
INTELLECTUAL INFORMATION TECHNOLOGIES

ІНТЕЛЕКТУАЛЬНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

<https://doi.org/10.15407/intechsys.2026.03.030>
УДК 004.8

С.Л. КРИВИЙ, д-р фіз.-мат. наук, професор,
каф. Інтелектуальних програмних систем, ф-т. комп'ютерних наук та кібернетики,
Київський національний університет імені Тараса Шевченка,
просп. Акад. Глушкова, 4Д, Київ, 03022, Україна
<https://orcid.org/0000-0003-4231-0691>
sl.krivoi@gmail.com

Л.М. ГРИДНЄВА, канд. філолог. наук, доцент,
Національний технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського»,
Берестейський просп., 37, Київ, 03056, Україна
grid_l@ukr.net

Т.К. СКРИПНИК, старш. викладач, каф. комп'ютерних наук,
Хмельницький національний університет,
вул. Інститутська, 11, Хмельницький, 29016, Україна
<https://orcid.org/0000-0002-8531-5348>
tkscripnik1970@gmail.com

ІТЕРАЦІЙНИЙ ЛОГІКО-ЛІНГВІСТИЧНИЙ МЕТОД АНАЛІЗУ ПРИРОДНОМОВНИХ ТЕКСТІВ

Пропонується триетапний ітераційний метод добування знань із природномовних текстів, використовуючи формально-логічний і синтактико-семантичний аналіз, тлумачні словники, енциклопедії, спеціальну літературу. Логічний підхід до такого ґрунтується на використанні дескриптивних логік та засобів, якими оперують ці логіки. Метою такого аналізу є побудова несуперечних онтологічних баз знань, поповнення баз новими фактами, та перевірка сумісності бази. Основним завданням в цьому процесі є перевірка виконуваності концептів, несуперечності фактів, поповнення множини фактів новими фактами, які неявно присутні в тексті з метою успішного проведення їхнього формально логічного аналізу. На першому етапі аналізу збирається вся інформація з тексту, яка в цьому тесті явно присутня. На другому етапі в разі необхідності (яка диктується роботою ризонерів) виконується додатковий аналіз тексту для пошуку неявно присутньої в ньому інформації. Для успішної роботи засобів логічного аналізу фактів виникає потреба у пошуку інформації, яка знаходиться за межами аналізованої області. Пошук потрібної (релевантної) інформації виконується

Цитування: Кривий С.Л., Гриднєва Л.М., Скрипник Т.К. Ітераційний логіко-лінгвістичний метод аналізу природномовних текстів. Information Technologies and Systems. 3 (9). 2026. 30—46. <https://doi.org/10.15407/intechsys.2026.03.030>

© Publisher PH “Akademperiodyka” of the NAS of Ukraine, 2025. This is an Open Access article under the CC BY-NC-ND 4.0 license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

на третьому останньому етапі аналізу. Зокрема, цей пошук може використовувати засоби систем штучного інтелекту, різного роду енциклопедичні словники тощо.

Ключові слова: аналіз тексту, вилучення знань, онтологічна база знань.

Вступ

У роботі розглянуто підхід до аналізу природномовного тексту шляхом ітераційного покрокового способу пошуку фактів та поповнення ними системи формально-логічного аналізу. На підставі результатів аналізу будується база знань з перевіркою виконуваності концептів і несуперечності бази знань на основі дескриптивних логік. У пропонуваному підході виділено три етапи побудови онтологічної бази знань (БЗ). На даний час побудовано декілька онтологічних баз знань, найвідомішими з них є *Bdpedia* [1], *Freebase* [2], *YAGO* [3] та інші.

Побудова цих БЗ ґрунтується на понятті графа знань (ГЗ) (*knowledge-graph*) [4], у якому вершини асоціюються з концептами реального світу або абстрактними концептами, а ребра — з відношеннями, які описують семантико-логічні зв'язки між цими концептами. Оскільки знання в часі можуть еволюціонувати і змінюватися, то були впроваджені динамічні графи знань [5], темпоральні графи знань [6], а також їх інтеграція за великими мовними моделями [7] тощо. Побудова БЗ на основі графів знань, який на даний час є домінуючим, та застосування онтологічних БЗ детально описано у огляді [8].

Порівняння перерахованих систем між собою ведеться за засобами логічної мови, яка використовується, її виразністю та доступністю до інших джерел інформації, зокрема до систем штучного інтелекту. З цього погляду онтологічні БЗ поділяються на пропозиційні (логіка висловлювань), модальні (модальна логіка), темпоральні (темпоральна логіка), ймовірнісні (рекомендаційні системи) тощо.

Одним з основних питань при роботі з онтологічною БЗ є питання видобування нових знань та поповнення ними БЗ з метою розширення можливостей самої БЗ. Оскільки задекларовані знання і видобуті нові знання не завжди можуть бути абсолютно істинними, то в роботі [9] розглядалися можливості залучення багатозначних і нечітких логік до генерації наслідків з можливістю оцінки ступеня їхньої логічної істинності. Стосовно логіки, яку пропонується використовувати в даній роботі, то це багатозначна (зокрема тризначна) логіка. Крім того, тут основна увага приділятиметься способам побудови БЗ, виконання аналізу отриманих фактів, несуперечності поповнення новими фактами (знаннями) існуючих та простий спосіб застосування тризначної логіки при перевірці виконуваності запитів. Як зазначалося вище такий процес реалізується у три етапи.

На першому етапі виконується побудова множин концептів, фактів і відношень між концептами, які явно фігурують у тексті, що аналізується (перше наближення результатів аналізу).

На другому етапі знайдені множини концептів поповнюються похідними концептами, фактами і відношеннями, які з'являються в

процесі перевірки виконуваності концептів та логічного аналізу знайдених фактів.

На третьому етапі до БЗ додаються концепти, факти і відношення, добути з тлумачних словників, енциклопедій, спеціалізованих словників.

З метою добування знань, які в них акумульовані, формально логічний аналіз природномовних текстів потребує логічної повноти множини знайдених фактів, оскільки в разі цієї неповноти методи такого аналізу стають незастосовними [10, 12]. Розв'язання цієї проблеми вимагає поповнення БЗ фактами, яких бракує для успішного проведення логічного аналізу. Результати аналізу залежать від точності перекладу знайдених фактів на мову тієї чи іншої математичної логіки (вибір математичної логіки залежить від складності тексту). У разі повноти множини фактів логічний аналіз виконує перевірку на суперечність/несуперечність цієї множини фактів. Проблема неповноти виникає практично завжди, оскільки добути факти, як правило, завжди неповні. Це природно, тому що в тексті можуть фігурувати посилання на інші тексти, архівні дані, аудіо записи, інтерв'ю тощо. Під час аналізу тексту можуть бути потрібні факти, які в ньому явно не фігурують, і тоді виникає потреба внесення їх у систему аналізу, чого вимагає формально-логічний аналіз. На допомогу приходить синтактико-семантичний аналіз тексту, що дає змогу використовувати тлумачні словники, словники аналізованого тексту, глибинні семантичні характеристики понять, термінів, словосполучень тощо.

Основним об'єктом, який нас цікавитиме в цій роботі, є **знання**. Поняття знання ми будемо розуміти згідно з таким загальним визначенням.

Визначення 1. Знання — це філософська категорія для позначення упорядкованого в систему і підтвердженого практикою результату пізнання дійсності навколо нас. Предметними формами існування знання, як відображення об'єктивної дійсності, є такі категорії, як поняття, судження (міркування), гіпотези, теорії, факти тощо [12].

Логіка лінгвістична та логіка математична у системах аналізу текстів

Для аналізу природно мовних текстів застосовують два типи логік: синтактико-семантичну логіку і формальну математичну логіку.

Синтактико-семантична логіка передбачає пошук алогізмів у текстах, повторень, тавтологій у визначеннях тощо, які є надлишковими. Видалення надлишковості спрощує (симпліфікує) текст. Синтактико-семантичний аналіз тексту (особливо наукового) пов'язано зі знанням логічних законів побудови тексту. Цей аналіз потребує уваги до термінологічного апарату і правил визначення термінів. Саме за допомогою визначення розкривається значення (семантика) термінів.

Визначення 2. Термін — це лексична одиниця для спеціальних цілей, яка позначає загальне, конкретне або абстрактне поняття (концепт) певної галузі знань [13].

У цьому визначенні наголошується на тому, що, **по-перше**, терміни мають ті семантичні й формальні ознаки слів і словосполучень, що і слова природної мови; **по-друге**, терміни функціонують у лексичній мові тільки для спеціальних цілей; **по-третє**, терміни служать для позначення спеціальних понять. Можна тільки додати, що до термінів ставляться нормативні вимоги, які розрізняють термін і не-термін. Основних рис терміна, які відрізняють його з-поміж інших лексичних одиниць, є декілька.

1. *Однозначність*: термін має позначати тільки одне наукове або технічне поняття, а поняттю повинен відповідати тільки один термін.

2. *Системність*: будь-який термін — член певної терміносистеми, отже, значення терміна фіксується в системі й підтримується параметрами, що використовуються в ній.

3. *Мотивованість*: значення терміна визначається всією системою понять і приписується йому за допомогою дефініції; мотивованість терміна сприяє його запам'ятовуванню, полегшує зв'язок з іншими термінами.

4. *Лінгвістична правильність*: точність терміна значною мірою забезпечується завдяки правильному використанню словотворчих засобів.

5. *Термінологічна упродовженість*: під час укладання термінологічних рекомендацій перевага надається терміну, який має ширше розповсюдження, більш тривалу традицію використання поколіннями спеціалістів.

На відміну від звичайної побутової лексики терміни мають соціально-побутовий характер, тому що в них відображаються факти, що їх спостерігає дослідник, а також їх теоретичне осмислення. Кожний крок дослідження закріплюється у термінах — продуктах мисленнєвої діяльності людини.

Однозначно визначити термін важко тому, що це залежить від предметної області, до якої він належить. В разі логічного визначення терміна на перше місце ставиться його зв'язок із поняттям: з одного боку, будь-який термін називає поняття, з іншого — виражає його, але характер цього зв'язку ще досі нез'ясований [2, 7]. Тут важливими є дві обставини:

1) поняття, яке позначається терміном, взаємопов'язане з іншими поняттями тієї самої предметної галузі та є елементом системи понять;

2) термін є елементом термінологічної системи, і є взаємопов'язаним з іншими термінами.

Математична логіка спершу потребує перекладу об'єктів (концептів), які фігурують у тексті, зв'язків об'єктів між собою (відношень), ознак самих об'єктів з природної мови на формальну мову математичної логіки. Тільки після цього методи логічного аналізу застосовуються до результатів такого перекладу. Завдання математичної логіки полягає в тому, щоб виявити в отриманій інформації формально-логічні суперечності/несуперечності, повноту/неповноту термінології та фактів, і знайти нове знання у вигляді наслідків.

Розглянемо різницю між визначенням у лінгвістиці і у формальній логіці.

Визначення 3. Визначення у лінгвістиці — це така логічна операція, за допомогою якої:

- розпізнається зміст якогось терміну;
- описується значення (семантика) якогось слова чи словосполучення;
- стає можливим відрізнити один термін від інших термінів [13, 15].

Результатом цієї логічної операції є об'єкт, який визначається (*дефінієндум*) і засіб визначення (*дефінієнс*). При цьому намагаються слідувати Аристотелевим правилами [14, 15].

Математична логіка максимально конкретизує ці закони і правила, оскільки в ній семантика визначається однозначно і правила, на підставі яких вона побудована, вже не залежать від природномовного тексту за умови точного перекладу всіх понять і зв'язків між ними на мову формальної логіки.

Далі в роботі використовується сім'я дескриптивних логік, зокрема логіка *ALC* та її теоретико-множинна семантика [11]. Дескриптивні логіки будують за допомогою введення конструкторів, які застосовують до множин первинних концептів (нотація *CN*) і відношень (нотація *RN*). Основними конструкторами у *ALC* логіці є заперечення концепту — $\neg C$, перетин концептів — $C \sqcap D$, об'єднання концептів — $C \sqcup D$, квантори $\exists R.C$ і $\forall R.C$, де R — атомарна роль, C, D — довільні концепти.

У мовах дескриптивної логіки під **концептом** розуміють клас індивідних об'єктів, що можуть мати деякі прості властивості, які називають *атрибутами* [16]. Усі концепти мають наповнення (обсяг). Один і той самий концепт, яким позначається певний клас об'єктів, нічого не говорить про його обсяг. Для того щоб добути цю інформацію, застосовують операцію поділу (обсягу, наповнення) концептів. *Поділ поняття* — це операція, за допомогою якої розкривається наповнення родового поняття через перелік його сортів (видів) або елементів.

Між концептами вводяться бінарні відношення $R_1, R_2, \dots, R_n$, які характеризують зв'язки між ними. Концепти і відношення поділяють на атомарні та похідні. Атомарний концепт (відношення) — це первинний клас об'єктів (відношень), за допомогою яких визначаються нові концепти і відношення, які називають похідними [11, 16]. Формалізація похідних концептів і їхні властивості описуються за допомогою термінологічних аксіом (*TBox* — *Terminological Box*) та фактів (*ABox* — *Assertion Box*).

Визначення 4. Термінологічною аксіомою називають вираз $C \sqsubseteq D$ або $C \equiv D$, де C, D — довільні концепти. Термінологією (*TBox*) називають скінченну множину термінологічних аксіом деякої предметної області.

Вираз $C \subseteq D$ означає скорочення концепт C «міститься» в концепті D , а вираз $C \equiv D$ – скорочення концепти C і D «еквівалентні». Системою фактів $ABox$ називають скінченну множину A тверджень такого вигляду $a:C$ і aRb , де a, b – деякі індивіди, C – довільний концепт, R – довільне відношення.

Твердження $a:C$ і aRb означають, що індивід a належить концепту C , і що індивіди a, b зв'язані відношенням R відповідно.

Семантику логіки ALC визначає інтерпретація $I = (\Delta, f)$, де Δ – область інтерпретації, f – функція інтерпретації. Функція інтерпретації ставить у відповідність

- кожному атомарному концепту A підмножину (його наповнення в області Δ) $A^f \subseteq \Delta$;
- кожній атомарній ролі R – (і її наповнення в області D) $R^f \subseteq \Delta \times \Delta$.

Інтерпретація I розповсюджується індукцією за побудовою концепту у такий спосіб:

- $T^f = D$, $\perp^f = \emptyset$, $\neg C^f = \Delta \setminus C^f$, (де $T(\perp)$ – істинний (хибний) концепт);
- $(C \cup D)^f = C^f \cup D^f$, $(C \cap D)^f = C^f \cap D^f$;
- $\exists(R.C)^f = \{e \in \Delta \mid \exists d \in D: (e, d) \in R^f \wedge d \in C^f\}$;
- $(\forall R.C)^f = \{e \in \Delta \mid \forall d \in D: (e, d) \in R^f \rightarrow d \in C^f\}$.

Визначення 5. У дескриптивній логіці визначенням називається введення нового концепту за допомогою конструкторів і операцій цієї логіки через атомарні концепти та відношення і раніше визначені (похідні) концепти та відношення.

Термінологію T називають системою визначень, якщо всі її аксіоми мають вигляд $A \equiv C$, де A – атомарний, а C – довільний концепт, причому в T немає двох різних аксіом з однаковою лівою частиною. Самі аксіоми вигляду $A \equiv \Box C$ називають визначеннями.

Аксіома $C \subseteq D$ ($C \equiv D$) істинна за інтерпретації I , якщо $C^I \subseteq D^I$ ($C^I = D^I$). I в такому разі називається моделлю цієї аксіоми, Інтерпретацію I називають моделлю термінології T , якщо вона є моделлю для всіх аксіом із T . Термінологія T називається **сумісною** або **виконуваною**, якщо вона має непорожню модель.

Концепт C виконується відносно термінології T , якщо існує модель I термінології T така, що $C^I \neq \emptyset$.

Нехай T – система визначень. Поділимо атомарні концепти, які зустрічаються в T , на дві категорії:

- **базові** (атомарні, первинні), які не належать до лівої частини жодної з аксіом із T ;
- **похідні** (ті, що визначаються), які належать до лівої частини хоча б однієї аксіоми із T .

Базовою називають інтерпретацію J , яка інтерпретує лише базові концепти.

Нехай J – базова інтерпретація, а I – довільна інтерпретація. I є розширенням інтерпретації J , якщо області їх інтерпретації збіга-

ються $D^I = D^I$ і вони однаково інтерпретують базові концепти, тобто для кожного базового концепту B справедливо $B^I = B^I$.

Систему визначень T називають **коректною**, якщо для довільної базової інтерпретації J існує єдине розширення I , яке є моделлю термінології T .

Термінологія дає можливість записувати загальні знання про концепти і ролі. Але часто існує необхідність записувати знання про конкретні індивіди: якому класу належить індивід, якими відношеннями (ролями) вони зв'язані між собою. Це виконується в тій частині бази знань, яку називають системою фактів про індивіди або $ABox$. Із цією метою, крім множини CN атомарних концептів і множини RN атомарних ролей, тобто імен для класів і відношень, вводять також **скінченну множину IN імен індивідів**. Факти про індивідів є двох типів:

- твердження про належність індивіда a до концепту C ($a:C$);
- твердження про зв'язок двох індивідів a і b роллю R (aRb).

Семантика $ABox$ визначається за допомогою інтерпретації. Візьмемо довільну інтерпретацію $I = (\Delta, f)$ і розширимо її на індивіди: кожного індивіда $a \in IN$ зіставляємо з елементом області інтерпретації $a^I \in \Delta$.

Факт $a:C$ або aRb справедливий в інтерпретації I , якщо $a^I \in C^I$ або $(a^I, b^I) \in R^I$. У такому разі інтерпретацію I називають **моделлю** цього факту. Інтерпретацію I називають *моделлю системи фактів A* , якщо вона є моделлю всіх фактів з A . $ABox$ називають *виконуваним* у термінології T , якщо A має модель (яка одночасно є моделлю термінології T).

Визначення 6. Базою знань (БЗ)} називають пару $K=(T, A)$, де T — довільна термінологія ($TBox$), а A — довільна система фактів ($ABox$). Інтерпретацію I називають моделлю бази знань K , якщо I є моделлю T і A . БЗ називають **виконуваною** або **несуперечною**, якщо вона має непорожню модель.

Факт α (вигляду $a:C$ або aRb) впливає з K , якщо α виконується в довільній моделі I бази знань K .

У логіці ALC може порушуватися лише Аристотелеве правило заборони циклу у визначеннях. Справа в тому, що можуть існувати термінологічні аксіоми вигляду

$$OSOBA \equiv CCAVEЦЬ \cap \forall \text{ має.батька } OSOBA,$$

у яких є цикл. Такого типу цикли розв'язуються за допомогою операції «блокування точки», яка розриває цикл [11]. Насправді ліва і права частини, де фігурує концепт $OSOBA$, мають різні семантичні значення (різне наповнення).

Перейдемо до ілюстрації ітераційного методу аналізу тексту, побудови бази знань та проблем, які виникають в процесі такого аналізу на конкретному прикладі.

Приклад аналізу тексту і побудови його бази знань

Розглянемо приклад тексту, на якому продемонструємо етапи аналізу, роль логіки в цьому аналізі та способи видобування понять, термінології, відношень та фактів, рисутніх у ньому.

Приклад 1. «СІМ'Я І ШКОЛА» (текст, взятий в інтернеті з оповідання невідомого автора, який не хотів відкривати своє прізвище).

«В тихому мальовничому селищі, яке притулилося одним боком до лісу, а другим до огорожі місцевої птахофабрики та невеличкого клаптика поля, жила сім'я Петренків. Сім'я складалася з батька **Семена Петровича**, матері **Олени Степанівни** та двох доньок **Іринки** та **Яни**. Сім'я була звичайна, як більшість українських сімей. Діти ходили до школи. Іринка була старша і вчилася уже в сьомому класі, а Яна була меншою в сім'ї і вчилася у третьому класі селищної школи-десятирічки. Загалом у школі навчалось **82 учні**, серед яких **42 учні** навчалися у початкових класах, а решта - у старших класах. Учителів у школі було небагато і тому деякі учителі викладали по декілька предметів у різних класах. Зокрема **математику і фізику** викладала **Марія Петрівна**, **біологію і хімію** — **Петро Іванович**, **Степан Іванович** — **фізкультуру та військову справу**. **Стеха Романівна** — **природознавство, екологію**, **Семен Петрович** — **працю та народознавство**, **Петро Юрійович** — **історію**, **Микола Миколайович** — **українську та зарубіжну літературу**, **решта вчителів** викладали **домоводство, етику, психологію, спів**. Всього у школі працювало **11 учителів**, які забезпечували навчально-виховний процес. Їм у цьому допомагали інженерно технічні працівники: електрик **Михайло Микитович**, сантехнік **Павло Павлович**, завідувач шкільним господарством **Микола Михайлович** та дві прибиральниці — **Марина Тихонівна** та **Тамара Степанівна**. На допомогу цьому колективу приходили **батьки учнів** та **представники птахофабрики** (в особі **директора** та **членів професійної спілки підприємства**), яка була основним роботодавцем для населення селища.

Іринка вчилася старанно на відміну від Яни, яка ще не зовсім розуміла, навіщо вивчати арифметику чи правопис, якщо сучасні комп'ютерні системи це зроблять краще, ніж вона. Тому Ірина часто допомагала молодшій сестрі у навчанні і не тільки. Іринка мала багато подруг у своєму класі і у старших класах. Яна теж дружила зі всіма у своєму класі, через те у неї і виникали проблеми з навчанням, оскільки вона віддавала перевагу іграм з друзями, а не читанню скучних підручників. Серед учнів були такі, які віддавали перевагу тим, чи іншим предметам і тому деякі **класи мали певну предметну орієнтацію**.»

Яку інформацію можна одержати з наведеного тексту і як її упорядкувати та поповнити так, щоб можна було прослідкувати життя школи, її сучасний стан, учнів, учителів, батьків та інших учасників навчально-виховного процесу у школі ?

Перший етап аналізу

Аналіз наведеного тексту дозволяє виділити концепти, які безпосередньо фігурують у наведеному тексті, а саме:

- селище, школа, птахофабрика,
- учні,
- учителі,
- інженерно-технічний склад працівників (ІТП),
- батьки, діти,
- статъ,
- представники підприємства;
- предмети навчання.

Отже, маємо таку множину концептів: $CN = \text{ОСОБА} \cup \text{ОБ'ЄКТ}$, де

ОСОБА $\equiv \{\text{УЧЕНЬ, УЧИТЕЛЬ, ІТП, БАТЬКИ, ДІТИ, ПРЕД-ПІДПР}\} \cup \{\text{ЖІН-СТАТЬ}\}$

ОБ'ЄКТ $\equiv \{\text{ПТАХОФАБРИКА, ШКОЛА, СЕЛИЩЕ, ПРЕДМЕТИ, ПРОФЕСІЯ}\}$,

які є *атомарними концептами*.

Цю множину потрібно класифікувати (поділити) і наповнити конкретними об'єктами та відношеннями між ними, оскільки різні об'єкти цієї множини мають різні професії, різні завдання і функції тощо. Ця класифікація в такому разі виконується шляхом введення відношень, на підставі яких створюється термінологія. Як це відбувається?

Вводяться відношення такого типу:

учень — це особа яка вивчає якісь предмети;

учитель — це особа, яка викладає якісь предмети;

ІТП — це особи, які забезпечують умови успішного проведення (навчального) процесу;

батьки — це особи, які мають дітей, і ці діти навчаються в школі;

працівники підприємства — це особи, які сприяють успішному про-
веденню (навчального) процесу у школі і працюють на підприємстві.

Отже, дістаємо *таку множину (атомарних) відношень* RN :

- R_1 = працює-у-школі,
- R_2 = представник -підприємства = працює-на-підприємстві,
- R_3 — статъ-особи,
- R_4 = має-дітей,
- R_4^1 — має-батьків,
- R_5 = викладає-предмет,
- R_6 = фах-особи,
- R_7 = вивчає-предмет,
- R_8 = навчається-у-школі.

Елементи множин CN і RN графічно пов'язані такими відношеннями (рис. 1).

Тепер за цими відношеннями будується термінологія ($TBox$):

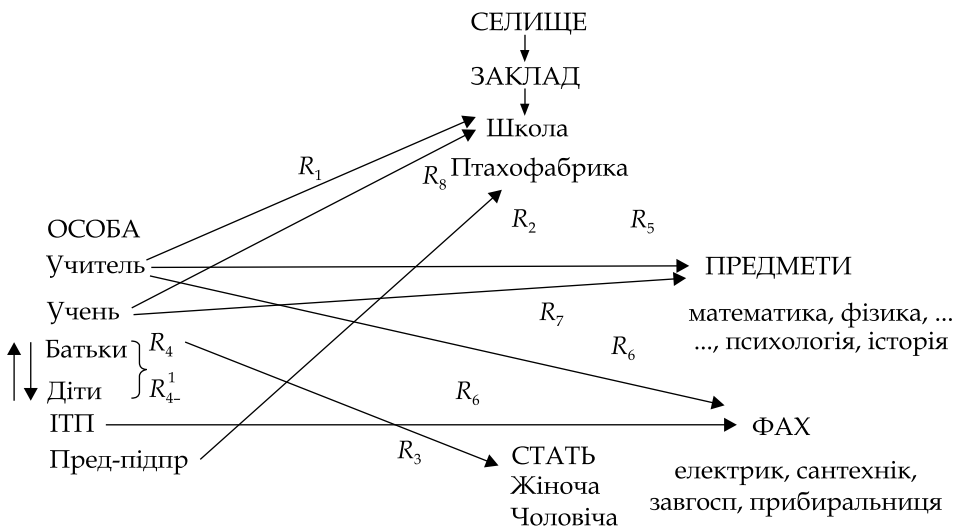


Рис. 1. Граф-схема концептів і відношень між ними

УЧЕНЫ $\sqsubset$ ОСОБА $\forall$ вивчає-предмет. **ТУЧИТЕЛЬ** $\equiv$ ОСОБА $\cap$ $\forall$ викладає-предмет. Т

ПРЕДМЕТИ $\equiv$ {математика, арифметика, фізика, біологія, хімія, фізкультура, військова справа, екологія, природознавство, праця, історія, домоводство, правопис, етика, психологія, українська-література, зарубіжна-література};

ІТП $\equiv$ ОСОБА $\cap$ $\forall$ професія (* електрик, сантехнік, завгосп, прибиральниця *) – це індивіди;

БАТЬКИ $\equiv$ ОСОБА $\cap$ $\exists$ УЧЕНЫ (має-дітей(ОСОБА, УЧЕНЫ));

ДІТИ $\equiv$ ОСОБА $\cap$ $\forall$ має-батьків. Т;

РОДИНА $\equiv$ БАТЬКИ $\cup$ ДІТИ;

ПРОФЕСІЯ $\equiv$ ФАХ;

ПРЕД-ПІДПР $\equiv$ ОСОБА $\cap$ $\exists$ ПІДПРИЄМСТВО (пред-підп(ОСОБА, ПІДПРИЄМСТВО)) – це теж можуть бути індивіди.

Далі з наведеного тексту знаходимо (часткову) інтерпретацію (наповнення) знайдених концептів. У даному випадку областю інтерпретації є така множина

Д { **ДІТИ** = {Іринка, Яна, ще принаймні ≥ 80 осіб}.

УЧЕНЫ = {Іринка, Яна, ще 80 індивідів},

УЧИТЕЛЬ = {Марія Петрівна, Петро Іванович, Степан Іванович, Стеха Романівна, Семен Петрович, Петро Юрійович, Микола Миколайович, ВЧ1, ВЧ2, ВЧ3, ВЧ4}.

ІТП = {Михайло Микитович, Павло Павлович, Микола Михайлович, Марина Тихонівна, Тамара Степанівна}.

БАТЬКИ = {Семен Петрович, Олена Степанівна, та ще не більше 80 осіб}.

ПРЕД-ПІДПР = [директор, профспілка].

ПІДПРИЄМСТВО = {птахофабрика, школа}.

Далі виконується наповнення (інтерпретація) відношень ($Abox$):
вивчає = {(Іринка, математика), (Іринка, українська мова), (Іринка, фізика), (Іринка, хімія), (Іринка, біологія), ..., (Яна, правопис), (Яна арифметика),... };

навчається-у-школі = {(Іринка, 7), (Яна, 3),... };

викладає-предмет = { (Марія Петрівна, математика), (Марія Петрівна, фізика), (Петро Іванович, біологія), (Петро Іванович, хімія), (Степан Іванович, фізкультура), (Степан Іванович, військова справа), (Стеха Романівна, англійська мова), (Стеха Романівна, екологія), (Семен Петрович, праця), (Семен Петрович, природознавство), (Петро Юрійович, історія), (Микола Миколайович, українська мова), (Микола Миколайович, зарубіжна література), (ВЧ1,домоводство), (ВЧ2, етика), (ВЧ3, психологія), (ВЧ4, співи)};

фах-особи = $УЧИТЕЛЬ \cup ІТП$ {електрик(Михайло Микитович), сантехнік (Павло Павлович), завгосп(Микола Михайлович), прибиральниця1(Марина Тихонівна), прибиральниця2(Тамара Степанівна)};

має-фах = **фах-особи**;

має-дітей = {(Семен Петрович, Іринка), (Семен Петрович, Яна), (Олена Степанівна, Іринка), (Олена Степанівна, Яна), (x_1, y_1) ...};

пред-підпр = {директор, профспілка},

Такий вигляд мають **результати першого етапу аналізу** без подальшої деталізації, які утворюють Базу Знань для цього фрагменту тексту.

Запитами до такої БЗ є формули вигляду $a:C$ або aRb , де C — концепт, а R — відношення, які відповідно означають що об'єкт a є елементом концепту C і що об'єкти a і b пов'язані відношенням R .

Наприклад, запит:

чи вивчає Яна математику і чи має вона батьків?

набуває такого вигляду у термінології, що наведена вище:

Яна: $(ОСОБА \cap \exists.вивчає(ОСОБА.математика)) \cap \exists.має-дітей(ОСОБА,УЧЕНЬ))$.

Застосуємо алгоритм, за допомогою якого отримуємо відповідь на запит, який називають алгоритмом семантичного табло (СТ) [11].

Яна: $(ОСОБА \cap \exists.вивчає(ОСОБА.математика)) \cap \exists.має-дітей(ОСОБА,УЧЕНЬ))$.

↓

Яна: $(ОСОБА \cap \exists.вивчає(ОСОБА.математика))$, Яна : $\exists.має-дітей(ОСОБА,УЧЕНЬ))$.

↓

Яна: $ОСОБА$, Яна: $вивчає(ОСОБА,математика)$, Яна : $\exists.має-дітей(ОСОБА,УЧЕНЬ))$.

↓

Яна: $ОСОБА \cap \emptyset \cap$ Яна : $\exists.має-дітей(ОСОБА,УЧЕНЬ)) = \perp$

↓

Яна: $\perp$

Отримали суперечність з тим, що Яна належить до порожньої множини. А це означає, що моделі, де йдеться про вивчення Яною математики, немає, оскільки Яна віддає перевагу іграм.

Розглянемо питання несуперечності отриманої бази знань. Згідно з означенням несуперечної БЗ потрібно показати, що вона має непорожню модель. З наведеного вище випливає непорожність моделі БЗ, тобто її несуперечність. Можна було б окремо переконатися у справедливості кожної термінологічної аксіоми і факту у цій моделі.

Другий етап аналізу

Зафіксувавши термінологію і знайдені факти у вигляді БЗ, можна виконувати пошук нових фактів (знань) як логічних наслідків зафіксованих фактів. Це завдання ускладнюється тією обставиною, що деяка кількість концептів і відношень можуть мати невизначену інтерпретацію. Отже, разом з відомою оцінкою ступеня істинності одних концептів, існують концепти, ступінь істинності яких за цієї інтерпретації невідомий. У нашому прикладі інтерпретація концепту

$$C = (\text{ОСОБА} \setminus \{\text{Іринка}\}) \cap \exists.\text{вивчає}(\text{ОСОБА.математика}),$$

який означає: *чи вивчає ще хто-небудь, крім Іринки, математику?* Нам точно відомо, що концепт $\exists.\text{вивчає}(\text{ОСОБА.математика})$ виконується, оскільки існує непорожня модель, яка міститиме особу Іринка, як і концепт $\exists.\neg\text{вивчає}(\text{ОСОБА.математика})$, яка має непорожню модель, яка містить особу Яна. Але виконуваність концепту C не визначена, оскільки хто ще з 80 учнів школи вивчає математику нам з тексту невідомо.

Алгоритм перевірки виконуваності концепту C в такій ситуації вимагає додавання нового індивіда y до множини індивідів IN , наприклад індивіда Ольга. Але *додавання нового індивіда чи концепту мусить супроводжуватися перевіркою несуперечності* такого поповнення. У такому разі необхідно переконатися в тому, що така особа як Ольга є серед учнів школи і ця особа вивчає математику. Глибший аналіз тексту дає підказку, яка виражається словами «*класи мали певну предметну орієнтацію*». Звідси можна зробити висновок, що разом з Іринкою в її класі (предметно орієнтованого на вивчення математики) є ще особи, які вивчають математику (наприклад, Ольга).

Але можлива ситуація, коли з яких-небудь причин ми не можемо перевірити несуперечність такого поповнення. Тоді ми змушені працювати в умовах невизначеності і тут у пригоді стає тризначна логіка (у даному випадку — логіка Лукасевича), яку позначають L_3 [11]. У цій логіці є три логічні значення: 1 — ALC -формула істинна, $1/2$ — значення (ступінь істинності) ALC -формули невизначено, 0 — ALC -формула хибна. Якщо можливий діалог з БЗ та її творцями, то такого типу невизначеності частково розв'язуються за допомогою прямого діалогу.

У протилежному випадку всі добуті факти градуюються значеннями логіки L_3 і залишається сформулювати правила градування наслідків, отриманих після застосування правил побудови моделі у алгоритмі семантичного табло [11, 16]. Це зробити неважко, оскільки якщо ступінь істинності деякої підформули ALC-формули невизначено, то цей ступінь зберігається й у тих підформулах, до яких ця формула належить.

Сформулюємо правила градування формул у процесі застосування L_3 -логіки. Позначимо $v(A)$ значення ALC-формули A . Тоді

- якщо $v(A) = x$, то $v(\neg A) = 1 - v(A) = 1 - x$;
- якщо $v(A \cap B) = 1$, то $v(B) = v(A) = 1$;
- якщо $v(A \cap B) = 0$, то замкнути гілку;
- якщо $v(A \cap B) = 1/2$, то $v(B) = x$, $v(A) = y$, де x, y – значення формул A і B ;

- якщо $v(A \cup B) = 0$, то $v(B) = v(A) = 0$;
- якщо $v(A \cup B) = 1/2$, то $v(B) = v(A) = 1 \setminus 2$;
- якщо $A \cup B = 1$, то $v(A) = 1$ коли A несуперечний і $v(B) = 1$ коли B несуперечний;

- $v(A) = 1$ коли A несуперечний і $v(B) = 1/2$ коли B невизначений;
- $v(A) = 1$ коли A несуперечний і замкнути гілку B коли $v(B) = 0$;
- $v(B) = 1$ коли B несуперечний і $v(A) = 1/2$ коли A невизначений;
- $v(A) = 1$ коли B несуперечний і замкнути гілку A коли $v(A) = 0$;
- якщо $v(\exists R.A) = x$, то ввести індивіда a до концепту A і aRb . Якщо a в A немає, то виконати перевірку несуперечності такого поповнення. Якщо поповнення суперечне, то замкнути гілку у СТ, інакше покласти $v(b:A) = v(aRb) = x$.

- якщо $v(\langle R.A \rangle) = x \neq 0$, то ввести всіх індивідів a до концепту A таких, що aRb , якщо a належить до A . Якщо a в A немає, то виконати перевірку несуперечності такого поповнення. Якщо поповнення суперечне, то замкнути гілку у СТ, інакше покласти $v(a:A) = v(aRb) = x$ для всіх таких $a \in D$.

Наприклад, концепт

$$C = (\text{ОСОБА} \setminus \{\text{Іринка}\}) \cap \exists. \text{вивчає}(\text{ОСОБА.математика})$$

матиме значення $1/2$. А алгоритм семантичного табло побудує дві формули

$$a:\text{ОСОБА} \wedge a \neq \text{Іринка} \text{ і } a:\text{вивчає}(\text{ОСОБА.математика}),$$

які набувають значень $1/2$, оскільки про індивіда a в аналізованому тексті явно нічого не говориться.

Третій етап аналізу

На цьому етапі уточнюється та розширюється множина концептів і відношень, з використанням тлумачних словників, енциклопедій тощо. Зокрема, із Тлумачного Словника української мови [17] дістаємо уточнення таких концептів:

- **ЗНАННЯ:** *знання* — це обізнаність у чому-небудь, наявність відомостей про кого небудь;
 - це сукупність відомостей з якої-небудь предметної галузі, набуті у процесі навчання, дослідження тощо;
 - пізнання дійсності в окремих її проявах та в цілому.
- **ШКОЛА:** *школа* — це навчальний заклад, в якому навчаються учні;
 - це навчальний заклад середньої освіти;
 - це приміщення, де навчаються учні та працюють учителі;
- **ОСВІТА:** *освіта* — це сукупність знань, здобутих у процесі навчання в навчальному закладі (школі, інституті, університеті тощо)
- **ШКОЛА+ДЕСЯТИРІЧКА≡СЕРЕДНЯ-ШКОЛА:** — це школа, в якій учні навчаються протягом десяти років;
- **ШКОЛА ПОЧАТКОВА:** — це школа, в якій учні навчаються протягом перших чотирьох років;
- **СІМ'Я:** *сім'я* — це батьки і діти цих батьків;
- **ІТП:** *інженерно-технічні працівники* — це електрик, сантехнік, завгосп, прибиральниця.
- **УЧИТЕЛЬ:** *учитель* — це особа, яка є авторитетом у своїй галузі,
 - це особа, яка передає свій досвід знання,
 - це особа, яка впливає на інших і є для них прикладом.
- **ДИРЕКТОР** — це особа, яка є керівником підприємства або навчального закладу.
- **НАВЧАЛЬНИЙ ЗАКЛАД** ≡ ШКОЛА.
- **ПРИМІЩЕННЯ** ≡ {приміщення-пта-ф, школа}:
- **СПЕЦІАЛЬНІСТЬ** ≡ ФАХ

Подальший аналіз тексту розширює множину концептів і відношень, які були отримані на попередніх етапах, так званими *смісловими концептами і відношеннями*. Наприклад, КІЛЬК(школа, учнів), ЛОК(птахофабрика, селище), ЛОК(школа, селище), СПЕЦ(учитель, предмет). СПЕЦ1(ітп, спеціальність), ПРИНАЛ(приміщення, школа), ЗВ'ЯЗ(школа, птахофабрика), СУМІСНИК(особа, підприємство). Ці відношення описують кількість учнів, які навчаються у школі, місце знаходження школи, спеціальність учителя, спеціальність інженерно-технічного працівника, яке приміщення належить школі, зв'язок школи і птахофабрики тощо.

Такі додаткові концепти і факти дають можливість точніше і повніше відповідати на запити до БЗ, формувати нові знання, які не фігурують у початковому тексті, поповнювати множини концептів і фактів, які необхідні для завершення роботи систем логічного аналізу, та формувати посилання на джерела, де відповідні об'єкти описуються або використовуються.

З огляду на сказане, сформулюємо проблеми, які виникають в процесі створення БЗ:

- 1) автоматизація повного або часткового поповнення БЗ (які мають бути правила, засоби і механізми);
- 2) автоматизації процесу пошуку фактів, яких бракує для успішного завершення перевірки виконуваності концептів;
- 3) поповнення БЗ метафоричними даними;
- 4) оброблення інформації, яка ї невизначеною (формули зі значеннями $1/2$).

Підсумки

Розглянуто застосування результатів семантико синтаксичного аналізу та мов математичної логіки до видобування знань та перевірку отриманих знань на суперечність/несуперечність. За результатами семантико синтаксичного аналізу тексту на основі мов дескриптивної логіки будуються бази знань, а за допомогою мов тризначної логіки отримуємо ступені істинності висновків в умовах невизначеності.

ПОВІДОМЛЕННЯ:

Внесок авторів: Кривому С.Л. належить ідея статті, розробка математичного апарату, Гриднєвій Л.М. – застосування синтактико-семантичної логіки для пошуку алгоритмів з метою симпліфікації тексту; Скрипник Т.К. – реалізація окремих частин алгоритму семантичного табло.

Застосування ІІІ: автори не використовували засоби ІІІ при підготовці статті.

Конфлікти інтересів: немає конфлікту інтересів.

ЛІТЕРАТУРА/ REFERENCES

1. Morsey M., Lehmann J., Auer S., Stadler C., Hellmann S. DBpedia and the live extraction of structured data from Wikipedia. Program: electronic library and information systems, 2012. 46 (2), pp. 157–168.
2. Bollacker K., Evans C., Paritosh P., Sturge T., and Taylor J. Freebase: a collaboratively created graph database for structuring human knowledge, in *Proc. of SIGMOD*, 2008. pp. 1247–1250.
3. Suchanek F.M., Kasneci G., and Weikum G. Yago: a core of semantic knowledge. In *Proc. of WWW*, 2007. pp. 697–678.
4. Hogan A., Blomqvist E., Cochez M., et. al. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 2021. V. 54. N 4. pp. 1–37.
5. Wang Q., Mao Z., Wang B., and Guo L. Knowledge graph embedding: A survey of approaches and applications, *IEEE Transactions on Knowledge and Data Engineering*, 2017. v. 29. pp. 724–2743.
6. Xu C., Nayyeri M., Alkhoury F., Yazdi H., and Lehmann J. Temporal knowledge graph completion based on time series gaussian embedding, in *Proc. of ISWC*, 2020. pp. 654–671.
7. Qiao Z., et al. A survey on augmenting knowledge graphs with large language models. *AI Open*, 5, 2024. pp. 230–252.
8. Ge H., et al. (2025). On the evolution of knowledge graphs: A survey and perspective. 125 p.
9. Kryvyi S., Hoherchak G. Analyzing Natural Language Knowledge in Uncertainty on the base of manyvalued Logic. *Cybernetics and Systems Analysis*, Vol. 60, № 1, Jan, 2024. pp. 32–47.

10. Kryvyi S., Darchuk N., Skrypnyk T. Analyzing natural languages: antimems, contradiction, ontologies. *Cybernetics and Systems Analysis*, 2023, Vol. 59 (2), 177–189. <https://doi.org/10.1007/s10559-023-00553-7>
11. Kryvyi S.L. *Elements of non-classical mathematical logic*. VTS «Kafedra», Kyiv, 2026. Publishing and Printing Center «Kyiv University», Kyiv, 2026 2024, 608 p. [In Ukrainian] [Кривий С.Л. Елементи неklasичної математичної логіки. Київ: ВПЦ «Київський університет». 2024. 608 с.]
12. Article “Knowledge”. *Ukrainian Soviet Encyclopedia (URE)*, Kyiv, 1979, Vol 4, 2-th ed., 286 p. [In Ukrainian] [Стаття «Знання». Українська Радянська Енциклопедія, Головна редакція УРЕ, Київ, 1979, Том 4, вид. 2-ге, с. 286.]
13. Darchuk N.P., Sorokin V.M. *Term in the linguistic information science*. Publishing and Printing Center «Kyiv University», Kyiv, 2012, 143 p. [In Ukrainian] [Дарчук Н.П., Сорокін В. М. Термін в лінгвістичній інформатиці. ВПЦ «Київський університет», Київ, 2012, 143 с.]
14. Toftul M.H. *Logic*. Textbook, 2-th ed., Academy, Kyiv, 2008, 400 p. [In Ukrainian] [Тофтул М.Г. Логіка: підручник, 2-ге вид., доповнене. К.: Академія, 2008. 400с.]
15. Gridneva L.M. Logical analysis of terminological dictionaries and scientific texts. *Ukrainian linguistics*. Taras Shevchenko National University of Kyiv, 2015. Issue 45/1. P. 111–117. [In Ukrainian] [Гриднева Л.М. Логічний аналіз термінологічних словників і наукових текстів. Українське мовознавство. Київський національний університет імені Тараса Шевченка, 2015. Вип. 45/1. С. 111–117].
16. Baader F., Calvanese D., McGuinness D.L. et al. *The Description Logic Handbook*. Cambridge: University Press, 2010, 601 p.
17. *Explanatory dictionary of the Ukrainian language in 11 volumes*. Naukova Dumka, Kyiv, 1970, 1978. [In Ukrainian] [Тлумачний словник української мови в 11-ти томах. Київ: Наукова думка, 1970, 1978]

Отримано/Received 15.01.2026

Прийнято/Accepted 18.06.2026

Опубліковано/Published XX.XX.2026

S.L. KRYVYI, Doctor of Physical and Mathematical Sciences, Professor,
Department of Intelligent Software Systems, Faculty of Computer Science and Cybernetics,
Taras Shevchenko National University of Kyiv,
40 D, Acad. Glushkova ave., Kyiv, Ukraine, 03022
<https://orcid.org/0000-0003-4231-0691>
sl.krivoi@gmail.com

L.M. GRYDNEVA, Ph.D Philologist Sciences, Associate Professor,
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”,
37, Beresteysky ave., Kyiv, 03056, Ukraine
grid_l@ukr.net

T.K. SKRYPNYK, Senior Lecturer,
Computer Science, Khmelnytskyi National University,
11 Instytutska St., Khmelnytskyi, 29016, Ukraine
<https://orcid.org/0000-0002-8531-5348>
tkscripnik1970@gmail.com

ITERATIVE LOGICAL-LINGUISTIC METHOD OF ANALYZING NATURAL LANGUAGE TEXTS

Introduction. A three-stage iterative method of extracting knowledge from natural language texts is proposed, using formal-logical and syntactic-semantic analysis, explanatory dictionaries, encyclopedias, and special literature. The logical approach to this is based on the use of descriptive logics and the tools that these logics operate with. The purpose of such analysis is to build consistent ontological knowledge bases, supplement the bases with new facts, and check the compatibility of the base.

The main task in this process is to check the feasibility of concepts, the consistency of facts, and supplement the set of facts with new facts that are implicitly present in the text in order to successfully conduct their formal logical analysis. At the first stage of analysis, all information from the text that is explicitly present in this test is collected. At the second stage, if necessary (which is dictated by the work of reasoners), additional text analysis is performed to search for information implicitly present in it. For the successful operation of the logical analysis of facts, there is a need to search for information that is outside the analyzed area. The search for the necessary (relevant) information is performed at the third and last stage of analysis. In particular, this search can use the tools of artificial intelligence systems, various types of encyclopedic, dictionaries etc.

Purpose. The main task in proposed paper are methods of building the feasibility of concepts, the consistency of knowledge base, and supplement the set of facts with new facts that are implicitly present in the text in order to successfully conduct their formal logical analysis.

Methods. A three-stage iterative methods of extracting knowledge from natural language texts is proposed, using formal-logical and syntactic-semantic analysis, using systems of AI (as oracle) explanatory dictionaries, encyclopedias, and special literature.

Results. Proposed an approach to used three-valued logic for estimation of truth-values conclusions. For the successful operation of the logical analysis of facts, there is a need to search for information that is outside the analyzed area. The search for the necessary (relevant) information is performed at the stage of analysis.

Conclusions. We proposed the three stage methods of extraction of new facts and completion of such new facts of knowledge base, checking its consistency.

Keywords: *Text analysis, knowledge extraction, ontological knowledge base.*