



СЕМАНТИЧНИЙ НЕТВОРКІНГ

НА ОСНОВІ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Д. Ланде, Л. Страшной

Інжиніринг
Київ
2025





Дмитро ЛАНДЕ
професор, доктор технічних наук,
завідувач кафедри Інформаційної безпеки
Навчально-наукового фізико-технічного
інституту Національного технічного
університету України «Київський
політехнічний інститут імені Ігоря
Сікорського», Лауреат Премії Кабінету
міністрів України за розроблення і
впровадження інноваційних технологій,
Премії НАН України імені В.М. Глушкова,
Премії НАПрН України імені Ярослава
Мудрого.
ORCID: 0000-0003-3945-1178



Леонард СТРАШНОЙ
Solution Architect for Computer Science Corp,
Amazon Inc, and Tata Consulting Services.
Certified Amazon AWS Data Architect, Google
Certified Digital Cloud Leader.
ORCID: 0009-0008-5575-0286

Семантичні мережі
Великі мовні моделі
Архітектура Трансформера
Реалізація LLM
Семантичний нетворкінг
Мережі понять
Каузальні мережі
Двонаправлений алгоритм
формування каузальних мереж
Рій віртуальних експертів
Формування сценаріїв діяльності
Аналіз ієрархій за допомогою
віртуальних експертів
Візуалізація мереж
Екстрагування прямої мови
Аналіз подій
Семантичне індексування
Інформаційний пошук і LLM
Реконструкція текстів
Перевірка достовірності новин
Перевірка достовірності біографій із
Вікіпедії
Реконструкція наративу
Семантичні мережі у сфері
кібербезпеки
Семантичні мережі у галузі права
Етичні аспекти застосування штучного
інтелекту
Загрози і виклики Black Hat AI
Концепція створення White Hat AI
Епоха безкодових систем

**Ланде Д.В., Страшной Л.Л. Семантичний нетворкінг на основі
великих мовних моделей : монографія. Київ: Інжиніринг, 2025.**

ISBN 978-617-8180-01-0



СЕМАНТИЧНИЙ НЕТВОРКІНГ

НА ОСНОВІ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Д. Ланде, Л. Страшной

Інжиніринг

Київ
2025

УДК 004. 89:004.9

*Рекомендовано до друку
Вченою радою Інституту проблем реєстрації інформації
НАН України
(Протокол № 2 від 21 січня 2025 р.)*

*Рецензенти:
Додонов О.Г. – доктор технічних наук, професор
Субач І.Ю. – доктор технічних наук, професор*

Ланде Д.В., Страшной Л.Л. Семантичний нетворкінг на основі великих мовних моделей : монографія. Київ: Інжиніринг, 2025. – 274 с.

ISBN 978-617-8180-01-0

Сучасна епоха інформаційних технологій потребує інноваційних підходів до аналізу та використання текстових даних. Штучний інтелект, зокрема великі мовні моделі (LLM), надає унікальні можливості для автоматизації створення й аналізу семантичних мереж. Книга пропонує міждисциплінарний підхід, що поєднує методи генеративного штучного інтелекту, математичну формалізацію та практичне застосування в галузях соціальної аналітики, політики, інформаційної безпеки та реклами. У книзі розглядаються нові методи створення, аналізу та модифікації семантичних мереж із використанням LLM. Особливу увагу приділено концепції «рою віртуальних експертів», що базується на багаторазовому виділенні понять і зв'язків для формування семантичних структур, а також модифікації цих мереж для отримання нових текстів або прогнозів. Також запропоновано підходи до вирішення таких проблем, як ефективне використання генеративних моделей для аналізу текстових даних, застосування розроблених підходів у сфері кібербезпеки, соціального моніторингу й політичного аналізу.

Монографію розраховано на фахівців у галузі штучного інтелекту, дослідників у сфері аналізу даних, інформаційної безпеки та соціальної аналітики, а також здобувачів освіти, які цікавляться сучасними підходами до обробки текстових даних. Вона також стане в пригоді практикам, які працюють над завданнями прогнозування, моделювання та інформаційного впливу.

ISBN 978-617-8180-01-0

© Дмитро Ланде, 2025

© Леонард Страшной, 2025

ЗМІСТ

ПЕРЕДМОВА	6
1 ВСТУП	8
1.1 Семантичні мережі	11
1.2 Великі мовні моделі	17
1.3 Модель шести сутностей	28
1.4 Реалізація LLM	30
1.5 Огляд сервісів для створення зображень	37
2 СЕМАНТИЧНИЙ НЕТВОРКІНГ: ФОРМУВАННЯ СЕМАНТИЧНИХ МЕРЕЖ	43
2.1 Застосування LLM для формування мереж	43
2.2 Ієрархічне формування причинно-наслідкових мереж	51
2.3 Концепція «рою віртуальних експертів» в рамках семантичного нетворкінгу	56
2.4. Кібербезпека в контексті каузальних мереж	74
2.5 Візуалізація мереж	82
3 ЕКСТРАГУВАННЯ ІНФОРМАЦІЇ ТА СЕМАНТИЧНЕ ІНДЕКСУВАННЯ	91
3.1 Автоматичне екстрагування прямої мови з текстів	91
3.2 Аналіз подій на основі застосування LLM	100
3.3 Семантичне індексування	104
3.4 Інтеграція інформаційного пошуку та LLM	117

4 РЕКОНСТРУКЦІЯ ТЕКСТІВ І ПЕРЕВІРКА ДОСТОВІРНОСТІ	128
4.1 Реконструкція текстів через семантичні мережі та LLM	128
4.2 Методологія перевірки достовірності даних	135
4.3 Перевірка достовірності біографій із Вікіпедії	148
4.4 Реконструкції текстового нарративу за допомогою LLM з урахуванням часу та без нього	163
5 ЗАСТОСУВАННЯ СЕМАНТИЧНОГО НЕТВОРКІНГ У РІЗНИХ СФЕРАХ	181
5.1 Формування мереж понять у правовій сфері	181
5.2 Формування семантичних мереж у сфері кібербезпеки	185
5.3 Формування і аналіз семантичної мапи понять в галузі парламентського контролю	189
5.4 Семантичне індексування у сфері інформаційних технологій в будівництві	195
5.5 Семантичне індексування у сфері енциклопедистики	201
6 ПІДТРИМКА ПРИЙНЯТТЯ РІШЕНЬ	208
6.1 Двонаправлений алгоритм формування каузальних мереж	208
6.2 Формування сценаріїв діяльності на базі семантичного нетворкінгу	216
6.3 Рої віртуальних експертів для систем підтримки прийняття рішень	221
6.4 Метод аналізу ієрархій із залученням віртуальних експертів	239
7 ЕТИЧНІ АСПЕКТИ ТА ВИКЛИКИ BLACK HAT AI	239
7.1. Етичні аспекти застосування великих мовних моделей	239
7.2 Виклики і загрози Black Hat AI	247
7.3 Принципи створення White Hat AI	250

7.4 Моделі існування White Hat AI	252
7.5 План створення White Hat AI	257
7.6 Епоха безкодових систем	260
8 ВИСНОВКИ	263
БІБЛІОГРАФІЧНІ ПОСИЛАННЯ	265

ПЕРЕДМОВА

Штучний інтелект та його генеративні моделі, зокрема великі мовні моделі (LLM), стали невід'ємною частиною сучасного суспільства, суттєво змінюючи підходи до аналізу даних, інформаційного пошуку та прийняття рішень.

Революційна архітектура трансформера, як нейронної мережі нового покоління, змінила правила гри в сфері штучного інтелекту і, взагалі, всіх інформаційних технологій. Вона дозволила створити моделі й методики, здатні ефективно обробляти великі обсяги даних, навчатися без вчителя, розуміти контекст і семантику тексту, а також генерувати осмислені відповіді, які майже не відрізняються від людських. Завдяки використанню механізму уваги (attention), трансформери можуть фокусуватися на найважливіших частинах даних, що робить їх надзвичайно потужними для задач, таких як машинний переклад, текстова аналітика, створення зображень, і навіть моделювання складних наукових процесів.

У центрі уваги цієї книги – семантичний нетворкінг, нова методика, яка передбачає використання LLM, для створення, аналізу та модифікації семантичних мереж. Головне її призначення – представити нові методи формування та аналізу семантичних мереж, зокрема, із використанням концепції «рою віртуальних експертів», що базується на багаторазовому опрацюванні запитів (промптів) до LLM. Такий підхід дозволяє виявляти концепти (поняття), встановлювати приховані зв'язки між ними, зважувати їхню значущість та генерувати нові знання на основі існуючих даних.

У книзі наведені методологічні та технологічні засади семантичного нетворкінгу, які включають:

1. Розробку методології формування семантичних мереж за допомогою LLM, враховуючи частотність та значущість концептів і зв'язків.
2. Використання сучасних інструментів для кластеризації й візуалізації мереж, зокрема Gephi.
3. Створення алгоритмів для модифікації семантичних мереж, які дозволяють інтегрувати нові знання або реконструювати тексти.
4. Впровадження математичних моделей для формального опису процесів формування та трансформації семантичних мереж.
5. Створення нових і удосконалення існуючих методик підтримки прийняття рішень на базі застосування LLM.
6. Практичне застосування у багатьох сферах, серед яких як кібербезпека, правові науки, парламентський контроль тощо.

Книга розкриває потенційні виклики та ризики, пов'язані з використанням генеративних моделей, з акцентом на етичні аспекти їх застосування.

До речі, в цій книзі майже немає алгоритмів, на їхньому місці знаходяться методики, немає програмних кодів і SQL-запитів, на їхньому місці стоять промпти. Чому це так можна дізнатися у пункті «Епоха безкодових систем».

Монографію виконано у рамках науково-дослідної роботи Інституту проблем реєстрації інформації НАН України за темою «Розробити теоретичні і методологічні засади формування онтологічних моделей на основі автоматичного екстагування знань із мережевих інформаційних ресурсів», номер державної реєстрації 0121U109077 та ініціативної науково-дослідної роботи Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» за темою «Обґрунтування та розроблення аналітично-прогностичних засобів в рамках технології OSINT», номер державної реєстрації 0123U100618.

Виражаємо подяку своїм співавторам по науковим роботам, з якими обговорювались окремі наведені тут результати, посилання на спільні статті з якими включені у цю монографію, а саме Ігорю Субачу, Олександрю Додонову, Олександрю Пучкову, Олексію Новікову, Олександрю Рибаку, Ігорю Свободі, Віктору Кузьмінському, Анатолію Фегеру, Олегу Гуменюку, Дмитру Бушу, Веймонду Роджерсу, Олегу Дрямову.

Книгу призначено для дослідників у галузях штучного інтелекту, інформаційної безпеки, соціальної аналітики та управління даними. Сподіваємося, що вона надихне читача на нові дослідження та впровадження інноваційних технологій для вирішення складних інтелектуальних завдань у різних галузях діяльності.

Дмитро Ланде

Леонард Страшной

P.S. Ця монографія у значній частині складається з матеріалів препринтів, написаних авторами впродовж трьох років революції штучного інтелекту різними мовами. Тому для перекладу на українську ми широко застосовували сервіси DeepSeek і ChatGPT, які ще й забезпечували додаткову структурування матеріалу, частину з якої ми зберегли. Також для ілюстрації обкладинки, що символізує злиття мережевих технологій і штучного інтелекту, ми використовували сервіс Bing Image Creator.

1 ВСТУП

На цей час, коли інформація є одним із ключових ресурсів практично у будь якій сфері, особливого значення набувають методи її аналізу, узагальнення та використання. Семантичні мережі, що розглядається як основа для моделювання знань, роблять наявними зв'язки між окремими концептами, створюючи структуроване представлення інформації. З іншого боку, великі мовні моделі відкрили нову еру в обробці текстових даних, дозволяють їх структурувати, виявляти сутності, генерувати нові й модифікувати наявні тексти.

Ця книга об'єднує ці два потужні напрямки інформаційних технологій, а саме лінгвістичні моделі та аналіз мереж. Запропоновано технологію семантичного нетворкінгу – підхід, що дозволяє будувати, аналізувати й модифікувати семантичні мережі на основі обробки текстових даних із використанням LLM. Цей підхід відкриває нові горизонти у вивченні зв'язків між концептами, прогнозуванні подій, автоматизації створення текстів і навіть моделюванні інформаційних впливів. Завданням цієї книги є представлення методів створення, аналізу та модифікації семантичних мереж із використанням LLM. Особливу увагу приділено концепції «рою віртуальних експертів», що базується на багаторазовому виділенні концептів і зв'язків для формування семантичних мереж, а також модифікації цих мереж для отримання сценаріїв, прогнозів або нових текстів.

Разом з цим, у книзі не лише описується методологія, але й показано, як поєднання семантичних мереж та LLM може бути застосоване в таких сферах, як соціальна аналітика, інформаційна безпека, політичне прогнозування, реклама та наука. Це відкриває можливості для дослідників і практиків, які прагнуть не лише розуміти, а й активно змінювати інформаційне поле на основі сучасних технологій.

Семантичні мережі та великі мовні моделі – це не просто інструменти, а потужні технології, які відкривають нові можливості для аналізу, моделювання та управління інформацією. Тому читачам пропонуються не лише теоретичні знання, а й практичні інструкції, які допоможуть застосувати ці технології у роботі та дослідженнях.

У цьому першому вступному розділі розглядаються наявні фундаментальні аспекти семантичних мереж та великих мовних моделей, закладається основа для розуміння семантичного нетворкінгу. Зокрема, у першому пункті вступу увага приділена історії, розвитку й принципам роботи семантичних мереж. У другому – розкриваються можливості LLM, їх еволюція та використання у сфері аналізу текстів. Ці знання стануть ключовими для подальшого занурення у основний

об'єкт вивчення. Формування і аналіз семантичних мереж передбачає ідентифікацію ключових понять та визначення зв'язків між ними. Цей процес може бути автоматизованим або автоматичним, залежно від обсягу та складності даних. Він включає виявлення ключових концептів, визначення їхньої важливості та дослідження взаємозв'язків між ними. Це дозволяє отримувати глибоке розуміння текстових даних та виявити приховані закономірності.

Другий розділ «Семантичний нетворкінг: формування семантичних мереж» містить інформацію про те, що інтеграція засобів аналізу текстової інформації та візуалізації дозволяє автоматизувати процес формування семантичних мереж. LLM використовуються для екстрагування ключових понять та встановлення зв'язків між ними, тоді як засоби візуалізації, зокрема програма Gephi, забезпечують представлення та аналіз отриманих мереж. Концепція «рою віртуальних» експертів базується на багаторазовому аналізі тексту за допомогою LLM. Кожен «віртуальний експерт» виділяє свої ключові поняття та зв'язки з різним смислом, що дозволяє створювати багатовимірні семантичні мережі. Формування причинно-наслідкових мереж шляхом ієрархічного уточнення дозволяє моделювати складні взаємозв'язки між подіями та їхніми наслідками. Цей підхід використовується для подальшого сценарного аналізу і прогнозування подій.

У третьому розділі розглядаються питання екстрагування інформації та семантичного індексування текстових документів. Серед інших задач, автоматичне екстрагування прямої мови дозволяє виділяти ключові висловлювання та цитати з текстових даних, що спрощує аналіз та інтерпретацію інформації. Формування мереж понять, зокрема, у правовій сфері допомагає у розумінні складних правових концептів та взаємозв'язків між ними. Це спрощує аналіз правових документів та допомагає в процесі прийняття рішень. Аналіз подій на основі генеративного штучного інтелекту дозволяє прогнозувати майбутні події та аналізувати їхній вплив на різні сфери життя. Семантичне індексування документів дозволяє створювати структуровані бази знань, що спрощує пошук та аналіз інформації в різних галузях. Візуалізація мереж допомагає у наочному представленні складних взаємозв'язків між поняттями, що полегшує навігацію в предметній галузі та аналіз інформації. У розділі також розглядаються питання поєднання технологій пошуку інформації та штучного інтелекту. Зокрема, показано, як ця інтеграція дозволяє виявляти та аналізувати кіберзагрози, що сприяє ефективнішому захисту інформації. Використання LLM для підвищення точності пошуку дозволяє знаходити релевантні результати та спрощує процес пошуку

інформації. Семантичне індексування масивів інформації дозволяє створювати структуровані бази знань, що спрощує пошук та аналіз великих обсягів даних.

Четвертий розділ присвячено зворотному побудові семантичних мереж процесу, а саме реконструкції текстів на основі вже сформованих семантичних мереж. Реконструкція текстів через семантичні мережі та LLM дозволяє генерувати нові тексти на основі існуючих мережевих даних. Також у четвертому розділі на основі семантичного нетворкінгу запропоновано методологію перевірки достовірності даних (новинних повідомлень, біографій із Вікіпедії). Методологія перевірки достовірності даних допомагає у визначенні точності та надійності інформації, що використовується.

Питанням застосування семантичного нетворкінгу в різних сферах діяльності присвячено п'ятий розділ книги. Семантичні мережі допомагають у моделюванні та аналізі різних типів кіберзагроз, що сприяє розробці ефективніших заходів захисту. Візуалізація та аналіз даних, зокрема, кіберзагроз за допомогою ШІ дозволяє наочно представляти складні взаємозв'язки між кіберзагрозами та виявляти їхні причини та наслідки. Формування мереж вразливостей і суб'єктів кібербезпеки допомагає у виявленні та аналізі вразливостей, що сприяє покращенню захисту інформації. Також розглядаються прикладні проблем застосування семантичного нетворкінгу у парламентському контролі, зокрема, розглядається аналіз мереж подій у цій сфері, що дозволяє моделювати та аналізувати різні події, що відбуваються при цьому, та їхній вплив на політичну ситуацію.

Шостий розділ присвячено питанням інтеграції технологій семантичного нетворкінгу та підтримки прийняття рішень на основі великої кількості критеріїв, що дозволяє враховувати різні фактори, залучати велику кількість «віртуальних експертів». Динамічне формування мереж через генеративний штучний інтелект дозволяє автоматично адаптуватися до змін у даних і контексті, що сприяє ефективнішому аналізу та моделюванню. Формування сценаріїв діяльності дозволяє моделювати чисельні варіанти розвитку подій та аналізувати їхній вплив на різні сфери життя.

Сьомий розділ присвячено етичним питанням застосування штучного інтелекту та охоплює розгляд викликів «шкідливого штучного інтелекту» Black Hat AI. Загрози виникнення Black Hat AI, змушують моделювати та аналізувати різні сценарії, формувати концепцію створення «корисного» White Hat AI, який має забезпечувати людство захистом від кіберзагроз, загроз Black Hat AI.

У останньому розділі наведені висновки, що відповідають основному змісту, також обговорюються перспективи розвитку методології, наведеної у книзі, та можливі напрямки подальших досліджень.

Таким чином, ця книга може бути корисним ресурсом для дослідників та практиків, які цікавляться інноваційними підходами до семантичних мереж та використанням LLM у різних сферах. Вона пропонує міждисциплінарний підхід, що поєднує методи генеративного штучного інтелекту, математичну формалізацію та практичне застосування в галузях соціальної аналітики, політики, інформаційної безпеки та реклами. Видання розраховане на фахівців у галузі штучного інтелекту, дослідників у сфері аналізу даних, інформаційної безпеки та соціальної аналітики, а також студентів і викладачів, які цікавляться сучасними підходами до обробки текстових даних. Вона також стане в пригоді практикам, які працюють над завданнями прогнозування, моделювання та аналізу інформаційних впливів.

Далі розглянемо основні теоретичні і практичні засади, які є фундаментом для подальшого розгляду семантичного нетворкінгу, основної теми цієї книги.

1.1 Семантичні мережі

Семантичні мережі є універсальним засобом для моделювання й організації знань у вигляді графів, де вершини символізують поняття, а зв'язки між ними відображають семантичні відносини. Такі графові структури легко адаптуються для розв'язання різноманітних завдань у багатьох галузях. Вони дають змогу не лише візуалізувати зв'язки між поняттями, а й глибше аналізувати їхній зміст та взаємодію.

Перші кроки у використанні семантичних мереж були зроблені ще у 1956 році, коли Річард Річенс [Lehmann, 1992] розробляв підходи до машинного перекладу в Кембриджському центрі вивчення мов. У цій ранній концепції семантичні мережі слугували допоміжною графовою моделлю, яка спрощувала переклад текстів між різними мовами, відображаючи значення та смислові зв'язки. З того часу ця ідея суттєво еволюціонувала, і була покладена в основу багатьох сучасних технологій, орієнтованих на обробку та представлення знань.

Особливого значення концепція семантичних мереж набула в 1980-х роках із розвитком веб-технологій. Вона стала базисом для ідеї Семантичного вебу, метою якого було створення структурованого середовища для більш глибокого розуміння інформації, доступної онлайн. Тім Бернерс-Лі, винахідник Всесвітньої павутини, запропонував бачення вебу, який був би зрозумілим не лише для людей, а й для машин. Завдяки цьому веб як мережева структура міг би