, повнотекстових документів, тобто в автоматичному режимі. Вони також можуть формуватися шляхом вилучення з тезаурусів, онтологій та інших баз знань або створюватися вручну, безпосередньо експертом предметної області.

Для створення якісного тезауруса потрібно дотримуватися низки практичних рекомендацій, які стосуються подання визначень термінів тезауруса, тобто представлення поняття в описовій формі, що дозволяє відрізнити його від інших понять, які пов'язані з ним.

Формальні вимоги до визначення термінів тезауруса.

Визначення повинне:

- 1) бути викладене в однині (виняток становлять ті поняття, які в однині не вживаються);
- 2) визначати, чим є поняття, а не тільки чим воно не є;
- 3) мати вигляд описової фрази або речення, оскільки реченням можна подати точне визначення, яке містить важливі характеристики поняття;
- 4) містити лише широко відомі скорочення, тому що розуміння значення скорочення (аббревіатур, ініціалів) зазвичай обмежується певною галуззю, в іншій галузі ті ж самі скорочення можуть мати інше розуміння, таким чином, для запобігання неоднозначності, у визначеннях доцільно використовувати тільки повні слова без скорочень;
- 5) бути сформульованим, не використовуючи визначення інших даних або базових понять: визначення другого елемента даних або пов'язаного з ним поняття не може з'являтися у визначенні первинного елемента даних [80].

Семантичні вимоги до визначень термінів тезауруса.

Визначення повинне:

- 1) відображати суттєвий зміст поняття: усі первинні характеристики поняття, що представлені, мають бути відображені у визначенні з відповідним рівнем специфічності залежно від контексту, потрібний рівень деталізації залежить від потреб користувача системи та середовища;
- 2) бути точним та однозначним: визначення терміна повинно бути досить зрозумілим для забезпечення його однозначної інтерпретації;
- 3) бути коротким, потрібно запобігати використанню додаткових конструкцій описового характеру;
- 4) мати можливість використовуватися окремо, тобто, щоб зрозуміти поняття не потрібно використовувати додаткові роз'яснення;

5) бути поданим без використання пояснювальної інформації, у разі потреби такі пояснення можуть бути розміщені в інших атрибутах метаданих;

6) запобігати циклічних посилань, два визначення не повинні бути визначені одне через одне;

7) використовувати ідентичну термінологію та логічну структуру для пов'язаних визначень, що спрощує їхнє розуміння [80].

Моделювання термінологічних систем за допомогою тезаурусів можливе завдяки таким властивостям термінів як системність, стійкість і регулярність зв'язків, відсутність експресії, установка на об'єктивність опису.

Основними ознаками терміна вважається однозначна відповідність терміна поняттю, системність, раціональна стислість, словотворча (дериваційна) здатність [194].

Термін існує лише як елемент терміносистеми: впорядкованої сукупності термінів, які тотожно представляють систему понять певної спеціальної сфери людського знання або діяльності. Таким чином, система термінів конкретної науки складає певну численність елементів, які є взаємопов'язаними та створюють стійку єдність. Точний зміст поняття, яке визначається терміном, стає зрозумілим лише через його дефініцію, що представляє собою лаконічне, логічне визначення, в якому окреслено суттєві ознаки предмета чи дії. Визначення, окреслюючи межі поняття, фіксує зміст терміна, що відповідно передається у значенні. «Необхідні й достатні для дефініювання ознаки визначають місце термінопоняття серед інших понять, створюють його специфічність» [195].

Терміни входять до лексичного складу мови, тому всі особливості лексичних одиниць поширюються й на терміни. У зв'язку з цим, на думку деяких вчених [196], потрібно відмовитись від ідеалізації термінів та визнати, що вони мають риси, властиві загальнонавчній лексиці: багатозначність, словотворчі відношення, закономірності функціонування (вживання),

сполучуваність, розподіл у тексті. Таким чином, вивчати терміни треба не тільки зі словників та довідників, а також у науково-технічній літературі: статті у наукових журналах, праці конференцій, симпозіумів, навчальній літературі та ін.

Тому принцип характерологічного тлумачення є важливим під час укладання тезауруса.

3.5 Технологічні фази формування тезауруса предметної області

Моделювання предметної області «Розумне місто» – складна задача, оскільки потрібно поєднати індивідуальні концептуальні моделі різних фахівців для побудови єдиної моделі, яка не викликала б розбіжностей. Модель предметної області «Розумне місто» подається сукупністю термінів, які слугують для визначення термінології цієї галузі, відношень та взаємозв'язків між ними. Саме такою моделлю досліджуваної предметної області «Розумне місто» має стати створений нами тезаурус, який повинен відігравати роль посередника між інформаційною системою й користувачем, між дослідниками зазначеної галузі у процесі міждисциплінарних наукових досліджень та сприяти інтелектуальному пошуку даних у інформаційних системах, включаючи й бібліотечні.

Аналізуючи літературу, можемо сказати, що опис термінологічних ідеографічних словників подається у значній кількості робіт [152, 197, 198, 199, 200], проте консолідованого підходу до загальних принципів роботи над створенням двомовних тезаурусів немає. Розглянемо технологічні фази (поетапний характер) укладання тезаурусів у роботах науковців.

Більшість науковців виділяють чотири етапи роботи над тезаурусом. Так, С. Вискушенко подає такі етапи: 1) аналіз існуючих словників, дослідження потреб адресата, визначення вимог і майбутніх характеристик, розроблення макро- та мікроструктури словника; 2) збирання лексичного матеріалу, укладання словника, групування слів за семантичними полями,

підбирання перекладних еквівалентів та дефініцій; 3) оформлення словникових статей, впорядкування словника згідно з розробленою структурою; 4) уточнення, редагування матеріалу та перевірка відповідності словника визначеній меті [201].

Такої ж позиції дотримується С. Гриньов. На його думку, процес розроблення тезауруса передбачає: 1) проєктування (визначається тип, читацький запит та головні характеристики); 2) відбір лексичного матеріалу й формування тезауруса; 3) аналіз та опис спеціальної лексики; 4) підготовка словника до видання (редагування, уточнення) [202].

За В. Дубичинським проєкт створення словника може бути реалізований при дотриманні проходження поетапного його здійснення. Він виділяє такі сім етапів: 1) формування авторського колективу; 2) створення проєкту словника; 3) створення словникової картотеки або комп'ютерного банку даних; 4) формування реєстру слів на основі принципів, закладених у проєкті словника, та систематизація лексичного матеріалу відповідно до методики, прийнятої авторським колективом; 5) розроблення авторської концепції структури словникової статті; 6) безпосередня лексикографічна (автоматизована, комп'ютерна) інтерпретація відібраних мовних одиниць; 7) підготовка словника до видання [203].

Працюючи над укладанням двомовного словника пожежної термінології, Б. Шуневич та інші науковці дотримувалися такої послідовності: 1) відбір термінів із словників, їх переклад та уточнення значення деяких термінів; 2) створення першого варіанту української частини словника й доповнення його термінами з інших джерел; 3) пошук англійських еквівалентів українських термінів; 4) кінцевий варіант побудови української частини словника та уточнення англійського варіанту [204].

Опираючись на роботи Б. Городецького [205], А. Баранова [152], В. Дубичинського [203], О. Румянцева пропонує три етапи роботи над словником: 1) аналіз словників та дослідження потреб адресата – кінцевого користувача словника; визначення вимог до словників на основі отриманих

даних; розроблення макроструктури та мікроструктури словника; 2) укладання словника (розподіл слів за семантичними полями, підбір перекладних еквівалентів, оформлення словникових статей та словника в цілому відповідно до розробленої раніше структури); 3) експериментальна перевірка відповідності словника поставленим завданням [206].

Репрезентуючи роботу над укладанням багатомовного електронного словника гідротурбінних термінів, Є. Купріянов пропонує такий алгоритм створення електронного термінологічного словника (рис.3.3): 1) пропедевтичний етап (аналіз розроблених спеціалізованих комп'ютерних словників для виділення в них лексикографічних параметрів та шляхи їхнього доповнення або уточнення; 2) формування реєстрової частини словника з термінів, що прямо стосуються досліджуваної галузі, та термінів, які відносяться до інших галузей, але часто використовуються в досліджуваній терміносистемі; 3) розроблення макроструктури, медіоструктури та мікроструктури словника (за Е. Агриколою, макроструктура – це організація словника в цілому, мікроструктура – це окрема словникова стаття, а медіоструктура – взаємозв'язки між цими статтями [207]; 4) створення комп'ютерного забезпечення словника [208].

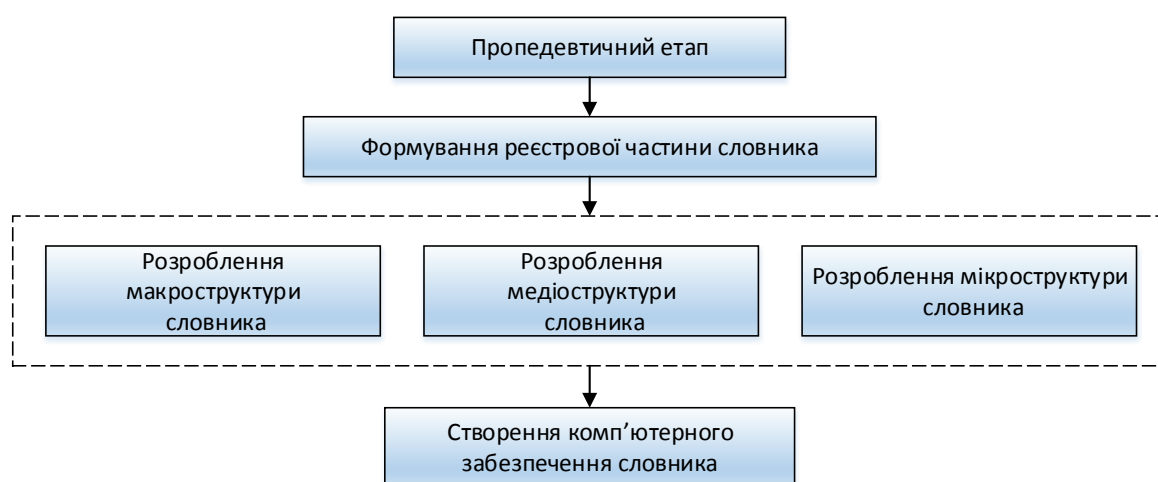


Рис. 3.3 Блок-схема алгоритму створення електронного термінологічного словника (за Є. Купріяновим)

У стандарті онтологічного дослідження IDEF5 [122] подано методологію побудови онтології предметної області, структуру якої можна

використати для формування алгоритму розроблення інформаційно-пошукового тезауруса «розумного міста». Схема (рис. 3.4) складається з п'яти основних етапів:

- 1) вивчення й систематизація початкових умов – мети й контексту розроблення тезауруса, визначення меж предметної області, яка цікавить користувача;
- 2) збирання й накопичення даних – підбір інформаційних ресурсів, що належать до цієї предметної області;
- 3) аналіз даних – вивчення відібраних інформаційних ресурсів, формування словника термінів предметної області, що містяться у відібраних інформаційних ресурсах;
- 4) початкове розроблення тезауруса – встановлення зв'язків між термінами предметної області;
- 5) уточнення та затвердження тезауруса – аналіз користувачем отриманого тезауруса та його коректування.



Рис. 3.4 Схема розроблення моделі тезауруса відповідно до стандарту IDEF5

А. Гладун, Ю. Рогушина основними етапами формування тезауруса вважають [95]:

- 1) виділення лексичних одиниць, тобто формування глосарія;

- 2) розроблення набору семантичних зв'язків;
- 3) актуалізація зв'язків – встановлення зв'язків між термінами.

Вони акцентують увагу на тому, що дуже важливо сформулювати принципи, за якими буде здійснюватися робота на кожному етапі. На першому етапі (формування глосарія) ключовими є два аспекти – джерело лексичних одиниць та критерій їхнього добування. На другому етапі (розроблення набору семантичних зв'язків) при розробленні набору семантичних зв'язків відношення можна отримувати двома способами: знаходити в тексті, що описує галузь (намагатися вичленувати й уніфікувати відношення, що існують у текстах між термінами) або безпосередньо аналізувати знання. На практиці звичайно використовується поєднання обох методик. І на третьому етапі (встановлення зв'язків між термінами) для актуалізації семантичних зв'язків між термінами тезауруса можна використовувати знання експертів, а також документи, призначені як для фіксації структури знань (словники, класифікатори тощо), так і самі знання, що відображають предметну область (статті, монографії тощо).

Проведений аналіз наукових джерел показує, що аспектам опису та параметризації інформаційно-пошукових тезаурусів присвячено значну кількість робіт, проте, як бачимо, у колі науковців існує розбіжність поглядів щодо поетапного характеру розроблення тезаурусів, що можемо пояснити наявністю в кожного науковця свого рівня деталізації. Причинами такої неоднорідності І. Кудашев вважає залежність від стартових умов, факультативність, тобто необов'язковість деяких етапів, перетин, одночасність, безперервність етапів роботи, різницю в поглядах щодо моменту закінчення роботи. Він робить висновок, що неможливо розробити й оформити у вигляді лінійної послідовності детального універсального набору етапів проектування [209]. Варто зазначити, що при цьому основна увага приділяється створенню тезауруса певної предметної області однією мовою, ми ж ставимо за мету створення двомовного інформаційно-пошукових тезауруса.

У результаті дослідження проблеми поетапного розроблення тезауруса на основі праць таких науковців як С. Вискушенко [201], С. Гриньов [202], В. Дубичинський [203], Б. Шуневич [204], А. Баранов [152], О. Рум'янцева [206], Є. Купріянов [208], І. Кудашев [209], Р. Робертс (R. Roberts) [210] та інших ми консолідували та узагальнили підходи до технології формування інформаційно-пошукового тезауруса предметної області «Розумне місто» та сформували алгоритм, якого дотримувалися, працюючи над розробленням тезауруса «розумного міста» [75, 211, 212].

Таким чином, вважаємо, що розроблення двомовного інформаційно-пошукового тезауруса предметної області «Розумне місто» передбачає такі етапи (рис. 3.5):

1) дослідження потреб адресата, визначення меж предметної області, підбір джерельної бази, формування вимог до визначення термінів тезауруса, а також його структури;

2) формування реєстрової частини тезауруса: виділення лексичних одиниць зі словників, наукових статей, матеріалів конференцій, стандартів тощо і їх переклад відповідно українською чи англійською мовами; складання списку ключових слів; побудова класів умовної еквівалентності (для усунення синонімії, полісемії і переходу до дескрипторної інформаційно-пошукової мови, тобто приведення лексичних одиниць до заданої стандартної форми);

3) побудова двомовних словникових статей (формулювання дефініцій термінів з опрацюванням вітчизняної та зарубіжної наукової літератури – монографій, посібників, статей, матеріалів конференцій тощо); розроблення набору семантичних зв'язків та їхня актуалізація – установлення відношень між термінами (при розробленні набору семантичних відношень ми знаходили їх у текстах, що описують цю галузь, тобто намагалися вичленувати й уніфікувати ті відношення, що існують у текстах між термінами, або безпосередньо аналізували знання); впорядкування тезауруса відповідно до розробленої структури;

4) уточнення й редагування отриманого матеріалу, узгодження з користувачем.

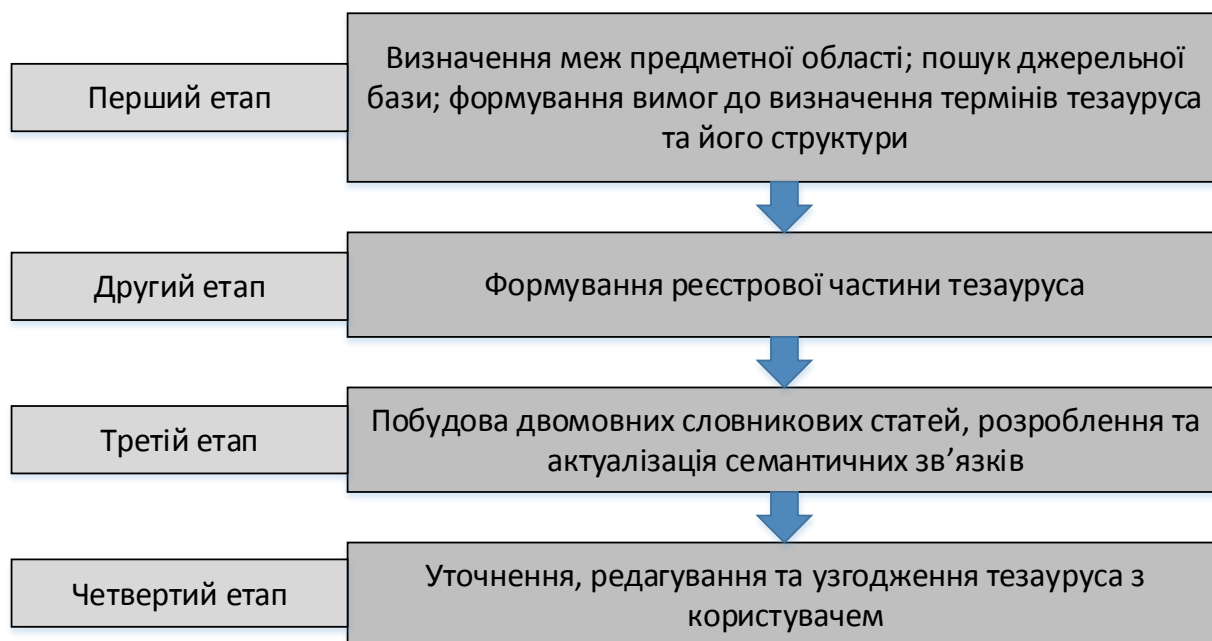


Рис. 3.5 Алгоритм укладання тезауруса предметної області «Розумне місто»

Отже, як бачимо з алгоритму, якщо перший етап роботи над тезаурусом носить, швидше, аналітико теоретичний характер, то другий та третій є повністю практичними. А четвертий етап можна розглядати як «feedback» («зворотний зв'язок») [212].

Варто акцентувати увагу на тому, що досить складно провести межі між відповідними етапами й чітко визначити кінець одного та початок іншого етапу, як зазначає Р. Робертс (R. Roberts) етапи розроблення словника не можуть бути лінійними й послідовними, навпаки, вони перетинаються та є неперервними, циклічними [210]. Це підтверджує думку про те, що розроблення тезауруса є досить складним і довготривалим процесом.

Для розроблення тезауруса предметної області «Розумне місто» формувався керований словник досліджуваної галузі, що складався з термінів, які описують поняття галузі й складають інтерес для вивчення. Керований словник – це спосіб організації знань з метою полегшення їхнього представлення й подальшого пошуку; список допустимих термінів, які

описують поняття певної галузі та можуть використовуватися для тих чи інших цілей: визначення, введення, індексації, пошуку. Всі терміни в керованому словнику повинні мати однозначне тлумачення (визначення). Керований словник складає основу класифікації, в якій передбачається використання заздалегідь визначених термінів, які були попередньо вибрані, на відміну від природних мов, де немає обмежень на словниковий запас.

Керовані словники використовуються в різних системах організації знань: схемах предметного індексування, предметних рубриках, таксономії, тезаурусах та ін. У бібліотечних та інформаційних системах керований словник є списком ретельно відібраних слів чи словосполучень, які використовуються для розмітки інформаційних одиниць (документів) для спрощення пошуку.

Особливостями керованого словника є:

- наявність множини понять, що закріпилися в досліджуваній галузі;
- терміни мають визначення та використовуються для індексації в системах організації знань (предметні рубрики, тезауруси та ін.);
- знімається питання неоднозначності поняття, що притаманне природній мові, коли одне поняття може мати різні назви і навпаки.

Систематизація понять у формі ієрархічної структури перетворює керований словник у таксономію.

Особливості таксономії:

- формально це деревоподібна структура класифікації певного набору об'єктів;
- зміст понять розкривається через взаємозв'язки (родо-видові відношення) з іншими поняттями;
- розширює семантику предметної області.

Додавання в таксономію інших відношень перетворює її у тезаурус. Тезаурус, особливо в електронному форматі, є одним з інструментів для опису предметної області.

Особливості тезауруса:

- крім відношення «рід-вид» у тезаурусі існують інші відношення, які утворюють складну структуру предметної області;
- використовуються асоціативні відношення;
- використовується відношення еквівалентності, яке дозволяє робити висновки чи є терміни семантично тотожними;
- семантика тезауруса розкривається через наявні в ньому відношення;
- електронний тезаурус є ресурсом для зберігання знань [213].

Схематичне представлення формування тезауруса предметної області «Розумне місто» подано на рис. 3.6.

Наведені нами твердження формують методологічні підходи до розгляду основних параметрів термінологічного тезауруса предметної області «Розумне місто». На вихідному етапі принциповими завданнями є визначення таких установок:

- попередній відбір лексичних одиниць для тезауруса;
- побудова класів умовної еквівалентності, тобто зведення лексичних одиниць до заданої стандартної форми;
- виявлення системи логіко-змістових відношень, які повинні встановлювати відповідні відношення між термінологічними одиницями тезауруса.

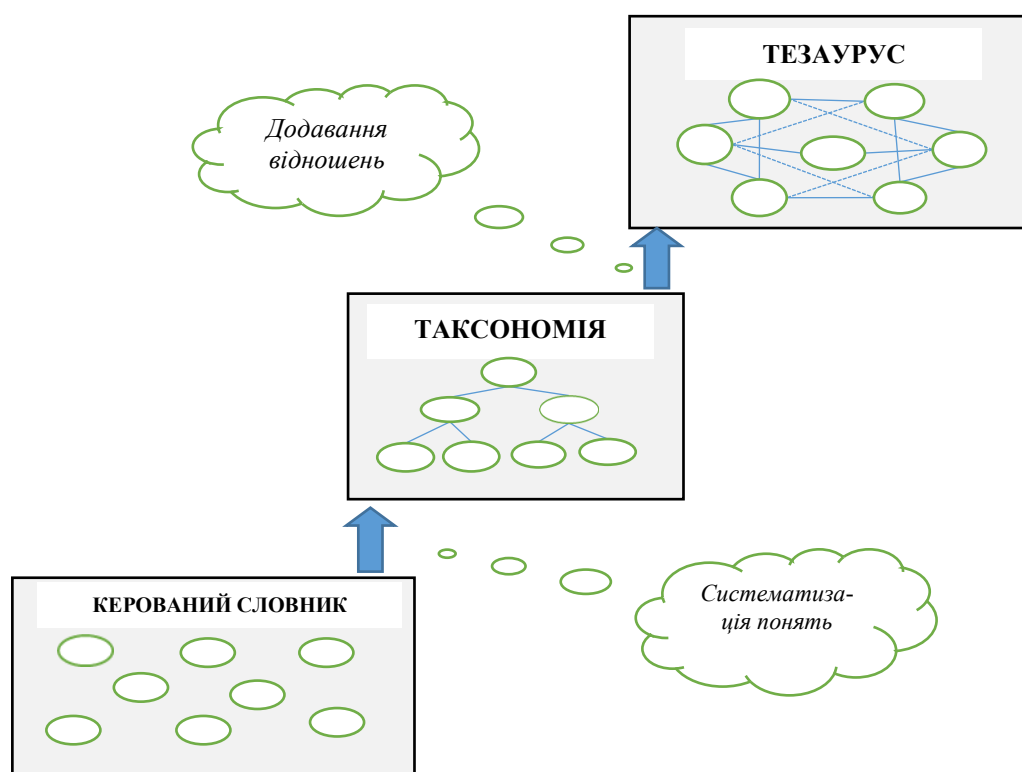


Рис. 3.6 Схематичне представлення формування тезауруса «розумного міста»

3.6 Отримання інформації для побудови тезауруса

Основним засобом спілкування є мова, яка складається із лексичних одиниць. Інтерес нашого дослідження формують лексичні елементи, що утворюють термінологічну базу предметної області аналітика та експерта. Спілкування аналітика та експерта включає два потоки: мова, якою розмовляє аналітик та мова, якою розмовляє експерт (рис. 3.7). Основу кожного потоку складає «внутрішня мова», попередній досвід, який має вплив на формування «внутрішньої мови» та невербальні компоненти спілкування.

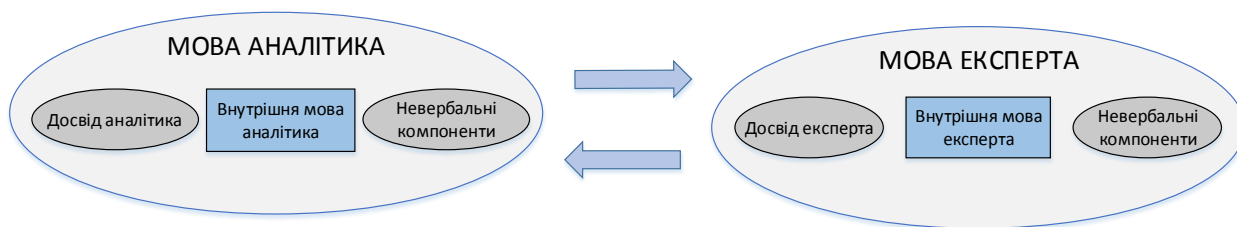


Рис. 3.7 Схема лінгвістичної взаємодії
аналітика та експерта

Відмінність між цими «цілісними системами мов» утворює «мовний бар'єр» у спілкуванні аналітика та експерта.

Таким чином, наш інтерес становлять дві мови – мова аналітика та мова експерта. Аналітик оперує термінами предметної області, які одержав з фахової літератури; загальнонауковими термінами, якими вже користується, а також побутовою мовою, якою розмовляє. Експерт володіє спеціальною термінологією предметної області, загальнонауковою термінологією, побутовою мовою, неологізмами, створеними ним за час роботи. Для успішної взаємодії аналітика та експерта необхідно виробити спільну штучну мову, яка буде складатися із фахових термінів з літератури, які отримує аналітик, та фахових термінів експерта. На цю штучну мову буде спостерігатися й вплив побутової мови та загальнонаукових термінів (рис. 3.8). У подальшому така штучна мова перетворюється в семантичну мережу понять, котра є прототипом знань предметної області [214].

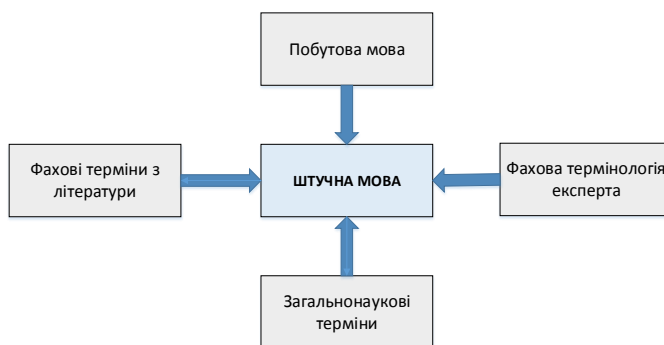


Рис. 3.8 Схема отримання штучної мови

Формування спільної штучної мови розпочинається зі створення аналітиком переліку всіх термінів, які використовує експерт, та, звичайно, уточнення їхнього значення. Практично це є формування словника предметної області. Пізніше терміни групуються та відшуковуються до них синоніми. Формування штучної мови закінчується складанням словника термінів предметної області з попереднім групуванням їх за близькістю понять (за змістом). Це перший крок структурування знань предметної області.

Доцільно акцентувати увагу на тому, що важливим на цьому етапі є те, що аналітик повинен уважно розглядати всі спеціальні терміни, намагаючись максимально зрозуміти термінологію та суть проблеми, що вирішується. Звичайно, що експертові не треба уникати вживання «складних» незрозумілих для аналітика слів та виразів, бо у такому випадку це призводить до втрати сенсу, корисно лише уточнювати їх. Таким чином, запорукою успішного отримання необхідної інформації є засвоєння аналітиком мови предметної області [214].

Спеціалісти в галузі штучного інтелекту та когнітивної психології основною особливістю природного інтелекту та пам'яті вважають зв'язаність усіх понять у певну систему. Тому для формування бази знань важливим є не словник, а тезаурус, що подає пояснення термінів як у словникових статтях, так і через зв'язки з іншими термінами. Отже, лінгвістична робота аналітика полягає в побудові таких відношень між термінами.

Лінгвістичні результати, що стосуються штучної мови та структури понять, спрямовані на створення відповідної бази знань. Проте професійний рівень деяких користувачів може не дозволити їм використати спеціальну мову предметної області в повному обсязі. Лінгвістичними причинами таких невдач може бути спілкування з допомогою різних понять (як результат слабкої підготовки аналітика); неправильна інтерпретація термінів (через відсутність зворотного зв'язку з експертом); не враховані відмінності між рівнем знань експерта й користувача [214].

Іншим способом отримання інформації є гносеологічний аспект спілкування з експертом.

Гносеологія – це розділ філософії, в якому розглядається теорія відображення реальності в свідомості людини (теорія пізнання). Вона подає таку інтерпретацію процесу отримання інформації: дійсність відображається у свідомості експерта, у свою чергу свідомість та досвід експерта пояснюються свідомістю аналітика, що є основою для побудови поля знань. Процес пізнання створює внутрішнє представлення навколишнього світу у свідомості людини. Таким чином, йде мова про перетворення експертного знання й теоретичного, тобто книжкового досвіду у поле знань, що є матеріалізацією моделі світу аналітика.

Якщо вважати, що аналітик видобуває лише частину знань експерта, то його завдання полягає в тому, щоб ця частина знань відповідала знанням експерта та якомога повніше відображала їх, оскільки неповнота знань не дає можливості повного описання предметної області.

Методологічна структура пізнання охоплює послідовність таких етапів: описування та узагальнення фактів; встановлення відношень, побудова моделі; пояснення й передбачення явищ з позиції аналітика.

Результатом спілкування аналітика з експертом є описування та узагальнення фактів. Найчастіше на цьому етапі просто збирають «факти» – визначення понять, відношень, необхідних для описування процесів вирішення завдань у певній предметній області.

Для побудови моделі предметної області, що формується в уявленні суб'єкта, необхідна спеціалізована мова, яка буде описувати та конструювати такі моделі світу, які виникають у процесі мислення. При створенні такої мови використовується категоріальний апарат певної предметної області та формально-знакові засоби математики та логіки. Таким чином, поле знань, яке аналітик опише напівформалізованим способом, може бути першим кроком до створення штучної мови.

Завершальним етапом структури пізнання є пояснення та передбачення моделей, це одночасно й частковий критерій істинності отриманого знання. Пояснення будь-яких явищ певної предметної області може відбуватися на ґрунті виявленої системи знань експерта за умови, що вона є повною та об'єктивною.

Для отримання інформації для побудови тезауруса було використано комунікативні та текстологічні методи. Комунікативні методи отримання знань включають різні види контактів аналітика з безпосереднім джерелом знань, яким є експерт. Текстологічні методи – це методи отримання інформації із документів (методик, посібників, інструкцій та ін.) та спеціальної літератури (статей, монографій, підручників тощо). Зазвичай комбінувалися різні методи, наприклад, студіювалася література, а потім відбувалося спілкування з експертами або ж навпаки – спілкувалися з експертами, після чого опрацьовували потрібну літературу для уточнення або вивчення необхідного поняття.

Комунікативні методи поділяються на активні та пасивні. В активних методах головна роль відводиться фахівцеві, котрий різними способами активно контактує з експертом – діалог, бесіда, анкетування і т. д. Активні методи в залежності від кількості експертів можна поділити на дві групи. Якщо експертів більше одного, то крім індивідуальних контактів із кожним з них доречно користуватися і методами групових обговорень предметної області, за допомогою яких активізується мислення учасників дискусії та виявляються своєрідні аспекти їхніх знань. На сьогодні індивідуальні методи залишаються провідними.

У пасивних методах (спостереження, протоколювання «думок вголос», лекції та ін.) основна роль у процесі отримання інформації передається експерту, аналітик лише записує роздуми експерта. Використовуючи пасивні методи, аналітику потрібно чітко аналізувати інформацію, яку подає експерт, та виявляти у ній важливі фрагменти знань.

На вибір методу отримання інформації впливають три фактори [215]: особистість фахівця, особистість експерта та характеристика предметної області. За психологічними характеристиками виділяють три типи людей: мислителі (пізнавальний тип), які орієнтовані на інтелектуальну роботу, навчання, теоретичні узагальнення; співрозмовники (емоційно-комунікативний тип) – товариські, відкриті люди, готові до співпраці; практики (практичний тип) надають перевагу діям, а не розмовам, добре реалізують задуми інших, спрямовані на результативність роботи.

Для характеристики предметних областей доцільно запропонувати таку класифікацію: добре документовані; середньо документовані; слабо документовані (такою є досліджувана предметна область «Розумне місто»). Цей поділ пов'язаний зі співвідношенням двох видів знань – експертні «особисті» знання та «загальні» знання в певній конкретній галузі, матеріалізовані у книжках.

Предметні області за критерієм структурованості знань можна розділити на добре структуровані (із чіткою усталеною термінологією); середньо структуровані (із термінологією, що визначилася, теорією, що розвивається, явними зв'язками між явищами); слабо структуровані (із розмитими визначеннями, багатою емпірикою, прихованими взаємозв'язками). Під структурованістю розуміється ступінь теоретичного осмислення і виявлення основних закономірностей та принципів, що діють у предметній області.

Методи отримання інформації, що базуються на вивченні спеціальних текстів (підручників, монографій, статей, методик та інших носіїв фахових знань) називають текстологічними методами.

Завдання отримання інформації із текстів полягає в розумінні змісту тексту. Текст природною мовою є інструментом для автора, провідником змісту, тобто текст не містить та не передає зміст. Задум та знання автора лежать у вторинній, змістовій, структурі. Можна виділити дві змістові структури: зміст, який подає автор (це його модель світу), та зміст, який

розуміє читач, тобто аналітик (рис. 3.9) у процесі інтерпретації. При цьому словесною формою стає текст природною мовою як результат вербалізації.



Рис. 3.9 Отримання інформації зі спеціальних текстів

Складність отримання знань зі спеціальних текстів полягає в тому, що не можуть співпасти знання, що утворюють зміст, який подає експерт, та зміст, який розуміє аналітик, тому що перший утворюється завдяки єдності уявлень, потреб, зацікавлень та досвіду автора й лише деяка частина яких відображається в тексті (книзі). Відповідно зміст, який розуміє аналітик, виникає в процесі інтерпретації тексту природною мовою. Отже, два аналітики із одного тексту отримують дві різні моделі змісту цього тексту.

Виникає завдання: з'ясувати, яким чином можна досягнути максимальної тотожності між змістом, який подає експерт, та змістом, який розуміє аналітик.

Будь-який науковий текст містить два основні компоненти: первинний матеріал спостережень та систему наукових понять під час створення тексту. Крім цього в тексті обов'язково є суб'єктивні погляди автора (результат його власного досвіду), деякі «загальні місця» та запозичення із інших джерел (статей, монографій і т. д.). При цьому всі компоненти заглиблені в мовне середовище.

Складність пояснення наукових текстів полягає в тому, що видобуваючи інформацію з тексту, аналітик, змушений вирішувати завдання декомпозиції для виділення важливих фрагментів та будь-який текст набуває змісту лише у контексті. Існує мікро- та макроконтексти. Мікроконтекст – це найближче оточення тексту. Наприклад, слово набуває змісту у контексті речення, речення у контексті абзацу, абзац у контексті глави і т. д.

Макроконтекст – це знання, пов'язані з предметною областю, які безпосередньо не вказані у тексті. Тобто, будь-які знання набувають змісту в контексті метазнання [215].

Головним у процедурі отримання інформації є розуміння тексту, тобто спроба відтворити семантичну структуру змісту, який подає експерт у процесі формування моделі змісту, що розуміє аналітик. Це є перший крок структурування інформації.

Основними етапами розуміння тексту є: формування попереднього положення про зміст тексту; визначення значень нових слів-термінів; виникнення загальної гіпотези про зміст тексту; уточнення термінів; формування змістової структури тексту, завдяки встановленню внутрішніх зв'язків між окремими ключовими словами та фрагментами, а також завдяки утворення абстрактних понять, що узагальнюють конкретні фрагменти знань; корегування загальної гіпотези відносно фрагментів знань, що містяться в тексті від часткового до цілого; прийняття основної гіпотези, тобто формування змісту, який розуміє аналітик [215].

Аналізуючи текст, важливо знаходити внутрішні відношення між окремими елементами тексту та поняттями. Виділяють експліцитні (явні) та імпліцитні (приховані, внутрішні) зв'язки у тексті. Експліцитні відображаються в зовнішньому поділі тексту та ділять його на параграфи за допомогою перерахування компонентів ввідних слів. Імпліцитні зв'язки викликають основне ускладнення при розумінні тексту [216]. У свідомості суб'єкта, що працює з текстом, семантична структура цього тексту утворюється за допомогою знань про мову, світ, а також загальних знань у тій предметній області, яку розкриває текст. Таким чином, для правильного прочитання спеціального тексту потрібна певна підготовка. Такою підготовкою є вибір базового переліку літератури, яка поступово введе аналітика в предметну область. Це можуть бути підручники, глави та фрагменти із монографій, популярні видання. Лише після ознайомлення із такою літературою доцільно ознайомлюватися зі спеціальними текстами.

Отже, на процес розуміння та зміст, який розуміє аналітик, впливають інформація, взята із тексту; попередні знання аналітика про дану галузь; його загальнонаукова ерудиція та власний досвід.

Основою індивідуального процесу розуміння тексту є поділ його на частини (змістові групи) та стиснення вмісту кожної змістової частини в ключове слово. Таке стиснення (компресія) тексту у формі системи ключових слів, які відображають головний зміст тексту, є зручною методичною основою для отримання інформації з текстів. Цим ключовим словом можуть виступати іменник, прикметник або словосполучення. Набір ключових слів складає набір опорних точок. Саме за цими опорними точками й розгортається текст при кодуванні в пам'яті й розуміється при декодуванні. Це буде семантичне ядро цілісності.

Існує три види текстологічних методів – аналіз спеціальної літератури, аналіз підручників та аналіз методик. Вони відрізняються за співвідношенням фахових та фонових знань. Найпростішим методом є аналіз підручників. Оскільки логіка викладення у підручниках зазвичай відповідає логіці предмета, тому макроструктура такого тексту буде більш важливою в порівнянні зі структурою тексту будь-якої спеціальної статті. Аналіз методик ускладнено відсутністю фонових знань, що полегшують розуміння для неспеціалістів, подано лише стислість викладення і практично відсутні коментарі. Тому для практичної роботи було використано комбінацію перерахованих методів.

Аналіз тексту для отримання та структурування інформації складається з таких етапів:

- формування базового списку літературних джерел для ознайомлення з предметною областю;
- вибір тексту для отримання інформації;
- ознайомлення з текстом, визначення значень незнайомих слів (консультації зі спеціалістами або використання довідкової літератури);
- висунення гіпотези про макроструктуру тексту;

- компресія тексту (уважне прочитання, виписуванням ключових слів);
- розроблення макроструктури тексту у формі графу (визначення зв'язків між ключовими словами);
- формування поля знань на основі макроструктури тексту.

На рис. 3.10 подано загальну схему отримання інформації для побудови тезауруса.

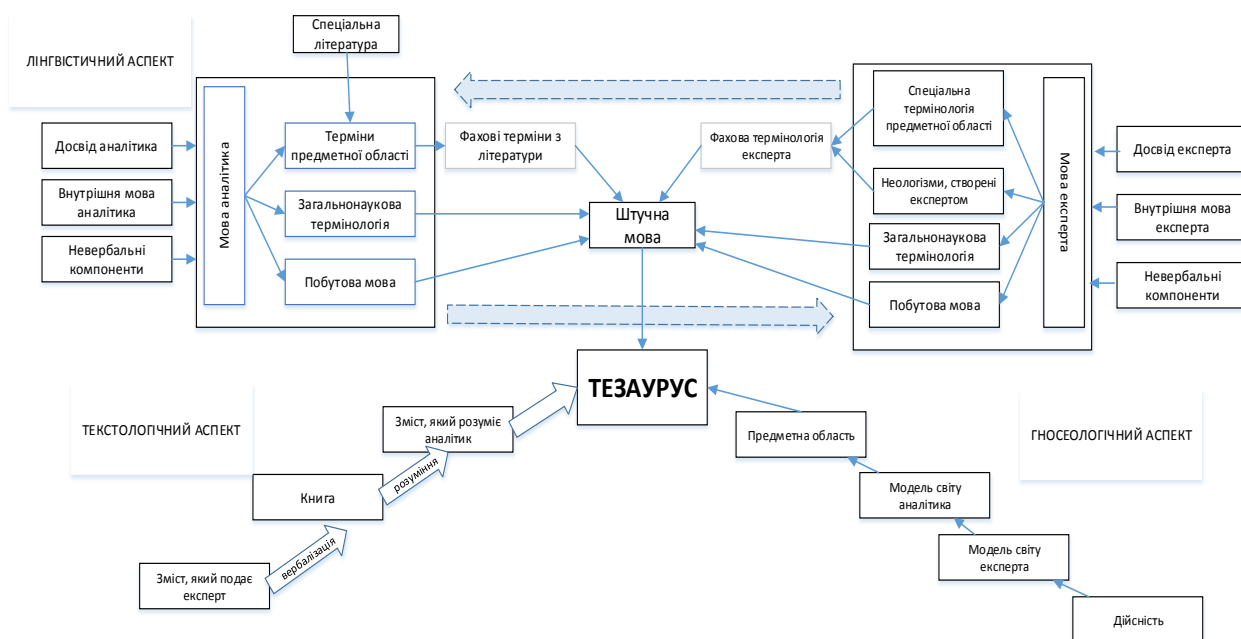


Рис. 3.10 Схема отримання інформації для побудови тезауруса

3.7 Тезаурус як засіб лінгвістичного забезпечення автоматизованих бібліотечних інформаційних систем

Зростання ефективності інформаційного пошуку в бібліотеках досягається низкою чинників, одним із яких формування адекватного пошукового образу. Технології створення пошукового образу документа ґрунтуються на сучасному лінгвістичному забезпеченні, використанні інформаційних та комунікаційних технологій, які є інструментами підвищення пошукових можливостей бібліотечних інформаційних систем. Лінгвістичне забезпечення складається із множини елементів: комплексу

мовних засобів, лінгвістичних процесорів, лінгвістичних банків даних. Лінгвістичне забезпечення в першу чергу призначене для опрацювання та пошуку документів, поданих природною мовою [217]. При цьому комплекс мовних засобів формується із інформаційно-пошукових мов, а лінгвістичні процесори – із алгоритмів і програм, що використовуються для автоматизованого опрацювання текстів. Лінгвістичне забезпечення для бібліотечних інформаційних систем є комплексом компонентів – інформаційно-пошукових мов різного типу, які є взаємопов'язаними та доповнюють один одного, а також правил та методів їхнього функціонування.

Формування та розвиток повноцінних інформаційних ресурсів та технологій пошуку в них є важливим завданням сучасних бібліотек та архівів. Для здійснення цих завдань використовуються автоматизовані інформаційні системи, які дозволяють значно розширити можливості пошуку документів та отримання релевантної інформації. Забезпечення доступу до інформаційних ресурсів виконується засобами лінгвістичного забезпечення, що є невід'ємною частиною інформаційно-пошукових систем. Від якості таких засобів залежить ефективність пошуку й, безумовно, імідж книгозбірні та архіву. Лінгвістичне забезпечення складає комплекс мовних засобів, які необхідні для опрацювання документів і запитів, проведення ефективного пошуку.

Значний внесок у розроблення теоретичних і прикладних питань автоматизації інтегрованої бібліотечної та архівної систем, лінгвістичного забезпечення бібліотечних пошукових систем, розвитку класифікаційних схем, електронних ресурсів бібліотек здійснили Ф. Воройський [218], Н. Гендіна [219], К. Зайцева [220], Е. Сукіасян [221], К. Бардієр [222], К. Лобузін [162, 223], О. Збанацька, Т. Вилегжаніна [224], Л. Дубровіна, О. Онищенко [225], В. Широков [190] та ін.

Лінгвістичні засоби є інтерфейсом між природною мовою та формальними пошуковими механізмами інформаційно-пошукової системи

бібліотеки. Вони включають мови представлення даних в інформаційно-пошукових системах, які визначають архітектуру, синтаксис і семантику подання інформації в базах даних інформаційно-пошукових систем, та інформаційно-пошукову мову, на якій користувач звертається до системи для отримання необхідної інформації [226].

Лінгвістичне забезпечення сприяє індексуванню документів та формуванню запитів, ефективному пошуку в базах даних, формуванню баз даних бібліотек та архівів, забезпечує сумісність електронних каталогів різних бібліотек щодо обміну даними. Саме тому виникають високі вимоги до лінгвістичного забезпечення бібліотечних та архівних систем.

Забезпечити потреби користувачів бібліотек та архівів у отриманні релевантної інформації на якісно новому рівні дозволяє використання автоматизованих інформаційно-пошукових систем, електронних каталогів, бібліографічних та архівних баз даних.

Одним із показників якості інформаційного бібліотечного ресурсу є не лише дотримання державних стандартів у процесі формування бібліографічного запису, а й можливість результативного пошуку документів. Швидкість, повнота, ефективність пошуку та отримання інформації залежить від якості формування пошукового образу документа та адекватного відображення змісту документа в інформаційно-пошуковій системі. Кожен документ забезпечується пошуковим образом, в якому коротко подається основний смисловий зміст документа, і який створюється фахівцем із застосуванням правил індексування. У такій самій формі формується й пошуковий образ запиту, який створюється споживачем інформації, здійснюючи пошук потрібного документа в пошуковому масиві та використовуючи правила пошуку. Таким чином, процес інформаційного пошуку може бути зведеним до простого порівняння пошукового образу документа із заданим пошуковим образом запиту, тобто запит порівнюється не з текстом документа, а з його пошуковим образом. Пошук релевантної інформації має забезпечуватися якістю змістовної обробки документа, тобто

зміст знайденого документа має відповідати сформульованому інформаційному запиту користувача й максимально задовольняти потреби користувача в релевантній інформації. Отже, пошуковий запит та пошуковий образ повинні задаватися стандартизованими термінами.

Важливим процесом у створенні бібліографічних записів є індексування, основним лінгвістичним засобом якого виступає інформаційно-пошукова мова – штучна мова, призначена для передачі змісту, формальних ознак документів та інформаційних запитів, лексичними одиницями якої є терміни індексування. Під час індексування «прогнозують майбутні запити користувача, а класифікаційне рішення ухвалюють з урахуванням можливих шляхів пошуку документа» [227]. Від якості індексування залежить ефективність використання інформаційних ресурсів бібліотек та архівів, оскільки індексування є основним засобом розкриття змісту документа і відповідно документних потоків, що складають фонди бібліотек та архівів. Новим та перспективним видом індексування є координатне індексування, що передбачає багатоаспектне подання основного змісту документа або змісту інформаційного запиту безліччю ключових слів і / або дескрипторів, сукупність яких складають пошуковий образ документа. Відповідно виділяється дескрипторна інформаційно-пошукова мова та мова ключових слів.

Дескрипторна інформаційно-пошукова мова використовується для координатного індексування документів та інформаційних запитів за допомогою дескрипторів. Основу дескрипторної інформаційно-пошукової мови складають лексичні одиниці в алфавітному порядку, які виражаються словами чи словосполученнями. Призначена ця мова для контрольованого індексування – контролю термінології (вилучення неоднозначних слів природної мови), якою послуговуються в процесі індексування. Тобто, використовуються лексичні одиниці вербальної інформаційно-пошукової мови, які отримуються не з текстів, а зі спеціальних словників-тезаурусів, а також лексичні одиниці, які є іменами класів синонімічних чи близьких за

змістом ключових слів. Дескрипторам, на відміну від ключових слів, притаманна смислова однозначність та наявність відношень між ними.

Основним лінгвістичним засобом індексування (при координатному індексуванні) можуть виступати інформаційно-пошукові тезауруси, саме дескриптори, що пов'язані між собою парадигматичними відношеннями, використовуються для індексування. Тезауруси полегшують відбір лексичних одиниць відповідно до запиту. Перевагою використання тезаурусів можна вважати їхню ієрархічну структуру. Побудова тезауруса від загального до конкретного дозволяє забезпечити значно більшу кількість різноманітних семантичних зв'язків між лексичними одиницями. Переваги тезауруса в порівнянні зі списком ключових слів чи термінів очевидні, оскільки в списках ключових слів чи термінів не відображені смислові зв'язки між термінами, при цьому алфавітний порядок розташування термінів посилює цю роз'єднаність: близькі поняття віддалені один від одного. Використання тезауруса в інформаційно-пошукових системах підвищує якість аналізу тексту і повноту пошуку інформації, дозволяючи розширити запит синонімічними поняттями.

Тезаурус є універсальним за змістом практичним інструментом [228], який можна використовувати для індексування. Завдання, що повинні бути вирішені з допомогою тезаурусів при індексуванні документів – це уніфікація ключових слів, які відбираються з документів. Іншими словами, потрібно створити пошукові образи даних документів у вигляді дескрипторів для подальшого їхнього пошуку. Кожен дескриптор та аскриптор (нондескриптор) мають словникові статті, які відрізняються від словникових статей тлумачних словників. В останніх подано пояснення терміна, тоді як у тезаурусах крім визначення подано парадигматичні відношення з іншими термінами.

Розглянемо технологію індексування документів у бібліотеці з використанням інформаційно-пошукового тезауруса, що подано на рис. 3.11 (алгоритм побудований на основі аналізу стандартів ДСТУ 2395-2000, ГОСТ

7.66-92 [229, 230])). При такому типі індексування потрібно вивчити документ і скласти перелік істотних для його змісту понять з урахуванням специфіки інформаційно-пошукової мови. Розглянути перше поняття. Віднайти у тезаурусі термін, який позначає це поняття. Якщо такий термін виявлено, то з'ясувати чим він поданий у тезаурусі. Якщо це аскриптор, то замінити його відповідним дескриптором, після чого розглянути усі його відношення. Якщо не виявлено кращого дескриптора, то записати його у пошуковий образ. При з'ясуванні, що аналізоване поняття буде краще розкрито за допомогою іншого дескриптора, тоді замінити вибраний дескриптор новим і записати вже його в пошуковий образ. Після чого потрібно розглянути інші поняття й звернутися до тезауруса, щоб віднайти їх у ньому та пройти ці ж самі етапи. Якщо у тезаурусі не віднайдено терміна, який позначає певне поняття, потрібно знайти близькі за значенням терміни, встановити термін, що означає поняття та ввести його в тезаурус.

Вузькогалузеві інформаційно-пошукові тезауруси сприяють наповненню наукових термінологічних баз. Вони є зручними як для індексаторів, так і для споживачів інформації. Використання координатного індексування (ключових слів і / або дескрипторів) дає можливість забезпечити повне (відношення кількості знайдених релевантних документів до загальної кількості документів у базах даних) й точне (співвіднесення релевантних документів, знайдених у інформаційно-пошукових системах, та знайдених документів) відображення опрацьованої інформації.

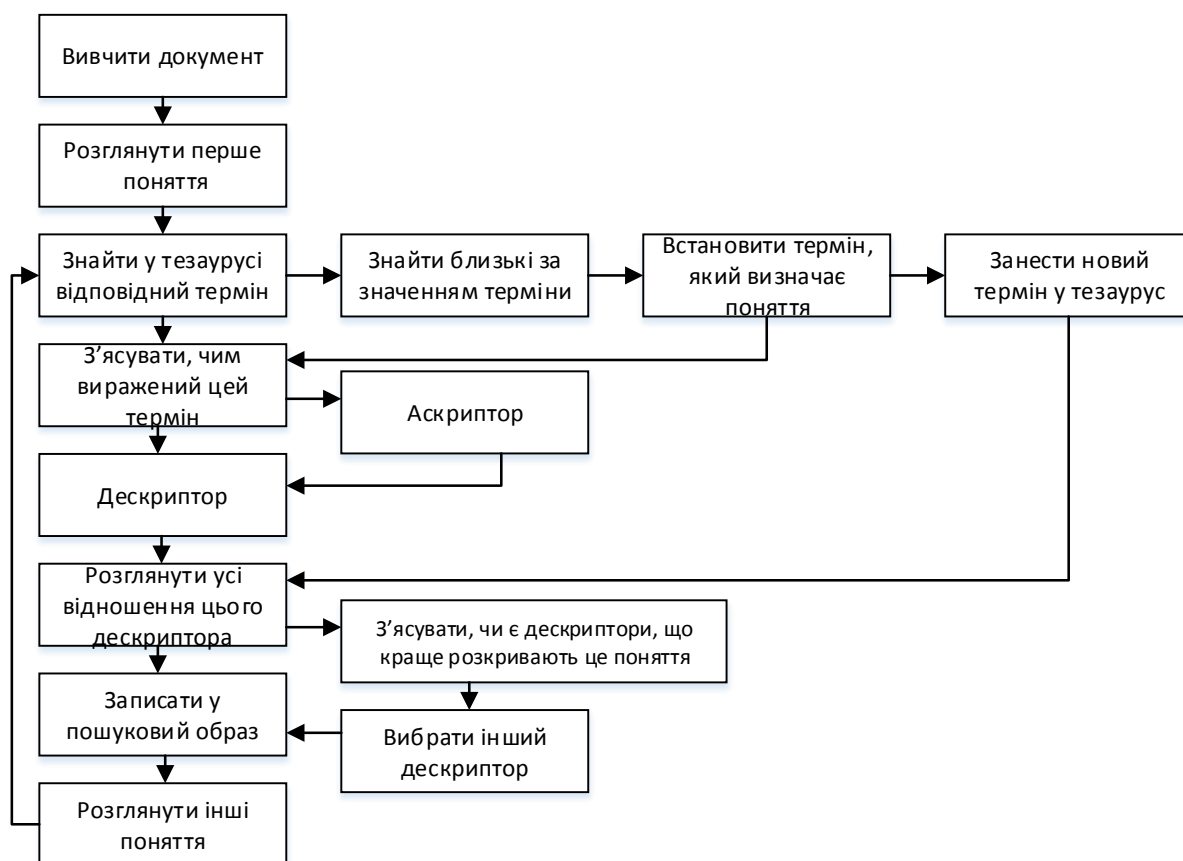


Рис. 3.11. Блок-схема алгоритму індексування документів з використанням інформаційно-пошукового тезауруса

І. Зарванська [231] координатне індексування вважає базовим засобом індексування та пошуку документів у електронному середовищі бібліотек. Вона висловлює думку про те, що об'єднати бібліотеки та інтернет спільною інформаційно-пошуковою мовою дескрипторів може дієвий, потужний, демократичний інструмент точного та повного відображення змісту документа. Предметні рубрики, які є одним із пошукових елементів електронного каталогу, не можуть бути ефективним способом предметного пошуку, якщо при їхньому формуванні не використовувалася спеціалізована лексика. Тому для формування предметних рубрик коректно використовувати тезауруси (поряд з класифікаційними системами), для забезпечення якісного тематичного пошуку в електронних каталогах. Предметні рубрики, сформовані з використанням тезаурусів, значно полегшують роботу систематизатора та користувача електронного каталогу,

дозволяють задовольняти його інформаційні потреби. Головне завдання предметизації полягає в тому, щоб визначити предмет як основну тему розгляду та розкрити основні аспекти його характеристики, зв'язки з іншими предметами, формулювання змісту документа у вербальній формі як предметні рубрики [232].

Часто документ з одною і тою ж назвою та змістом може мати різні предметні рубрики, що ускладнює тематичний пошук, який є для користувача основним, оскільки максимально наближений до природної мови. Тому важливою стає уніфікація електронної предметизації: створення результативних засобів навігації в базах даних бібліотеки чи архіву – лінгвістичного забезпечення інформаційно-пошукової мови, предметних рубрик. Пошук за предметними рубриками зводиться до того, що кожна предметна рубрика є закінченою думкою та подає точну інформацію про документ, а система предметних рубрик, включаючи усі підзаголовки, розкриває більшість аспектів предмета. Саме предметні рубрики найточніше розкривають зміст документа, коли його назва не є інформативною. Складаючи предметні рубрики важливо правильно визначити тему документа й правильно її сформулювати відповідно до вимог інформаційно-пошукових мов, тобто за найменшої кількості слів вона повинна якнайточніше та якнайглибше розкривати документ.

Правильно та грамотно складені предметні рубрики дають адекватне відображення змісту документа, забезпечують результативний тематичний пошук навіть у тому випадку, коли користувач не володіє необхідними знаннями з певної тематики, що призводить до недостатньої чіткості визначення інформаційних потреб. Предметні рубрики повинні бути цілісною системою, що складається з елементів, зв'язаних між собою. Тому неусунені синоніми, омоніми та полісемія значно знижують ефективність пошуку необхідної інформації. Використання тезаурусів при формулюванні предметних рубрик дає можливість гарантувати однозначність пояснення термінів індексування, мати чітку семантику. Коректно складені предметні

рубрики забезпечують простоту, зручність та швидкість тематичного пошуку в електронних каталогах, простоту індексування, відображають сучасний стан терміносистеми певної галузі, дають можливість повно й точно подати зміст документа.

Таким чином, побудовані авторитетні файли предметних рубрик з використанням тезаурусів, можуть бути необхідним процесом індексування документів, що підвищує якість баз даних, розширення можливостей пошуку документів, засобом обміну інформацією та доступу до баз даних бібліотек та архівів.

Надзвичайно важливим фактором предметизації є дотримання чіткого формулювання термінів для представлення змісту документа й для забезпечення максимальної результативності тематичного пошуку, що має важливе значення для ефективної взаємодії книгозбірень.

Прийнявши виклик часу, бібліотеки та архіви переходять на on-line обслуговування користувачів, відчуючи гостру конкуренцію з боку інтернету. «Вони ще не завжди можуть запропонувати користувачу ранжування, кешування, багатомовність одного й того самого джерела, фолксономію, оперативне обробляння документів, простоту пошуку (в основному використання ключових слів у вигляді токенів), великі масиви оцифрованої інформації, інтерактив, наочність. Інтернетівська інформація має й інші характеристики: повтори, інформаційний шум, випадання (не все із запропонованого вдається зчитати і правильно вибрати), механічність індексування, недостовірність, комерційні витрати, неточність, невідповідність тексту перекладу тощо» [231].

Тому правильно організовані електронні каталоги бібліотек та архівів можуть зекономити час, використаний користувачем-науковцем на пошук необхідних документів, забезпечити релевантність отриманої інформації. Саме використання тезаурусів дозволяє уникнути властивостей природної мови, що не дозволяють їй виступати ефективною інформаційно-пошуковою мовою: синонімії, омонімії, багатозначності, використання скорочених

варіантів словосполучень, абревіатур на основі вибудованих відношень між лексичними одиницями. Бібліотекарі можуть створити якісні бібліографічні записи, певні метадані та пошукові образи документів.

Укладання в бібліотеках та архівах електронних каталогів та баз даних формує потребу обміну інформацією, що накопичилася, для поширення своїх баз знань і використання баз знань інших установ; ведення зведених каталогів на регіональному, національному чи міжнародному рівнях; видання спільних бібліографічних покажчиків. Проте усі бібліотеки використовують різні автоматизовані бібліотечні системи, різні формати представлення баз знань з різними мовами індексування, різні методики змістового опрацювання документів у бібліотеках. Використання різних лінгвістичних засобів знижують ефективність обміну інформацією [233] та отримання релевантних даних. Тому введення міжнародних стандартів для укладання одномовних та багатомовних тезаурусів повинно уніфікувати лінгвістичне представлення метаданих у базах даних. Як зазначається у роботі Ю. Селіванової, В. Войскунського [233] одним із основних шляхів вирішення вищезазначених проблем є створення національної системи тезаурусів. Так, наприклад, у роботі О. Збанацької [234] пропонується концепція системи тезаурусів, згідно з структурою Національного архівного фонду України: макротезаурус (Національний архівний фонд України) – мідітезауруси (архівні установи) – мінітезауруси (архівні фонди) – мікротезауруси (одиниці зберігання). З яких макротезаурус повинен містити загальні поняття усіх тематичних галузей, відношення між цими галузями, підтримуватися одною установою, бути в одному вигляді та базовим для усіх бібліотек та архівів України. Лексика макротезауруса є основою для мідітезаурусів, що розробляються окремою архівною установою на свої фонди. У свою чергу обидва типи тезаурусів є основою для побудови мінітезаурусів, що розробляються на окремі архівні фонди. Мінітезауруси, що розробляються на окрему справу, будуються на лексиці трьох попередніх тезаурусів. Побудова

такої системи тезаурусів задовольняє вимогу сумісності завдяки ланцюговому використанню лексичних одиниць тезауруса вищого рівня.

Використання інформаційно-пошукових тезаурусів у бібліотечних установах сприяє якісному індексуванню й пошуку документів, веденню авторитетних файлів та їхнього використання для каталогізації та пошуку документів, функціонуванню електронних каталогів і баз даних книгозбірень, архівів та організації ефективного пошуку через інтернет.

Висновки до Розділу 3

У третьому розділі досліджено розвиток тезаурусів у історичному аспекті їхнього формування. Акцентовано увагу на тому, що історія виникнення тезаурусів сягає коренями праць великих мислителів стародавніх часів і пов'язані тезауруси були з фундаментальними проблемами пізнання, відображення картини світу та закономірностями сприйняття світу. Тезауруси в інформаційному аспекті почали функціонувати у XX столітті, коли стали розглядатися як засіб для формування пошукових образів документів і пошукових запитів. Однак сьогодні розроблення тезаурусів є актуальним, так як у зв'язку з процесами глобалізації та розширення інформаційного простору виникає необхідність оптимізації пошукових процесів, їхнього удосконалення.

Досліджено поняття «тезаурус», яке зумовлено полярністю поглядів дослідників на його феномен. У сучасній науці подаються різні підходи до визначення тезауруса. Перший підхід подає тезаурус як спосіб відображення термінологічної системи лінгвістики, при другому підході тезаурус розглядається як інструмент індексування та інформаційного пошуку, який забезпечує формалізоване представлення інформації в інформаційних системах.

Проаналізовано поняття «інформаційно-пошуковий тезаурус» у різних дослідженнях та обґрунтовано трактування терміна «інформаційно-

пошуковий тезаурус предметної області» як сукупність структурованих знань, поданих у вигляді набору дескрипторів та аскрипторів, які пов'язані між собою певними відношеннями, що відображають семантичні зв'язки між поняттями певної предметної області. Наявність зв'язків між термінами створює передумови для використання тезаурусів у бібліотеках у процесах індексування документів та інформаційного пошуку.

Подано міжнародні та вітчизняні стандарти, що регламентують формат побудови інформаційно-пошукових одномовних та багатомовних тезаурусів.

Розглянуто формальні та семантичні вимоги до визначень термінів тезауруса, оскільки при побудові тезауруса маємо справу не із загальновживаною лексикою, а з термінами. Тому для створення якісного тезауруса потрібно дотримуватися низки практичних рекомендацій, щодо подання визначень термінів у описовій формі.

Важливою особливістю тезауруса є його здатність завдяки своїй спеціальній структурі відображати існуючі відношення між включеними в нього терміноодинацями. На відміну від тлумачного словника тезаурус дозволяє з'ясувати значення одиниці не стільки за допомогою визначення, скільки за допомогою співвіднесення змісту концепту з іншими одиницями тезауруса, завдяки цьому тезаурус може використовуватися в системах штучного інтелекту.

Головним призначенням інформаційно-пошукового тезауруса є сприяти пошуку та відбору релевантної інформації. Таким чином тезаурус предметної області виступає інструментом для аналізу й конструювання лексичної системи, класифікації й зберігання термінологічних даних, опрацювання інформації в пошукових системах. За допомогою тезауруса можна сформувати уніфікований пошуковий образ документа й пошуковий образ запиту користувачів, що підвищує ефективність пошукових процесів у автоматизованих інформаційних системах, зокрема в автоматизованих бібліотечних інформаційних системах.

Консолідовано підходи до технології формування тематичного інформаційно-пошукового тезауруса та сформовано алгоритм, якого дотримувалися, працюючи над розробленням тезауруса «розумного міста». Процес побудови тезауруса проходить наступні основні етапи: відбір лексичних одиниць, побудова класів умовної еквівалентності, встановлення семантичних відношень між термінами. Ця концепція розроблення двомовного тезауруса ґрунтується на найбільш суттєвих поглядах різних науковців на це питання. Варто зазначити, що більшість дослідників указують на поетапний характер укладання будь-якого словника, включаючи і тезаурус, але з різною кількістю етапів. На нашу думку, розбіжність щодо трактування кількості етапів укладання тезаурусів пояснюється наявністю в кожного науковця свого рівня деталізації.

Розглянуто особливості отримання інформації для побудови інформаційно-пошукового тезауруса для використання в бібліотеці, а саме лінгвістичні та гносеологічні аспекти спілкування з експертом, що сприяють отриманню інформації за допомогою комунікативних та текстологічних методів. Основним при цьому є розуміння змісту документа, що дозволяє здійснити компресію тексту у вигляді ключових слів, які передають зміст тексту, та слугує зручною методичною основою для проведення текстологічних процедур отримання інформації.

Розглянуто тезаурус як засіб лінгвістичного забезпечення бібліотечних та архівних інформаційних систем. Саме лінгвістичні засоби є інтерфейсом між природною мовою та інформаційно-пошуковою системою. Від лінгвістичних засобів тематичного пошуку залежить ефективність використання інформаційних ресурсів бібліотек та архівів у контексті інформаційного супроводу науково-дослідної роботи користувачів. Створення інформаційних ресурсів бібліотек генерує потребу використання коректного лінгвістичного забезпечення для АБІС, що сприятиме ефективному формуванню метаданих для пошукового образу документа. Тезауруси слугують одним із засобів лінгвістичного забезпечення в процесах

індексування, формування предметних рубрик для електронних каталогів бібліотек. Саме від їхньої якості залежить ефективність інформаційного пошуку, а в подальшому і релевантність інформаційного забезпечення міждисциплінарних наукових досліджень.

У третьому розділі дисертації подано основні результати наукових праць, включених до Списку публікацій здобувача за темою дисертації (пп. 1, 2, 4, 7, 12, 16, 17, 19, 21, 22) та представлених у Списку використаних джерел: [134], [212], [163], [75], [211], [162], [164], [177], [154], [214]).

РОЗДІЛ 4

ДВОМОВНИЙ ТЕЗАУРУС ПРЕДМЕТНОЇ ОБЛАСТІ «РОЗУМНЕ МІСТО» ЯК ІНСТРУМЕНТ ВИРІШЕННЯ ІНФОРМАЦІЙНО-ПОШУКОВИХ ЗАДАЧ У БІБЛІОТЕКАХ

4.1 Підготовка масиву термінів для побудови двомовного інформаційно-пошукового тезауруса «розумного міста»

Сучасний світ з його швидкостями, великою кількістю інформації, глобальною комп'ютеризацією, розвитком міжкультурної комунікації диктує нові умови для пошуку, отримання, перероблення і зберігання знань, що особливо актуально в час формування нових соціально-економічних структур – соціополісів на основі смарт-рішень. Однією з концепцій такого розвитку є «розумне місто» – модель розбудови міст на основі використання сучасних інформаційних та комунікаційних технологій.

Зі збільшенням потоку інформації в «розумних містах» з'являється необхідність пошуку нових способів її зберігання, подання, формалізації, пошуку, систематизації та автоматичного опрацювання. Таким чином, зростає інтерес до інформаційних ресурсів, одним із типів подання яких є розроблений нами інформаційно-пошуковий тезаурус предметної області «Розумне місто», оскільки, для науковців (і не тільки), що працюють у цій галузі, однією з необхідних умов є оволодіння відповідною термінологією, вміння оперувати нею, і в підсумку, використовувати концептуальні термінологічні знання.

Одним з джерел знань є тезаурус як складний тип словника, в якому всі терміни пов'язані між собою семантичними відношеннями, що відображають основні співвідношення понять предметної області знань, що описується.

Тезауруси набувають все більшого значення в процесі структурування та моделювання знань, систематизації та впорядкування наукової термінології, формування термінологічних банків даних, електронних

бібліотек. У науковій літературі існує поняття «тезаурусний метод», який широко застосовується у теорії та практиці інформаційного пошуку. Основний акцент цього методу робиться на змістовий аналіз призначених для відбору термінів із текстів, які є важливими для запиту користувача [162].

Процеси впорядкування та стандартизації міжгалузевої термінології, в результаті чого виділяються підсистеми одиниць, що включають базові поняття предметної області, виконують певні функції: оптимізуючу, нормативну, систематизуючу, практичну, прогностичну, комунікативну, оцінну, регулятивну, контрольну.

Розглянемо подані функції:

- оптимізуюча полягає в пошуку найбільш оптимальних термінологічних одиниць з точки зору їхньої форми й змісту, способів опису й розташування;
- нормативна проявляється у формулюванні норм, правил і рекомендацій з використання термінів;
- систематизуюча полягає в об'єднанні термінологічних одиниць у систему за певними ознаками;
- за практичною функцією стандарт виступає прикладним інструментом у процесі комунікації між членами професійної спільноти;
- прогностична забезпечує можливості використання термінологічного стандарту як засобу перетворення терміносистеми, як матеріалу для розроблення можливих альтернатив її розвитку;
- комунікативна полягає в передаванні за допомогою термінів певної змістовної інформації між суб'єктами і вироблення зворотного зв'язку;
- оцінна дозволяє стандартам виступати як критерій правильності при формуванні та використанні термінів;
- регулятивна виражається у впливі стандарту на комунікативні відношення шляхом визначення правил використання термінів;

– контрольна здійснюється на основі спостереження за станом терміносистеми з метою забезпечення її оптимального функціонування.

Для реалізації цих функцій необхідне виявлення предметної області, яку представляють термінологічні одиниці, а також визначення зв'язків між термінами, що входять у суміжні понятійні класи. Так предметна область «Розумне місто» є відкритою системою, пов'язаною з іншими соціальними системами: економікою, політикою, правом, історією та іншими, і ця взаємодія знаходить своє відображення в термінології даної галузі знань. При вирішенні питання про включення нового елемента в корпус лексики, яка описує цю предметну область, важливо враховувати ключові поняття цієї галузі, виявлені внаслідок порівняння декількох досліджуваних джерел.

Зважаючи на той факт, що інноваційна галузь «Розумне місто» набирають інтенсиву, потрібне уважне відношення до одиниць, що утворюють її базу метамови. Неповнота термінологічних даних, їхня фрагментарність і недостатній рівень опису для дослідницьких цілей роблять актуальним завдання створення нових термінологічних стандартів предметної області «Розумне місто», а також розроблення інноваційних за своїм змістом інструментів стандартизації. Найбільш ефективними формами таких інструментів можуть бути термінологічні словники, лексикографічні частини стандартів, тезауруси, що об'єднують базові одиниці предметної області. Саме тому важливі оптимізація та систематизація понятійно-термінологічного апарату й вдосконалення методики роботи з реалізацією обов'язкового етапу – формуванням експериментального термінологічного масиву, який повинен мати властивість якісної й кількісної репрезентативності. Формування та подання такого масиву може здійснюватися двома основними способами: перший полягає у використанні надійного готового набору одиниць, представленого в термінологічному словнику або попередньому стандарті; другий – в обробці текстів документів для отримання впорядкованого словника термінів, вибраних з текстів. Проте

необхідно враховувати відносні переваги й недоліки цих способів, оскільки вони можуть проявитися на наступних етапах дослідження.

Розроблення методики, відбір лексичного матеріалу й подальша робота над укладанням тезауруса «розумного міста» передбачають значний обсяг роботи. У процесі цієї роботи проводився аналіз наукових текстів предметної області «Розумне місто» для того, щоб виявити терміни-претенденти для включення в тезаурус та їхні еквіваленти англійською мовою. При визначенні основних принципів упорядкування та стандартизації галузевої термінології визначальною є пара «поняття – термін» і, на більш високому рівні, «система понять – система термінів».

Побудова тезауруса «розумного міста» – це досить багатоопераційний складний процес. Кожен етап роботи пов'язаний з аналізом багатьох варіантів. Для того, щоб отримати новий інтелектуальний продукт, необхідно забезпечити певні умови: чітке розуміння мети та постановки завдання; достатність інформаційної бази; повний опис об'єктів галузі; чітку співпрацю групи дослідників.

Робота з укладання тезауруса розпочиналася з відбору лексичного матеріалу. Певні труднощі на цьому етапі пов'язані перш за все з необхідністю розмежувати термінологічну й загальновживану лексику. Далі, у міру накопичення термінологічного матеріалу, що видобувається з текстів, визначаються основні термінологічні поля досліджуваної галузі [163].

У працях багатьох науковців розглядається питання формування об'єктивних критеріїв для однозначного виділення термінологічних одиниць, що виражають поняття й зв'язки з тою чи іншою галуззю знань, і тим самим визначення обсягу й межі термінології. Дослідники по-різному дивляться на вирішення проблеми виявлення термінів у певній підмові, користуючись при цьому різними визначеннями термінів, пропонуючи свої методики та критерії відбору термінів, у випадку, коли понятійна система ще не склалася і немає чіткої межі між терміном і нетерміном [193].

При побудові інформаційно-пошукового тезауруса «розумного міста» першим завданням було відбір термінів для включення до тезауруса. Існує декілька можливих джерел для відбору термінів при розробленні тезауруса. Насамперед повинні бути вивчені уже існуючі тезауруси чи тезауруси близьких предметних областей, що можуть містити значну кількість термінів, які можуть увійти в структуру нового ресурсу. Терміни-кандидати на внесення в тезаурус можуть бути запропоновані експертами предметної області. Ще одним вагомим джерелом формування термінологічного фонду предметної галузі є наукові тексти, електронні каталоги наукових бібліотек та реферативні бази даних [163].

Терміни, які включають у тезаурус «розумного міста», повинні відповідати таким правилам:

- представляти поняття, які присутні в наукових текстах, і мають відбиратися з міркувань ефективності їхнього використання в пошуку документів;
- частотність вживання терміна в текстах;
- урахування вже наявних термінів: потрібно перевіряти, чи представляє термін-кандидат окреме поняття, якому нема відповідників серед існуючих термінів тезауруса; необхідно уникати включення термінів, значення яких перетинаються зі значеннями вже існуючих термінів у тезаурусі [134].

Предметна область «Розумне місто» перебуває на етапі встановлення. З урахуванням її міждисциплінарного характеру понятійний апарат цих галузей використовується шляхом його адаптації та застосування для побудови термінологічної бази «розумного міста».

Для побудови тезауруса «розумного міста» потрібно було створити список лексичних одиниць (керований словник), тобто необхідно було визначити терміносистему предметної області, проаналізувавши корпус різних джерел, у яких розкриваються терміни досліджуваної галузі. Виділення термінів з контекстів і проведення семантичного аналізу термінів,

а також дефініційного аналізу ключових термінів робить можливим представити всю термінологію галузі у вигляді термінологічного поля [235].

Основними джерелами відбору термінів для тезауруса є автентичні фахові тексти (монографії, публікації результатів теоретичних та експериментальних досліджень, статті в періодичних фахових виданнях, доповіді конференцій та ін.), перевага яких обумовлена їхньою особливістю – наявністю термінологічної лексики, що розкриває наукові поняття досліджуваної галузі та подає їхнє однозначне пояснення.

Основні принципи, які були використані при відборі масиву документів, джерел термінології, полягають у наступному:

- найбільш точна відповідність документів цій тематиці;
- термінологічна наповненість інформаційних документів і важливість інформації, що міститься в них;
- висвітлення тематики з охопленням різних аспектів.

Матеріалом для пошуку термінів «розумного міста» стали українські, російські та англійські автентичні фахові джерела, в яких зафіксовані поняття галузі.

На першому етапі було досліджено наукові видання, в яких розглядаються питання реалізації проєктів «розумне місто» та були опубліковані українською мовою.

Головну роль у становленні понятійного апарату предметної області відіграють словники, як першоджерело термінів та знань про предмет, тому перші розвідки були спрямовані на пошук таких словникових баз, які відіграли б велике значення для осмислення понятійного комплексу предметної області «Розумне місто».

У результаті аналізу було з'ясовано, що відсутні термінологічні словники з предметної області «Розумне місто», що опубліковані українською мовою. Проте вдалося дослідити аналогічну базу англійських видань, у результаті було виявлено британський стандарт PAS 180:2014 Smart cities – Vocabulary [236], укладений Британським інститутом стандартів

– BSI (British Standards Institution), який містить терміни предметної області «Розумне місто». Цей словник вийшов друком у 2014 році. Необхідно зазначити, що він був розроблений для застосування у Великій Британії і містить як загальну частину, яка може застосовуватися практично в будь-якій країні, так і має особливості, що вимагають адаптації в інших країнах [235].

PAS 180 подає англomовне тлумачення більше ста словникових одиниць, саме вони стали основою лексичного складу реєстру двомовного тезауруса «розумного міста». PAS 180 є першою версією словника «розумних міст», початком процесу збору різноманітних термінів та виразів, що використовуються в щоденних дискусіях про «інтелектуальні» міські системи. Словниковий запас спрямований на забезпечення узгодженого комплексу робочих умов, щоб практики могли мати краще загальне розуміння [236].

Таким чином, не існує джерела, в якому вся термінологія «розумного міста» була б подана єдиною системою. Цю відчутну прогалину міг би заповнити двомовний тезаурус, що містить українські та англійські терміни досліджуваної галузі та їхнє тлумачення обома мовами. Двомовність тезауруса дасть можливість українським ученим і фахівцям швидше й ефективніше орієнтуватися у світовій ситуації в цій галузі, надасть їм допомогу при написанні та перекладі статей, пошуку релевантних документів, інформаційному забезпеченні користувачів, інформаційні потреби яких належать до предметної області «Розумне місто».

Розроблений тезаурус дозволить структурувати й накопичувати інформацію, релевантну для цієї галузі знань, проводити інтелектуальний пошук даних у інформаційних сховищах і мережах, формувати пошукові образи документів, створювати інформаційні ресурси бібліотек. Крім того, він сприятиме підвищенню рівня професійної підготовки фахівців не тільки в галузі «розумного міста», а й у галузі інформаційних технологій [235].

Під час виявлення термінів предметної області «Розумне місто» джерелами дослідження слугували такі автентичні фахові тексти, які наповнені термінологічною лексикою:

- стандарти для забезпечення технологій «розумного міста», розроблені Британським інститутом стандартів (BSI): PAS 181 Smart City Framework – Guide to establishing strategies for smart cities and communities [237] (Керівництво щодо створення стратегій для розумних міст та населених пунктів); PAS 182 Smart city concept model [238] (Концепція моделі розумного міста); PAS 183 Smart cities – Guide to establishing a decision-making framework for sharing data and information services [239] (Керівництво щодо створення основи прийняття рішень для спільного використання даних та інформаційних послуг); PAS 8101:2014 Smart cities. Guide to the role of the planning and development process [240] (Розумне місто. Керівні принципи планування та розвитку);

- низка стандартів, розроблених Міжнародною організацією зі стандартизації ISO: An indicator reporting standard for quality of life and services in cities (BS ISO 37120) [241] (Показник стандарту звітності якості життя та послуг у містах); Research on smart infrastructure projects (PD ISO/TR 37150) [242] (Дослідження проєктів інтелектуальних інфраструктур); A specification for KPIs for smart infrastructure projects (PD ISO/TS 37151) [243] (Специфікація ключових показників ефективності для інтелектуальних інфраструктурних проєктів); Smart community infrastructures – Common framework for development and operation (ISO/PRF TR 37152) [244] (Розумні громадські інфраструктури – загальні рамки для розвитку та функціонування);

- матеріалом дослідження була англійська, українська та російська науково-технічна література (технічні журнали, праці конференцій, наукові статті), в яких розглядається тематика «розумне місто».

Опрацювання статей є досить кропіткою роботою та займає надзвичайно багато часу. Проте, варто зазначити, що наукові статті, праці

конференцій на відміну від словників, довідників, енциклопедій та ін., дають можливість виявити актуальні терміни, які знаходяться в науковому обігу, прослідкувати їхні взаємозв'язки, побачити реальну картину їхнього функціонування в текстах.

Отже, вивчення термінів, для включення в тезаурус «розумного міста», відбувалося не лише на основі словників і довідників, а також на основі науково-технічних публікацій, які засвідчують факт їхнього реального функціонування [235].

Оскільки об'єктом аналізу, на етапі відбору термінів та формування тезауруса «розумного міста», була спеціальна лексика досліджуваної галузі, предметом аналізу були глосарій PAS 180:2014 Smart cities – Vocabulary та наукова література, таким чином ми можемо говорити про кваліфікованість тезауруса предметної області «Розумне місто», тому що перевагами цих джерел є надійність, високий рівень загальноприйнятої інформації, зрозуміла мова подання, доступність (можливість отримати вільний доступ до цих джерел), висока насиченість термінологічною лексикою.

Наступним етапом роботи над тезаурусом «розумного міста» було формування переліку ключових слів (керованого словника), вилучених із вищезазначених наукових джерел. При наповненні керованого словника ставилося завдання з найбільшою повнотою подати термінологію, яка використовується для опису понять предметної області «Розумне місто». З цього переліку було відібрано поняття, кандидати в терміносистему тезауруса, і подано їх як складові частини системи та репрезентацію системності досліджуваної галузі через термінологічні концепти.

Відбір термінів здійснювався на основі таких критеріїв:

- інформаційність – можливість терміна-кандидата позначати мовне явище та одночасно бути основою для інформаційного запиту;
- частотність – зверталася увага на терміни, які найбільше використовуються в досліджуваній предметній області;

- відповідність темі – відбиралися терміни, які відображають цю проблему;
- актуальність – перевага віддавалася термінам, що відображають сучасні розробки в цій галузі або мають велике значення для її розуміння;
- практична значущість – виділялися терміни, розуміння яких має суттєвий вплив на з'ясування текстів.

Терміни, що формують корпус одиниць тезауруса «розумного міста», не існують ізольовано від інших, а навпаки, тісно співвідносяться як зі загальнонауковими, так і з галузевими поняттєвими категоріями. Тому тезаурус містить терміни із суміжних галузей. Серед його термінів вирізняються загальнонаукові, міжгалузеві та вузькоспеціальні терміни, зробити чітке розмежування між якими нелегко, оскільки вони є дифузними завдяки постійній взаємодії та перехідності [135].

Оскільки терміни «розумного міста» складаються з одного або декількох слів, тому подамо їхні структурні класифікації.

У роботах учених знаходимо різні структурні класифікації термінів. Вісім видів термінів подає Т. Кияк [178]: 1) терміни-кореневі слова, 2) похідні терміни, 3) терміни-складні слова, 4) терміни-словосполучення, 5) терміни-аббревіатури, 6) буквені умовні позначення, 7) символи-знаки, 8) номенклатура.

На прості, складні та термінологічні словосполучення з прийменниковим або безприйменниковим сполученням елементів поділяє терміни Л. Ткачева [245].

У ході аналізу термінологічного матеріалу предметної області «Розумне місто» за даними лексикографічних джерел було з'ясовано, що термінологія цієї галузі, складається з лексичних одиниць двох типів: термінів-слів (однокомпонентних) та термінів-словосполучень (багатокомпонентних).

Однокомпонентні терміни називає ознаку поняття, що дає можливість віднести це поняття до певного класу та диференціювати його від інших

класів понять. Додавши до однокомпонентного терміна означальні компоненти, які називають додаткові ознаки, отримуємо багатоконпонентний термін (термінологічне словосполучення), який, завдяки саме цим додатковим ознакам, диференціюється від інших понять того ж класу. Тобто, термінологічні словосполучення утворюються в результаті ускладнення синтаксичної конструкції, наприклад: *виробництво (production)* – *безвідходне виробництво (non-waste production)*, *зелене виробництво (green manufacturing)*, *низьковуглецеве виробництво (low carbon production)*, *розумне виробництво (smart manufacturing)*, *стале виробництво (sustainable production)*, *чисте виробництво (clean production)*; *відео (video)* – *відеодані (video data)*, *відеоконсультація (video-consultation)*, *відеоконференція (videoconference)*, *відеотелефонія (videotelephony)*, *відеоспостереження; система (system)* – *система автоматичного керування (automatic control system)*, *система відновлення тепла (heat recovery system)*, *система електронних платежів (e-payment system)*, *система спільного використання велосипедів (bicycle-sharing system)*, *система керування енергією (energy management system)* тощо.

Однокомпонентні терміни виражаються однією лексичною одиницею – іменником (*біоенергетика, виробництво, відходи, гаджет, дані, давач, інтернет, мережа, прозорість, система, слід, управління, урбанізація, утилізація, хмара* та ін.) чи прикметником (*зелений, інтегрований* та ін.). До цієї моделі входить конструкція, яка виражена складним елементом з двох слів (*біоенергетика, відеодані, відеотелефонія, відеоконференція, браунфілд, екобудинки, електромобіль, електромобільність, біогаз, велодоріжка, екопарковка, електроавтобус, крауд-технологія* та ін.)

Багатоконпонентні терміни досліджуваної галузі «Розумне місто» складаються з двох та більше одиниць і є найпоширенішим видом термінів:

– двокомпонентні терміни – *розумне місто, зелена інфраструктура, альтернативна енергетика, енергозберігаючий будинок, електронні гроші, електронний консалтинг, великі дані, віртуальний*

колектив, зелене виробництво, Інтернет речей, новий урбанізм, пасивний будинок, сталий розвиток, хмарні обчислення та ін.;

– терміни, що включають три та більше компонентів – *альтернативні джерела енергії, розумне вуличне освітлення, розумний громадський транспорт, мультимодальні пасажирські перевезення, цифрове управління будинком, будинок з плюсовим енергопостачанням, електронне управління документами, п'яте покоління мобільного зв'язку, пасивна система сонячного теплопостачання, система спільного використання велосипедів, клінічна система підтримки прийняття рішень та ін.*

Багатокомпонентні терміни ще називають термінами-словосполученнями, багатокомпонентними термінологічними словосполученнями, складеними термінами, термінологічними словосполученнями. Лексикографи визнають допустимим наявність у заголовному слові трьох-чотирьох слів, навіть п'яти (як правило, це стосується багатокомпонентних моделей, в яких слова поєднані за допомогою прийменників, займенників, сполучників). Останні в досліджуваній термінології не виявленні.

Аналіз структурних форм термінів предметної області «Розумне місто» показав, що в досліджуваній галузі переважають термінологічні словосполучення, які виражають більш складні поняття.

Термінологічне словосполучення – це багатокомпонентний термін, в якому один компонент виступає носієм родового, узагальнюючого й систематизуючого значення, а інші виражають видову, конкретну назву [246]; це словосполучення, що має певні властивості і терміна, і словосполучення. Результати досліджень словосполучення як лінгвістичного явища характерні й для термінологічних словосполучень. Словосполучення – це лексичні одиниці вищого рівня, до складу яких входять одиниці нижчого рівня – слова [247].

Термінологічне словосполучення як термін має такі ознаки: однозначність, чітка дефініція, системність, стилістична нейтральність, короткість.

Основна функція термінологічного словосполучення – номінативна (функція називання), за допомогою якої певні поняття отримують назви. Вони мають здатність конкретизувати значення за допомогою додаткових уточнювальних характеристик, які виражаються загальновживаними словами.

Оскільки термінологічне словосполучення є синтаксичною конструкцією, що складається з кількох пов'язаних між собою компонентів, таким чином, воно утворюється в процесі синтаксичного способу термінотворення – найпродуктивнішого способу творення термінів.

За кількістю структурних компонентів термінологічні словосполучення предметної області «Розумне місто» поділяються на:

- а) прості термінологічні словосполучення, що складаються з двох слів, одне з котрих головне, а інше – залежне;
- б) складні термінологічні словосполучення, які складаються більше ніж з двох компонентів.

Це слугує підставою розрізняти двокомпонентні та багатоконпонентні термінологічні словосполучення. Багатоконпонентні термінологічні словосполучення утворюються розширенням та уточненням значень двокомпонентних термінів.

За морфологічною ознакою основного слова або за граматичними категоріями стрижневого компонента термінологічні словосполучення диференціюємо на:

- а) субстантивні іменні словосполучення;
- б) ад'єктивні термінологічні словосполучення зі стрижневим прикметником чи дієприкметником;
- в) вербальні термінологічні словосполучення зі стрижневим дієсловом.

Найрезультативнішою структурою серед моделей термінологічних словосполучень «розумного міста» є двокомпонентні та трьохкомпонентні термінологічні словосполучення. Структура таких словосполучень різна. Найпоширенішими є структурні моделі базового характеру. Двокомпонентні словосполучення містять головний компонент, що називає основне поняття, та підпорядковане йому означення. Головним компонентом переважно виступає іменник. Залежним компонентом виступає прикметник або дієприкметник, роль якого зводиться до уточнення чи конкретизації бази.

Розглянемо моделі дво- та багатоконпонентних термінологічних словосполучень предметної області «Розумне місто».

У групі дво- та багатоконпонентних термінологічних словосполучень наявні субстантивні (іменникові) словосполучення, утворені за структурними типами «іменник + іменник», «іменник + іменник + іменник», моделі з відсутнім граматико-морфологічним оформленням атрибутивного препозитивного іменника (*автоматизація будівель, база даних, база знань, давач газу, давач світла, управління відходами, центр опрацювання даних та ін.*). Це означає, що наступний компонент (чи наступні компоненти), маючи синтаксичні властивості іменника, виступає як означення до першого компонента – означуваного, тому нелегко зрозуміти граматичні відношення між компонентами такого типу термінологічних словосполучень.

Іншою формою субстантивних словосполучень є моделі: прикметник + іменник, прикметник + прикметник + іменник та ін., що широко використовуються для творення термінологічних словосполучень (*розумні люди, розумна парковка, інтелектуальне місто, зелені дахи, жорстка інфраструктура, дистанційна медицина, електронна платіжна система, розумна автобусна зупинка, локальна комп'ютерна мережа та ін.*).

Ад'єктивні (прикметникові) термінологічні словосполучення та вербальні (дієслівні) термінологічні словосполучення у нашому тезаурусі відсутні.

Лінгвістична структура термінів тезауруса, подана в таблиці 4.1.

Лінгвістична структура термінів «розумного міста»

№	Структура шаблону	Приклади
1.	Іменник	<i>Велодоріжка, відеоконсультація, відеоспостереження, краудсорсинг, хмара, бенчмаркінг, біодизель, метробус, геліоенергетика.</i>
2.	Іменник + іменник	<i>Автоматизація будівель, база даних, давач диму, щільність забудови.</i>
3.	Іменник + іменник + іменник	<i>Система відновлення тепла, центр опрацювання даних.</i>
4.	Іменник + іменник + іменник + іменник	<i>Системи керування базами даних.</i>
5.	Іменник + іменник + прикметник + іменник	<i>Камера безпеки дорожнього руху, контролер заряду сонячної батареї, камера безпеки дорожнього руху.</i>
6.	Іменник + іменник + іменник + прикметник + іменник	<i>Система збору даних дорожнього руху.</i>
7.	Іменник + прикметник	<i>Будівля енергозберігаюча, відходи небезпечні, слід водний.</i>
8.	Іменник + прикметник + іменник	<i>Система електронних платежів, система автоматичного керування, дані дистанційного зондування.</i>
9.	Іменник + прикметник + іменник + іменник	<i>Система спільного використання велосипедів.</i>
10.	Іменник + сполучник + іменник	<i>Місто, як сервіс.</i>
11.	Іменник + прийменник +	<i>Будівля з плюсовим енергопостачанням.</i>

	прикметник + іменник	
12.	Іменник + прийменник + числівник + іменник + іменник	<i>Будівля з нульовим викидом вуглецю.</i>
13.	Іменник + числівник	<i>Місто 2.0.</i>
14.	Прикметник + іменник	<i>М'яка інфраструктура, розумна економіка, хмарні технології, екологічний будинок, зелене будівництво.</i>
15.	Прикметник + іменник + іменник	<i>Бездротова точка доступу, автоматична система керування, спільне використання автомобілів, розумна система безпеки.</i>
16.	Прикметник + іменник + іменник + іменник	<i>Електронна система охорони здоров'я.</i>
17.	Прикметник + іменник + іменник + іменник + іменник	<i>Клінічна система підтримки прийняття рішень.</i>
18.	Прикметник + іменник + прикметник + іменник	<i>Автоматичне розпізнавання номерного знака, активна система сонячного теплопостачання.</i>
19.	Прикметник + прикметник + іменник	<i>Швидкісний автобусний транзит, розумне вуличне освітлення, вільні земельні ділянки.</i>
	Прислівник + прикметник	<i>Екологічно чистий.</i>
20.	Прислівник + прикметник + іменник	<i>Екологічно чисте місто, екологічно чистий транспорт.</i>
21.	Числівник + іменник + прикметник + іменник	<i>Шість ознак «розумного міста», п'яте покоління мобільного зв'язку, четверте покоління мобільного зв'язку.</i>

Проведений аналіз термінів предметної області «Розумне місто» дає підстави стверджувати, що в цій термінології переважають термінологічні словосполучення. За кількісними показниками серед багатокомпонентних термінологічних словосполучень двокомпонентні словосполучення займають перше місце. Найпродуктивнішою структурною моделлю двокомпонентних термінологічних словосполучень є тип «прикметник + іменник».

Проведені класифікаційні операції свідчать, що побудова категоріально-понятійного апарату предметної області «Розумне місто» містить складні моделі субстантивних іменних словосполучень, виражених схемою «узгоджене слово + іменник». У цій структурі іменник є головним словом, а узгоджене – залежним і може виражатися прикметником, іменником чи прислівником. Кількість слів у термінологічній сполуці може складати від двох до п'яти одиниць.

Наступним етапом нашого дослідження було укладання тезауруса «розумного міста» на основі зібраного матеріалу. Тезаурус є реєстром логіко-семантичних відношень між термінами. Він охоплює не лише множину термінів, що подані в алфавітному списку з їхніми поясненнями, а й моделями зв'язків між термінами [163].

Робоча структура тезауруса предметної області «Розумне місто» формується із низки тематичних комплексів різних галузей:

– загальні поняття – *управління / governance, урбанізація / urbanization, сталий розвиток / sustainable development, давач / sensor, інфраструктура / infrastructure, жорстка інфраструктура / hard infrastructure, м'яка інфраструктура / soft infrastructure, інформаційно-комунікаційні технології / information and communication technologies, Інтернет / Internet, мережа / network, система / system;*

– інформаційні системи «розумного міста» – *розумне місто / smart city, розумний індикатор міста / smart city indicator, шість ознак «розумного міста» / six characteristics of the Smart City, розумна економіка / smart economy, розумне життя / smart living, розумна мобільність / smart mobility,*

розумне довкілля / smart environment, розумне управління / smart environment, розумні люди / smart people;

– управління ресурсами – *слід / footprint, життєвий цикл / lifecycle, відновлювальна енергія / renewable energy, автоматизація будівель / building automation, зміна поведінки / behavioural change, міські обчислення / urban computing;*

– технології – *технологія OLEV / online electric vehicle, великі дані / big data, управління доступом / access control, віддалений доступ / broadband access, широкосмуговий / broadband, широкосмуговий зв'язок / broadband connectivity, Інтернет / Internet, локальна комп'ютерна мережа / local area network, глобальна мережа / wide area network, зелені інформаційні технології / green information technology, міські обчислення / urban computing); віртуальний асистент / virtual assistant;*

– інфраструктури – *вільні виробничі площі / brownfield land, глобальна інформаційна інфраструктура / global information infrastructure, жорстка інфраструктура / enabling infrastructure, м'яка інфраструктура / soft infrastructure, зелена інфраструктура / green infrastructure;*

– мобільність – *стала мобільність / sustainable mobility, інтегрований транспорт / integrated transport, розумна мобільність / smart mobility, екомобільність / ecomobility); сталий транспорт / sustainable transport, управління дорожнім рухом / traffic management, розумна автобусна зупинка / smart bus stop, спільне використання автомобілів / car sharing;*

– енергетика – *активний будинок / active house, розумний лічильник / smart meter, розумне вуличне освітлення / smart street lights, розумне освітлення будинку / smart home lighting, система сонячного теплопостачання / solar heating system, управління енергоспоживанням / energy management, сонячний колектор / solar collector, зелений тариф / green tariff, енергія відновлювальна / renewable energy, енергозберігаючий будинок / energy-efficient house;*

– телекомунікації – телекомунікаційні технології / *telecommunication technologies*, телекомунікаційна мережа / *telecommunication network*, четверте покоління мобільного зв'язку / *fourth generation long term evolution*, широкосмуговий зв'язок / *broadband connectivity*, глобальна інформаційна інфраструктура / *global information infrastructure*;

– транспорт – електромобіль / *electric car*, зелений автомобіль / *green car*, зелений транспорт / *green vehicle*, екомобіль / *ekomobil*, екологічно чистий транспорт / *eco-friendly vehicle*, розумна навігація / *smart navigation*, розумна парковка / *smart parking*, розумна транспортна система / *intelligent transportation system*, мультимодальні пасажирські перевезення / *multimodal passenger transport*, гібридний електромобіль / *hybrid electric vehicle*, електронний квиток / *e-ticket*;

– фінанси й економіка – широкосмугова економіка / *broadband economy*, система мікроплатежів / *micropayment system*, віртуальна економіка / *virtual economy*, віртуальне підприємство / *virtual corporation*, екодизайн / *ecodesign*, економічна життєздатність / *economic vitality*, електронне управління документами / *electronic document management*, електронний гаманець / *e-wallet*, електронна платіжна система / *electronic payment system*, зелене виробництво / *green manufacturing*); розумне виробництво / *smart manufacturing*, стале виробництво / *sustainable production*, чисте виробництво / *clean production*, віртуальна вода / *water footprint*;

– місцеві та центральні органи влади – управління / *governance*, соціальне включення / *social inclusion*, відкритий уряд / *open government*, електронна послуга / *electronic service*, електронний уряд / *electronic government*, розумний уряд / *smart government*, електронна демократія / *e-democracy*, прозорість / *transparency*, розумна громада / *smart community*;

– спільноти, освіта – зміна поведінки / *behavioural change*, освіта для сталого розвитку / *education for sustainable development*,

персональне навчальне середовище / personal learning environment, електронне навчання / e-learning, мобільне навчання / mobile learning, розумна громада / smart community;

– охорона здоров'я – *телесервіс / telecare, телемедицина / telemedicine, розумне життя / smart living, відеоконсультація / video-consultation, мобільне здоров'я / m-health, дистанційна медицина / remote medicine, дистанційна діагностика / remote diagnostics, електронна система охорони здоров'я / e-health, клінічна система підтримки прийняття рішень / clinical decision support system;*

– державні й приватні моделі надання послуг – *бізнес-модель / business model, протокол міста / city protocol, цифрове включення / digital inclusion, цифрова інфраструктура / digital infrastructure, відкриті дані / open data, електронний магазин / e-shop, електронний туризм / e-tourism, екологічний туризм / ecotourism;*

– система безпеки – *інформаційна безпека / information security, кібербезпека / cybersecurity, відеоспостереження / closed-circuit television, камера безпеки дорожнього руху / traffic enforcement camera, розумна система безпеки / smart alarm system;*

– навколишнє середовище – *екологічна стійкість / environmental sustainability, якість повітря / air quality, якість ґрунту / soil quality, якість води / water quality, низьковуглецева технологія / low carbon technology, екологічний коридор / greenway, зелений пояс / green belt, низьковуглецеве виробництво / low carbon production, будинок з нульовим викидом вуглецю / zero carbon building, вторинна переробка відходів / recycling of garbage, вуглецевий слід / carbon footprint.*

Таке тезаурусне міждисциплінарне моделювання знання дає можливість науковцям виходити за межі термінологічної системи однієї наукової галузі знань у порівнянні з іншою. Таким чином відбувається обмін інформацією між науковцями різних галузей досліджень.

Тезаурус включає всі важливі розділи предметної області «Розумне місто» та відображає сучасний стан досліджуваної галузі знань.

Ці категорії, як і визначення, відкриті для змін і розширень у майбутньому, але досить широкі, щоб охопити перший етап відбору термінів [235].

Тезаурус подає широкий спектр термінів і виразів, які використовуються при дослідженні інтелектуальних систем міста, систем комунікації, виступаючи термінологічним інструментом для розробників, дизайнерів, виробників. Він може бути використаний керівниками на всіх рівнях і у всіх секторах «розумних міських громад». Може також використовуватися і як багатопланова довідкова система, і як база для подальших лінгвістичних досліджень.

Розроблення тезауруса предметної області «Розумне місто» характеризується необхідністю опису понять, що зустрічаються в текстах цієї предметної області. Тезаурус містить не тільки терміни, які характеризують важливі поняття в текстах цієї галузі, але також охоплює широке коло більш специфічних термінів, виявлення яких у конкретному тексті зробить його релевантним запитом за поняттями більш високого рівня. Релевантність полягає у відповідності між бажаною інформацією і тою, яка дійсно одержується. Це міра близькості між реально одержаними документами й тими, що варто було б отримати із системи. Таким чином, ми описуємо не лише дескриптор «розумне довкілля» і його основні підрозділи, такі як «відновлювальна енергетика», «екологічний моніторинг» та інші, але й конкретні екземпляри «розумного довкілля», щоб текст, у якому обговорюється, наприклад, відновлювальна енергія, міг би бути отриманий при пошуку за терміном «розумне довкілля».

Вагомим аспектом укладання тезауруса «розумного міста» був спосіб розмежування та перекладання значення термінів однієї мови засобами іншої, оскільки до термінів повинні добиратися еквівалентні слова-відповідники для забезпечення передачі усіх аспектів значення слова.

Переклад науково-технічної та ділової літератури ближче знаходиться до галузі науки на відміну від перекладання художніх творів, що тяжіє до сфери мистецтва; саме тому всі питання її перекладу не можна пояснювати безпосередньо лінгвістичним шляхом, їх треба розв'язувати разом із фахівцем галузі науки і техніки, вважає С. Головащук [248]. Тому термінологічна база тезауруса була обговорена колективом науково-дослідної лабораторії «Smart city» Тернопільського національного технічного університету імені Івана Пулюя із залученням викладача англійської мови.

Чимало проблем при перекладі термінологічних одиниць виникає в розкритті та переданні іншомовних реалій засобами української мови. На жаль, сучасна українська науково-технічна мова не завжди містить відповідні терміни для перекладу нових понять або понять, що набули нового значення в іншій предметній області. Обов'язкова умова повноцінного перекладу будь-якого спеціального тексту – це повне розуміння його перекладачем. Грубі помилки в перекладі та перекручування семантики термінів при перекладі можуть виникнути при дослівному перекладі термінів, без проникнення в їхню сутність, без знання самих явищ, процесів та механізмів, про які йдеться в оригіналі [249].

При перекладі термінів «розумного міста» використовувалися такі трансформації: транслітерація, калькування, описовий переклад.

Транслітерація або транскрибування – це передання звукової форми лексичної одиниці вихідної мови засобами абетки мови перекладу. Шляхом транслітерації були запозичені такі терміни: *Internet* – *інтернет*, *гаджет* – *gadget*, *краудфандинг* – *crowdfunding*, *краудсорсинг* – *crowdsourcing*, *каршерінг* – *каршерінг* та ін.

Калькування припускає існування двосторонніх міжмовних відповідностей між елементарними лексичними одиницями, які використовуються як «будівельний матеріал» для відтворення внутрішньої форми слова, яке перекладається. Калькування, як прийом створення еквівалента, подібне до буквального перекладу – еквівалент цілого

створюється шляхом простого складання еквівалентів його складових частин [250]. Наприклад: *computer technology* – комп'ютерні технології, *building information modeling* – інформаційне моделювання будівель, *smart transportation system* – розумна транспортна систем та ін.

Експлікація (від лат. *explicatio* – роз'яснення) – це заміна лексичної одиниці мови-оригіналу словосполученням, яке пояснює або дає визначення одиниці, що перекладається. Експлікація використовується, коли словосполучення-термін складається зі слів, які ще не використовуються в певній галузі на мові перекладу і потребують свого тлумачення. За допомогою експлікації перекладають багатокомпонентні термінологічні словосполучення. Цим способом було перекладено багатокомпонентні термінологічні словосполучення: *zero carbon building* – будівля з нульовим викидом вуглецю, *energy-plus house* – будівля з плюсовим енергопостачанням, *charge controller* – контролер заряду сонячної батареї, *health monitoring device* – прилад для моніторингу здоров'я, *bicycle-sharing system* – система спільного використання велосипедів, *passive solar system* – пасивна система сонячного теплопостачання та ін.

Якщо значення англійської одиниці повністю співпадає зі значенням української, то це – еквівалентний переклад. Можна навести такі приклади еквівалентного перекладу: *memory* – пам'ять, *mode* – режим, *screen* – екран, *network* – мережа та ін.

Як бачимо семантичні, синтаксичні та структурні відмінності терміносистем англійської та української мов викликали певні труднощі при виборі способу перекладу.

При укладанні тезауруса предметної області «Розумне місто» було вилучено з текстів чимало синонімів до певних термінів, які Сидні Лендау називає «варіантами» та вважає термін «синоніми» недопустимим для позначення різних термінологічних назв одного об'єкта й замість того пропонує термін «варіанти», тому що на відміну від слів звичайної мови, наукова номенклатура залежить від процесу означення та від зовнішніх

фізичних обставин чи властивостей означення речей [251]. Тому виникло питання перехресних посилань. Ми позначаємо перехресне посилання на варіант за допомогою скорочення *див.* (дивіться). Якщо поняття має два чи більше синонімів, лише один з них умовно вибирається як головний термін, усі інші терміни-синоніми мають посилання на цей головний термін.

Наступне питання, яке виникло при роботі з термінологією – це аббревіатури. Офіційні скорочення внесено до реєстру тезауруса. Їхні аббревіатури наведено в дужках після повної назви. Проте не завжди ми маємо аббревіатури в українській та англійській мовах, так терміни: *bus rapid transit (BRT)*, *home area network (HAN)*, *electronic document management (EDM)*, *open access (OA)*, *artificial intelligence (AI)*, *near field communication (NFC)*, *energy management system (EMS)*, *air quality index (AQI)* та ін. не мають скорочень в українських відповідниках, а терміни: *інтелектуальний аналіз даних (ІАД)*, *система електронних платежів (СЕП)*, *центр опрацювання даних (ЦОД)* та ін. у свою чергу мають скорочення лише в українському варіанті. Проте більшість термінологічних одиниць мають аббревіатури як в українській, так і в англійській мовах, наприклад, *геоінформаційна система (ГІС)* – *geographic information system (GIS)*, *глобальна інформаційна інфраструктура (ГІІ)* – *global information infrastructure (GII)*, *інтелектуальна транспортна система (ІТС)* – *intelligent transportation system (ITS)*, *інтернет речей (ІР)* – *Internet of things (IoT)*, *інтерфейс користувача (ІК)* – *user interface (UI)*, *інформаційне моделювання будівель (БІМ)* – *building information modelling (BIM)*, *системи керування базами даних (СКБД)* – *database management systems (DBMS)* та ін.

4.2 Визначення відношень між термінами інформаційно-пошукового тезауруса «розумного міста»

Термінологію предметної області «Розумне місто» описує створений інформаційно-пошуковий тезаурус, побудова якого передбачає розкриття

різних типів відношень між поняттями, виражених термінами з набором диференційованих семантичних відношень. Основна ідея такого опису полягає у використанні не тільки універсальних, стандартних відношень, таких як «рід-вид», «частина-ціле» та інших, але й специфічних для цієї галузі відношень, що самі по собі несуть значне смислове навантаження, визначають різні аспекти семантики терміна.

У результаті дослідження було доповнено класифікацію типів семантичних відношень, з урахуванням міждисциплінарних тенденцій сучасного наукового знання.

Побудова класифікації відношень між термінами досліджуваної галузі виконувалося у декілька етапів. На першому етапі було проаналізовано типи відношень, встановлені стандартами (IDEF5, ДСТУ 4032-2001), а також відношення, якими оперують у інших тезаурусах. Після чого на зібраному матеріалі було виділено проблемно-специфічні типи відношень, характерні для термінів «розумного міста». Отримана інформація була об'єднана в класифікацію відношень для тезауруса предметної області «Розумне місто».

Класифікація відношень між дескрипторами інформаційно-пошукового тезауруса «розумного міста» містить два типи: традиційні, а саме ієрархічні й еквівалентні, та проблемно-специфічні (неієрархічні або асоціативні) відношення.

Назва відношення є двомісним предикатом $R(A, B)$, що з'єднує заголовне слово статті – A з введеним ним терміном – B [252].

На першому етапі використовувалися традиційні відношення.

Основними типами понятійних зв'язків є ієрархічні відношення «вище-нижче», що охоплюють три логічно різні типи ситуацій, які розрізняються таким чином:

– «рід-вид» (generic) (таксономічні відношення, «is-a», гіперонімія і гіпонімія – такі різні назви маємо в залежності від термінологічних традицій використання ресурсу) (A є родовим до B);

- «частина-ціле» («whole-part») (мереологія, композиційні відношення) ($B \in$ частиною A)
- «екземпляр-клас» (instance) ($B \in$ екземпляром A).

Ці базові відношення найбільшою мірою відрізняють тезаурус від неструктурованого списку термінів, наприклад від глосарію або традиційного словника.

Ієрархічні відношення – зв'язують два дескриптори, якщо обсяг поняття, що відповідає одному з дескрипторів, містить у собі обсяг поняття іншого дескриптора. Ці відношення розкривають логічне підпорядкування дескрипторів: вищий дескриптор, що визначає поняття роду або цілого, включає в себе нижчий дескриптор, що визначає поняття виду або частини цілого. Відношення ієрархічного типу в інформаційно-пошукових тезаурусах встановлюються лише в тому випадку, якщо вони є істинними незалежно від контексту. Кожне з них утворює ієрархії, які можна піддати логічній перевірці через посилання на базові типи понять, виражені термінами.

Родове відношення ідентифікує зв'язок між класом чи категорією та їхніми членами або різновидами. Родо-видові відношення охоплюють велику групу термінів, утворюючи мікросистеми: мінімальну кількість термінологічних одиниць – видових, які об'єднані родовим терміном. Саме класифікація термінів, утворена на відношеннях «рід-вид», дає можливість об'єктивно виокремити головні поняття в предметній області «Розумне місто», згрупувати навколо них інші, підлеглі їм периферійні поняття, створити загальну класифікаційну схему понять предметної області.

Семантичний простір термінів «розумного міста» є складною ієрархічною структурою, елементи якої зв'язані родо-видовими відношеннями (вище-нижче). Тезаурусна модель предметної області «Розумне місто» представляється у вигляді схеми з вершиною *розумна країна*, видовим до якої є термін *розумне місто*, який розщеплюється на шість головних тематичних термінополів: *розумне життя*, *розумні люди*, *розумна економіка*, *розумне довкілля*, *розумна мобільність*, *розумне*

управління. Відповідні терміни є видовими по відношенню до родового *розумне місто*. У свою чергу, кожен цей термін вступає в родо-видові відношення з іншими терміноодиницями. Наприклад, термін *розумна мобільність* є родовим до термінів *розумна транспортна система* та *екомобільність*, а термін *розумна транспортна система* – родовим до *розумний транспорт* та *розумні громадські перевезення*, останній, у свою чергу, є родовим до терміна *розумний громадський транспорт* та ін. Приклад родо-видових відношень предметної області «Розумне місто» подано на рисунку 4.1.

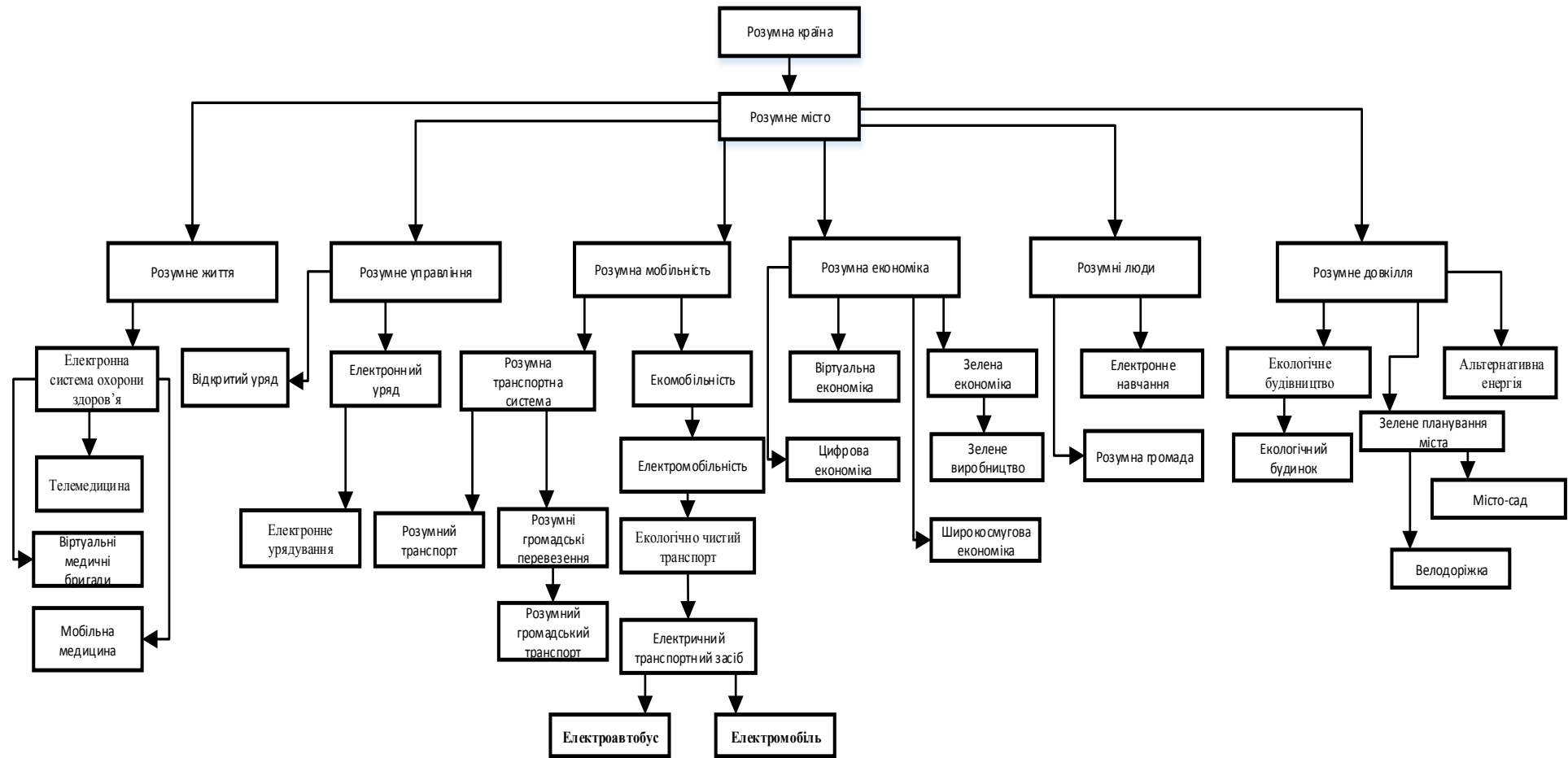


Рис. 4.1 Приклад родо-видових відношень у тезаурусі «розумного міста»

Відношення «частина-ціле» доповнює ієрархії, сформовані на основі родо-видових відношень, тим самим утворюючи складні системи. Цей тип відношень встановлюється між поняттями, коли залежне поняття частково володіє ознаками головного й входить до його складу. Зауважимо, що відношення «частина-ціле» не завжди можна описати структурою «є частиною / має частину». У нашому тезаурусі ми виділили ще такі типи меронімії, які позначили як «є представником», «є елементом» (рис 4.2), наприклад, *сонячний колектор «є елементом» активної системи сонячного теплопостачання.*

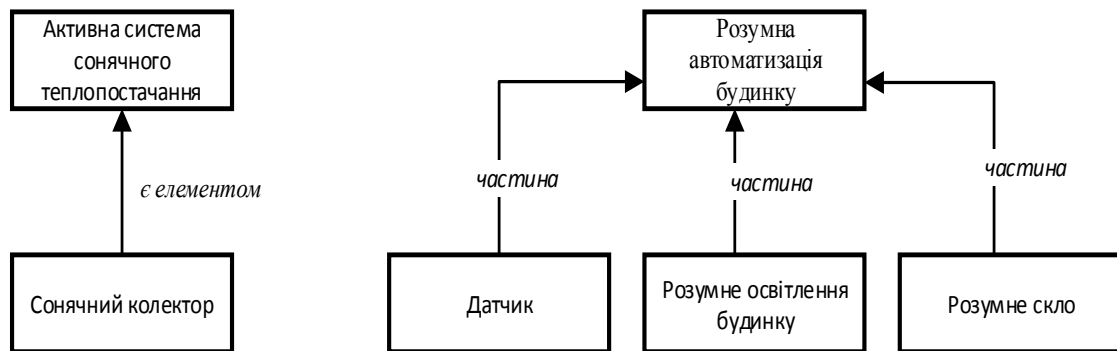


Рис. 4.2 Приклад відношення «частина-ціле»
в тезаурусі «розумного міста»

Одним із видів ієрархічних відношень є зв'язок «екземпляр-клас» (рис.4.3), який будується між поняттями, коли підпорядковане поняття (екземпляр) є окремим предметом із множини таких же предметів, що входять до одної групи (класу).

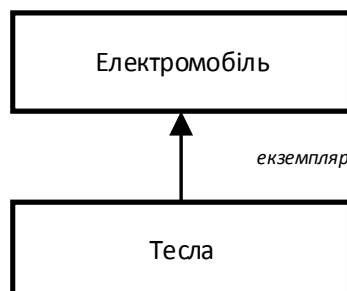


Рис. 4.3 Приклад відношення «екземпляр-клас»
у тезаурусі «розумного міста»

До традиційних відношень належить і відношення еквівалентності (синонімії), які встановлюються між термінами, що позначають одне і теж поняття (подібні терміни знаходяться у відношенні синонімії). Відношення синонімії дозволяє встановити рівність між різними термінами чи стверджувати, що два терміни є семантично тотожними чи ні. Такі терміни належать до однієї категорії, мають тотожне значення і взаємозамінні в усіх контекстах. Стандартний вираз – *A* тотожне *B*. Приклади еквівалентних відношень між термінами: *розумне місто (smart city) – інтелектуальне місто (intelligent city)*, *стійкі джерела енергії (sustainable energy sources) – відновлювальні джерела енергії (renewable energy sources) – альтернативні джерела енергії (alternative energy sources)*, *зелений транспорт (green vehicle) – екологічно чистий транспорт (eco-friendly vehicle)*, *екомісто (eco-city) – екологічно чисте місто (eco-friendly city)*, *мобільне здоров'я (mobile health (m-health) – дистанційна медицина (remote medicine) – телемедицина (telemedicine)*, *відкрита влада (open government) – відкритий уряд (open government)*, *зелене будівництво (green construction) – екологічне будівництво (ecological construction)* та ін.

Для більш повного розкриття специфіки досліджуваної предметної області «Розумне місто» разом із визначенням традиційних відношень між термінами, було виділено ряд особливих, характерних для цієї галузі, проблемно-специфічних (асоціативних) відношень.

Деякі науковці дотримуються думки, що асоціативні відношення встановлюються між термінами один з яких визначає предмет, а інший вказує на галузь його застосування, також між термінами, що визначають близькі поняття [253]. Американський стандарт подає асоціативний зв'язок як відношення між термінами, коли один з них обумовлює використання іншого [99].

Головним завданням встановлення асоціативних відношень між дескрипторами інформаційно-пошукового тезауруса «розумного міста» є те,

що побудова таких зв'язків може вказати додаткові дескриптори, корисні при індексуванні або пошуку.

З одного боку, наявність традиційних відношень дозволяє будувати онтології з можливістю об'єднання і повторного використання, а з іншого боку, для того щоб інтелектуальна система, побудована на онтології, працювала адекватно, необхідна досить детальна фіксація проблемно-специфічних відношень між поняттями предметної області. Існує певний баланс між детальністю визначення проблемно-специфічних відношень і суб'єктивністю конкретної групи експертів.

У процесі дослідження терміносистеми «розумного міста» була розроблена класифікація проблемно-специфічних відношень між дескрипторами тезауруса. До цієї групи відносяться такі зв'язки:

- «виробляє» (*A «виробляє» B*): *відновлювальна енергетика – відновлювальна енергія;*
- «є джерелом утворення» (*A «є джерелом утворення» B*): *відновлювальні джерела енергії – відновлювальна енергія;*
- «є результатом програми» (*A «є результатом програми» B*): *зелені будівлі – зелене будівництво;*
- «використовується в» (*B «використовується в» A*): *активна система сонячного теплопостачання та пасивна система сонячного теплопостачання – розумний будинок; альтернативне паливо – екологічно чистий транспорт, великі дані – розумне місто; віртуальний асистент – розумний прилад;*
- «оцінюється» (*B «оцінюється» A*): *якість води, якість повітря, якість ґрунту – екологічний моніторинг; сталий розвиток – індикатор сталого розвитку;*
- «визначає» (*A «визначає» B*): *індекс якості повітря – якість повітря;*

- «включає» (А «включає» В): відновлювальна енергія – вітроенергетика, гідроенергетика, сонячна енергетика; інтегрована система – інтегрований транспорт;
- «використовує» (А «використовує» В): будівля енергозберігаюча – відновлювальна енергія, електронний уряд – інформаційно-комунікаційні технології, активний будинок – альтернативні джерела енергії, розумний уряд – краудсорсинг;
- «контролює» (А «контролює» В): система контролю за збором сміття – відходи;
- «ґрунтується на принципах» (А «ґрунтується на принципах» В): сталий туризм – сталий розвиток;
- «є напрямком» (В «є напрямком» А): екологічний туризм – сталий туризм;
- «галузь» (В «галузь» А): інформаційні технології – високі технології;
- «підключений до» (В «підключений до» А): віддалена робоча станція – локальна мережа, розумний дім – інтернет речей;
- «сприяє» (В «сприяє» А): відкриті інновації – сталий розвиток, розумна економіка – сталий розвиток
- «потребує» (А «потребує» В): відходи – утилізація, екомісто – вуглецевий слід;
- «основа створення» (В «основа створення» А): геоінформаційні технології – географічні інформаційні системи;
- «може бути» (В «може бути» А): екологічно чистий – екологічно чисте місто, екологічно чистий транспорт, екологічний будинок;
- «зберігає» (А «зберігає» В): електронний гаманець – електронні гроші;
- «надходить з» (В «надходить з» А): зелена енергія – відновлювальні джерела енергії;

- «спеціальний тариф на» (В «спеціальний тариф на» А): *зелений тариф – зелена енергія*;
- «впроваджується в» (В «впроваджується в» А): *інтернет речей – розумний будинок*;
- «фінансує» (А «фінансує» В): *громадське фінансування – стартап-компанія*;
- «описує» (В «описує» А): *міський протокол – розумне місто*;
- «пов'язаний з» (В «пов'язаний з» А): *мобільне навчання – дистанційне навчання, розумна навігація – розумна парковка*;
- «приймається в» (В «приймається в» А): *розумні рішення – розумне місто*;
- «здійснюється в» (В «здійснюється в» А): *електронна комерція – інтернет*;
- «забезпечує підключення до» (В «забезпечує підключення до» А): *широкосмуговий доступ – Інтернет*;
- «є принципом» (В «є принципом» А): *щільність забудови, екологічно чистий транспорт, сталий розвиток, енергоефективність – новий урбанізм*;
- «оцінює» (А «оцінює» В): *протокол міста – екологічна стійкість, якість життя*;
- «транзакція» (В «транзакція» А): *мікроплатіж – електронна комерція та ін.*

На рисунку 4.4 представлено зв'язки дескриптора *розумна мобільність*. Подана структура включає ієрархічні «родо-видові» відношення (*розумна мобільність – розумна транспортна система, екомобільність, екомобільність – електромобільність, розумна транспортна система – розумний транспорт, розумний транспорт – розумний громадський транспорт, електричний транспортний засіб – електроавтобус, електромобіль та ін*), «частина-ціле» (*розумна парковка – давач, гібридний автомобіль – бортовий комп'ютер*) та «екземпляр-клас» (*електромобіль –*

тесла). Подано й відношення еквівалентності (на рисунку позначено штриховою лінією) між дескриптором *екологічно-чистий транспорт* та аскрипторами *екомобіль*, *зелений транспорт*, дескриптором *біопаливо* та аскриптором *альтернативне паливо*. Групу проблемно-специфічних відношень складають такі зв'язки: «використовує» (*електромобільність – технологія OLEV, екологічно чистий транспорт – біопаливо*), «включає» (*розумна транспортна система – система управління дорожнім рухом, система спільного використання велосипедів, розумна парковка*), «керує» (*система управління дорожнім рухом – розумний світлофор, камера безпеки дорожнім рухом*), «пов'язаний з» (*розумна мобільність – стала мобільність*) та ін.

Певну систему відношень має дескриптор *розумне довкілля*, частину яких подано на рисунку 4.5. Цей дескриптор є родовим до дескриптора *екологічно чисте місто*. Ієрархічні родо-видові відношення задані між дескрипторами: *відходи – відходи небезпечні, відходи тверді; утилізація – зелена утилізація, утилізація замкнута; альтернативна енергетика – геліоенергетика, геотермальна енергетика, вітроенергетика, біоенергетика, гідроенергетика*. Класи еквівалентності встановлені між термінами: *екологічно чисте місто – екомісто, екологічне будівництво – екобудівництво, зелене будівництво; екологічний будинок – екобудинок, зелений будинок; альтернативна енергетика – відновлювальна енергетика; альтернативні джерела енергії – відновлювальні джерела енергії; альтернативна енергія – відновлювальна енергія*. Асоціативні зв'язки представлено такими типами: «контролює» (*екологічно чисте місто – відходи*), «використовує» (*екологічно чисте місто – зелена інфраструктура, альтернативна енергетика, екологічний будинок – альтернативні джерела енергії*), «здійснює» (*зелена інфраструктура – екологічний моніторинг*), «дає» (*альтернативні джерела енергії – альтернативна енергія*), «керується» (*відходи – управління відходами*) та ін.

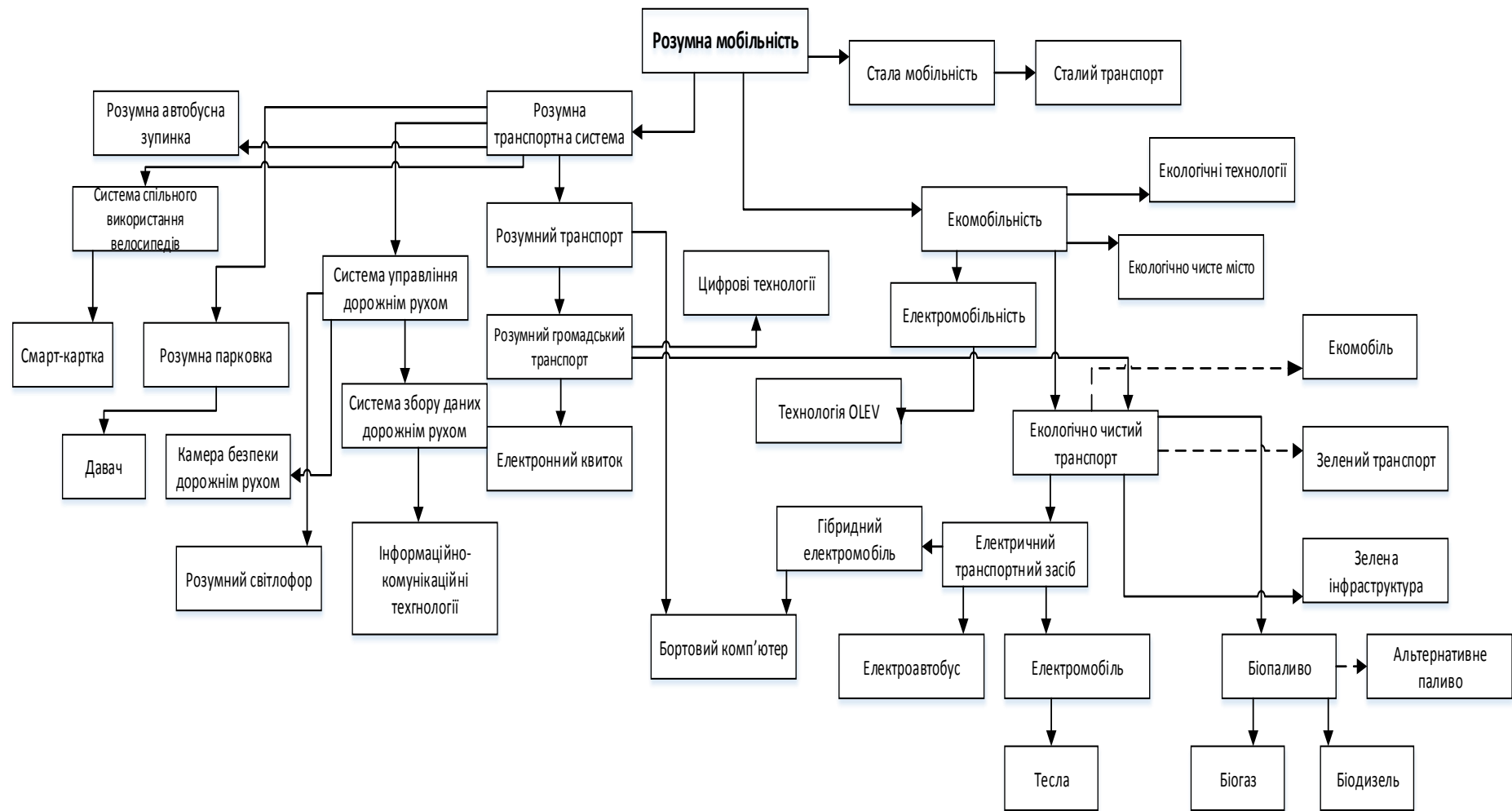


Рис. 4.4 Приклад відношення дескриптора *розумна мобільність* у тезаурусі «розумного міста»

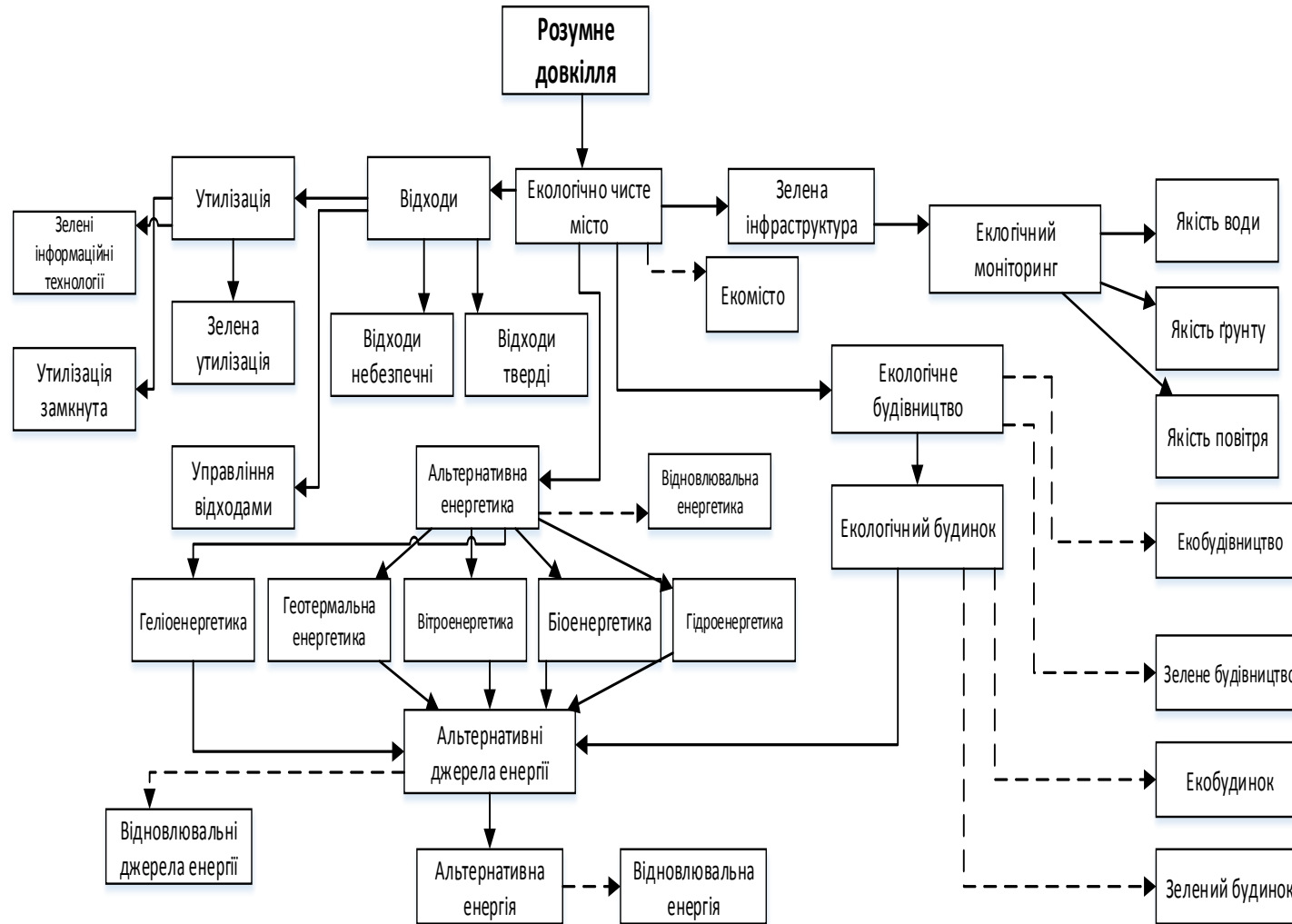


Рис. 4.5 Приклад відношення дескриптора *розумне довкілля* у тезаурусі «розумного міста»

Розроблена класифікація асоціативних відношень між дескрипторами інформаційно-пошукового тезауруса «розумного міста» дозволяє зробити опис кожного його терміна більш повним. Детальна класифікація цих зв'язків дозволяє більш точно вираховувати функцію семантичної близькості двох понять, що може істотно підвищити ефективність автоматичних систем, побудованих тезаурусним методом.

Досліджуючи систему термінів предметної області «Розумне місто», можна зробити висновок, що родо-видові відношення є основним фактором побудови тезауруса цієї області. Ці відношення розглядаються як тип семантичних відношень, в якому лексичні одиниці, що означають вид семантично підпорядковуються лексичним одиницям, що відповідно означають рід. Таким чином було виявлено велику кількість термінів, для яких характерні родо-видові відношення: *утилізація – утилізація замкнута, утилізація зелена; відходи – відходи тверді, відходи небезпечні; система сонячного теплопостачання – активна система сонячного теплопостачання, пасивна система сонячного теплопостачання, слід – вуглецевий слід, екологічний слід, водний слід; активний будинок – будинок енергозберігаючий, будинок з нульовим викидом вуглецю* та ін.

Отже, для розробленого інформаційно-пошукового тезауруса міждисциплінарної галузі «Розумне місто» характерна семантика, яка визначає зв'язки між термінами (ієрархія, синонімія, асоціація). Ці відношення не є фіксованими. Така особливість тезауруса міждисциплінарного дослідження дозволяє представляти внутрішню організацію кожного окремого тематичного напрямку та використовувати тезаурус у вирішенні завдань наукового пошуку. Для формального представлення термінологічної бази міждисциплінарного наукового напрямку досліджень можна використовувати прості зв'язки між основними термінами, а також відношення синонімії. Крім цього, у структурі подання враховуються окрім основної галузі й інші галузі, які беруть участь у формуванні міждисциплінарного наукового напрямку.

З допомогою формування тезауруса забезпечується побудова моделі даного міждисциплінарного напрямку наукових досліджень способом класифікації та ранжирування термінів-концептів.

4.3 Структура інформаційно-пошукового тезауруса «розумного міста»

Тезаурус «розумного міста» складається зі вступу; алфавітного лексико-семантичного покажчика термінів; пермутаційного покажчика.

Тезаурус «розумного міста» включає 649 словникових статей (з них 560 дескрипторів і 89 аскриптори), що відображають різні аспекти «розумного міста».

У вступі подано такі дані:

1. Мета створення і галузь застосування інформаційно-пошукового тезауруса;
2. Посилання на джерела, які використовувалися для збору лексичних одиниць тезауруса (монографій, публікацій результатів теоретичних та експериментальних досліджень, статей у періодичних фахових виданнях, доповідей конференцій та ін.);
3. Опис структури тезауруса;
4. Опис алфавітного лексико-семантичного покажчика;
5. Перелік класів відношень;
6. Перелік посилань для дескрипторів;
7. Перелік посилань для аскрипторів;
8. Опис пермутаційного покажчика.

Основною частиною тезауруса «розумного міста» є алфавітний лексико-семантичний покажчик. У ньому представлено всі дескриптори та аскриптори зі словниковими статтями. Кожен термін має статус дескриптора, який використовується при індексуванні документів та запитів, чи аскриптора, який не рекомендовано використовувати в системі й замінюється

відповідним дескриптором. Елементом алфавітного лексико-семантичного покажчика є словникова стаття. Усі словникові статті впорядковані за алфавітом.

Для вибору дескрипторів у класах еквівалентності бралася одна лексична одиниця як представник всього класу, і критеріями вибору були:

- повнота вираження значень класу;
- відповідність стандартам і рекомендаціям з науково-технічної термінології;
- стислість і зрозумілість;
- при рівних умовах – частота використання (поширеність) в документах і запитах.

Статус аскриптора визначався в таких випадках:

- лексична одиниця входить у клас еквівалентності, але не є представником класу;
- значення лексичної одиниці при пошуку інформації доцільно представити як перетин значень інших лексичних одиниць, які отримали статус дескриптора;
- значення лексичної одиниці при пошуку інформації доцільно подати як об'єднання значень інших лексичних одиниць, несумісних в одному пошуковому образі.

В інших випадках лексична одиниця отримує статус дескриптора.

Словникова стаття дескриптора або дескрипторна стаття складається із заголовного дескриптора, зв'язаного парадигматичними відношеннями з іншими дескрипторами чи аскрипторами, а також дескрипторів та умовних синонімів, пов'язаних із заголовним дескриптором за змістом.

Усі види зв'язків у тезаурусі «розумного міста» відображаються в словникових статтях дескрипторів і аскрипторів.

До відношень дескрипторної статті входять три класи :

- ієрархічні відношення – базове відношення, де родовий термін подає клас або ціле, а видовий термін стосується його членів або частин.

Взаємні відношення виражаються такими скороченнями: *B* – ширший (вищий) термін, *H* – вузьчий (нижчий) термін;

– відношення синонімії заголовного дескриптора з аскриптором, що означає, що представником аскриптора в системі є заголовний дескриптор, який використовується замість нього при координатному індексуванні документів і, відповідно, запитів;

– асоціативні відношення – це будь-який вид смислових відношень між дескрипторами, можливий у цій галузі, крім відношень синонімії; посилання на семантичні (неієрархічні) зв'язки з іншими дескрипторами. Вони охоплюють відношення між парами дескрипторів, що не є членами множини еквівалентності і не можуть бути впорядковані за ієрархією, в якій один з дескрипторів підпорядкований іншому; однак вони уявно пов'язані один з одним у такій мірі, що цей зв'язок має бути явним, для того щоб виявити в тезаурусі альтернативні одиниці, які можна було б використати з метою індексування чи пошуку. Одним з критеріїв встановлення асоціативних відношень між термінами є досить часте їхнє використання в літературі разом. Асоціативний зв'язок доречно сприймати як рекомендацію звернути увагу на терміни, які мають якесь відношення до індексованого поняття, тобто пересічні з ним у якомусь аспекті відображення. Ці відношення є взаємними і позначаються за допомогою скорочення *A* (асоціативний термін). У таблиці 4.2 подані позначки, які використовуються в словникових статтях дескрипторів та аскрипторів для позначення парадигматичних зв'язків між термінами.

Табл. 4.2

Значення позначок у словникових статтях дескрипторів і аскрипторів інформаційно-пошукового тезауруса «розумне місто»

Позначка	Значення позначки	Статус заголовного терміна статті	Вид зв'язку
<i>B</i>	вищий термін	дескриптор	ієрархічний
<i>H</i>	нижчий термін	дескриптор	ієрархічний

<i>С</i>	термін-синонім	аскриптор	синонімія
<i>А</i>	асоціативний	дескриптор	асоціативний
<i>див.</i>	відсилання до українського дескриптора	аскриптор-синонім	синонімія
<i>англ.</i>	аскриптор англійською мовою	аскриптор	синонімія
<i>USE</i>	відсилання до англійського аскриптора	аскриптор-синонім	синонімія

Відсилання від одного терміна до іншого позначають зв'язок цієї терміноодиниці з іншими і є результатом виконання таких операцій:

- усунення неоднозначності лексичної одиниці;
- встановлення відношень еквівалентності;
- вибору дескриптора, що представляє клас еквівалентності при індексуванні;
- встановлення ієрархічних та асоціативних відношень дескрипторів.

Класифікаційна схема окремого поняття складається з ієрархії дескрипторів, відношень синонімії та асоціативних зв'язків.

У дескрипторній статті терміни розподілені в такому порядку:

- заголовний дескриптор українською мовою;
- визначення заголовного дескриптора українською мовою;
- представлення заголовного дескриптора англійською мовою;
- визначення заголовного дескриптора англійською мовою;
- умовні синоніми до заголовного дескриптора;
- вищі дескриптори (ширші);
- нижчі дескриптори (вужчі);
- асоціативні дескриптори.

Заголовний дескриптор є обов'язковим елементом словникової статті дескриптора й супроводжується тлумачною частиною (визначенням) в обох мовах. Інші дескриптори та аскриптори є факультативними. Використавши відповідний термін-дескриптор можна не лише здійснити пошук потрібного

інформаційного об'єкта, але й дізнатися про значення терміна, що символізує розшукувану інформаційну одиницю.

У дескрипторній статті, лексичні одиниці, які пов'язані з заголовним дескриптором одним із видів зв'язку, розміщуються в алфавітному порядку. Наприклад, нижчі та асоціативні дескриптори розташовуються за алфавітом (окремо) у словниковій статті дескриптора *сталий розвиток*, яка має такий вигляд:

сталий розвиток

розвиток, який задовольняє потреби теперішнього часу, не ставлячи під загрозу здатність майбутніх поколінь задовольняти свої власні потреби, включаючи їхню потребу в безпечному й здоровому довкіллі

англ. sustainable development
development that meets the needs of the present without compromising the ability of future generations to meet their own needs, including their need for a safe and healthy environment

В розумне місто
Н безвідходне виробництво
Н зелена економіка
Н низьковуглецеве виробництво
Н стале виробництво
Н стале споживання
Н сталий спосіб життя
А індикатор сталого розвитку
А інформаційні технології

Словникова стаття дескриптора містить посилання (умовні позначки), які проставляються перед термінами і використовуються для відображення взаємозв'язків між термінами тезауруса. Кожне посилання вказує відношення чи функцію терміна, що йдуть за ним і розташовуються у вказаному нижче порядку для позначення парадигматичних відношень з іншими термінами:

Заголовний дескриптор (українською мовою)

визначення заголовного дескриптора (українською мовою)
англ. англійський відповідник заголовного дескриптора
визначення заголовного дескриптора (англійською мовою)
С умовний синонім
В вищий дескриптор
Н нижчий дескриптор
А асоціативний дескриптор

У дескрипторній статті заголовний дескриптор виділяється жирним шрифтом та малими літерами.

Побудова словникових статей інформаційно-пошукового тезауруса й укладання алфавітного лексико-семантичного покажчика зводиться до дескрипторизації ключових слів та визначення парадигматичних відношень між дескрипторами. При формуванні словникових статей тезауруса термінам приписуються відповідні посилання. Дескрипторизація здійснюється для побудови дескрипторного словника, а задання парадигматичних відношень використовується для збільшення семантичної сили мови. Дескрипторизація дозволяє вилучати неоднозначність (омонімія та полісемія) ключових слів, за допомогою приписування ключовому слову відповідного пояснення, що відносить його до визначеної категорії. Неоднозначність (синонімія) усувається через групування ключових слів у класи умовної еквівалентності, з яких вибирається один термін – дескриптор, що є іменем класу синонімічних або близьких за змістом ключових слів.

Позначка *C* (синонім) – це дескриптор, що замінюється при індексуванні на заголовний дескриптор. Цією позначкою вводяться в дескрипторну статтю умовні синоніми заголовного дескриптора, розташовані за алфавітом. Разом із заголовним дескриптором вони утворюють клас умовної еквівалентності.

Лексичні одиниці є еквівалентними в інформаційно-пошуковому тезаурусі, якщо при заміні однієї лексичної одиниці при пошуку документа на іншу видача документів буде такою ж як і за першою одиницею. Такі лексичні одиниці утворюють один клас еквівалентності, додаються до відповідної дескрипторної статті й позначаються символом «*C*».

Еквівалентність встановлюється між:

– абсолютними синонімами, зокрема – між аббревіатурами і повними формами лексичних одиниць: *4G* = *четверте покоління мобільного зв'язку*, *відкритий уряд* = *відкрита влада*, *дистанційне навчання* = *електронне навчання*;

– стилістичними синонімами (якщо відмінність у сфері вживання несуттєва для задач інформаційно-пошукового тезауруса): *екологічне будівництво* = *зелене будівництво*, *вільні виробничі площі* = *браунфілд*, *розумна автоматизація* = *цифрове управління будинком*.

Позначкою *B* (вищий) у словниковій статті дескриптора споряджається дескриптор, який може означати або родовий термін, який передає більш загальне, ширше за обсягом поняття, або ціле по відношенню до цього (заголовного) дескриптора, який, у свою чергу, є видовим або означає частину цього цілого, і називається вищий або ширший дескриптор.

Позначкою *H* (нижчий) уводиться в дескрипторну статтю дескриптор, який може означати або видове поняття з більш конкретним значенням, або якусь частину по відношенню до поняття, яке представлене у тезаурусі вищим дескриптором, і має назву нижчий чи вузький дескриптор, тобто видовий дескриптор.

Позначкою *A* (асоціативний), включається до тезаурусної дескрипторної статті дескриптор, який пов'язаний із заголовним дескриптором іншими видами семантичних відношень, і має назву асоціативний. Цей термін, певним чином пов'язаний, споріднений з цим поняттям, але не є його еквівалентом (синонімом), родовим чи видовим поняттям.

Приклад дескрипторної словникової статті дескриптора *розумне місто*:

розумне місто

ефективна інтеграція фізичних, цифрових і людських систем у штучному середовищі для забезпечення сталого, благополучного майбутнього своїх громадян

англ. smart city

effective integration of physical, digital and human systems in the built environment to deliver a sustainable, prosperous and inclusive future for its citizens

C інтелектуальне місто

B розумна країна

H розумна економіка

H розумна мобільність

H розумне довкілля

H розумне життя

H розумне управління

H розумні люди

<i>A</i>	зелене місто
<i>A</i>	Інтернет речей
<i>A</i>	місто майбутнього
<i>A</i>	цифрове місто

Аналіз словникової статті дескриптора *розумне місто* показує, що синонімічним до терміна *розумне місто* є термін *інтелектуальне місто*, який у тезаурусі подається як аскриптор; родовим поняттям є поняття *розумна країна*, а видовими поняттями виступають *розумна економіка*, *розумна мобільність*, *розумне довкілля*, *розумне життя*, *розумне управління*, *розумні люди*. Крім цього є поняття: *зелене місто*, *Інтернет речей*, *місто майбутнього*, *цифрове місто*, що якимось чином пов'язані з поняттям *розумне місто*, але не впорядковані з ним за ієрархією. Саме вони пов'язані асоціативними зв'язками з дескриптором *розумне місто*.

На кожну лексичну одиницю, подану в словниковій статті дескриптора *розумне місто* розробляється й подається в тезаурусі окрема словникова стаття. Наприклад, словникові статті на дескриптори *розумна мобільність*, *розумне довкілля*, *розумне управління* та аскриптор *інтелектуальне місто* мають такий вигляд:

розумна мобільність

	продуктивне переміщення людей, товарів та інформації
<i>англ.</i>	smart mobility
	productive movement of people, information and goods
<i>B</i>	розумне місто
<i>H</i>	екологічно чистий транспорт
<i>H</i>	розумна транспортна система

розумне довкілля

	зменшення забруднення, спричиненого промисловістю та людською діяльністю, з тим щоб зменшити парниковий ефект; раціоналізація споживання енергії у всіх секторах; а також управління насадженнями зелених рослин
<i>англ.</i>	smart environment
	a reduction of the pollution produced from industries and people in order to reduce the greenhouse effect, the rationalization of the consumptions of energy in all of the sectors and the production and management of green spaces
<i>B</i>	розумне місто
<i>H</i>	альтернативна енергетика
<i>H</i>	відходи
<i>H</i>	екологічне будівництво
<i>H</i>	екологічний моніторинг
<i>A</i>	сталий розвиток

розумне управління

	прозорість управління й участь громадян у процесі прийняття рішень
англ.	smart-governance
	transparency of management and citizens' participation in decision-making processes
Н	відкритий уряд
А	інформаційно-комунікаційні технології
А	хмарні технології

інтелектуальне місто

див.	розумне місто
англ.	intelligent city
USE	smart city

У дескрипторній статті вказуються всі синоніми заголовного дескриптора. Для кожного синонімічного дескриптора вказується один дескриптор.

екологічно чистий транспорт

транспортний засіб, що має менш шкідливий вплив на навколишнє середовище, ніж транспорт з двигуном внутрішнього згоряння, що працює на бензині або дизельному паливі

англ.	eco-friendly vehicle
	vehicle produces less harmful impacts to the environment than comparable conventional internal combustion engine vehicles running on gasoline or diesel
С	зелений транспорт
С	сталий транспорт
В	розумний транспорт
А	екологічно чистий
А	екомобільність
А	зелена інфраструктура
А	розумна інфраструктура

зелений транспорт

див.	екологічно чистий транспорт
англ.	green vehicle
USE	eco-friendly vehicle

сталий транспорт

див.	екологічно чистий транспорт
англ.	sustainable transport
USE	eco-friendly vehicle

Приклад показує, що дескриптор *екологічно чистий транспорт* має два синоніми: *зелений транспорт* та *сталий транспорт*, що виступають дескрипторами й кожен з яких має свою дескрипторну статтю.

Словникова стаття аскриптора, як і стаття дескриптора, містить заголовне слово – аскриптор та відповідне відсилання до дескриптора, що є його умовним еквівалентом. Більшості аскрипторів однозначно відповідає визначений дескриптор. Якщо в дескрипторній статті заголовний дескриптор виділяється жирним шрифтом, то в аскрипторній статті аскриптор виділяється курсивом. Від кожного аскриптора дається відсилання до дескриптора, що заміняє його при індексуванні. У цьому випадку використовується відсилання *див.* (дивіться) (в українській частині) та *USE* (в англійській частині). Аскрипторна словникова стаття включає англійський відповідник аскриптора.

Словникова стаття аскриптора містить такі посилання:

Заголовний аскриптор (українською мовою)

див. дескриптор українською мовою

англ. аскриптор англійською мовою

USE дескриптор англійською мовою

Приклади аскрипторної статті:

екомотобіль

див. екологічно чистий транспорт

англ. eco-car

USE eco-friendly vehicle

інтелектуальне місто

див. розумне місто

англ. intelligent city

USE smart city

цифрове управління будинком

див. автоматизація будівель

англ. digital home control

USE building automation

Алфавітний лексико-семантичний покажчик словникових статей використовується для пошуку потрібного дескриптора та його парадигматичних зв'язків. Цей покажчик забезпечує можливість адекватного формулювання запиту, а при незадовільних результатах пошуку – зміни

стратегії пошуку, зокрема використовуючи дескриптори, що входять у словникову статтю.

Допоміжним показником до алфавітного лексико-семантичного показника в інформаційно-пошуковому тезаурусі «розумного міста» є пермутаційний показник. Він містить дескриптори й аскриптори без словникових статей. Дескриптори виділені жирним шрифтом, аскриптори – курсивом.

Особливістю пермутаційного показника є формування словникових гнізд за словами, що входять до складу термінів-словосполучень.

Упорядкування проводиться за алфавітом кожного окремого слова, що входить у словосполучення, яке позначає дескриптор чи аскриптор, і вказані всі дескриптори та недескриптори, до складу яких входить це слово. Відтак, кожен термін зустрічається в пермутаційному показнику стільки разів, зі скількох слів він складається. Таким чином, для кожного такого слова формується словникове гніздо. Залежно від кількості вхідних у них слів, дескриптори й аскриптори можуть дублюватися в декількох словникових гніздах.

Пермутаційний показник використовується для пошуку потрібних дескрипторів у відповідному словниковому гнізді, за окремими словами, що входять у терміноодиницю, зокрема і за тими, які не стоять на початку словосполучення. Наприклад, якщо користувача цікавить термін «мережа», який зазвичай входить до складу різних словосполучень, то йому потрібно звернутися до відповідного словникового гнізда:

	мережа
бездротова	мережа
глобальна	мережа
<i>інтелектуальна</i>	<i>мережа (ІМ)</i>
кампусна	мережа
локальна комп'ютерна	мережа
міська	мережа
розумна	мережа водопостачання
телекомунікаційна	мережа

Кожен термін словникового гнізда може бути включений у інші гнізда, наприклад, за першим словом словосполучення:

цифрова економіка
цифрова інфраструктура
цифрова карта
цифрове включення
цифрове місто
цифрове управління будинком
цифровий уряд
цифрові технології

Включення пермутаційного покажчика в інформаційно-пошуковий тезаурус дозволяє уникнути повторень, отримати високий рівень накопичення термінів, які пов'язані родо-видовими й асоціативними відношеннями, виявити групи однорідних понять, які не були представлені повністю, і доповнити їх, увести пропущені поняття. У пермутаційному покажчику всі терміни, які містять однакові слова, зібрані разом, що дає можливість побачити та використати ті терміни, які потенційно можуть виявитися необхідними при пошуку.

Вибрані в пермутаційному покажчику терміни доцільно перевіряти в алфавітному лексико-семантичному покажчику, де представлені словникові статті дескрипторів з семантичними зв'язками, а також синоніми (аскриптори) з дескрипторами, які замінюють їх при пошуку.

Інформаційно-пошуковий тезаурус «розумного міста» представляє систему термінів предметної області «Розумне місто», з урахуванням складності й багатовимірності семантичних взаємозв'язків, що дозволяє користувачу або комп'ютеру орієнтуватися в ній.

Тезаурус є зручним способом ієрархічної організації сукупності понять предметної області «Розумне місто» з усіма лексичними одиницями, що належать до цієї галузі, які утворюють особливі множини – семантичні поля. За його допомогою можна не лише здійснювати пошук, а й вивчати терміни (поняття), отримувати семантичні зв'язки з іншими термінами (поняттями),

з'ясовуючи їхню роль у системі знань даної галузі чи в ході розв'язання конкретної задачі [162].

Завдання інформаційно-пошукового тезауруса «розумного міста» – уніфікувати ключові слова, відібрані зі змісту документів, тобто створити пошукові образи цих документів у вигляді дескрипторів і аскрипторів, які замінюються відповідними дескрипторами, для подальшого їхнього пошуку. Тезаурус забезпечує такі можливості пошуку термінів при формулюванні запиту: пошук словникових статей за заголовними дескрипторами та пошук дескрипторів і аскрипторів у словниковому гнізді.

Необхідно звернути увагу на той факт, що тезаурус не може вважатися повністю сформованим, закінченим, оскільки його зміст, обсяг та форма постійно будуть змінюватися з розвитком науки, техніки та виробництва. На внесення змін у тезаурус можуть впливати кількісні та якісні зміни документопотоку, що поступає в систему; зміни режимів пошуку в системі; зміни засобів реалізації автоматизованих інформаційних систем; перехід автоматизованих інформаційних систем від незалежного функціонування до роботи в режимі мережі.

Підтримання інформаційно-пошукового тезауруса «розумного міста» в сучасному стані потребує корегування його термінологічного складу, що передбачає поповнення новими лексичними одиницями, зміни статусу певних лексичних одиниць; зміни парадигматичних відношень у тезаурусі.

Упорядкування понятійно-термінологічної системи пов'язано зі структуризацією галузі, якою є предметна область «Розумне місто». У ході наукового дослідження відбувається тезаурусне моделювання знань, визначення понять, які направлені на формування моделі нового наукового знання в площині дослідження «розумних міст». Сформований тезаурус є моделлю знань предметної області «Розумне місто».

Тезаурусне моделювання знання предметної області «Розумне місто» ґрунтується на системі понять та зв'язків між ними. Проаналізувати концепції та теорії, означає порівняти понятійні бази різних галузей наук та

виділити з них спільне, яке породжується перетином понять галузей знань, що опрацьовуються. Це дає можливість здійснювати міждисциплінарне перенесення частини теоретичного знання між цими галузями. Таким чином, вибудовується теоретична модель досліджуваної предметної області «Розумне місто».

Так, наприклад, термінами: *віртуальний колектив*, *геоінформаційні технології* оперують науковці таких галузей як соціальні комунікації, інформаційні технології, телекомунікації, економіка та ін. (звичайно, що вони входять у термінологічну систему й інших галузей). Усі вони повинні мати однакове семантичне навантаження для цих суміжних галузей.

Віртуальний колектив – групи людей, які працюють разом у різних географічних місцях і покладаються на інформаційно-комунікаційні технології, такі як електронна пошта, факс, відео або голосові послуги конференц-зв'язку для того, щоб співпрацювати.

англ. *Virtual team* – a group of individuals who work together from different geographic locations and rely on communication technology such as email, FAX, and video or voice conferencing services in order to collaborate.

Геоінформаційні технології – технологічна основа створення географічних інформаційних систем, що дозволяють реалізувати їхні функціональні можливості.

англ. *Geoinformation technologies* – technological basis for the creation of geographic information systems that allow them to realize their functionality

Терміни: *екологічний будинок*, *зелена інфраструктура* належать до галузей охорона навколишнього середовища, містобудування, інформаційні технології, економіка та ін.

Екологічний будинок – будинок, розроблений для того, щоб бути екологічно чистим і стійким та спрямованим на ефективне використання енергії, води і будівельних матеріалів.

англ. *Ecological house* – a house designed to be environmentally friendly and sustainable and aimed at the efficient use of energy, water and building materials.

Зелена інфраструктура – мережа об'єктів, що забезпечують «компоненти» (управління якістю води, поліпшення якості повітря, заходи по адаптації до змін клімату, стабільне вироблення екологічно чистої енергії, виключення втрат тепла, будівництво енергозберігаючих споруд, підвищення біорізноманіття, виробництво екологічно чистих продуктів харчування, забезпечення екологічно збалансованого використання водних ресурсів і ресурсів ґрунтів) для вирішення міських і кліматичних проблем, при цьому обов'язковою умовою є виконання принципу «будівництва з природою».

англ. *Green infrastructure* – a network of facilities providing "components" (water quality management, air quality improvement, measures to adapt to climate change, sustainable development of clean energy, elimination of heat losses, construction of energy saving constructions, biodiversity enhancement, production of environmentally friendly food products, provision of ecologically balanced use of water resources and soil resources) to address urban and climatic problems, while the necessary condition is the implementation of the principle of "building with nature".

Термін *великі дані* є спільним для галузей інформаційних технологій, медицини, енергетики, економіки, соціальних комунікацій, телекомунікації, містобудування та ряду інших сфер. Тлумачення його повинне бути ідентичним для них.

Великі дані – великий обсяг, висока швидкість та / або велика різноманітність інформаційних ресурсів, що потребують нових форм обробки, що дає змогу розширювати процес прийняття рішень, виявляти розуміння та оптимізувати процес.

англ. *Big data* – high volume, high velocity, and or high variety of information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.

Отже, інформаційно-пошуковий тезаурус «розумного міста» виступає інструментом інформаційного забезпечення міждисциплінарних наукових досліджень. Він забезпечує однакове семантичне навантаження певного терміна для різних галузей, сприяючи передусім однозначному розуміння його у колі науковців цих сфер науки. Одночасно тезаурусне представлення знань галузі «Розумне місто» дозволяє працівникам бібліотеки сформулювати однозначне трактування певного терміна, забезпечує якість роботи.

Тезаурус «розумного міста» – це дослідження понять і побудова зв'язків між ними, що точно моделює структуру знання даної галузі.

Міждисциплінарний обмін знаннями у тезаурусі «розумного міста» представляється як проблема інтеграційних відношень на таких рівнях: логічному (рід-вид), системно-структурному (частина-ціле), еквівалентності (синоніми), асоціативному. Такий підхід створює синергетичний ефект, що проявляється шляхом поєднання лексичних одиниць інформаційно-пошукового тезауруса, пов'язаних між собою певними відношеннями, що відображають семантичні зв'язки між поняттями «розумного міста».

Специфіка структури інформаційно-пошукового тезауруса міждисциплінарної предметної області «Розумне місто» визначається набором наукових напрямів, що беруть участь у її формуванні. Тезаурус PAS 180 Smart cities – Vocabulary є основним і задає основні терміни всього міждисциплінарного наукового дослідження. На нашу думку, для галузі «Розумне місто» системоформуючим науковим напрямом є «цифрові технології» («інформаційні технології», «комунікаційні технології»). Виходячи з цього, ми вважаємо, що терміни цього напрямку є ключовими, а терміни, які використовуються в інших наукових дисциплінах (економіка, медицина, управління, інженерія, охорона навколишнього середовища та ін.) – допоміжними. При аналізі цифрових інформаційних ресурсів було з'ясовано, що найбільш дотичними до галузі «розумного міста», яка лише формується, є такі наукові галузі: економіка, інформаційні системи, управління, охорона навколишнього середовища, освіта, медицина,

містобудування, соціальні комунікації телекомунікації та ін. Підтвердженням цього слугує виділення видових понять: *розумне життя, розумні люди, розумна економіка, розумне довкілля, розумна мобільність, розумне управління* до родового *розумне місто*.

Співвіднесення зазначених термінів з різними галузями знання дозволяє зробити висновок про міждисциплінарний характер самого поняття «розумне місто» зокрема й даного наукового дослідження.

4.4 Функції двомовного інформаційно-пошукового тезауруса «розумного міста»

Термінологія будь-якої предметної області, зокрема і предметної області «Розумне місто», входить до загальнолітературної мови як підсистеми лексичної системи мови. Нормативний характер термінології визначається приналежністю до мовної норми, а також окремими обмеженнями та особливими вимогами (стислістю, однозначністю, системністю та інші).

Терміни однієї предметної області співвідносяться з термінами суміжних чи пов'язаних галузей, взаємодіють з ними. Закріплення термінології в тезаурусах, словниках передбачає в перспективі можливість її подальшого розвитку, поповнення та зміни. Проте ці процеси можливі тільки тоді, коли вони реалізуються в системі з вихідними поняттями, відношеннями між цими поняттями і наявністю стійкого змісту, закріпленого за цими поняттями.

Найбільш важливими й лінгвістично вагомими комплексними функціями терміносистеми є її здатність закріпити найважливішу частину галузевої термінології й представити конкретну термінологічну систему.

Двомовний інформаційно-пошуковий тезаурус «розумного міста» створювався для здійснення науково-практичних задач, таких як:

- 1) фіксації мовних проекцій сучасного рівня наукових знань і технічного розвитку;
- 2) забезпечення формалізованого подання інформації в інформаційно-пошукових системах;
- 3) вираження парадигматичних та системних відношень, що існують між термінологічними одиницями;
- 4) підвищення повноти й точності пошуку інформації в базах даних природною мовою;
- 5) створення авторитетних / нормативних файлів предметних рубрик або визначення ключових слів;
- 6) індексування документів та запитів;
- 7) уточнення інформаційних запитів;
- 8) встановлення гіпертекстових зв'язків між подібними документами чи фрагментами документів, що дає можливість користувачеві, знайшовши один релевантний документ, запросити аналогічні документи;
- 9) автоматичне виділення з текстів документів найбільш змістовних фрагментів, що забезпечує користувачеві можливість швидко переглядати масиви документів великих обсягів;
- 10) використання в ролі інструмента організації та координації терміносистеми предметної області «Розумне місто»;
- 11) аналізу термінологічного складу цієї галузі знань;
- 12) створення нормативної бази та забезпечення єдності підходів до тлумачення та використання термінологічних одиниць при формуванні єдиного наукового простору в досліджуваній предметній області;
- 13) забезпечення правильності використання термінів, що встановлює строго однозначну систему понять і сприяє тим самим упорядкуванню термінології;
- 14) реферування й анотації текстів;
- 15) редагування текстів;

16) забезпечення взаєморозуміння між сторонами, що беруть участь у соціальній комунікації та прискорення цих процесів комунікації.

У моделі запропонованого інформаційно-пошукового тезауруса «розумного міста» були реалізовані універсальні лексикографічні функції терміносистем:

1) інформаційна – завдання тезауруса полягає в зберіганні та передачі інформації; виникнення труднощів у використанні терміна часто спонукає користувачів звертатись до тезауруса в пошуках відповідної інформації, що й сприяє поглибленню їхніх знань у цій галузі;

2) комунікативна – тезаурус виступає засобом зв'язку між членами мовного колективу в процесі створення та передачі інформації;

3) систематизуюча (термінологія структурується за допомогою попередньо сформованої системи понять галузі), мета якої полягає в забезпеченні користувачів структурованими знаннями, систематизованою та упорядкованою терміносистемою предметної області «Розумне місто»; в інформаційно-пошуковому тезаурусі реалізація цієї функції здійснюється завдяки відсиланням, які відображають парадигматичні відношення між термінологічними одиницями;

4) довідкова функція вбачає завдання тезауруса в допомозі вирішення труднощів, пов'язаних з використанням лексикографічних одиниць галузі, завдяки іншомовним відповідникам, що дозволить користувачу краще орієнтуватися в специфіці мовних одиниць «розумного міста»;

5) нормативна функція дасть користувачам норми вживання мовних одиниць цієї галузі, тому що будь-який словник (зокрема й тезаурус) має усувати розбіжності між індивідуальним і колективним знанням; а оскільки наш тезаурус є двомовним, то ця функція також подає форми перекладу терміносистем;

6) навчальна функція реалізується через перекладні відповідники, що слугуватиме засвоєнню іншомовної термінології та знайомству з особливостями побудови «розумних міст» у інших країнах;

7) кумулятивна – тезаурус має здатність накопичувати та упорядковувати інформацію з метою її зберігання для різних поколінь, тобто він служить сховищем і засобом передачі позамовленнєвого колективного досвіду;

8) допоміжна функція – тезаурус складає основу веб-застосунків, які формуються в проєктах «Розумне місто»;

9) організаційна – вплив на групи людей з метою організації та координації їхньої діяльності за допомогою тезауруса;

10) функція впорядкування – тезаурус скерований на подолання необґрунтованого різноманіття понять, їхню уніфікацію й обмеження межами цієї предметної області;

11) джерела інформації, оскільки тезаурус є одним із засобів нагромадження результатів пізнавальної діяльності людства;

12) соціальна – тезаурус виступає об'єктом соціального значення, оскільки певні соціальні потреби (потреба обміну інформацією та її пошуку) сприяли його формуванню;

13) культурна – тезаурус слугує засобом закріплення й передачі культурних традицій у часі, етапів розвитку суспільної свідомості, оскільки в термінології знаходить відображення рівень розвитку системи понять;

14) пошукова – тезаурус призначений для автоматичного оброблення й пошуку інформації, що відбувається на основі зв'язків між дескрипторами.

Крім цього тезаурус «розумного міста» виконує функцію лексикографічної основи, на базі якої відбуваються процеси, які ведуть до однозначності термінів, позбавлення від термінологічного різноманіття, полісемії, омонімії і семантичних суперечностей в тлумаченні терміносистем.

Таким чином, тезаурус, з одного боку, виконує комунікативну, інформаційну, систематизуючу, довідкову, нормативну, навчальну,

кумулятивну, допоміжну, організаційну, впорядкування, джерела інформації, соціальну, культурну, пошукову функції, і з іншого – нормативну лексикографічну функцію та експлікацію стану предметної області «Розумне місто» [212].

4.5 Здійснення інформаційного пошуку релевантної інформації за допомогою тезауруса «розумного міста»

Величезні обсяги інформації, поданої у текстовій формі й до того ж різними мовами, яка розкриває напрями розвитку «розумних міст», що стали доступні в останні кілька років у зв'язку з розвитком різноманітних інформаційних та телекомунікаційних технологій, ставлять принципово нові проблеми перед постачальниками й споживачами цієї інформації, генерують потребу використання ефективних засобів автоматичного опрацювання та пошуку документів.

Ефективним інструментом формування запитів у пошукових системах мережі «Інтернет» є двомовний інформаційно-пошуковий тезаурус «розумного міста», що представляє термінологію досліджуваної галузі знань. З допомогою тезауруса можна значно підвищити якість інформаційного пошуку релевантної інформації в АБІС, завдяки зафіксованим парадигматичним відношенням між його одиницями.

Користувач бібліотеки, звертаючись до тезауруса, формує пошуковий запит документа згідно інформаційного запиту, яка видає йому результати пошуку. Йому зазвичай потрібно знайти не ті документи, в яких ужито певне слово, а ті, в яких розглядається відповідне поняття або відповідний об'єкт. Тобто, пошук повинен здійснюватися не за словами – лексичний пошук, а за їхнім змістом – семантичний пошук, який реалізується на підставі знань про відношення між поняттями предметної області «Розумне місто», що виражені природною мовою. Кожне поняття може розкриватися через набір інших понять, утворюючи семантичні поля. Носієм такого роду інформації є

інформаційно-пошуковий тезаурус «розумного міста», оскільки він є словником термінів, у якому перш за все подано відсилання від термінів до їхніх синонімів (еквівалентів), до загальних (родових) та нижчих (видових) понять. Саме при наявності в системі такого інструмента, система має можливість у відповідь на запит приєднати до видачі ті документи, які містять терміни-синоніми до пошукового терміна. Як правило приєднуються і документи, у яких зафіксовані ієрархічні поняття. При спеціальній вказівці користувача для вичерпного збору інформації сюди можуть бути додані також документи по асоціативних поняттях.

Головне призначення використання тезауруса – це інформаційний пошук, отримання релевантних документів. Пошук з використанням тезауруса дозволяє розгорнути пошукове слово в необхідних аспектах, рухаючись на один рівень нижче чи вище в структурі пошукового тезауруса, ніж це притаманне процедурі простого інформаційного пошуку. З використанням тезауруса запит може бути автоматично розширено й пошук буде повнішим.

У пошуковому запиті кожному інформаційному ресурсу можуть бути співставленні одне або більше понять, які описуються термінами в тезаурусі, а користувач, що здійснює пошук, може, використовуючи тезаурус, знайти необхідні поняття даної галузі, а також усі терміни, що їх характеризують. Тобто на основі зв'язків кожного терміну, прописаних у тезаурусі, відбувається розширення пошукового запиту (розширення слів запиту синонімічними, більш загальними чи вужчими за змістом термінами). Навігація по зв'язках тезауруса допомагає чіткіше сформулювати сам запит.

Розглянемо приклади використання ієрархічних відношень між дескрипторами, зафіксовані в інформаційно-пошуковому тезаурусі «розумного міста».

Наприклад, у базі даних є документи, пошукові образи яких містять ключові слова: *давач газу, давач диму, давач полум'я, давач розбиття скла, давач*. Користувач задає пошукове слово *давач* і система може видати

документи, які містять ключові слова *давач газу, давач диму* та інші, які містять у своїй структурі слово «*давач*», але не має в цьому впевненості. Для того, щоб отримати усю релевантну інформацію за даним запитом необхідно встановити відношення між цими лексичними одиницями, що зроблено у тезаурусі. Саме задання ієрархічних, у даному випадку родо-видових зв'язків, дає можливість отримати правильні документи.

Зовсім іншу ситуацію маємо у наступному випадку. Наприклад, у базі даних є документи, пошукові образи документів яких містять ключові слова: *автоматизація будівель, давач руху, розумне скло, сонячний колектор, енергозберігаючі технології*.

Користувач задає пошукове ключове слово *енергозберігаючі технології*. Якщо в інформаційно-пошуковій мові не зафіксовано відношення рід-вид, буде знайдено тільки документи, у пошукових образах документів яких є лексична одиниця *енергозберігаючі технології*. Документи, в пошуковий образ документів яких індексатори включили ключові слова з більш вузьким значенням (*автоматизація будівель, давач руху, розумне скло, сонячний колектор*), за даним запитом не будуть віднайдені, що призводить до втрати інформації. Користувач може й не передбачати, що система не відносить перераховані поняття до енергозберігаючих технологій.

Такі знання ієрархічних відношень в системі задаються у формі інформаційно-пошукового тезауруса: для дескрипторів *автоматизація будівель, давач руху, розумне скло, сонячний колектор* задається дескриптор вищого рівня *енергозберігаючі технології*.

Наприклад, у пошуковому образі документа використано ключове слово *розумний лічильник*. Для того, щоб даний документ було отримано за ключовими словами *розумний прилад, розумний будинок* потрібно для дескриптора *розумний лічильник* побудувати зв'язки з двома вищими дескрипторами: *розумний прилад* та *розумний будинок*.

Логічне підпорядкування дескрипторів у тезаурусі, що утворюють ієрархічні відношення між лексичними одиницями інструмента, дозволяють

користувачу розширювати свій запит за родо-видовими відношеннями (відношенням вище-нижче), наприклад, *утилізація – зелена утилізація*; відношеннями частина-ціле, наприклад, *екологічне будівництво – екологічний будинок*.

Ієрархічна організація термінів у інформаційно-пошуковому тезаурусі дозволяє автоматично приєднувати вузьким термінам вищі, більш широкі терміни, чи навпаки, тобто допомагає користувачу знайти більш широке поняття, переміщаючись угору по ієрархічному дереву, або більш вузьке, переміщаючись униз по схемі. Це дозволяє не перераховувати в запитах, наприклад, по давачах, усі конкретні типи давачів. Ієрархічні термінологічні структури відображають відношення «рід-вид», «частина-ціле» та «екземпляр-клас», які підвищують точність та повноту пошуку.

Таким чином, якщо дескриптор Д1 згадується в тексті й Д1 є видом Д2, це означає, що в тексті подається і Д2. Якщо даний текст релевантний за запитом Д1, то він буде й релевантний і за запитом Д2.

Розглянемо як за ключовими словами, наприклад, *краудсорсинг*, *інформаційно-комунікаційні технології* віднайти документи, яким приписано ключове слово *розумний уряд*. Поняття «краудсорсинг» належить до категорії «бізнесу», поняття «інформаційні та комунікаційні технології» до категорії «технології», а поняття «розумний уряд» до категорії «управління». Усі ці поняття не перебувають в ієрархічних відношеннях. Тому в інформаційно-пошуковому тезаурусі будуються асоціативні зв'язки між відповідними дескрипторами: *краудсорсинг*, *інформаційні та комунікаційні технології*, *розумний уряд*. Маючи інформацію про асоціативні відношення між даними дескрипторами, автоматизована інформаційна система використовує всі асоціативні дескриптори, що задані в списку дескрипторної статті головного дескриптора, у даному випадку – *розумний уряд*, і включить у результат усі знайдені документи.

Таким чином, використання смислових зв'язків предметів, об'єктів, дій та інших одиниць, виражених дескрипторами у тезаурусі, які утворюють

асоціативні відношення, допомагають користувачу відшукати додаткові дескриптори, інформативні при пошуку необхідних документів. Іншими словами, якщо два дескриптори зв'язані відношенням асоціації, то при пошуку за одним заданим дескриптором система запропонує користувачу здійснити пошук і за асоціативним дескриптором. Наприклад,

активна теплопостачання	система	сонячного	пасивна теплопостачання	система	сонячного
<i>A</i>	пасивна система теплопостачання	сонячного	<i>A</i>	активна система теплопостачання	сонячного
<i>A</i>	розумний будинок		<i>A</i>	розумний будинок	
розумний будинок					
<i>A</i>	активна система теплопостачання	сонячного			
<i>A</i>	пасивна система теплопостачання	сонячного			
<i>A</i>	Інтернет речей				

Асоціативні тезаурусні відношення аналогічні асоціативним зв'язкам у свідомості людини, коли виникає уявлення про один об'єкт воно викликає уявлення про інші. Наприклад, при пошуку за дескриптором *вуглецевий слід* (сукупність викидів усіх парникових газів, які утворилися в наслідок діяльності окремо людини, організації, міста, країни тощо) користувач автоматизованої інформаційної системи може здійснити додатковий пошук за дескрипторами: *сталий розвиток, зелена економіка, альтернативна енергетика* та ін.

Таким чином, якщо два дескриптори інформаційно-пошукового тезауруса Д1 і Д2 зв'язані асоціативними відношеннями, то використання цих неієрархічних відношень для розширення запиту полягає в тому, що, якщо запит містить Д1, то можна розширити запит, включивши в нього Д2 та отримати додатково релевантні документи.

Безліч понять у інформаційно-пошуковому тезаурусі «розумного міста» представлено не одним, а декількома термінами, які є синонімічними. Саме наявність у тезаурусі синонімів має велике значення, проте неправильне використання цих одиниць тезауруса призводить до втрати

інформації. Розглянемо це на прикладі. У предметній області «Розумне місто» функціонують терміни: *екологічно чистий транспорт*, *зелений транспорт* і *сталий транспорт*, які означають одне і те ж поняття. Але одне поняття не може бути представлено в тезаурусі різними дескрипторами. Це означало б, що при пошуку документів на запит, що містить синонім С1 пошукова система не видала б документи, що містять у пошуковому образі С2 чи С3, хоча вони підлягають видачі, так як відповідають запиту. Саме завдяки наявності в тезаурусі синонімічних зв'язків між термінами, пошук можливий за будь-яким даним терміном-синонімом. У запиті може бути використаний кожен з цих термінів, при цьому пошук буде здійснюватися за основним терміном. Наприклад, у пошуковому образі документа задано один із синонімів – *екологічно чистий транспорт*. Користувач уведе словосполучення *сталий транспорт*, зрозуміло, що даний документ не буде знайдено в автоматизованих інформаційних системах. Пошук документів за лексичною одиницею *екологічно чистий транспорт*, у пошуковий образ документа, якого вона введена, будуть видані на екран, проте, документи, індексовані ключовим словом *сталий транспорт*, не будуть віднайдені. Унаслідок отримується висока точність пошуку при прихованих втратах інформації. Уникнути цього дозволяє використання інформаційно-пошукового тезауруса з фіксацією відношення синонімії.

Використання термінів-синонімів призводить до часткової втрати інформації, тому ключові слова, що знаходяться у відношенні синонімії, об'єднуються в групу умовної еквівалентності й одне з них задається дескриптором. Отже, щоб запобігти втраті інформації з даних синонімів термін *екологічно чистий транспорт* задається як дескриптор, а інші супроводжуються відсиланнями до першого (позначкою С): синоніми *зелений транспорт* та *сталий транспорт* є дескрипторами. Така позначка означає, що замість одного терміна при складанні пошукових образів документів або запитів слід використовувати інший, який є дескриптором. Тобто, при пошуку за лексичними одиницями: *зелений транспорт* чи *сталий*

транспорт будуть видані автоматизованою інформаційною системою й документи, які індексовані дескриптором *екологічно чистий транспорт* та навпаки. Саме так ліквідується в тезаурусі синонімія, тобто вибір аскриптора веде до заміни його в формуванні пошукового запиту відповідним дескриптором.

Наявність у тезаурусі аскрипторів з відсиленнями до відповідного дескриптора забезпечує швидкий пошук потрібного терміна та сприяє підвищенню якості функціонування інформаційно-пошукової системи.

Отже, встановлені в тезаурусі відношення синонімії дозволяють провести пошук по одному з синонімів, знайшовши в базі даних автоматизованої інформаційної системи ті документи, з якими як ключові слова прив'язані інші синоніми. Наприклад, *екологічний будинок* = *екобудинок* = *зелений будинок*.

Для кожного терміна-дескриптора, крім списку дескрипторів, пов'язаних з ним парадигматичними відношеннями, можна отримати його синоніми, еквіваленти англійською мовою.

Важливим елементом статті тезауруса є відсилення до інших термінів чи інших синонімів того ж терміна. Вони дають можливість знайомитися зі змістом тезауруса методом «навігації» за відсиленнями та уточнювати значення термінів саме через його зв'язки з іншими термінами.

Інформаційно-пошуковий тезаурус «розумного міста» може використовуватися бібліотечними працівниками для пошуку в корпусі текстів предметної області «Розумне місто». Пошукова система отримує запит природньою мовою з формулюванням теми пошуку. Запит аналізується через тезаурус, у результаті чого будується семантичний образ запиту. Цей образ активує семантичний підпростір, що покриває всі можливі семантичні складові, релевантні даному запиту. А семантичний підпростір, у свою чергу, синтезує на основі тезауруса всі текстові реалізації цих складових, які використовуються пошуковим механізмом як зразки в процесі пошуку.

Здійснюючи пошук з використання тезауруса, користувач бібліотеки має можливість провести точний пошук з мінімальними інформаційними втратами завдяки включення в пошуковий образ запиту всіх релевантних термінів, які зв'язані з тим дескриптором, за яким було здійснено «вхід» у тезаурус.

Застосування інформаційно-пошукового тезауруса для інформаційного пошуку документів, як уже зазначалося, полягає в описі змісту документів і запитів за допомогою дескрипторів, що входять до його складу. На практиці ця ідея реалізується в такий спосіб. Текст документа, що вводиться в інформаційно-пошукову систему чи електронну бібліотеку, зменшується до обсягу реферату, в якому виділяються слова, що несуть основне смислове навантаження, так звані ключові слова. За допомогою ключових слів досить точно передається зміст документа. Після цього кожне слово замінюється близьким йому за змістом дескриптором інформаційно-пошукового тезауруса.

Здійснення пошуку в інформаційних системах проходить наступні етапи: формування пошукового образу запиту та передача його в інформаційну систему (за допомогою користувацького інтерфейсу у вигляді текстового поля); передача запиту в тезаурус; отримання з тезауруса списку ідентифікаторів понять, релевантних даному пошуковому запиту; формування запиту інформаційною системою на основі початкового запиту та його виконання; отримання списку знайдених релевантних документів у результат пошуку користувача бібліотеки.

Сукупність дескрипторів тезауруса, утворює пошуковий образ документа. Дескрипторною мовою задається і запит пошуку. Пошук документів після цього вже доручається пошуковій системі, яка з властивою їй швидкістю переглядає масиви пошукового образу документа й порівнює їх з пошуковим образом запиту. Мета такого порівняння – виявити дескриптори, що належать одночасно пошуковому образу документа й пошуковому образу запиту, тобто знайти перетин множин дескрипторів. Чим більша зона

перетину, тим точніша відповідність пошукового документа сформованому запиту. Таке зіставлення виправдане лише тоді, коли пошуковий образ і пошуковий запит формулюються термінами однієї мови та допускають тільки одне тлумачення, що реалізується за допомогою інформаційно-пошукового тезауруса.

Такими є принципи роботи інформаційно-пошукової системи з інформаційно-пошуковими мовами дескрипторного типу, якість роботи яких багато в чому залежить від точності побудови інформаційно-пошукового тезауруса, що використовується.

Отже, інформаційно-пошуковий тезаурус «розумного міста» надає користувачу інформаційно-пошукової системи структуровані фрагменти знань через експлікацію парадигматичних відношень між лексичними одиницями і, відповідно, між поняттями, що вони позначають.

Висновки до Розділу 4

Розвиток сучасних міст на основі використання інформаційних та комунікаційних технологій потребують нових способів опрацювання інформації. Це породжує інтерес до інформаційних технологій, пов'язаних зі створенням інформаційних систем, базованих на знаннях будь-якої предметної області. Важливо, щоб знання, які отримані при створенні однієї інформаційної системи, були придатними для роботи інших інформаційних систем, тобто інтероперабельні. Саме такою формою подання знань предметної області «Розумне місто» є розроблений інформаційно-пошуковий тезаурус «розумного міста», який досить повно представляє термінологічну систему галузі з урахуванням складності й багатовимірності семантичних зв'язків, що дозволяє людині або комп'ютеру орієнтуватися в ній. За допомогою тезауруса можна здійснювати пошук релевантної інформації в автоматизованих бібліотечних системах, створивши уніфікований пошуковий образ документа та пошуковий образ запиту користувачів.

У розділі подано хід роботи над формуванням масиву термінів для побудови двомовного інформаційно-пошукового тезауруса «розумного міста». Для організації терміносистеми предметної області «Розумне місто» було використано автентичні фахові тексти, перевага яких обумовлена їхньою особливістю – наявністю термінологічної лексики, що розкриває наукові поняття досліджуваної галузі та представляє їхнє однозначне пояснення. Проаналізовані джерела (а саме: наукову статті, праці конференцій, технічні журнали та ін.) дають змогу виявляти терміни, які знаходяться в науковому обігу та встановити зв'язки між ними. Це дає можливість говорити про кваліфікованість тематичного тезауруса, оскільки перевага таких джерел – це надійність, високий рівень загальноприйнятої інформації, зрозуміла мова подання, доступність (можливість отримати вільний доступ до цих джерел), висока насиченість термінологічною лексикою.

Відбір термінів здійснювався на основі критеріїв: інформаційність, частотність, відповідність темі, актуальність, практична значущість. Серед термінів тезауруса є загальнонаукові, міжгалузеві та вузькоспеціальні терміни, так як досліджувана галузь розвивається в тісному взаємозв'язку з іншими сферами.

Сформовано структурні класифікації термінів «розумного міста». У ході аналізу термінологічного матеріалу було з'ясовано, що термінологія цієї галузі, складається з лексичних одиниць двох типів: термінів-слів та термінів-словосполучень, останні у свою чергу поділяються на двокомпонентні та багатоконпонентні. У терміносистемі «розумного міста» переважають термінологічні словосполучення, які виражають більш складні поняття в порівнянні з термінами-словами.

Тезаурус подає широкий спектр термінів і виразів, які використовуються при дослідженні інтелектуальних систем міста, систем комунікації, виступаючи термінологічним інструментом для розробників, дизайнерів, виробників. Він може бути використаний керівниками на всіх

рівнях і в усіх секторах «розумних міських громад». Може також використовуватися і як багатопланова довідкова система, і як база для подальших лінгвістичних досліджень.

Розроблено класифікація відношень між одиницями тезауруса «розумного міста», яка включає два типи: традиційні (ієрархічні та еквівалентні) та проблемно-специфічні (асоціативні) відношення.

Подано структуру інформаційно-пошукового тезауруса «розумного міста», який складається зі вступу, алфавітного лексико-семантичного покажчика термінів та пермутаційного покажчика. Елементом алфавітного лексико-семантичного покажчика є словникова стаття. Усі види розроблених зв'язків у тезаурусі «розумного міста» відображаються в словникових статтях дескрипторів і аскрипторів. Подано структуру дескрипторної та аскрипторної статті та проілюстровано їх прикладами.

Визначено науково-практичні задачі, які здійснює розроблений інформаційно-пошуковий тезаурус «розумного міста» та універсальні лексикографічні функції терміносистем, які були реалізовані у моделі тезауруса.

Представлено здійснення інформаційного пошуку релевантної інформації за допомогою тезауруса «розумного міста».

У четвертому розділі дисертації подано основні результати наукових праць, включених до Списку публікацій здобувача за темою дисертації (пп. 1, 5, 16, 17) та представлених у Списку використаних джерел: [134], [164], [235], [162].

ВИСНОВКИ

У процесі роботи над дисертаційним дослідженням досягнуто його мету, виконано всі поставлені завдання та отримано такі основні науково-практичні результати:

1. Проведений аналіз наукових джерел за темою дослідження щодо формування пошукових образів та запитів документів, що містять результати міждисциплінарних досліджень показав, що незважаючи на значну кількість досліджень щодо побудови й функціонування інформаційно-пошукових тезаурусів та наявності значної кількості робіт, пов'язаних з предметною областю «Розумне місто», відсутній тезаурус, який нормалізував би лексику досліджуваної галузі й був би інструментом для індексації та пошуку документів у автоматизованих бібліотечно-інформаційних системах.

2. Обґрунтовано вибір моделі представлення знань предметної області «Розумне місто» для підвищення ефективності інформаційного пошуку в бібліотеках за допомогою використання методу аналізу ієрархій. Для парного порівняння було запропоновано такі альтернативи: логічні моделі, продукційні моделі, фреймові моделі, семантичні мережі та онтології.

Застосування методу аналізу ієрархій дозволило на основі науково-обґрунтованих підстав обрати найзручніший та найінформативніший спосіб подання концептів – онтологічний підхід як найкращий інструмент для формування інформаційно-пошукового тезауруса, що сприяє ефективній процедурі пошукових образів документів у автоматизованих бібліотечних інформаційних системах.

3. Простежено розвиток тезаурусів у процесі історії їхнього розвитку. Акцентовано увагу на тому, що історія виникнення тезаурусів ґрунтується на працях мислителів стародавніх часів, їхня побудова пов'язана з фундаментальними проблемами пізнання, відображення картини світу та закономірностями їхнього сприйняття. У інформаційному аспекті тезаурус почав розглядатися в XX ст.

Проаналізовано роль тематичних інформаційно-пошукових тезаурусів у процесах систематизації термінології предметних областей та можливість їхнього застосування як засобу підвищення ефективності пошуку в сучасних автоматизованих бібліотечних інформаційних системах. Використання інформаційно-пошукових тезаурусів сприяє підвищенню ефективності змістового опрацювання документів, зокрема при розробленні уніфікованого підходу до формування пошукового образу документа й пошукового образу запиту користувачів бібліотек для реалізації процедур пошуку в автоматизованих бібліотечних інформаційних системах для інформаційного забезпечення міждисциплінарних наукових досліджень. Використовуючи тезаурус, бібліотеки можуть формувати авторитетні файли, предметні рубрики або визначати ключові слова, індексувати документи й запити засобами дескрипторної мови, замінюючи ключові слова відповідними дескрипторами. При якісному індексуванні документів використання уніфікованої інформаційно-пошукової мови в бібліотеках та архівах дозволяє досягти високої точності пошуку та оперативно реагувати на появу нових термінів. Інформаційно-пошуковий тезаурус сприяє уніфікації довідкового апарату бібліотек, оскільки чітко фіксує місце терміна в понятійній системі певної предметної області, відображає його структурні зв'язки.

4. Запропоновано засадничий підхід до формування двомовного тезауруса предметної області, на прикладі предметної області «Розумне місто», який є одним із важливих практичних інструментів інформаційної діяльності бібліотек. Сформовано алгоритм його розроблення на основі консолідації принципів побудови тезаурусів, який складається з чотирьох етапів: 1) визначення меж предметної області, пошук джерел та накопичення галузевої інформації; 2) формування реєстрової частини тезауруса; 3) побудова словникових статей; 4) уточнення та редагування отриманого тезауруса. Запропонований підхід сприяє вирішенню актуальних завдань у досліджуваній галузі в контексті інформаційного забезпечення міждисциплінарних наукових досліджень, автоматизації процесів

розроблення двомовних пошукових образів; створення алгоритмів формування пошукового запиту у вузькоспеціальній предметній області.

5. Побудова тезауруса передбачала реалізацію низки послідовних кроків. Було виокремлено терміни з досліджуваних джерел на основі критеріїв: інформативність, частотність, відповідність темі, актуальність, практична значущість, що описують поняття предметної області. З'ясовано, що концепти предметної області «Розумне місто» складається з двох типів: термінів-слів (однокомпонентних) та термінів-словосполучень (багатокомпонентних). Переважають термінологічні словосполучення, які виражають більш складні поняття. Система термінів охоплює поняття, які належать до таких категорій: загальні поняття, інформаційні системи «розумного міста», управління ресурсами, технології, інфраструктури, мобільність, енергетика, телекомунікації, транспорт, фінанси й економіка, місцеві та центральні органи влади, спільноти, освіта, охорона здоров'я, системи безпеки, навколишнє середовище.

Простежено систему відношень між термінами «розумного міста». Класифікація зв'язків містить два типи: традиційні (ієрархічні й еквівалентні) та проблемно-специфічні (неієрархічні або асоціативні).

Розроблено двомовний інформаційно-пошуковий тезаурус «розумного міста», який максимально наповнений «робочими» термінами, що відповідають існуючому документному потокові, і складається зі вступу, алфавітного лексико-семантичного покажчика термінів та пермутаційного покажчика. В алфавітному лексико-семантичному покажчику термінів представлено всі дескриптори та аскриптори зі словниковими статтями. Він використовується для пошуку дескриптора та його парадигматичних зв'язків і забезпечує можливість адекватного формулювання запитів. Допоміжним покажчиком до алфавітного лексико-семантичного є пермутаційний покажчик термінів, який містить дескриптори та аскриптори без словникових статей і використовується для пошуку за окремими словами, що входять до складу термінів-словосполучень, у відповідних словникових гніздах.

Інформаційно-пошуковий тезаурус «розумного міста» дозволяє формувати уніфіковані пошукові образи документів та запитів з метою подальшого пошуку в автоматизованих бібліотечно-інформаційних системах; індексувати документи; формувати авторитетні файли, які сприяють забезпеченню якості опрацювання документів; уніфікувати введення інформації, що дозволяє отримати одноманітність вводу та скорочення чи уникнення помилок; здійснювати релевантний пошук у електронних каталогах, базах даних бібліотек; організовувати обмін авторитетними записами між бібліотеками; заощаджувати час на опрацювання документів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Мацюк Г. Інформаційно-пошукові тезауруси : світовий та вітчизняний досвід формування. *Бібліотекознавство. Документознавство. Інформологія*. 2019. № 2. С. 106–115.
2. WordNet. URL: <https://wordnet.princeton.edu/> (last accessed: 01.01.2020).
3. EuroWordNet. URL: <http://projects.illc.uva.nl/EuroWordNet/> (last accessed: 01.08.2020).
4. GermaNet. URL: <http://www.sfs.uni-tuebingen.de/GermaNet/> (last accessed: 02.01.2020).
5. DanNet. URL: <https://cst.ku.dk/projekter/dannet/> (last accessed: 02.01.2020).
6. MultiWordNet. URL: <http://multiwordnet.fbk.eu/english/home.php> (last accessed: 02.01.2020).
7. RussNet. URL: http://project.phil.spbu.ru/RussNet/index_ru.shtml (last accessed: 04.01.2020).
8. PyТез. URL: <http://www.labinform.ru/pub/ruthes/index.htm> (last accessed: 04.01.2020).
9. Алексеев А. А., Добров Б. В., Лукашевич Н. В. Лингвистическая онтология–тезаурус PyТез. *Open Semantic Technologies for Intelligent Systems – OSTIS* : материалы 3 междунар. науч.-техн. конф., 2013. С.153–158. URL: <https://istina.msu.ru/publications/article/4707107/> (дата звернення: 10.01.2017).
10. Кульчицький І. М., Романюк А. Б., Харів Х. Б. Розроблення Wordnet-подібного словника української мови. *Вісник Національного університету «Львівська політехніка»*. 2010. № 673. С. 306–318.
11. UNESCO Thesaurus. URL: <http://skos.um.es/unescothes/> (last accessed: 14.01.2020).
12. EuroVoc. URL: <https://op.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc> (last accessed: 14.01.2020).

13. EuroVoc 4.10. URL: <https://publications.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc> (last accessed: 17.01.2020).
14. Антоненко С. Актуальні питання створення та впровадження української версії тезаурусу «EUROVOC». *Наукові записки Інституту законодавства Верховної Ради України*. 2014. № 5. С. 79–84.
15. SNOMED. URL: <http://www.snomed.org/snomed-ct/snomed-in-action> (last accessed: 10.01.2020).
16. MeSH. URL: <https://www.ncbi.nlm.nih.gov/mesh> (last accessed: 15.01.2020).
17. Web-ИРБИС. URL: http://library.gpntb.ru/cgi-bin/irbis64r/62/cgiirbis_64.exe?C21COM=F&I21DBN=MESH&S21FMT=web_mesh_wn&S21All=%3C.%3ER=0%3C.%3E&P21DBN=IBIS&Z21ID (last accessed: 15.01.2018).
18. The Getty Research Institute. URL: http://www.getty.edu/research/conducting_research/vocabularies/ulan/ (last accessed: 20.01.2018).
19. Art & Architecture Thesaurus. URL: <http://www.getty.edu/research/tools/vocabularies/aat/index.html> (last accessed: 20.01.2018).
20. Getty Thesaurus of Geographic Names. URL: <http://www.getty.edu/research/tools/vocabularies/tgn/index.html> (last accessed: 20.01.2018).
21. Union List of Artist Names. URL: <http://www.getty.edu/research/tools/vocabularies/ulan/index.html> (last accessed: 20.01.2018).
22. Cultural Objects Name Authority. URL: <http://www.getty.edu/research/tools/vocabularies/cona/index.html> (last accessed: 20.01.2018).

23. AGROVOC Multilingual Thesaurus. URL: <http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/index?clang=uk> (last accessed: 20.01.2019).
24. ИНИОН РАН. URL: <http://inion.ru/resources/bazy-dannykh-inion-ran/> (дата звернення: 20.11.2019).
25. Информационно-поисковый тезаурус Социология. URL: http://inion.ru/site/assets/files/2641/tezaurus_sotciologiiia_versiiia_5-1.pdf (дата звернення: 20.11.2019).
26. Информационно-поисковый тезаурус Этнология. Антропология. URL: http://inion.ru/site/assets/files/2980/tezaurus_etnologiiia.pdf (дата звернення: 20.11.2019).
27. Экономика. Демография: информационнопоисковый тезаурус. URL: http://inion.ru/site/assets/files/3799/tezaurus_ekonomika_demografiia.pdf (дата звернення: 25.05.2019).
28. Философия: информационно-поисковый тезаурус. URL: http://inion.ru/site/assets/files/3605/tezaurus_filosofiiia.pdf (дата звернення: 25.05.2019).
29. Религиоведение: информационно-поисковый тезаурус. URL: <http://inion.ru/publishing/publications/religiovedenie-informatsionno-poiskovy-tezaurus/> (дата звернення: 25.05.2019).
30. NASA Thesaurus. URL: <https://www.sti.nasa.gov/nasa-thesaurus/#.XeV2PVUzbDc> (last accessed: 13.01.2019).
31. Тезаурус по педагогике и образованию. URL: <http://www.gnpbu.ru/publications/thesaurus/> (дата звернення: 25.08.2019)
32. Информационно-поисковый тезаурус по сохранности документов. URL: http://www.rasl.ru/e_resours/tezaurus/lat-A.htm#A (дата звернення: 18.01.2019).
33. Тезаурус з комп'ютерної лексикографії. URL: <http://www.mova.info/Page3.aspx?l1=188&vocid=1> (дата звернення: 08.08.2020).

34. Тезаурус з лінгвістичної термінології. URL: http://www.mova.info/mov_thes.aspx?l1=68 (last accessed: 08.08.2020).
35. Національна бібліотека України імені Ярослава Мудрого, інформаційно-пошуковий тезаурус. URL: <http://ipt.nlu.org.ua/> (дата звернення: 28.08.2020).
36. ВилегжанінаТ., Збанацька О. Інформаційно-пошуковий тезаурус. Донецьк : Альфа-Прес, 2004. – 1104 с.
37. Збанацька О. М. Інформаційно-пошуковий тезаурус Національної бібліотеки України імені Ярослава Мудрого: сучасний стан та перспективи. URL: <https://elib.nlu.org.ua/view.html?id=8998> (дата звернення 28.12.2019).
38. Губренко І. Ю. Створення галузевого тезауруса в лінгвістичному забезпеченні електронної бібліотеки Харківського державного університету харчування та торгівлі. Всеукраїнська наук.-практ. конф. : [до 20-річчя створення АСІБТ та 45-річчя заснування ХДУХТ : матеріали]. Харків : ХДУХТ, 2012. С. 74–78.
39. Бержанский В., Нестеров Д. Разработка многоязычного словаря предметной области «Нанотехнологии». URL: <http://megaling.ulif.org.ua/tezi-2012-rik/berzhanskiy-vladimir-naumovich-nesterov-dmitriy-sergeevich-razrabotka-mnogoyazyichnogo-slovary-a-predmetnoy-oblasti-nanotehnologii.html> (дата звернення: 10.01.2017).
40. Андрощук О. С. Модель побудови тезауруса предметної сфери управління підрозділами в особливих ситуаціях. *Системи озброєння і військова техніка*. 2010. № 4. С. 196–200.
41. Філюк Л. М. Моделювання тезауруса української будівельної термінології. *Мова*. 2014. №14. С. 127–131.
42. Асмукович І. В. Принципи побудови двомовного тезауруса фахової мови авіації. *Науковий вісник Східноєвропейського національного університету імені Лесі Українки. Філологічні науки. Мовознавство*. 2013. № 20. С. 39–44.

43. Колодницький В., Кулаковский В., Лебедева А., Чистяков Е. Электронный информационно-поисковый тезаурус предметной области «сверхтвердые материалы» как формальная система знаний. URL: <http://dspace.nbuv.gov.ua/handle/123456789/23442> (дата звернення: 10.01.2017).

44. Татарчук Л. Лінгвістичне забезпечення формування єдиного інформаційного простору в мережі сільськогосподарських бібліотек. *Бібліотечний вісник*. 2011. № 3. С. 3–8.

45. Колот А. М. Міждисциплінарний підхід як домінанта розвитку економічної науки та освітньої діяльності. *Социальная экономика*. 2014. № 1–2, С. 76-83.

46. Палагін О., Кургаєв О. Міждисциплінарні наукові дослідження : оптимізація системно-інформаційної підтримки. *Вісник НАН України*. 2009. № 3. С. 14-25.

47. Філіпова Л. Інформаційне забезпечення наукових досліджень : тенденції розвитку (на прикладі медичної науки). *Вісник Книжкової палати*. 2009. № 11. С. 24–31.

48. Гришко О. П. Теоретичні засади інформаційного забезпечення регулювання ринку праці. *Науковий вісник Ужгородського національного університету*. 2018. Вип. 19. С. 88–92.

49. Про інформацію: Закон України. URL: <https://zakon.rada.gov.ua/laws/show/2657-12#Text> (дата звернення: 19.07.2020).

50. Гетманцева Н. Д. Методологія наукових досліджень : навч. посіб. для магістер. програм «Менеджмент корпорацій (АТ)» та «Менеджмент малого бізнесу». Київ : КНЕУ, 2009. 517 с.

51. Горбатюк С. Є. Інформаційне забезпечення наукового дослідження: інституційно-правові аспекти. *Вісник НАДУ*. 2014. №4. С.130–138.

52. Інформаційно-пошукові системи. URL: <https://sites.google.com/site/socialnafnformatika/home> (дата звернення: 01.11.2019).

53. Костєва Н. М. Тезаурусне підґрунтя інформаційно-інтелектуального забезпечення наукових досліджень. URL: <http://inb.dnsgb.com.ua/2006-1/06knmznd.html> (дата звернення: 01.11.2019).
54. Лукич Р. Методология права. Москва, 1981. 304 с.
55. Вебер М. Критическое исследования в области наук о культуре. Культурология. XX век : Антология. Москва : Юрист, 1995. С. 9–15.
56. Кельман М. Юридична наука : проблеми методології : монографія. Тернопіль : Терно-граф, 2011. 492 с.
57. Черчик Л. М. Електронний посібник з дисципліни «Методологія наукових досліджень». Луцьк : Луцький національний технічний університет, 2012. URL: <http://elib.lutsk-ntu.com.ua/book/fb/pesp/2012/12-31> (дата звернення: 11.11.2016).
58. Тарелкін Ю. П., Цикін В. О. Методологія наукових досліджень : навч. посіб. Суми : СумДПУ ім. А. С. Макаренка, 2010. 196 с.
59. Кунанець Н. Е. Інформаційно-бібліотечне обслуговування користувачів з особливими потребами : історія та сучасність : монографія. Львів : Галицька видавнича спілка, 2013. 440 с.
60. Черчик Л. М. Електронний посібник з дисципліни «Методологія наукових досліджень». Луцьк : Луцький національний технічний університет, 2012. URL: <http://elib.lutsk-ntu.com.ua/book/fb/pesp/2012/12-31> (дата звернення: 11.11.2016).
61. Катренко А. В. Системний аналіз об'єктів та процесів комп'ютеризації. Львів : Новий світ, 2003. 368 с.
62. Кунанець Н. Э. Библиографический указатель «Библиотечное дело на западноукраинских землях : консолидированный продукт. Библиотеки национальных академий наук : проблемы функционирования, тенденции развития: науч.-практ. и теорет. сб. / НАН Украины, Нац. б-ка Украины им. Вернадского, МААН, Совет директоров науч. б-к и информ. центров. Киев, 2011. Вып. 9. С. 285–295.

63. Кунанець Н. Е., Пасічник В. В. Вступ до спеціальності: «Консолідована інформація» : навч. посіб. Львів : Видавництво Національного університету «Львівська політехніка», 2010. 196 с.
64. Різун В. Начерки до методології досліджень соціальних комунікацій. *Психолінгвістика*. 2012. Вип. 10. С. 305–314.
65. Холод О. Сучасні теорії та тенденції розвитку соціальних комунікацій. Київ - Кривий Ріг : Київський національний університет культури і мистецтв. 2016. С. 213.
66. Мацюк Г. Р., Кунанець Н. Е. Двомовний тезаурус як соціокомунікаційна складова проекту «Розумне місто». *Современные информационные технологии в экономике и управлении предприятиями, программами и проектами* : материалы XV междунар. науч.-практ. конф., г. Харьков, 29 сен. 2017 г. Харьков, 2017. С. 85–87.
67. Сахнюк П. А. Интеллектуальные системы и технологии : учеб. пособ. Ставрополь: Агрус, 2012. 228 с.
68. Уэно Х., Кояма Т. Окамото Т., Мацуби Б. Представление и использование знаний. Москва : Мир, 1989. 220 с.
69. Поспелов Д. А. Представление знаний. Опыт системного анализа. Системные исследования. *Методологические проблемы. Ежегодник*. 1985. № 17. С.83–102.
70. Невмержицький О. В. Аналіз сучасних моделей, орієнтованих на знання, та методів прийняття рішень. *Інформаційні технології проектування*. 2013. № 13. С. 119–125.
71. Шаров С. В., Лубко Д. В., Осадчий В. В. Вибір моделі представлення знань у системі ІСІКС. *Системи обробки інформації*. 2015. № 11 (136). С. 108 –111.
72. Федорук П. І., Масловський С. М. Модель представлення знань в адаптивній системі дистанційного навчання та контролю знань «EduPRO». *Штучний інтелект*. 2011. № 3. С. 463–472.

73. Палагин А. В., Петренко Н. Г., Малахов К. С. Методика проектирования онтологий предметной области. *Комп'ютерні засоби, мережі та системи*. 2011. № 10. С. 5–12.
74. Петренко А. І. Методологія розробки онтологій. *Вісник університету «Україна»*. 2010. № 8. С. 56–71.
75. H. Matsiuk, A. Rzheuskyi, N. Kunanets, N. Veretennikova Formation of Search Queries Based on Thesaurus of Narrowly Specialized Subject Areas *Econtechmod. An international quarterly journal*. 2019. Vol. 08. (No. 3). P. 11–14.
76. Герман О. В. Введение в теорию экспертных систем и обработку знаний. Минск, 1995. 255 с.
77. Рассел С., Норвиг П. Искусственный интеллект. Современный подход, 2-е изд. Москва : Вильямс, 2007. 1410 с.
78. Барышев М. В., Гатчин И. Ю., Гатчин Ю. А. Модели представления знаний экспертных систем. *Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики*. 2006. № 29. С. 14–18.
79. Субботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навч. посіб. Запоріжжя : ЗНТУ, 2008. 341 с.
80. Рогушина Ю. В., Гладун А. Я. Методика розробки термінології інформаційних ресурсів як базису формування онтологій та тезаурусів для семантичного пошуку. *Інженерія програмного забезпечення*. 2014. № 1. С. 41–51.
81. ДСТУ 2481-94 Системи оброблення інформації. Інтелектуальні інформаційні технології. Терміни та визначення. [Чинний від 1995–01–01]. Вид. офіц. Київ, 1994. 162 с.
82. Саати Т., Керис К. Аналитическое планирование. Организация системы : пер. с англ. Москва : Радио и связь, 1991. 224 с.

83. Шаповалова О. О., Бурменський Р. В. Розробка програмного додатка для реалізації методу аналізу ієрархій. *Системи обробки інформації*. 2017. № 3 (149). С. 45–48.

84. Шумська Н. В., Науменко Н. Ю. Застосування методу аналізу ієрархій до задачі прийняття рішення про оптимальний вибір країни стажування. *Геоінформаційні системи, комп'ютерні технології еколого-економічного моніторингу, ГІСКТЕЕМ - 2016* : зб. тез доп. міжнар. наук.-техн. конф., м. Дніпропетровськ, 13-15 квіт. 2016 р. URL: http://gis.dp.ua/conf2016-publications/sections/iad/18_SHUMSKAJA_NAUMENKO.pdf (дата звернення: 05.08.2017).

85. Щербань А. А., Макарова В. М. Метод аналізу ієрархій у плануванні та прийнятті управлінських рішень при запровадженні інновацій у нафтосервісних підприємствах. *Економіко-математичне моделювання соціально-економічних систем* : зб. наук. пр. Київ : МННЦІТС НАН та МОН України, 2013. Вип. 18. С. 286–299.

86. Трунова О. В. Застосування методу Сааті при прийнятті управлінських рішень. *Вісник Чернігівського національного педагогічного університету. Педагогічні науки*. 2013. Вип. 108.1. URL: http://nbuv.gov.ua/UJRN/VchdpuP_2013_1_108_34 (дата звернення: 06.08.2017).

87. Гладун Т. М. Застосування методу аналізу ієрархій для вибору франчайзингової мережі. URL: http://ena.lp.edu.ua/bitstream/ntb/36011/1/19_109-115.pdf (дата звернення: 06.08.2017).

88. Лехан В. М., Волчек В. В., Горбань А. Є., Павленко М. В. Комплекс способів підвищення якості та ефективності стаціонарної медичної допомоги. Київ, 2015. С. 40.

89. Миронець С. М. Застосування методу аналізу ієрархій в системі заходів психологічного забезпечення фахівців ризиконебезпечних професій.

Вісник Чернігівського національного педагогічного університету. Чернігів, 2012. Т. 2. Вип. 105. С. 3–9.

90. Миронова Н. О., Табунщик Г. В. Використання групових методів прийняття рішень в задачах управління навчальним процесом. *Системи обробки інформації*. Харків, 2011. Вип. 7 (97). С. 76–79.

91. Донюк О. А., Буркіна Н. В. Про застосування МАІ для прийняття оптимального рішення вибору вишу в умовах багатокритеріальності. *Вісник студентського наукового товариства ДонНУ імені Василя Стуса*. Вінниця, 2018. Том 2. № 10. С. 179–183.

92. Кунанець Н., Кут В., Ковалюк Т. Вибір способу представлення знань у інформаційній системі дистанційного бібліотечного обслуговування осіб із особливими потребами. *Вісник Національного університету «Львівська політехніка»*. 2017. № 864. С. 257–262.

93. Kunanets Nataliia, Matsiuk Halyna Application of Saaty Method while Choosing Thesaurus View Model of the “Smart city” Subject Domain for the Improvement of Information Retrieval Efficiency. 13th International Scientific and Technical Conference on *Computer Sciences and Information Technologies (CSIT)* (11-14 September 2018). Lviv, 2018. P. 21–25.

94. Кунанець Н. Е., Ржеуський А. В., Мацюк Г. Р. Застосування методу аналізу ієрархій для вибору моделі представлення предметної області «Розумне місто». *Математика. Інформаційні технології. Освіта* : тези доп. VII міжнар. наук.-практ. конф., Луцьк-Світязь, 3-5 червня 2018 р. Луцьк-Світязь, 2018. С. 68–70.

95. Гладун А. Я., Рогушина Ю. В. Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу. *Искусственный интеллект*. 2008. № 5. С. 112–124.

96. Berners-Lee T., Handler J., Lassila O. The Semantic Web. *Scientific American*. May. 2001. Pp. 34–43.

97. Gruber T. R. A translation approach to portable ontologies. *Knowledge Acquisition*. 1993. № 5 (2). Pp. 199–220.

98. Noy N. F., McGuinness D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (last accessed: 13.05.2018)
99. Соловьев В. Д., Добров Б. В., Иванов В. В., Лукашевич Н. В. *Онтологии и тезаурусы : учебное пособие*. Казань—Москва, 2006. 158 с.
100. Дубров С. Н., Нечаев Ю. И. Интегрированная обучающая система на базе онтологии и формального концептуального анализа. URL: <http://www.ict.edu.ru/vconf/files/9222.pdf> (дата звернения: 06.05.2018).
101. Соколова Н. В., Голубева М. В. Информационная модель дисциплины как средство эффективного обучения. URL: http://wayforyou.naukapro.ru/ot2006/1_045.htm (дата звернения: 06.05.2018).
102. Батищев С. В., Виттих В. А., Волхонцев Д. В. Сетевая мультиагентная модель системы управления здравоохранением региона и система контроля эффективности и качества работы врачей поликлиники. URL: http://www.ssc.smr.ru/media/ipuss_conf/06/6_09.pdf (дата звернения: 10.05.2018).
103. Mazo C., Salazar L., Corcho O., Trujillo M., Alegre E.: A histological ontology of the human cardiovascular system. URL: 10.1186/s13326-017-0158-5 (last accessed: 13.05.2018).
104. Shaker El-Sappagh, Francesco Franda, Farman Ali, Kyung-Sup Kwak: SNOMED CT standard ontology based on the ontology for general medical science. URL: <https://doi.org/10.1186/s12911-018-0651-5> (last accessed: 13.05.2018).
105. Брагинская Л. П., Григорюк А. П., Загорулько Ю. А., Боровикова О. И. Онтологический подход к систематизации контента интернет-ресурса «Активная сейсмология». URL: <http://cyberleninka.ru/article/n/ontologicheskiiy-podhod-k-sistematizatsii-kontenta-internet-resursa-aktivnaya-seysmologiya> (дата звернения: 11.05.2018).

106. Моделювання і прогнозування стану довкілля : підручник / В. І. Лаврик, Б. М. Боголюбов, Л. Б. Полетаєва та ін. Київ : ВЦ «Академія», 2010. 400 с.
107. Рідкокаша А. А., Шевченко Б. М. Онтологічне моделювання пошуку інформації з проектування в електронній бібліотеці. *Вісник ЧДТУ*. 2013. № 4. С. 24–30.
108. Fensel D. *Ontologies : Silver Bullet for Knowledge Management and Electronic Commerce*. Berlin: Springer-Verlag, 2001. 102 p.
109. Кагаловский М. Р., Паринов С. И. Семантическое структурирование контента научных электронных библиотек на основе онтологий. *Электронная библиотека : современные технологии интеграции информационных ресурсов* : сб. науч. тр. Санкт Петербург, 2011. С. 26–45
110. Лобузін К. В. Технології організації знаннєвих ресурсів у бібліотечно-інформаційній діяльності : монографія / за ред. О. С. Онищенко; НАН України, Нац. б-ка України ім. В. І. Вернадського. Київ, 2012. 249 с.
111. Литвин В., Пасічник В., Яцишин Ю. Інтелектуальні системи : підручник / за наук. ред. В. Пасічника. Львів : Новий Світ-2000, 2008. 405 с.
112. Gruber T. R. A translation approach to portable ontologies. *Knowledge Acquisition*. 1993. Vol. 5 (2). P. 199-220. URL: <https://pdfs.semanticscholar.org/5120/f65919f77859a974fcc1ad08f72b2918b8ec.pdf> (last accessed: 13.12.2017).
113. Guarino N. *Formal Ontology and Information Systems*. FOIS'98, 6–8 June 1998, Trento, Italy: IOSPress, Amsterdam, 1998. P. 3–15. URL: <https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98Formal%20Ontology%20and%20Information%20Systems.pdf> (last accessed: 13.12.2017).
114. Gruber T. R. Toward principles for the design of ontologies used for knowledge sharing. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.5775&rep=rep1&type=pdf> (last accessed: 13.12.2017).

115. Hovy E. Ontologies: lecture 1, lecture 2, Issues of Content, lecture 3. Methods for Automated Ontology Building. *Information Sciences*. London, 2004. P. 136 – 148.
116. Ланде Д. В. Елементи комп'ютерної лінгвістики в правовій інформатиці. Київ : НДІП НАПрН України, 2014. 168 с.
117. Гладун А. Я., Рогушина Ю. В. Онтологічний підхід до проблем підвищення якості розроблення національних стандартів України. *Стандартизація, сертифікація, якість*. 2016. № 2 (99). С.19–28.
118. Кудряшова Э. Е. Методы и модели проектирования информационных систем (монография). *Международный журнал прикладных и фундаментальных исследований*. 2010. № 6. С. 117–118.
119. Nataliya Kunanets, Halyna Matsiuk Use of the Smart City Ontology for Relevant Information Retrieval. Proceedings of the 3rd International Conference on *Computational Linguistics and Intelligent Systems (COLINS-2019)*. Kharkiv, April 18-19. Kharkiv, 2019. Pp. 322–333.
120. Литвин В. В., Гопяк М. Я., Оборська О. В., Вовнянк Р. В. Метод побудови інтелектуальних агентів на основі адаптивних онтологій. *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. 2015. № 829. С. 186-200.
121. Бержанский В. Н., Потапова Е. В. Классификация связей между понятиями в онтологии по физике магнитных явлений. *MegaLing 2010* : сб. научн. трудов. Київ : Довіра. 2010. С.12–21.
122. IDEF5 – Ontology Description Capture Method. URL: file:///C:/Users/Admin/Downloads/Ontology_Capture_Method_IDEF5.pdf (last accessed: 10.04.2018).
123. Кунанець Н. Е., Мацюк Г. Р. Тезаурус предметної області «Розумне місто». *Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі»*. 2017. № 872. С. 179–189.
124. Кунанець Н. Е., Пасічник В, В., Мацюк Г. Р., Мацюк О. В., Левченко О, П. Онтологічне дослідження предметної області «Розумне

місто». *Математика. Інформаційні технології. Освіта* : VI міжнар. наук.-практ. конф., Луцьк-Світязь, 5-7 червня 2017 р. Луцьк-Світязь, 2017. С. 48–52.

125. Гладун А. Я., Рогушина Ю. В. Онтологии в корпоративных системах. *Корпоративные системы*. 2006. № 1. С. 13–26.

126. Amarakocha. URL: <https://archive.org/details/amarakochaouvoca02amaruoft/page/n5> (last accessed: 21.05.2018).

127. Livres dou trésor, Paris, 1863. URL: <https://archive.org/details/lilivresdoutreso00latiuoft/page/182> (last accessed: 15.05.2018).

128. Peter Mark Roget Thesaurus of English Words and Phrases (Fourth ed.). London: Longman, Brown, Green, and Longmans. 1856. URL: <https://books.google.com.ua/books?id=9nYCAAAAQAAJ&printsec=frontcover&hl=uk#v=onepage&q&f=false> (last accessed: 15.05.2018).

129. Голуб Ю.І Проблема відображення системної організації лексикону в рамках алфавітної лексикографії. *Вісник Запорізького державного університету*. 2001. № 4. С. 1-5.

130. Boissiere P. Dictionnaire analogique de la langue francaise. Paris, 1862. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k22361w/f3.item.shift.zoom> (last accessed: 15.05.2018).

131. Dornseiff F. Der deutsche Wortschatz nach Sachgruppen. Berlin und Leipzig, 1934. URL: <https://www.google.com/search?q=F.+Dornseiff.+Der+deutsch+e+Wortschatz+nach+Sachgruppen.+Berlin+und+Leipzig,+1934,+S.+1.&ei=yINQXKD9CI7NrgTM5bq4DQ&start=10&sa=N&ved=0ahUKEwiggdP2uJPgAhWOp0sKHcyuDtcQ8tMDCMEB&biw=1366&bih=636> (last accessed: 16.05.2018).

132. Casares J. Diccionario ideologico de la lengua espanola. Barcelona, 1959.

133. Морковкин В. В. Идеографические словари. Москва, 1970. 71 с.

134. Мацюк Г., Кунанець Н. Представлення знань у вузькоспеціальних предметних областях за допомогою тезауруса. *Бібліотечний вісник*. 2019. № 1. С. 19–23.
135. Андреев А. М., Березкин Д. В., Симаков К. В. Особенности проектирования модели и онтологии предметной области для поиска противоречий в правовых электронных библиотеках. URL: <http://www.inteltec.ru/publish/articles/textan / RCDL2004. pdf> (дата звернення: 19.05.2018).
136. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. Москва : издво МГУ, 2011. 495 с.
137. Гендина Н. И. Информационно-поисковые тезаурусы: основные виды и области применения. *Научные и технические библиотеки*. Москва, 2008 С. 5–14.
138. Браславский П. И., Гольдштейн С. Л., Ткаченко Т. Я. Тезаурус как средство описания системы знаний. *Научно-техническая информация*. Сер. 2. 1997. № 11. С. 16–22.
139. Щерба Л. В. Опыт общей теории лексикографии. Языковая система и речевая деятельность. Москва : Наука, 1974. С 266.
140. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка. Москва : Наука, 1981. 363 с.
141. Gorodishcheva A. N., Gorodishchev A. V. Standards, specifications and features of the logicallinguistic terminology management systems, communications. *Science Prospects* 2011. № 5 (20). Pp. 62–65. URL: <https://elibrary.ru/item.asp?id=17803177> (last accessed: 19.05.2018).
142. Мимиранова М. Проблема тезауруса в психологии : дисс. канд. психол. наук : 19.00.01. Москва, 1984. 194 с.
143. Величко Л. Тезаурус як засіб семантизації наукових понять. *Біологія і хімія в рідній школі*. 2014. № 2. С. 40-42.
144. Чурсин Н. Н. Популярная информатика. Київ : Техника, 1982. 158 с.

145. Бісікало О. В., Яхимович О. В. Автоматизоване визначення лексичних онтологій з тезаурусу технічного спрямування. *Опτικο-електронні інформаційно-енергетичні технології*. 2016. Том 31. № 1. С. 26–38.
146. Андрусенко Т. Б. Лингвистические структуры в компьютерных учебных средах. Киев : Наук. думка, 1994. 160 с.
147. ISO 2788-1986. Documentation – Guidelines for the establishment and development of monolingual thesauri [Publication date : 1986-11]. URL: <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=7776> (last accessed: 12.08.2018).
148. ДСТУ 4032-2001. Інформація та документація. Одномовний тезаурус. Методика розроблення (ISO 2788:1986) [ГОСТ 7.25-80; чинний від 2002-01-01]. Вид. офіц. Київ, 2001. URL: http://online.budstandart.com/ua/catalog/doc-page.html?id_doc=70748 (дата звернення: 12.06.2018).
149. Шульгіна В. Електронний тезаурус як різновид електронного видання. *Вісник Львівського університету*. 2004. Вип. 34. Ч. I. С. 323–326.
150. Рогушина Ю. В., Гладун А. Я., Штонда В. Н. Розробка онтологічних терміносистем інформаційних ресурсів інтернет та їх когнітивних моделей у наукових дослідженнях. *Проблеми програмування*. 2010. № 2–3. С. 390–394.
151. Гладун А. Я., Рогушина Ю. В. Онтологии и мультилингвистические тезаурусы как основа семантического поиска информационных ресурсов Интернет. *The Proc. of XII-th Intern. Conf. KDS'2006*, Varna, Bulgaria. 2006. Pp.115–121.
152. Баранов А. Н. Введение в прикладную лингвистику : учебн. пособ. изд 2-е испр. Москва, 2003. 306 с.
153. Дарчук Н. Текстозорієнтована ІПС з літературознавчої термінології. *Людина. Комп'ютер. Комунікація* : зб. наук. пр. Львів, 2017. С. 7–10.

154. Кунанець Н. Е., Мацюк Г. Р. Лексикографічна систематизація галузевої термінології за допомогою тезауруса. *X Львівський міжнародний бібліотечний форум «Нова бібліотека – відповідальність кожного»* : збірник матеріалів. Київ : УБА, 2019. С. 64–66.
155. Пшеничная Л. Е. Термин в документальных ИПС. Киев : Наукова Думка, 1977. 79 с.
156. Купцова А. К., Стерлигова А. Н. Современные вопросы развития терминологии логистики. *Логистика и управление цепями поставок*. 2004. № 2. С. 112–121.
157. Воронцова И. А. Принципы классификации тезаурусов английского языка. URL: http://vestnik.yspu.org/releases/novye_Issledovaniy/17_7/.
158. Широков В. А. Семантичні стани мовних одиниць та їх застосування в когнітивній лексикографії. *Мовознавство*. 2005. № 3–4. С. 47–62
159. Сірук О. Б. Тезаурус дієслів української мови: лінгвістичні проблеми та методика конструювання : дис. ... канд. філол. наук : 10.02.01 / Київський національний ун-т ім. Тараса Шевченка. Інститут філології. Київ, 2006. 201с.
160. Гладун А. Я., Рогушина Ю. В. Методика розробки термінології області як базису для формування онтологій та тазаурусів. *Проблеми інформатизації та управління*. 2009. № 3 (27). С. 26–34.
161. Лобузін К. В. Бібліотечні класифікації в сучасних інформаційних сервісах : монографія. НАН України, НБУВ. Київ : НБУВ, 2010. 132 с.
162. Кунанець Н. Е., Мацюк Г. Р. Формування тезауруса для виконання інформаційного пошуку у вузькоспеціальних предметних областях. *Бібліотека. Наука. Комунікація : 100-річчя Національної бібліотеки України імені В. І. Вернадського* : матеріали міжнар. наук. конф. (Київ, 6-8 лист. 2018 р.) / Нац. б-ка України ім. В. І. Вернадського, Асоц. б-к

України, Рада дир. б-к та інформ. центрів – членів МААН. Київ, 2018. С. 366–369.

163. Мацюк Г. Удосконалення технологій інформаційного пошуку на основі тезауруса предметної області. *Бібліотечний вісник*. 2019. № 5. С. 20–25.

164. Мацюк Г., Кунанець Н. Стандарти, що регламентують формат подання інформаційно-пошукового тезауруса. *Інформаційні моделі, системи та технології* : матеріали VI наук.-техніч. конф. Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 12-13 грудня 2018 р.). Тернопіль, 2018. С. 39.

165. ISO 5964:1985. Documentation – Guidelines for the establishment and development of multilingual thesauri. [Publication date : 1985-02]. URL: <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=12159> (last accessed: 12.08.2018).

166. ANSI/NISO. Z39.19: 2005 Guidelines for the construction, format and management of monolingual controlled vocabularies. [Publication Date: May 13, 2010]. URL: <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>.

167. ISO 25964-1:2011 Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval, 2011. 152 pp.

168. ISO 25964-2:2013 Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies, 2013. 99 pp.

169. ISO/IEC13250:2003. Information technology. SGML applications – Topic maps. 2003. 41 pp.

170. ISO/IEC 13250-2:2006. Information technology – Topic Maps. Part 2: Data model. 2006. 21pp.

171. ISO/IEC 13250-3:2013. Information technology – Topic Maps. Part 3: XML syntax. 2013. 24 pp.

172. ГОСТ 7.25-2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. [Взамен ГОСТ 7.2580; введ. 2002-07-01]. Москва : ИПК Изд-во стандартов, 2001. 16 с.

173. ГОСТ 7.24-2007 Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению. [Взамен ГОСТ 7.24-90; введ. 2008-07-01]. Межгосударственный совет по стандартизации, метрологии и сертификации. Москва : Стандартинформ, 2008. 7 с.

174. ГОСТ Р 7.0.91-2015 (ИСО 25964-1:2011) Система стандартов по информации, библиотечному и издательскому делу. Тезаурусы для информационного поиска. URL: <http://docs.cntd.ru/document/1200129056> (дата звернення: 25.08.2018).

175. ДСТУ 4031–2001. Інформація та документація. Багатомовний тезаурус. Методика розроблення (ISO 5964:1985, IDT): вид. офіц. [Чинний від 2002–01–01]. Київ : Держспоживстандарт України, 2001. 162 с.

176. ДСТУ 4032–2001. Інформація та документація. Одномовний тезаурус. Методика розроблення (ISO 2788:1986, IDT) : вид. офіц. [Чинний від 2002–01–01]. Київ : Держспоживстандарт, 2001. 69 с.

177. Kunanets N., Matsiuk H. Application of the “Smart City” data domain thesaurus as the tool for representing knowledge while improving the problem-oriented Web search effectiveness. *14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)* (Lviv, 17-20 September 2019). Lviv, 2019. P. 31–34.

178. Кияк Т. Р. Лингвистические аспекты терминоведения : учебн. пособ. Киев : УМКВО, 1989. 103 с.

179. Скороходько Е. Ф. Термін у науковому тексті (до створення терміноцентричної теорії наукового дискурсу) : монографія. Київ : Логос, 2006. 100 с.

180. Вознюк Г. Л., Булик-Верхола С. З. Термінологічна лексика в системі професійного мовлення : методичні вказівки та завдання для самостійної роботи студентів з курсу «Українська мова (за професійним спрямуванням)». Львів : Вид-во Нац. ун-ту «Львівська політехніка», 2010. 92 с.
181. Наукові засади та теоретико-методологічні принципи створення сучасних енциклопедій : колективна монографія / за ред. д-ра іст. наук, проф. Киридон А. М. Київ : Державна наукова установа «Енциклопедичне видавництво», 2015. 160 с.
182. ГОСТ 7.74-96 Межгосударственный стандарт система стандартов по информации, библиотечному и издательскому делу информационно-поисковые языки : термины и определения. Минск, 1997. 38 с.
183. Суперанская А. В., Подольская К. В., Васильева Н. В. Общая терминология: Вопросы теории. Москва : Наука, 1989. 246 с.
184. Burkhanov I. Linguistic foundations of ideography. Semantic Analysis and Ideographic Dictionaries. Rzeszow: Wydawnictwo Wyzszej Szkoły Pedagogicznej, 1999. 388 pp.
185. Широков В. А. Комп'ютерна лексикографія. Київ : Наукова думка, 2011. 351 с.
186. Рогова Н. А. Тезаурус как средство повышения эффективности современных информационно-поисковых систем. *Труды Академии управления МВД России*. 2011. № 1 (17). С 113-118.
187. Лобузін К. Корпоративна бібліотечна база знань : наук.-метод. посіб. НАН України, Нац. б-ка України ім. В. І. Вернадського. Київ, 2016. 103 с.
188. Широков В. А. Феноменологія лексикографічних систем. Київ : Наукова думка, 2004. 327 с.
189. Широков В. А. Елементи лексикографії. Київ : Довіра, 2005. 304 с.

190. Широков В. А. та ін. Лінгвістичні та технологічні основи тлумачної лексикографії : монографія. Київ : Довіра, 2010. 295 с.
191. Кучеренко О. Розбудова сучасної національної термінології цивільного захисту. *Вісник Нац. ун-ту «Львівська політехніка»*. Серія «Проблеми української термінології». 2009. № 648. С. 92–94.
192. Симоненко Л. О. Біологічна термінологія : формування та функціонування : навч. посіб. Умань : РВЦ «Софія», 2006. 103 с.
193. Шуневич Б. Сучасні способи відбору термінів та укладання перекладних словників нових терміносистем. *Вісник Житомирського держ. ун-ту ім. Івана Франка*. 2008. № 38. С. 90–93.
194. Нікуліна Н. В. Термінознавство : лінгвістичний опис української термінологічної системи автомобілебудування : навч. посіб. Харків: Вид-во ХНАДУ, 2006. 144 с.
195. Радзієвська С., Чекман І. Принцип укладання англо-українського словника-довідника з нанонауки. *Термінологічний вісник* : зб. наук. праць. Київ : Інститут української мови НАНУ, 2013. Вип. 2 (1). С. 171–179.
196. Петренко О. В. Етапи організації вибірки науково-технічної термінології в галузі робототехніки. *Наукові записки* [Ніжинського державного університету ім. Миколи Гоголя]. Сер. : Філологічні науки. 2014. Кн. 1. С. 168–171.
197. Апресян Ю. Д. Интегральное описание языка и системная лексикография. Москва : Языки русской культуры, 1995. Т. 2. 767 с.
198. Караулов Ю. Н. Общая и русская идеография. Москва : Наука, 1976. 355 с.
199. Табанакова В. Д. Идеографическое описание научной терминологии в специальных словарях : дис. ... д-ра филол. наук : 10.02.21. Тюмень, 2001. 284 с.
200. Шмелев Д. Н. Проблема семантического анализа лексики (на материале русского языка). Москва : Наука, 1973. 280 с.

201. Вискушенко С. А. До проблеми розробки двомовного термінологічного тематичного словника англійської фахової мови тваринництва. *Науковий вісник Волинського національного університету ім. Лесі Українки. Філологічні науки. Мовознавство* : зб. наук. праць. Луцьк : Волинський національний ун-т ім. Лесі Українки, 2011. № 3. Ч. 2. С. 194–197.
202. Гринев С. В. Введение в терминоведение : учеб. пособ. Москва, 1993. 309 с.
203. Дубичинский В. В. Теоретическая и практическая лексикография. Wien-Charkov: Wiener Slawistischer Almanach, 1998. 160 с.
204. Шуневич Б. І. Проблеми укладання нових словників нормалізованих терміносистем (на прикладі пожежної термінології). *Вісник Житомирського держ. ун-ту ім. Івана Франка*. 2009. Вип. 45. С. 138–142.
205. Городецкий Б. Ю. Проблемы и методы современной лексикографии. Новое в зарубежной лингвистике. Вып. XIV. Москва : Прогресс, 1983. С. 5–22.
206. Румянцева Е. А. Аспекты разработки двуязычных учебных электронных словарей. URL: <http://www.dialog-21.ru/dialog2006/materials/pdf/Rumyanceva.pdf> (дата звернення: 02.03.2019).
207. Агрикола Э. Микро-, медио- и макроструктура как содержательная основа словаря. *Вопросы языкознания*. 1984. Вып. 2. С. 72 – 82.
208. Купріянов Є. В. Теоретичні аспекти укладання багатомовного електронного словника гідротурбінних термінів. *Науковий вісник Волинського національного ун-ту імені Лесі Українки*. 2009. № 17. С. 132–136.
209. Кудашев И. С. Проектирование переводческих словарей специальной лексики. Helsinki : Helsinki University Print, 2007. 443 с.

210. Roberts R. P. Methods of Bilingual Dictionary-making: the Canadian experience. Proceedings of the Fifth International Symposium on Lexicography. Tübingen : Niemeyer, 1992. Pp. 91–115.
211. Кунанець Н. Е., Мацюк Г. Р. Вербальне моделювання предметної області засобами побудови тезауруса. *Інформаційні моделі, системи та технології* : матеріали V наук.-техн. конф., м. Тернопіль, 1-2 лют. 2018 р. Тернопіль, 2018. С. 36.
212. Кунанець Н., Мацюк Г. Алгоритм формування двомовного тезауруса інформаційного масиву предметної області «Розумне місто». *Бібліотекознавство. Документознавство. Інформологія*. 2019. № 1. С. 110–117.
213. Яковлев И. Г. Тезаурусный искусственный интеллект : машинное обучение. *Управление мегаполисом*. 2014. № 5. С. 18–33
214. Мацюк Г. Р., Кунанець Н. Е. Лінгвістичний аспект спілкування з експертом при формуванні пошукового тезауруса. *Інформаційні моделі, системи та технології* : матеріали VII наук.-техн. конф. Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 11-12 грудня 2019 р.). Тернопіль, 2019. С. 70.
215. Катренко А. В. Системний аналіз. Львів, 2011. 396 с.
216. Литвин В. В. Технології менеджменту знань : навч. посібник. Львів : Видавництво Львівської політехніки, 2010. 260 с.
217. ДСТУ 7448:2013. Інформація та документація. Бібліотечно-інформаційна діяльність. Терміни та визначення понять. Київ, 2014. 46 с.
218. Воройский Ф. С. Индексирование документов в АБИС. *Библиотека*. 1996. № 9. С. 42–44.
219. Гендина Н. И. Лингвистическое обеспечение информационно-поисковых систем : теория и технологии. *Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества*. 1996. Т. 2. С. 25–27.

220. Зайцева Е. М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем : современные требования и направления развития. *Научные и технические библиотеки*. 2004. № 7. С. 15–27.

221. Сукиасян Э. Р. Логика развития информационнопоисковых языков. *Научные и технические библиотеки*. 2004. № 7. С. 15–27.

222. Бардієр К. Структура лінгвістичного забезпечення сучасних інформаційних сервісів. *Бібліотечний вісник*. 2002. № 1. С. 55–61.

223. Лобузіна К. В. Класифікаційні схеми в лінгвістичному забезпеченні електронних бібліотек : Автореф. дис... канд. техн. наук : 05.13.06. НАН України. Нац. б-ка України ім. В. І. Вернадського. Київ, 2007. 18 с.

224. Вилегжаніна Т., Збанацька О. Лінгвістичне забезпечення для створення уніфікованого пошукового образу та пошукового запиту документа. *Електронні ресурси бібліотек* : за підсумками Всеукр. наук.-практ. конф. директорів держ. та обл. універсал. б-к (14–17 жовт. 2003 р.). Кіровоград, 2003. С. 126–134.

225. Онищенко О. С., Дубровіна Л. А. Електронні інформаційні ресурси бібліотек у піднесенні інтелектуального і духовного потенціалу українського суспільства. НАН України, Нац. б-ка України ім. В. І. Вернадського. Київ : НБУВ, 2011. 235 с.

226. Олифер В., Олифер Н. Основы компьютерных сетей. СПб. : Питер, 2009. 352 с.

227. Лобановська І. Г., Зоріна Н. Є. Предметизація документів з питань освіти, педагогіки, психології. URL: [file:///C:/Users/Admin/Downloads/npdnpbc_2014_4_47%20\(1\).pdf](file:///C:/Users/Admin/Downloads/npdnpbc_2014_4_47%20(1).pdf) (дата звернення: 25.08.2018)

228. Збанацька О. М. Документна евристика в системі соціальних комунікацій: бібліотекознавчий аспект : дис. ... д-ра соц. ком. : 27.00.03. Київ, 2019. 536 с.

229. ДСТУ 2395-2000. Інформація та документація. URL: http://ksv.do.am/GOST/DSTY_ALL/DSTU1/dstu_2395-2000.pdf (дата звернення: 25.03.2019).

230. ГОСТ 7.66-92 (ИСО 5963-85). Индексирование документов. Общие требования к координатному индексированию. [Введен в действие 01.01.93]. Москва, 1993. 16 с.

231. Зарванська І. Координатне (посткоординатне) індексування в електронному просторі як ефективний засіб створення пошукового образу документа. *Вісник Книжкової палати*. 2012. № 1. С. 27–31.

232. Подошова Н. В. Значення лінгвістичного забезпечення автоматизованих інформаційно-бібліотечних систем та проблеми уніфікації ПІМ для розвитку корпоративної каталогізації. *Актуальні питання каталогізації в умовах впровадження нових інформаційних технологій* : матеріали наук.-практ. семінарів / Держ. закл. «Харк. держ. наук. б-ка ім. В. Г. Короленка». Харків, 2011. С. 42-49.

233. Селиванова Ю. Г., Войскунский В. Г. Проблемы создания национального авторитетного файла предметных рубрик в России. *Библиотеки и ассоциации в меняющемся мире : новые технологии и новые формы сотрудничества* : 3-я Междунар. конф. «Крым-96» : материалы конф. 1996. Т. 2. С. 146–148.

234. Збанацька О. Розкриття змісту архівних документів: тезаурусний підхід. *Бібліотекознавство. Документознавство. Інформологія*. 2018. № 4. С. 52-64.

235. Мацюк Г. Р. Підготовка масиву термінів для побудови двомовного тезауруса предметної області «Розумне місто». *Social and Human Sciences. PolishUkrainian scientific journal*. 2018. № 3(19). URL: https://sp-sciences.io.ua/s2634980/matsiuk_halyna_2018_.stages_of_preparation_of_terms_array_for_creation_of_bilingual_thesaurus_of_the_smart_city_data_domain._social_and_human_sciences._polish-ukrainian_scientific_journal_03_19

236. PAS 180 Smart city technology. URL: <http://www.bsigroup.com/smart-cities/Smart-Cities-Standards-and-Publication/PAS-180-smart-cities-terminology/> (last accessed: 20.10.17).
237. PAS 181 Smart city framework. URL: <http://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-181-smart-cities-framework> (last accessed: 23.04.17).
238. PAS 182:2014 Smart city concept model. Guide to establishing a model for data interoperability. URL: <http://shop.bsigroup.com/ProductDetail/?pid> (last accessed: 23.04.17).
239. PAS 183 Smart cities – Guide to establishing a decision-making framework for sharing data and information services. URL: www.cibse.org/getmedia/96c2ff0e-73b6-4544-aebd-7debabe211d4/Draft-PAS-183-on-Smart-cities.pdf.aspx (last accessed: 23.04.17).
240. PAS 8101:2014 Smart cities. Guide to the role of the planning and development process. URL: <http://shop.bsigroup.com/ProductDetail/?pid=000000000030294642> (last accessed: 23.04.17).
241. ISO 37120:2014 Sustainable development of communities – Indicators for city services and quality of life (2014). URL: <https://www.iso.org/obp/ui/#iso:std:62436:en> (last accessed: 10.03.17).
242. ISO/TR 37150:2014(en) Smart community infrastructures – Review of existing activities relevant to metrics. URL: <https://www.iso.org/obp/ui/#iso:std:iso:tr:37150:ed-1:v1:en> (last accessed: 20.03.17).
243. ISO/TS 37151:2015(en) Smart community infrastructures – Principles and requirements for performance metrics. URL: <https://www.iso.org/obp/ui/#iso:std:iso:ts:37151:ed-1:v1:en> (last accessed: 20.03.17).
244. ISO/PRF TR 37152 Smart community infrastructures – Common framework for development and operation – Ad hoc group report. URL:

http://www.iso.org/iso/catalogue_detail.htm?csnumber=66898 (last accessed: 20.03.17).

245. Ткачева Л. Б. Основные закономерности развития английской терминологии. Томск, 1987. 198 с.

246. Панько Т. І., Кочан І. М., Мацюк Г. П.. Українське термінознавство : підручник. Львів, 1994. 216 с.

247. Семко Н. М. Англійські терміни-словосполучення у галузі механізації процесів сільськогосподарського виробництва. *Іноземна філологія*. Львів, 1990. № 98. С. 17–22.

248. Головащук С. І. Перекладні словники і принципи їх укладання. Київ : Наукова думка, 1976. 30 с.

249. Грицик Н. Комп'ютерна термінологія та основні способи її перекладу : матеріали XI міжнар. наук.-практ. інтерн.-конф. «Проблеми та перспективи розвитку науки на початку третього тисячоліття у країнах СНД», (Переяслав-Хмельницький, 30-31 трав. 2013 р.). URL: <http://oldconf.neasmo.org.ua/node/2280>.

250. Карабан В. І. Переклад з української мови на англійську мову : навч. посіб.-довід. для студентів вищих закладів освіти. Вінниця : Нова книга, 2003. 608 с.

251. Сидні І. Лендау Словники : мистецтво та ремесло лексикографії / [пер. з англ. Ольги Кочерги]. Київ : К. І. С., 2012. 480 с.

252. Лейчик В. М. Терминоведение : предмет, методы, структура. Москва : КомКнига. 2006. 256 с.

253. Демидова С. Е. Краеведческий документ: библиотечная обработка : учеб.-практ. пособие Москва : Либерия-Бибинформ, 2006. 104 с.

ДОДАТКИ

Додаток А

**Список публікацій здобувача за темою дисертації та відомості
про апробацію основних положень дисертаційної роботи**

Наукові праці, в яких опубліковано основні наукові результати дисертації

Статті в наукових фахових виданнях України:

1. Мацюк Г. Представлення знань у вузькоспеціальних предметних областях за допомогою тезауруса / Г. Мацюк, Н. Кунанець // Бібліотечний вісник. – 2019. – № 1. – С. 19–23.

Розкрито роль галузевого інформаційно-пошукового тезауруса в процесах систематизації термінології предметної області.

2. Мацюк Г. Алгоритм формування двомовного тезауруса інформаційного масиву предметної області «Розумне місто» / Н. Кунанець, Г. Мацюк // Бібліотекознавство. Документознавство. Інформологія. – 2019. – № 1. – С. 110–117.

Проаналізовано технологічні засади створення двомовного тезауруса предметної області «Розумне місто».

3. Мацюк Г. Інформаційно-пошукові тезауруси: світовий та вітчизняний досвід формування / Г. Мацюк // Бібліотекознавство. Документознавство. Інформологія. – 2019. – № 2. – С. 106–115.

4. Мацюк Г. Р. Удосконалення технологій інформаційного пошуку на основі тезауруса предметної області / Г. Р. Мацюк // Бібліотечний вісник. – 2019. – № 5. – С. 20–25.

Статті в наукових періодичних виданнях інших держав:

5. Мацюк Г. Р. Підготовка масиву термінів для побудови двомовного тезауруса предметної області «Розумне місто» / Г. Р. Мацюк // Social and Human Sciences. PolishUkrainian scientific journal. – 2018. – № 3

(19). – URL : https://sp-sciences.io.ua/s2634980/matsiuk_halyna_2018_.stages_of_preparation_of_terms_array_for_creation_of_bilingual_thesaurus_of_the_smart_city_data_domain.soci_al_and_human_sciences.polish-ukrainian_scientific_journal_03_19.

6. Matsuik H. Selective Dissemination of Information – Technology of Information Support of Scientific Research / A. Rzhеuskyi, H. Matsuik, N. Veretennikova, R. Vaskiv // Advances in Intelligent Systems and Computing III. – 2018. – Vol. 871. – P. 235–245. (*Індексовано в Scopus*).

Подано технологію вибіркового поширення інформації.

7. Matsuik H. Formation of Search Queries Based on Thesaurus of Narrowly Specialized Subject Areas / H. Matsuik, A. Rzhеuskyi, N. Kunanets, N. Veretennikova // Econtechmod. An international quarterly journal – 2019. – Vol. 08. – No. 3. – P. 11 – 14.

Проаналізовано роль галузевого інформаційно-пошукового тезауруса для підвищення ефективності інформаційного забезпечення наукових досліджень, що здійснюється з використанням технологій вибіркового розповсюдження інформації.

8. Halyna Matsuik Open online training courses for engineering purpose / Ihor Baran, Nataliya Kunanets, Halyna Matsuik, Mykola Mytnyk, Khristina Shunevich, Yuriy Skorenky and Volodymyr Jaskilka // Workshop Proceedings of the VIII International Scientific Conference «Mathematics. Information Technology. Education», MoMLеT&DS'2019 (Ukraine, Shatsk, June 2-4, 2019). – Shatsk, 2019. – Vol. 2386. – Pp. 331-339. – ISSN 1613-0073. (*Індексовано в Scopus*).

Запропоновано використовувати тезаурус при створенні онлайн-курсів дистанційного навчання інженерного спрямування.

9. Matsuik H. Use of the Smart City Ontology for Relevant Information Retrieval / Nataliya Kunanets, Halyna Matsuik // Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Systems

(COLINS-2019) (Ukraine, Kharkiv, April 18-19, 2019). – Kharkiv, 2019. – Рр. 322–333. – ISSN 1613-0073. (*Індексовано в Scopus*).

Подано алгоритм побудови онтології предметної області «Розумне місто».

Наукові праці, які засвідчують апробацію матеріалів дисертації:

10. Мацюк Г. Р. Онтологічне дослідження предметної області «Розумне місто» / Н. Е. Кунанець, В. В. Пасічник, Г. Р. Мацюк, О. В. Мацюк, О. П. Левченко // Математика. Інформаційні технології. Освіта : VI Міжнар. наук.-практ. конф. (Україна, Луцьк-Світязь, 5-7 червня 2017 р.). – Луцьк-Світязь, 2017. – С. 48–52.

Подано онтологічне дослідження предметної області «Розумне місто як основи для побудови бази знань про «інтелектуальне» місто, запропоновано класифікаційне дерево онтології «Розумне місто» на етапі представлення першого рівня.

11. Мацюк Г. Р. Двомовний тезаурус як соціокомунікаційна складова проекту «Розумне місто» / Г. Р. Мацюк, Н. Е. Кунанець // Современные информационные технологии в экономике и управлении предприятиями, программами и проектами: материалы XV Международной научно-практической конференции (Украина, Харьков, 29 сентября 2017 г.) – Харьков : Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт», 2017. – С. 85–87.

Обґрунтовано актуальність створення комплексного двомовного тезауруса як соціокомунікаційної складової проекту «Розумне місто».

12. Мацюк Г. Р. Вербальне моделювання предметної області засобами побудови тезауруса / Н. Е. Кунанець, Г. Р. Мацюк // Матеріали V науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 1-2 лютого 2018 р.). – Тернопіль, 2018. – С. 36.

Подано особливості вербального моделювання предметної області засобами побудови тезауруса, консолідовано підходи до технологій формування тезауруса.

13. Мацюк Г. Метадані інформаційних ресурсів у контексті інтеграції різнотипних даних у соціокомунікаційних інформаційних системах / Г. Липак, О. Липак, Н. Кунанець, Г. Мацюк // Матеріали V Міжнародної науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 1–2 лютого 2018 р.). – Тернопіль, 2018. – С. 38.

Здійснено порівняльний аналіз стандартів опису метаданих інформаційних ресурсів різних за профілем установ.

14. Мацюк Г. Р. Застосування методу аналізу ієрархій для вибору моделі представлення предметної області «Розумне місто» / Н. Е. Кунанець, А. В. Ржеуський, Г. Р. Мацюк // Математика. Інформаційні технології. Освіта : тези доп. VII Міжнар. наук.-практ. конф. (Україна, Луцьк – Світязь, 3-5 червня 2018 р.). – Луцьк – Світязь, 2018. – С. 68 – 70.

Розглянуто вибір моделі представлення знань предметної області «Розумне місто» у рамках створення інформаційно-пошукового тезауруса.

15. Matsiuk Halyna Application of Saaty Method while Choosing Thesaurus View Model of the “Smart city” Subject Domain for the Improvement of Information Retrieval Efficiency / Kunanets Nataliia, Matsiuk Halyna // 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Ukraine, Lviv, September 11-14, 2018). – Lviv, 2018. – P. 21–25. (Індексовано в Scopus, Web of Science).

Подано методику вибору моделі представлення концептів для побудови тезауруса предметної області «Розумне місто» методом Сааті.

16. Мацюк Г. Р. Формування тезауруса для виконання інформаційного пошуку у вузькоспеціальних предметних областях /

Н. Е. Кунанець, Г. Р. Мацюк // Бібліотека. Наука. Комунікація: 100-річчя Національної бібліотеки України імені В. І. Вернадського: матеріали міжнар. наук. конф. (Україна, Київ, 6-8 листопада 2018 р.) / Нац. б-ка України ім. В. І. Вернадського. – Київ, 2018. – С. 366–369.

Розкрито поняття тезауруса як інструмента інформаційного пошуку ресурсів у автоматизованих інформаційно-бібліотечних системах для створення пошукових образів для вузькоспеціальних предметних областей.

17. Мацюк Г. Стандарти, що регламентують формат подання інформаційно-пошукового тезауруса / Г. Мацюк, Н. Кунанець // Інформаційні моделі, системи та технології: матеріали VI науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 12-13 грудня 2018 р.). – Тернопіль, 2018. – С. 39.

Проаналізовано міжнародні та вітчизняні стандарти, що регламентують побудову інформаційно-пошукових тезаурусів.

18. Мацюк Г. Побудова онтології як інтероперабельної моделі подання знань у процесі наукового дослідження предметної області «Розумне місто» / Г. Мацюк, Н. Кунанець // Інформаційні моделі, системи та технології: матеріали VI науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 12-13 грудня 2018 р.). – Тернопіль, 2018. – С. 40.

Обґрунтовано обрання онтологічного підходу для методологічного забезпечення застосування системного підходу до аналізу та вивчення предметної області.

19. Halyna Matsiuk Application of the “Smart City” data domain thesaurus as the tool for representing knowledge while improving the problem-oriented Web search effectiveness / Halyna Matsiuk, Nataliya Kunanets // 2019 IEEE 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Ukraine, September 17-20, 2019). – Lviv, 2019. – P. 31–34. (Індексовано в Scopus, Web of Science).

Розкрито використання тезауруса предметної області «Розумне місто» як інструмента термінологічного контролю в процесі аналізу, опрацювання документів та пошуку інформації.

20. Мацюк Г. Р. Тезаурус як засіб семантизації понять предметної області / Г. Р. Мацюк, Н. Е. Кунанець // Бібліотека. Наука. Комунікація: актуальні тенденції у цифрову епоху: матеріали міжнар. наук. конф. (Україна, Київ, 8-10 жовтня 2019 р.) / Нац. б-ка України ім. В. І. Вернадського. – Київ, 2019. – С. 94–98.

Розкрито поняття тезауруса як засобу семантизації понять предметної області.

21. Мацюк Г. Р. Лексикографічна систематизація галузевої термінології за допомогою тезауруса / Н. Е. Кунанець, Г. Р. Мацюк // Х Львівський міжнародний бібліотечний форум «Нова бібліотека – відповідальність кожного» : збірник матеріалів. – Київ : УБА, 2019. – С. 64–66.

Розкрито роль галузевого інформаційно-пошукового тезауруса як лексикографічного систематизатора термінології певної галузі.

22. Мацюк Г. Р. Лінгвістичний аспект спілкування з експертом при формуванні пошукового тезауруса / Г. Р. Мацюк, Н. Е. Кунанець // Інформаційні моделі, системи та технології: матеріали VII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя (Україна, Тернопіль, 11-12 грудня 2019 р.). – Тернопіль, 2019. – С. 70.

Розглянуто особливості лінгвістичного аспект спілкування з експертом під час отримання інформації для побудови інформаційно-пошукового тезауруса для використання в бібліотеці.

Публікації, що додатково відображають наукові результати дисертації:

23. Мацюк Г. Р. Тезаурус предметної області «Розумне місто» / Н. Е. Кунанець, Г. Р. Мацюк // Вісник Національного університету

«Львівська політехніка» «Інформаційні системи та мережі». – 2017. – № 872. – С. 179–189.

Окреслено особливості формування тезауруса предметної області «Розумне місто», подано приклад побудови фрагменту онтології з предметної області «Розумне місто».

Відомості про апробацію основних положень дисертаційної роботи

- VI Міжнародна науково-практична конференція «Математика. Інформаційні технології. Освіта» (Луцьк-Світязь, 5-7 червня 2017 р., форма участі – публікація тез);
- XV Международная научно-практическая конференция «Современные информационные технологии в экономике и управлении предприятиями, программами и проектами» (Харків, 29 вересня 2017 р., форма участі – публікація тез);
- V Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 1-2 лютого 2018 р., форма участі – публікація тез);
- VII Міжнародна науково-практична конференція «Математика. Інформаційні технології. Освіта» (Луцьк-Світязь, 3-5 червня 2018 р., форма участі – публікація тез);
- 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Lviv, September 11-14, 2018p., форма участі – публікація тез);
- Міжнародна наукова конференція «Бібліотека. Наука. Комунікація: 100-річчя Національної бібліотеки України імені В. І. Вернадського» (Київ, 6-8 листопада 2018 р., форма участі – публікація тез);
- VI Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 12-13 грудня 2018 р., форма участі – публікація тез);

- The 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019) (Харків, 18-19 квітня 2019 р., форма участі – публікація тез);
- The VII International Scientific Conference «Mathematics. Information Technology. Education», MoMLeT&DS'2019 (Shatsk, June 2-4, 2019., форма участі – публікація тез);
- 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Lviv, September 17-20, 2019, форма участі – публікація тез);
- X Львівський міжнародний бібліотечний форум «Нова бібліотека – відповідальність кожного» (Львів, 18-21 вересня 2019 р., форма участі – публікація тез);
- Міжнародна наукова конференція «Бібліотека. Наука. Комунікація: актуальні тенденції у цифрову епоху» (Київ, 8-10 жовтня 2019 р., форма участі – публікація тез);
- VII Науково-технічна конференція «Інформаційні моделі, системи та технології» (Тернопіль, 12-13 грудня 2019 р., форма участі – публікація тез).

Онлайн-інтерфейс тезауруса WordNet

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options: (Select option to change)

Вікно тезауруса WordNet

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **table**, [tabular array](#) (a set of data arranged in rows and columns) "see table 1"
- [S:](#) (n) **table** (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) "it was a sturdy table"
- [S:](#) (n) **table** (a piece of furniture with tableware for a meal laid out on it) "I reserved a table at my favorite restaurant"
- [S:](#) (n) [mesa](#), **table** (flat tableland with steep edges) "the tribe was relatively safe on the mesa but they had to descend into the valley for water"
- [S:](#) (n) **table** (a company of people assembled at a table for a meal or game) "he entertained the whole table with his witty remarks"
- [S:](#) (n) [board](#), **table** (food or meals in general) "she sets a fine table"; "room and board"

Verb

- [S:](#) (v) [postpone](#), [prorogue](#), [hold over](#), [put over](#), **table**, [shelve](#), [set back](#), [defer](#), [remit](#), [put off](#) (hold back to a later time) "let's postpone the exam"
- [S:](#) (v) **table**, [tabularize](#), [tabularise](#), [tabulate](#) (arrange or enter in tabular form)

Вікна тезауруса ЮНЕСКО англійською мовою

SKOS

Homepage

UNESCO Thesaurus

Overview

Browse thesaurus

Alphabetical browsing

Hierarchical browsing

SPARQL Endpoint

Download

Statistics

Credits and legal notice

UNESKOS Vocabulary

Español

English

Français

Русский

Text to search

RDF/XML | N-Triples | N3/Turtle | JSON | JSON-LD

UNESCO Thesaurus

<http://skos.um.es/unescothes/CS000>

Other languages

- Tesoro de la UNESCO (Español)
- Thésaurus de l'UNESCO (Français)
- Тезаурус ЮНЕСКО (Русский)

Major subject domains

- 1 Education
- 2 Science
- 3 Culture
- 4 Social and human sciences
- 5 Information and communication
- 6 Politics, law and economics
- 7 Countries and country groupings

SKOS

UNESCO Thesaurus

Overview

Browse thesaurus

Alphabetical browsing

Hierarchical browsing

SPARQL Endpoint

Download

Statistics

Credits and legal notice

UNESKOS Vocabulary

Español

English

Français

Русский

art

83 concepts with labels containing "art"

African art (en)

<http://skos.um.es/unescothes/C00072>
Arte africano (es), Art africain (fr), Африканское искусство (ru)
3.45 Art

Ancient art (en)

<http://skos.um.es/unescothes/C00151>
Arte antiguo (es), Art antique (fr), Древнее искусство (ru)
3.45 Art

Arab art (en)

<http://skos.um.es/unescothes/C00198>
Arte árabe (es), Art arabe (fr), Арабское искусство (ru)
3.45 Art

Art books (en)

<http://skos.um.es/unescothes/C00248>
Libro de arte (es), Livre d'art (fr), Книги по искусству (ru)
5.30 Information sources

Art collections (en)

<http://skos.um.es/unescothes/C00249>
Colección de arte (es), Collection d'art (fr), Художественные коллекции (ru)
3.60 Museums

Вікна тезауруса ЮНЕСКО російською мовою

SKOS

Тезаурус ЮНЕСКО

Введение
Просмотреть тезауруса.
Алфавитный презентации
Иерархическое представление
SPARQL-точек доступа
Скачать
Статистика
Кредиты и официальное уведомление
Словарь UNESKOS

Español
English
Français
Русский

Текст для поиска

RDF/XML | N-Triples | N3/Turtle | JSON | JSON-LD

Тезаурус ЮНЕСКО

<http://skos.um.es/unescothes/CS000>

другие языки

- UNESCO Thesaurus (English)
- Tesauro de la UNESCO (Español)
- Thésaurus de l'UNESCO (Français)

Главных предметных областей

- 1 Образование
- 2 Естественные науки
- 3 Культура
- 4 Социальные и гуманитарные науки
- 5 Информация и коммуникация
- 6 Политика, право и экономика
- 7 Страны и группы стран

SKOS

Тезаурус ЮНЕСКО

Введение
Просмотреть тезауруса.
Алфавитный презентации
Иерархическое представление
SPARQL-точек доступа
Скачать
Статистика
Кредиты и официальное уведомление
Словарь UNESKOS

Español
English
Français
Русский

искусство

31 концепций с этикетками содержащих "искусство"

Авангардное искусство (ru)

<http://skos.um.es/unescothes/C00315>
 Avant-garde art (en), Arte de vanguardia (es), Art d'avant-garde (fr),
 3.45 Искусство

Азиатское искусство (ru)

<http://skos.um.es/unescothes/C00272>
 Asian art (en), Arte asiático (es), Art asiatique (fr),
 3.45 Искусство

Арабское искусство (ru)

<http://skos.um.es/unescothes/C00198>
 Arab art (en), Arte árabe (es), Art arabe (fr),
 3.45 Искусство

Африканское искусство (ru)

<http://skos.um.es/unescothes/C00072>
 African art (en), Arte africano (es), Art africain (fr),
 3.45 Искусство

Вікно тезауруса EUROVOC

 **EU Vocabularies**

EuroVoc

Search

EuroVoc

Search

EUROPA > Publications Office of the EU > EU Vocabularies > Controlled vocabularies > Thesauri > EuroVoc > EuroVoc

[Home](#) [Controlled vocabularies](#) [Models](#) [Business collections](#) [Releases](#)

 **EuroVoc**

Concept scheme

EuroVoc

URI: <http://eurovoc.europa.eu/100141>

Version: 4.9.1

[About](#) [Browse content](#) [Documentation](#) [Links](#) [Releases](#) [SKOS](#) [Web services](#)

Filter by:

EXPAND ALL

+ 04 POLITICS

+ 08 INTERNATIONAL RELATIONS

+ 10 EUROPEAN UNION

+ 12 LAW

+ 16 ECONOMICS

+ 20 TRADE

+ 24 FINANCE

+ 28 SOCIAL QUESTIONS

+ 32 EDUCATION AND COMMUNICATIONS

+ 36 SCIENCE

+ 40 BUSINESS AND COMPETITION

+ 44 EMPLOYMENT AND WORKING CONDITIONS

+ 48 TRANSPORT

+ 52 ENVIRONMENT

+ 56 AGRICULTURE, FORESTRY AND FISHERIES

+ 60 AGRI-FOODSTUFFS

+ 64 PRODUCTION, TECHNOLOGY AND RESEARCH

+ 66 ENERGY

+ 68 INDUSTRY

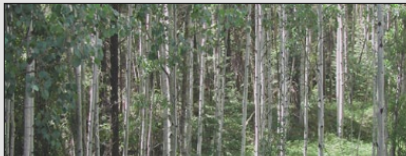
+ 72 GEOGRAPHY

+ 76 INTERNATIONAL ORGANISATIONS

Вікна тезауруса MeSH

NCBI Resources How To Sign in to NCBI

MeSH MeSH Search Limits Advanced Help



MeSH

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

Using MeSH

[Help](#)

[Tutorials](#)

More Resources

[E-Utilities](#)

[NLM MeSH Homepage](#)

You are here: NCBI > Literature > MeSH Database Support Center

GETTING STARTED NCBI Education NCBI Help Manual NCBI Handbook Training & Tutorials	RESOURCES Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures	POPULAR PubMed Bookshelf PubMed Central BLAST	FEATURED Genetic Testing Registry GenBank Reference Sequences Gene Expression Omnibus	NCBI INFORMATION About NCBI Research at NCBI NCBI News & Blog NCBI FTP Site
--	--	---	---	---

NCBI Resources How To Sign in to NCBI

MeSH MeSH flu vaccines Search Create alert Limits Advanced Help

Full ▾ Send to: ▾

Influenza Vaccines

Vaccines used to prevent infection by viruses in the family ORTHOMYXOVIRIDAE. It includes both killed and attenuated vaccines. The composition of the vaccines is changed each year in response to antigenic shifts and changes in prevalence of influenza virus strains. The flu vaccines may be mono- or multi-valent, which contains one or more INFLUENZAVIRUS A and INFLUENZAVIRUS B strains.

Year introduced: 2005 (1963)

PubMed search builder options

[Subheadings](#)

☐ administration and dosage	☐ economics	☐ pharmacology
☒ adverse effects	☐ etiology	☐ physiology
☐ analysis	☐ genetics	☐ radiation effects
☐ antagonists and inhibitors	☐ history	☐ standards
☐ biosynthesis	☐ immunology	☐ statistics and numerical data
☐ blood	☐ isolation and purification	☐ supply and distribution
☐ chemical synthesis	☐ metabolism	☐ therapeutic use
☐ chemistry	☐ organization and administration	☐ toxicity
☐ classification	☐ pharmacokinetics	

☐ Restrict to MeSH Major Topic.

☐ Do not include MeSH terms found below this term in the MeSH hierarchy.

Tree Number(s): D20.215.894.899.302

MeSH Unique ID: D007252

Entry Terms:

PubMed Search Builder

Add to search builder AND ▾

Search PubMed

[YouTube Tutorial](#)

Related information

[PubMed](#)

[PubMed - Major Topic](#)

[Clinical Queries](#)

[NLM MeSH Browser](#)

[dbGaP Links](#)

[MedGen](#)

Recent Activity

[Influenza Vaccines](#)

Сторінка Getty Vocabularies

The Getty: Explore ▾ Getty360 | Calendar | Blog | Connect with Us | Shop | Support Us

The Getty Research Institute

Exhibitions & Events | Special Collections | Library | Search Tools & Databases | Scholars & Projects | Publications | About the GRI

Print

Search Tools & Databases

- Library Catalog
- Getty Research Portal
- Collection Inventories & Finding Aids
- Photo Archive
- Research Guides & Bibliographies
- Digital Collections
- Article & Research Databases
- Collecting & Provenance Research
- BHA & RILA
- ▼ **Getty Vocabularies**
 - Art & Architecture Thesaurus (AAT) ®
 - Cultural Objects Name Authority (CONA) ®
 - Getty Thesaurus of Geographic Names (TGN) ®
 - Union List of Artist Names (ULAN) ®
 - Contribute
 - Editorial Guidelines
 - Getty Vocabularies as Linked Open Data
 - Frequently Asked Questions
 - Obtain the Getty Vocabularies
 - Training Materials

Getty Vocabularies

The Getty Vocabularies are available as Linked Open Data, XML, Relational Tables, and through APIs. [Learn more.](#)

Catherine wheel or rose window? Mona Lisa or La Gioconda? Gaochang or Qara-Hoja? The Getty Vocabularies contain structured terminology for art, architecture, decorative arts, archival materials, visual surrogates, conservation, and bibliographic materials. Compliant with international standards, they provide authoritative information for catalogers, researchers, and data providers. The Getty Vocabularies strive to be ever more multilingual, multi-cultural, and inclusive. The vocabularies grow through contributions from institutions and projects comprising the expert user community. In the new linked, open environments, the Getty Vocabularies may provide a powerful conduit for research and discovery for digital art history. To search the vocabularies, click on the links below. Explore options for contributing or obtaining the Getty Vocabularies data by clicking on links in the left navigation bar.



The Art & Architecture Thesaurus (AAT)®

The AAT is a thesaurus containing generic terms, dates, relationships, sources, and notes for work types, roles, materials, styles, cultures, techniques, and other concepts related to art, architecture, and other cultural heritage (e.g., *amphora, oil paint, olieverf, acetolysis, sintering, orthographic drawings, Olmeca, Rinascimento, Buddhism, watercolors, asa-no-ha-toji, sralais*).



The Cultural Objects Name Authority (CONA)®

CONA compiles titles/names and other metadata for works of art, architecture, and other cultural works, current and historical, documented as items or in groups, whether works are extant, destroyed, or never built; in development, may be used to record works depicted in visual surrogates and for other purposes (e.g., *Florentine Codex, Codice Fiorentino, Guernica, Girl with a Pearl Earring, Chayasomesvara Temple, Hagia Sofia, The Great Wave, Kanagawa oki nami ura, Le déjeuner sur l'herbe, Venus de Milo*).



The Getty Iconography Authority (IA)™

The Getty IA is a thesaurus that



The Getty Thesaurus of Geographic Names (TGN)®

TGN focuses on places relevant to art, architecture, and related disciplines, recording names, relationships, place types, dates, notes, and coordinates for current and historical cities, nations, empires, archaeological sites, lost settlements, and physical features; it is a thesaurus, may be linked to GIS and maps (e.g., *Thebes, Diopolis, Ottoman Empire, Mogao Caves, Ch'ien-fu-tung, Ganges River*).



The Union List of Artist Names (ULAN)®

ULAN contains names, relationships, notes, sources, and biographical information for artists, architects, firms, studios, repositories, patrons, sitters, and other individuals and corporate bodies, both named and anonymous (e.g., *Titian, Tiziano Vecellio, Mark Rothko, Cai Xiang, Crevole Master, Altobelli & Molins, Rajaraja Museum, Jenaer Glaswerk Schott*).



The Categories for the Description of Works of Art (CDWA)

CDWA is a set of guidelines and cataloging rules for the description of art, architecture, and other cultural works. CDWA advises which fields and categories in art metadata are



See What's New

- Growing through contributions: AAT conservation terms: 10,500; translations: Dutch: 63,300; Spanish: 56,200; Chinese: 92,700; German: 21,000

Have a Question?

- ✉ Contact the Vocabulary Program

- 📧 Subscribe to GRI News

Активация Windows

Активация Windows
Перейдіть до налаштувань ПК,

Сторінка тезауруса AGROVOC


Food and Agriculture Organization
of the United Nations

VocabulariesAboutFeedbackHelp

AGROVOC Multilingual Thesaurus

Content languageEnglishSearch

Alphabetical

Hierarchy

- features
 - genomic features
 - physiographic features
 - soil morphological features
 - soil crusts
 - soil horizons
 - A horizons**
 - B horizons
 - C horizons
 - diagnostic horizons
 - E horizons
 - hardpans
 - organic horizons
 - spodic horizons
 - soil micromorphological features
 - soil profiles

features > soil morphological features > soil horizons > A horizons

PREFERRED TERM

① **A horizons**

DEFINITION

① Mineral toprak tabakasının en üst katı veya O horizonunun altında olup organik maddenin karıştığı ve bu nedenle de genellikle koyu renkli olduğu toprağın yıkanma horizonudur.
(tr)

① The A horizon is the top layer of the soil horizons, often referred to as 'topsoil'. This layer has a layer of dark decomposed organic materials, which is called "humus"
(en)

BROADER CONCEPT

soil horizons (en)

IN OTHER LANGUAGES

① **A horizons** French

① **A-Horizonte** German

① **horizontes A** Portuguese

① **horizontes A** Spanish

① **A katmanı** Turkish

① **A horizonu**

URI

http://aims.fao.org/aos/agrovoc/c_47a98f68

Активация Windows

Панель задач и меню «Пуск» Windows не работают. Проверьте настройки ПК, чтобы активировать Windows.

Сторінки ІНІОН РАН


ІНІОН РАН

ОБ ИНСТИТУТЕ • НАУКА И ОБРАЗОВАНИЕ • БИБЛИОТЕКА • ИЗДАТЕЛЬСТВО • РЕСУРСЫ •

Поиск по сайту

Главная / Ресурсы / Базы данных ИНИОН РАН

Библиографические базы данных ИНИОН РАН

Библиографические базы данных ИНИОН РАН по социальным и гуманитарным наукам ведутся с начала 1980-х годов. Общий объем массивов составляет **более 3 млн. 500 тыс. записей** (данные на 1 января 2012 г.). **Ежегодный прирост — около 100 тыс. записей.**

В базы данных включаются аннотированные описания книг и статей из журналов и сборников на 140 языках, поступающих в Фундаментальную библиотеку ИНИОН РАН.

Описания статей и книг в базах данных снабжены шифром хранения и ссылками на полные тексты источников из Научной электронной библиотеки (<http://www.elibrary.ru>).

Активация Windows
Перейдите до настроек ПК, чтоб активировать Windows


ІНІОН РАН

ОБ ИНСТИТУТЕ ▾ НАУКА ▾ БИБЛИОТЕКА ▾ ИЗДАТЕЛЬСТВО ▾ ЖУРНАЛЫ ▾ РЕСУРСЫ ▾ EN

Рубрикаторы и тезаурусы					
24.	Рубрикатор ИНИОН РАН		4 901		
25.	Список периодических изданий, обрабатываемых в ИНИОН РАН		20 000		
26.	Электронный каталог ИНИОНа	2000	57 103	21.02.2002	4 868
27.	Bibliographia Studiorum Uralicorum	1917–1987	75 361		Не пополняется
28.	Тезаурус Правоведение	2000–2003	3 822		
29.	Тезаурус Экономика и демография (PDF-версия)	2001	6 600	5.4.2019	3 600
30.	Тезаурус Философия (PDF-версия)	2003	6 500	21.02.2019	3 700
31.	Тезаурус Социология (PDF-версия)	2018	4 758	29.06.2018	
32.	Тезаурус Гендерные исследования	2003	1 881		
33.	Тезаурус Языкознание	2010	1 890		
34.	Тезаурус Политология	2010	3 706		
35.	Тезаурус Религиоведение	2010	1 240		
36.	Тезаурус Этнология. Антропология (PDF-версия)	2018	6 698	26.10.2018	

Вікно тезауруса з комп'ютерної лексикографії



MOVA.info
ЛІНГВІСТИЧНИЙ ПОРТАЛ

[Вхід](#)


[TRANSLITERACIJA](#)
[СЛОВНИКИ](#)
[ПРОЕКТИ](#)
[ЧИТАЛЬНА ЗАЛА](#)
[ПОСИЛАННЯ](#)

[Головна](#) / [Словники](#) / [Тезаурус з комп'ютерної лексикографії](#)

- Комп'ютерна лексикографія (КЛ)
 - Комп'ютерна ідеографія (КІ)
 - Відношення між одиницями КТ
 - Антонімія
 - Антонім
 - Асоціативні відношення
 - Асоціативний ряд
 - Слово-стимул
 - Гіпонімія
 - Гіпернім
 - Гіпонім
 - Дериваційна гіпонімія
 - Ендоцентричний ряд
 - Тропонімія
 - Каузація
 - Меронімія
 - Пресупозиція
 - Рольові відношення
 - Синонімія
 - Комп'ютерний тезаурус (КТ)
 - Одиниці КТ
 - Укладання КТ

Комп'ютерна лексикографія (КЛ) - розділ лексикографії, напрям комп'ютерної лінгвістики, у межах якого проводиться розробка та впровадження комп'ютерних методів лексикографічних досліджень. Одним із розділів КЛ є комп'ютерна ідеографія.

Шукати:

☒ За початковими літерами
☐ За словом у середині терміна

Активация Win
Перейдіть до настї
Windows.

Вікно тезауруса з лінгвістичної термінології



[Вхід](#)

[ТРАНСЛІТЕРАЦІЯ](#)
[СЛОВНИКИ](#)
[ПРОЕКТИ](#)
[ЧИТАЛЬНА ЗАЛА](#)
[ПОСИЛАННЯ](#)

[Головна](#) / [Словники](#) / [Тезаурус з лінгвістичної термінології](#)

Термін:

українською

термін

Знайти

термін

термінальний вузол

термінальний словник

термінознавство

термінологізація

термінологічне значення

термінологічне словосполучення

термінологічний

термінологічний банк даних

термінологічний вузол

термін /Eng: term/ /Рус: термин/

слово або словосполучення, що позначає поняття певної галузі науки, техніки тощо

Носій певної характеристики

однозначність /Eng: monosemy/ /Рус: однозначность/

логічний принцип побудови знака як основний принцип загальної семіології, згідно з яким кожній одиниці змісту повинна відповідати одна певна одиниця вираження і навпаки

Видові поняття

науковий термін /Eng: / /Рус: научный термин/

Видові поняття

науково-технічний термін /Eng: / /Рус: научно-технический термин/

належить до класу

пермутаційний показчик термінів (дескрипторів) /Eng: permutation indicator of terms (descriptors)/ /Рус: пермутационный указатель терминов (дескрипторов)/

упорядкований за алфавітом список всіх значущих слів, що входять до складу термінів тезауруса: після кожного значущого слова наводяться всі терміни тезауруса, що містять це значуще слово як терміноелемент, таким чином, кожний термін зустрічається в пермутаційному показчику стільки разів, скільки значущих слів він містить

функція, яку виконує об'єкт

поняття /Eng: concept/ /Рус: понятие/

одна з форм мислення, результат узагальнення суттєвих ознак об'єкта дійсності

Операція

стандартизація /Eng: standartization/ /Рус: стандартизация/

Операція

впорядкування /Eng: arrangement, regulation/ /Рус: упорядочение/

Перебуває в асоціативному зв'язку з

професійна лексика /Eng: professional lexicon/ /Рус: профессиональная лексика/

Родове поняття

АБВГДЕ... /Eng: / /Рус: АБВГДЕ...{/}

Мовний рівень

лексичний рівень /Eng: lexical level/ /Рус: лексический уровень/

рівень мовної структури, одиниця якого - лексема

Частина, елемент поняття, названого терміном

термінологічний банк даних /Eng: term databank/ /Рус: терминологический банк данных/

спосіб представлення термінологічної інформації вхідною та вихідною мовами у вигляді словникової бази даних

Вікна інформаційно-пошукового тезауруса Національної бібліотеки України імені Ярослава Мудрого

Термін:

покритонасінні

Пошук

ІПТ

Національна бібліотека України імені Ярослава Мудрого ІНФОРМАЦІЙНО-ПОШУКОВИЙ ТЕЗАУРУС

• [Довідка](#)

Кількість термінів: **35890**

Останні зміни: **2019-08-15**

Правила пошуку

- наберіть термін або його частину (одне чи кілька слів, перші літери слів, як-от **еколог аудитор**) і натисніть клавішу Enter або кнопку [Пошук](#)
- шукається термін з усіма словами запиту, у разі його відсутності перше слово запиту
- знаки пунктуації ігноруються, регістр літер не має значення

© Національна бібліотека України імені Ярослава Мудрого, 1997–2019

Термін:

покритонасінні

Пошук

ІПТ

покритонасінні

тема / предмет –НП (не може мати етнічного поділу) Дата створення 2000-07-03
–ГП (не може мати географічного поділу) Дата ост. ред. 2003-10-28

лат. *angiospermae*
рос. *покрытосеменные*

НВ квіткові рослини
НВ магноліофіти
НВ покритонасінні рослини

Ш [вищі рослини](#)

Н [дводольні](#)
Н [однодольні](#)

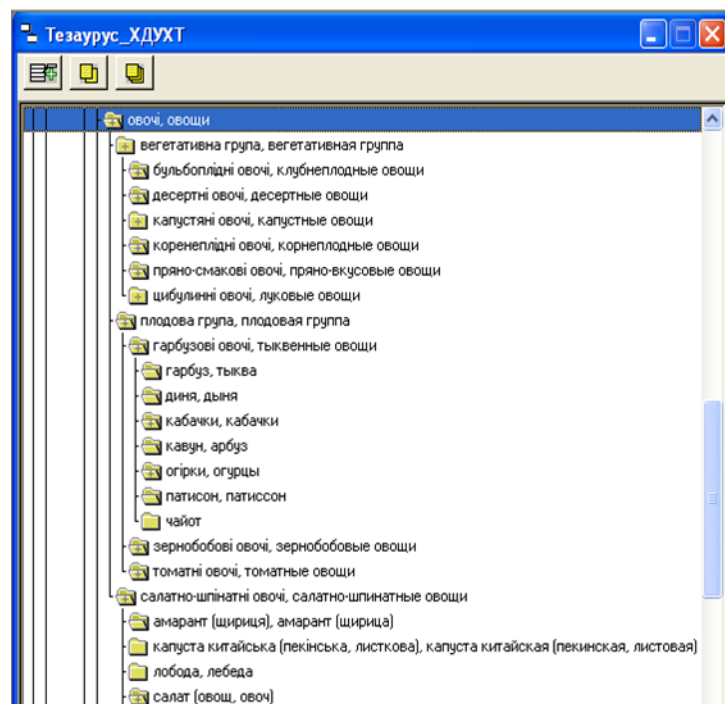
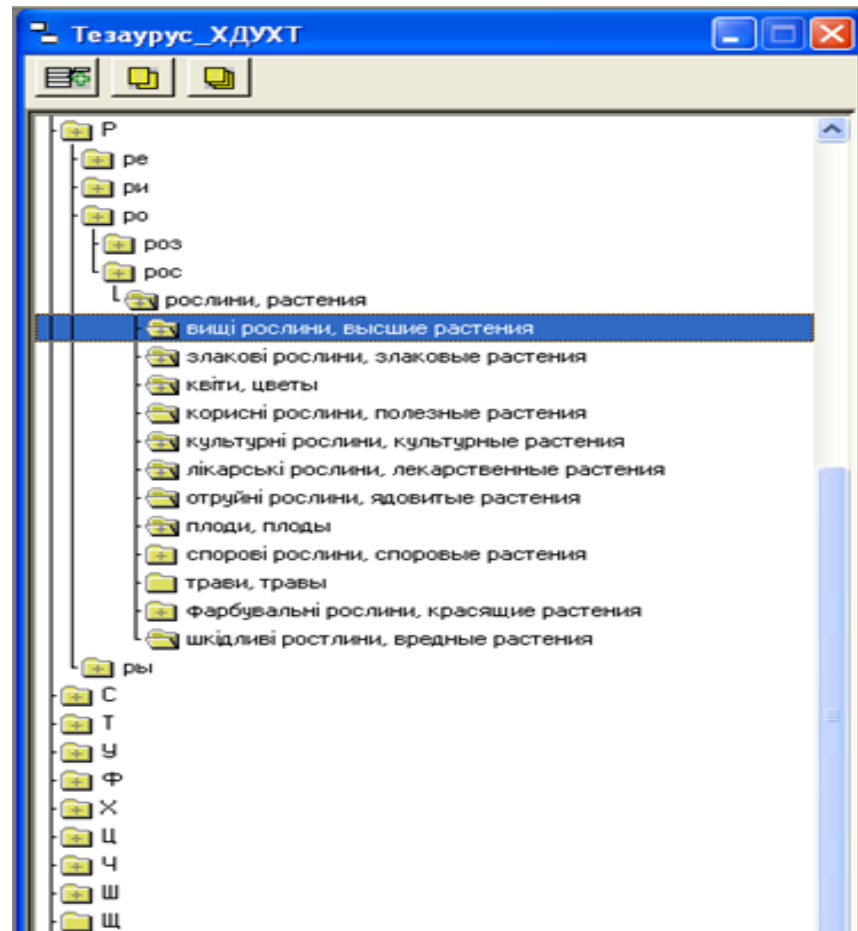
А [квіти](#)

Джерела

1. Биология. Большой энцикл. слов. / Гл. ред. М. С. Гиляров. — 3-е изд., репринт. изд. 1986 г. — М.: Большая Рос. энцикл., 1999. — 863, [1] с.

© Національна бібліотека України імені Ярослава Мудрого, 1997–2019

**Вікна галузевого тезауруса лінгвістичного забезпечення
електронної бібліотеки ХДУХТ**



Шкала відносної важливості

Ступінь важливості	Визначення	Пояснення
1	Однакова важливість	Обидва об'єкти вносять однаковий внесок у досягнення мети
3	Слабка важливість	Досвід і судження надають першому об'єкту легку перевагу перед другим
5	Суттєва або сильна важливість	Досвід і судження надають першому об'єкту сильну перевагу перед другим
7	Дуже сильна та очевидна важливість	Перевага першого об'єкту над другим дуже сильна. Його перевага практично явна.
9	Абсолютна важливість	Незаперечні переконливі свідчення на користь переваги першого об'єкту над другим.
2, 4, 6, 8	Проміжні значення сусідніми судженнями	Ситуація, коли необхідно компромісне рішення
Зворотні величини наведених вище чисел	Якщо при порівнянні одного об'єкта з іншим отримано одне з вищевказаних чисел (наприклад 3), то при порівнянні другого об'єкта з першим отримаємо зворотну величину (тобто 1/3)	

Обчислення пріоритетів

1) Обчислення вектора локальних пріоритетів.

Розрахунки вектора локальних пріоритетів та нормування значень елементів цього вектора проведено за формулою:

$$x_i = \frac{\sqrt[n]{\prod_{j=1}^n a_{ij}}}{\sum_{i=1}^n \sqrt[n]{\prod_{j=1}^n a_{ij}}}, i = \overline{1, n}$$

Спочатку обчислюються компоненти локального вектора пріоритетів за формулою:

$$x_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}, i = \overline{1, n},$$

де a_{ij} – i -й елемент j -го стовпця матриці попарних порівнянь критеріїв, n – кількість критеріїв.

2) Нормалізація векторів локальних пріоритетів.

Нормалізуються вектори локальних пріоритетів таким чином – обчислюється сума стовпця вектора локальних пріоритетів та ділиться кожне число вектора локальних пріоритетів на суму того стовпця:

$$x_i = \sum_{i=1}^n \sqrt[n]{\prod_{j=1}^n a_{ij}}, i = \overline{1, n}.$$

3) Визначення узгодженості локальних пріоритетів.

Для визначення узгодженості локальних пріоритетів потрібно визначити максимальне власне значення матриці, індекс узгодженості та відношення узгодженості.

Максимальне власне значення матриці $\lambda_{\max}$ визначаємо за формулою:

$$\lambda_{\max} = \sum_{i=1}^n U_j \cdot X_j,$$

де $U_j = \sum_{j=1}^n a_{ij}$ – сума елементів j -стовпця матриці; X_j – нормований вектор пріоритетів.

$$\lambda_{\max} = 12,5*0,092+6,83*0,151+13,5*0,087+5,53*0,218+2,12*0,452 = 5,515.$$

Індекс узгодженості (ІУ) думок експертів – кількісне оцінювання суперечливості результатів порівнянь – оскільки оцінки відносної важливості елементів, які порівнюються, повинні бути узгоджені:

$$IU = \frac{\lambda_{\max} - n}{n - 1},$$

$$IU = \frac{5,515 - 5}{5 - 1} = 0,13.$$

Відношення узгодженості (ВУ) – відношення індексу узгодженості до середньостатистичного значення індексу узгодженості при випадковому виборі коефіцієнтів матриці порівнянь. Якщо отримане значення є в межах 10%, то рівень узгодженості може вважатися задовільним. У деяких випадках можна обмежитися 20%, але не більше. Тому цей показник потрібно відстежувати на кожному етапі, щоб по завершенні роботи отримати оптимальні результати. Якщо відношення узгодженості виходить за ці межі, то учасникам потрібно перевірити свої судження.

ВУ обчислюється за формулою:

$$BU = \frac{IU}{BIU},$$

де BIU – випадковий індекс узгодженості, розрахований для квадратної n -вимірної позитивної зворотно-симетричної матриці. У табл. 2.5 представлені значення випадкового індексу (BI) для різних матриць порядку від 2 до 15 [73].

Значення випадкового індексу

Кількість об'єктів	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BIU	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49	1,51	1,54	1,56	1,57	1,59

$$BU = \frac{IU}{BIU} = \frac{0,13}{1,12} = 0,12$$

Матриця за критерієм «простота розуміння»

$$\lambda_{\max} = 5,269; \quad \text{IY} = 0,07; \quad \text{BY} = 0,06$$

Матриця за критерієм «відображення семантики предметної області»

$$\lambda_{\max} = 5,288; \quad \text{IY} = 0,07; \quad \text{BY} = 0,06$$

Матриця за критерієм «системність»

$$\lambda_{\max} = 5,431; \quad \text{IY} = 0,11; \quad \text{BY} = 0,10$$

Матриця за критерієм «універсальність»

№	Моделі представлення знань	Номер моделі представлення знань					Вектор локальних пріоритетів	Нормований вектор цих пріоритетів
		1	2	3	4	5		
1	Логічні моделі	1	5	0,2	0,14	0,13	0,449	0,055
2	Продукційні моделі	0,2	1	0,25	0,13	0,11	0,235	0,029
3	Фреймові моделі	5	4	1	0,33	0,25	1,105	0,137
4	Семантичні мережі	7	8	3	1	0,2	2,020	0,250
5	Онтології	8	9	4	5	1	4,282	0,529
	Сума	21,2	27	8,45	6,6	1,69	8,091	1
$\lambda_{\max} = 5,656; \quad IV = 0,16; \quad BV = 0,15$								

ПРАКТИЧНЕ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

ЗАТВЕРДЖУЮ

проректор з наукової роботи
Тернопільського національного технічного
університету імені Івана Пулюя



д.т.н., проф. П.О.Марущак

АКТ

впровадження результатів дисертаційного дослідження

Мацюк Галини Ростиславівни

«Тезаурус як інструмент інформаційного забезпечення

міждисциплінарних наукових досліджень»

у науково-технічній бібліотеці Тернопільського національного технічного
університету імені Івана Пулюя

Даний акт підтверджує, що основні теоретичні та практичні результати дисертаційного дослідження Г. Р. Мацюк «Тезаурус як інструмент інформаційного забезпечення міждисциплінарних наукових досліджень» використовуються в науково-технічній бібліотеці Тернопільського національного технічного університету імені Івана Пулюя.

Зокрема, запропонована технологія побудови інформаційно-пошукових тезаурусів у бібліотеках, що описує категоріально-понятійний апарат предметної області, сприяє вирішенню актуальних завдань у контексті інформаційного забезпечення користувачів, що проводять наукові дослідження на міждисциплінарних засадах: автоматизації процесів розроблення двомовних пошукових образів; створення алгоритмів формування пошукового запиту у вузькоспеціалізованій предметній області.

Двомовний інформаційно-пошуковий тезаурус «розумного міста» використовується для формування уніфікованого пошукового образу документа й пошукового образу запиту користувачів; індексування документів; здійснення релевантного пошуку в електронних каталогах, базах даних бібліотеки.

Директор бібліотеки
Тернопільського національного технічного
Університету імені Івана Пулюя

Г. Я. Онисько

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційного дослідження Мацюк Г. Р.
«Тезаурус як інструмент інформаційного забезпечення міждисциплінарних
наукових досліджень» у Державному архіві Тернопільської області

Результати, отриманні при виконанні дисертаційного дослідження
Мацюк Г. Р. «Тезаурус як інструмент інформаційного забезпечення
міждисциплінарних наукових досліджень», впровадженні в роботу
архівної установи, сприяючи ефективному розкриття архівних фондів за
допомогою інформаційно-пошукового тезауруса. Розроблені технології
формування тезауруса використовуються при наповненні баз даних
метаданими одиниць зберігання.

Впровадження результатів дисертаційної роботи щодо застосування
тезаурсного підходу до опрацювання архівних фондів дозволило
підвищити якість формування пошукових образів та інформаційного
пошуку релевантних документів у базах даних, що наповнюються в архіві
та покращити інформаційне обслуговування користувачів.

Даний акт не є підставою для фінансових розрахунків.

Директор Державного архіву
Тернопільської області
кандидат історичних наук, доцент



Ф. І. Полянський

27.05.2020 р.

ЗАТВЕРДЖУЮ

Перший проректор
Чернівецького національного
технічного університету
імені Івана Пулюя

М. М. Митник

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційного дослідження Мацюк Г. Р.

«Тезаурус як інструмент інформаційного забезпечення міждисциплінарних
наукових досліджень» у навчальний процес

Ми, що нижче підписалися: декан факультету комп'ютерно-інформаційних систем і програмної інженерії (ФІС) Баран Ігор Олегович, завідувача кафедри комп'ютерних наук (КН) Боднарчук Ігор Орестович – склали цей акт про те, що результати наукових досліджень за темою дисертації Мацюк Г. Р. використовуються в навчальному процесі на факультеті комп'ютерно-інформаційних систем і програмної інженерії

Найменування впровадженого результату	Форма впровадження матеріалів науково-дослідної роботи
1. Застосовано метод аналізу ієрархії для вибору способу представлення знань предметної галузі.	Впроваджено в методичні вказівки з виконання лабораторних робіт з дисципліни «Консолідовані інформаційні ресурси баз даних і знань» (спеціальність 124 Системний аналіз)
2. Технології формування двомовного інформаційно-пошукового тезауруса певної предметної галузі	Впроваджено в лекційний матеріал з дисциплін «Інформаційні ресурси та сервіси в публічному адмініструванні» та «Електронний документообіг та захист інформації» (спеціальність 281 Публічне адміністрування)
3. Використання двомовного інформаційно-пошукового тезауруса «розумного міста», з використанням якого подаються інтерфейси соціального комунікування в контексті інформаційного забезпечення електронної наукової діяльності	Впроваджено як додаткові теоретичні відомості в методичних вказівках до виконання лабораторних робіт з дисципліни ««Електронне місто та регіон» (спеціальність 122 Комп'ютерні науки)

Декан ФІС

Завідувач кафедри КН

І. О. Баран

І. О. Боднарчук