

НОВІТНІ ПІДХОДИ ОБРОБКИ ТЕХНОЛОГІЧНОЇ ІНФОРМАЦІЇ ПРОМИСЛОВИХ СИСТЕМ АВТОМАТИЗАЦІЇ НА ОСНОВІ БАЗ ЗНАНЬ

Р. В. Шевченко, здобувач освітнього ступеня доктора філософії

Н. М. Луцька, доктор технічних наук, професор

Національний університет харчових технологій

E-mail: darthhalk@gmail.com, lutkanm2017@gmail.com

Анотація. Статтю присвячено аналізу існуючих та синтезу новітніх підходів до обробки технологічної інформації промислових систем автоматизації в контексті моделювання цифрових двійників різнотипних взаємопов'язаних ділянок та систем. Метою дослідження є аналіз проблематики моделювання баз знань та підвищення ефективності обробки інформації з врахуванням гетерогенних даних. Об'єктом дослідження обрано системи автоматизації трьох ділянок цукрового виробництва: дифузія, дефекосатурація та випарювання. У рамках дослідження розроблено загальну асоціативну карту, що дозволяє чітко структурувати дані промислових систем автоматизації, утворювати зв'язки та насичувати дані контекстом для подальшої обробки. Розроблено семантичну модель бази знань, яка визначає зони впливу даних та дозволяє отримати схему даних, з врахуванням різнотипності даних. Створена семантична модель дозволяє проводити швидкий пошук по контексту та за критеріями, аналізувати вплив підсистем одна на одну, розраховувати ступінь схожості у інформації та аномаліях, що важливо при математичному моделюванні та синтезі мульти-шарового або когнітивного цифрового двійника процесів. На базі семантичної моделі розроблено модель візуалізації ділянок виробництва, яка включає: технологічну документацію, дані з SCADA систем, промислові архіви даних, зовнішні джерела та бази знань через WEB за допомогою метазнань. Використовуючи мову Clojure, було згенеровано запити з пошуку за зонами впливу, контексту, згадкам, розраховувано схожість в знаннях, що дозволяє поєднувати моделі декількох систем без значного збільшення витрат, що в результаті визначає структуру цифрового двійника з урахуванням можливих змін в системах знань, що в сумі гарантує систематизований та ефективний підхід до збору, обробки та збереження даних, а також їхню інтеграцію з іншими технологічними підсистемами. Додатковою перевагою є можливість поєднання окремих вже описаних цифрових двійників у багатошарову комплексну модель, що дозволяє отримати детальне представлення технологічного процесу та підвищити ефективність ухвалення рішень.

Ключові слова: онтологія, цифровий двійник, база знань, дані, інтелектуальні системи управління знаннями, автоматизація

Актуальність. Розвиток технологій зумовлює видозміну методів та принципів обробки інформації. Концепції Індустрії 4.0 та 5.0 надають доступ до різноманітних способів оптимізації технологічних процесів, зокрема: штучний інтелект, машинне навчання, нейронні мережі та моделі, системи підтримки прийняття рішень. Все це викликає потребу у підготовці даних та структурному моделюванні систем технологічного контролю виробництва, що включають як технологічні змінні з промислової лабораторії (дані LIMS) та автоматизованих систем керування (дані зі SCADA), так і документацію, експлуатаційні матеріали, табличні звіти, описи... Це в свою чергу зумовлює необхідність створювати бази знань, розробляти методологію ефективного менеджменту даних (DM – Data Management) і формування систем менеджменту знань (KM – Knowledge Management). Враховуючи, що ефективний менеджмент зобов'язує врахування моделей та стратегій обробки знань [1], а також побудову семантичних зв'язків між об'єктами знань, необхідно мати уніфіковану структуру зберігання даних, яка розріджена в одній площині. Але найчастіше в сучасних системах технологічного контролю, існує різнотипна інформація – звіти, документи, таблиці, відеозвіти, картинки, технологічні дані і так далі, що сильно ускладнює моделювання і примушує створювати складну інформаційну інфраструктуру для їх обробки. Як зазначають Чавез В. і Воларт Д. [2] для ефективного концептуального моделювання потрібно стандартизувати типи даних, що в свою чергу збільшує час розробки інформаційної моделі і навантаження на процес синтезу. Однак, цей крок необхідний для створення цифрових двійників і баз знань, що підтримують системи прийняття рішень. Різноманітність джерел даних додає складнощів у маніпулюванні ними, знижує швидкість обробки та ускладнює початкові етапи підготовки і синтезу рішень у процесах цифровізації виробництва.

Аналіз останніх досліджень та публікацій. В умовах постійного розширення інформаційних систем підприємств все нагальнішим постає питання узгодження інформаційної моделі та баз знань, особливо це стосується випадків, коли підприємство тільки розпочинає етапи цифровізації у відповідності до концепцій Industry 4.0 та “Smart Factory”. Саме перші етапи потребують обстеження, збору інформації з різних джерел, визначення сильнозв'язних та слабкозв'язних систем,

концептуальне моделювання з використанням технологічних даних, оцінки впливу різних підсистем, та формування первинної бази знань з залученням експертних знань, які можна ефективно використовувати як основу для майбутніх моделей, цифрових двійників. На етапі підготовки, визначення впливів на процеси різних чинників ускладнюється гетерогенністю джерел даних, оскільки це можуть бути звіти, таблиці від систем реального часу, картинки з систем аналітики, чи навіть документи формату PDF від систем контролю якості та лабораторних підсистем. Згадані чинники не дозволяють отримати єдине середовище, яке зв'язує різні джерела інформації і дозволяє поєднувати, створювати зв'язки та посилання між ними. Сінгх, С., Шехаб, Е., Хіггінс, Н., Фаулер, К., Рейнольдс, Д., Еркоунду, Дж.А. і Гадд, П. у своїй статті [3] чітко відобразили основи, за якими можна виділити необхідність менеджменту даних, побудови бази знань, створення зв'язків та попередніх математичних моделей. Таким чином виникає необхідність збору усіх можливих даних, які відображають інформацію, зв'язки, вплив на усі елементи системи для побудови математичних моделей цифрових двійників. Кафеза, Е., Дракпопулос, Г. і Майлонас, П. відобразили у [4] можливості використання бібліотек машинного навчання з використанням графів з математичних моделей, які першочергово можна синтезувати з графів моделей баз знань. Чіміно А., Лонго Ф., Мірабеллі Г., Соліна В. та Вертерамо С. у прикладній статті [5] розробили окрему платформу, з використанням рушія Unity для фізичної моделі, окремих 3D-компонентів, та дерева онтологій, яка має усі відсортовані та зв'язані знання за чіткими класифікаціями. Це надає можливість повного відтворення та цифровізації майже будь-якого процесу, втім, підготовка даних вимагає жорсткої ієрархії побудови платформи і вказує на використання в подальшому основ та методів онтологій, як оптимальних. Д'Аміко, Р.Д., Саркар, А., Каррей, М.Х., Аддепаллі, С. та Еркоунджу, Д.А. у своїй статті [6] яскраво висвітлюють нові підходи до використання базових, схожих онтологій, які лягають в основу роботи когнітивних цифрових двійників. Ці підходи дозволяють ефективно генерувати запити, проте потребують підготовки бази знань з урахуванням не тільки даних реального часу, а й створення простору понять, знань та зв'язків. Окрім цього, Чжен, С., Петралі, П., Лу, Дж., Туррін, К. та Кіріціс,

Д. у своїй статті [7] розглянули використання одночасно декількох підходів онтологій, провівши їх декомпозицію на декілька рівнів, та представили підхід до розробки моделей RMPFQ з підтримкою онтологій. Такий підхід дозволяє виносити цифрові двійники на новий рівень, включивши не поодинокі системи автоматизації, а об'єднуючи декілька систем. Цей підхід також додає можливостей подальшого розвитку з використанням даних про розробку та менеджменту виробництва з використанням графових баз даних. Проте такий підхід показав себе не оптимізованим з точки зору використаних ресурсів, і може бути не ефективним при використанні у виробництві малих та середніх масштабів. Окрім того, в статті вказані і складнощі з врахуванням усіх можливих зв'язків, оскільки задумане з'єднання різнотипних систем є високозатратною задачею. Ідею розширення функціоналу також розкрили і Ронген, С.; Ніколова, Н.; ван дер Пас, М. у статті [8], де стало можливим інтегрування системи адміністрування об'єктів та онтології на базі RDF-моделі, однак також зазначаючи важливість менеджменту знань. Коротко вищезгадані факти висвітлені на рис. 1



Рис 1. Менеджмент з використанням онтології з строгою типізацією

За статтею Вальдес-Валенсуела, А. та Гомес-Адорно, Х. [9] частину рішення можна вивести, якщо представляти різнотипні дані у вигляді структур описів з контекстом, і аналізувати їх за алгоритмами NLP (Інтелектуальні алгоритми обробки мови), проте такий підхід не надає можливостей обробки інформації від систем реального часу, і не дає можливостей обробляти специфічні документи, хоч і надає можливість зв'язувати джерела даних у мережу за їх текстовими сутностями. Через орієнтовність на частоту використання без залучення мовної моделі знань підхід не забезпечить достатньої точності. Якщо ж розглядати дані та їх властивості з наданням контексту, як документи NonSQL бази даних і формувати на її основі базу знань, як описують Бланкенберг К., Гебель-Зауер Б. і Шуберт П. [10], можна досить ефективно обробляти дані та звітності ERP (Система управління підприємством), відділень контролю якості та інші сутності, які можуть бути збагачені контекстом, і надалі обробляти їх з використанням методів, які зазвичай застосовують до онтології. Проте сам процес надання контексту вимагає строгої типізації, що зменшує ефективність використання, оскільки різні системи, особливо системи технологічного контролю мають багато різнотипних сутностей і порушують принципи поліглоту та єдиного інформаційного простору, а отже додатково необхідно узагальнювати інформацію. Важливо відмітити і використання мовних моделей та машинного навчання для формування первинного середовища бази знань, як це описують Сунь, К., Луо, Ю., Чжан, В., Лі, С., Лі, Дж., Ніу, К., Конг, Х. та Лю, В. [11]. Оскільки в публікації описується саме ситуація формування зв'язків з урахуванням гетерогенезу джерел інформації, а обробка зображень та документів при приведенні до спільного синтаксису надають можливість уникнути необхідності вручну строго типізувати знання та сутності, призводить до використання текстового синтаксису, аналогічного до Markdown з формуванням зв'язків. Такі сутності можна обробляти за схожими методами, як в онтології, проте даний підхід не зможе достатньою мірою задовільняти потребу у ефективній побудові бази знань, оскільки орієнтується на дуже складну та велику інфраструктуру з залучення LLM (великих мовних моделей), яка не є доцільною для обробки технологічної інформації. Основні критерії та вимоги, які повинні зберігатись були відображені в [12] за авторством

Абоні, Я.; Надь, Л.; Тамаш, Т. На рис. 2 зображено виклики та основну проблематику синтезу бази знань з використанням технологічної інформації з систем автоматизації. Технологічна інформація різних систем у більшості випадків мають різні формати даних, схеми їх побудови, а джерела можуть бути гетерогенні та різнотипові, що призводить до надскладної обробки. І в результаті отримуємо або дуже високі втрати часу та ресурсів в ефективності обробки всієї інформації, або ж зниження точності моделей, які враховують менше впливів. В той же час, якщо концентрувати ресурси на обробці даних тільки реального часу, виникає проблема обмеженого контексту даних, важкості інтеграції документів, які не потрапляють під стандартизовані схеми і в свою чергу погіршують точність. Важливо також розуміти і перспективи розвитку. У статті [13] Гебреаб, С.; Мусаміх, А.; Салах, К.; Джаяраман, Р.; Боскович, Д. Показали приклад розвитку цифрового двійника з залученням фреймворків генеративної моделі штучного інтелекту для аналізу даних.

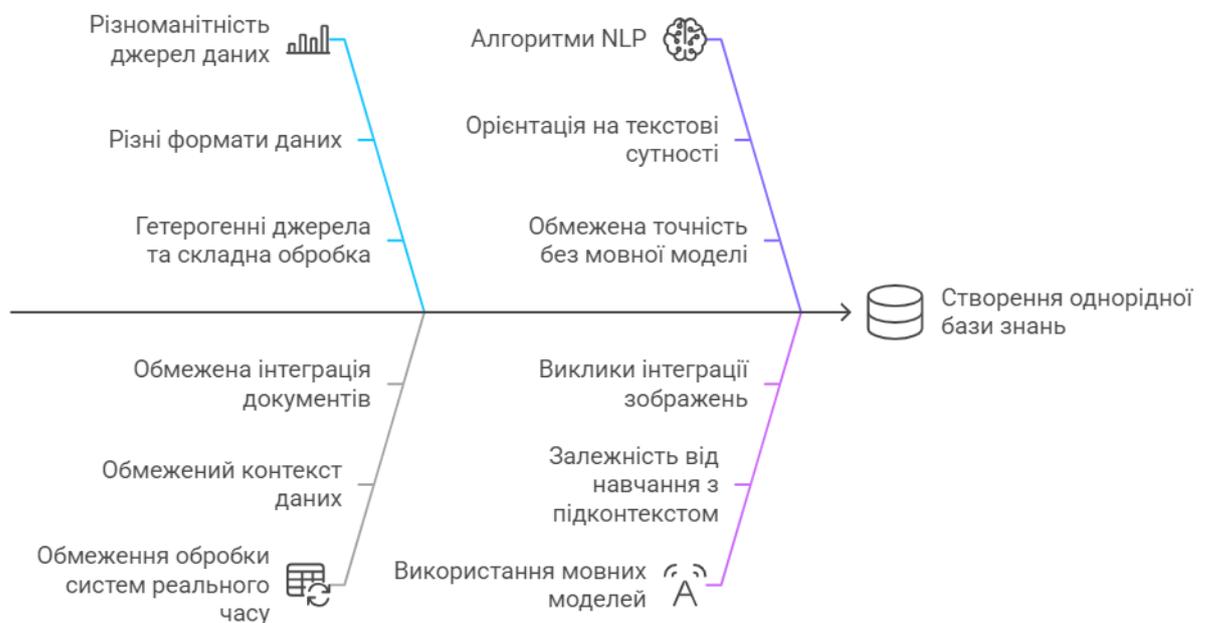


Рис 2. Виклики створення баз знань технологічної інформації

Мета дослідження – підвищення ефективності обробки технологічної інформації шляхом створення єдиного багатозв'язного середовища формування знань. Це дозволить ефективно проводити концептуальне моделювання виробництва, знаходити зв'язні системи виробництва та підвищити точність

синтезованих моделей технологічних процесів, і забезпечити збереження різнотипної інформації в єдиному середовищі.

Матеріали й методи дослідження. Для досягнення поставленої мети було досліджено можливі комбінації принципів та підходів, які дають змогу задовільнити вимоги до середовища бази знань виробництва. Методи, які застосовуються у онтологіях залишаються дієздатними для покращення ефективності, разом з орієнтацією на датацентричність, тобто формування єдиного середовища знань, яке надає доступ до знань різних типів інформації. Як матеріал, було запропоновано вибір між двома аналогічними один до одного середовищами - Logseq та Obsidian. Обидва середовища орієнтовані на збір та зберігання знань, і дозволяють зберігати різнотипні виробничі дані, які при цьому пов'язуються у зв'язки, формуючи семантичну модель знань. При цьому об'єкти організуються у вигляді контекстуальних зв'язків, не потребуючи суворої типізації. Це дозволяє працювати з інформацією, представленою в різних форматах, і створювати зв'язки за контекстом та властивостями, які потім можна обробляти за схожими методами, як обробляються онтології, але з більшою гнучкістю та без потреби в складній інфраструктурі.

У ході дослідження було обрано Logseq, оскільки це середовище має відкритий вихідний код, що дає можливість розробки функціональних інструментів, надає високу гнучкість побудови. На відміну від Obsidian, Logseq заснований з підтримкою мови Clojure, яка представляє собою мультипоточковий діалект мови LISP (Функціональна мова програмування, орієнтована на обробку даних), що дозволяє використовувати, як методи, аналогічні до онтологій, так і будувати зв'язки між сутностями. Окрім цього, дана мова дозволяє надавати об'єктам властивості, наприклад, різна інформація може мати свою область впливу, процес чи тип процесу, який вона характеризує.

На рис 3. показано приклад зв'язування об'єктів через властивості, використано прості сутності, які між собою пов'язані процесом. Так планування та виробництво

входить до самого виробництва, в той час, як тестування напряму до нього не відноситься.

```
○ #+BEGIN_QUERY
{
  :title ["Етапи процесу"]
  :query [:find (pull ?p [*])
    :where
    [?p :block/name]
    [?p :block/properties ?props]
    [(get ?props :process) ?process]
    [(= ?process "Процес виробництва")]
  ]
}
#+END_QUERY
```

Етапи процесу	
page	process
Планування	Процес виробництва
Виробництво	Процес виробництва

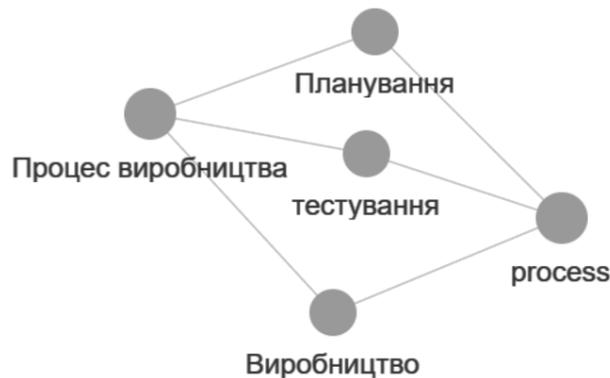


Рис 3. Приклад застосування методів онтології у Logseq, що відображає зв'язок знань через властивість «process»

За об'єкт дослідження було обрано стандартизований технологічний комплекс вироблення цукру, а саме три ділянки: дифузійна, дефекосатураційна та випарювання. База знань має об'єднати технологічну інформацію з документами, базами даних історії з можливістю генерації зв'язків, а також іншої інформації, включно з зовнішніми веб-сторінками. Це дозволяє ефективно формувати основу для моделі процесів, яка необхідна для подальшої розробки цифрових двійників. Збираючи дані про процес, інформація, незалежно від формату, формується в блоки, кожен з яких являє собою окрему сутність, проте ці блоки можуть посилатись один на одного, утворюючи зв'язки, створюючи контекст та генеруючи разом загальний

документ. На рис 4. Зображена асоціативна карта, де кожен елемент – окрема сторінка, що містить блоки, які в подальшому передаються та зв'язуються між собою, утворюючи базу знань з відстежуваними зв'язками.

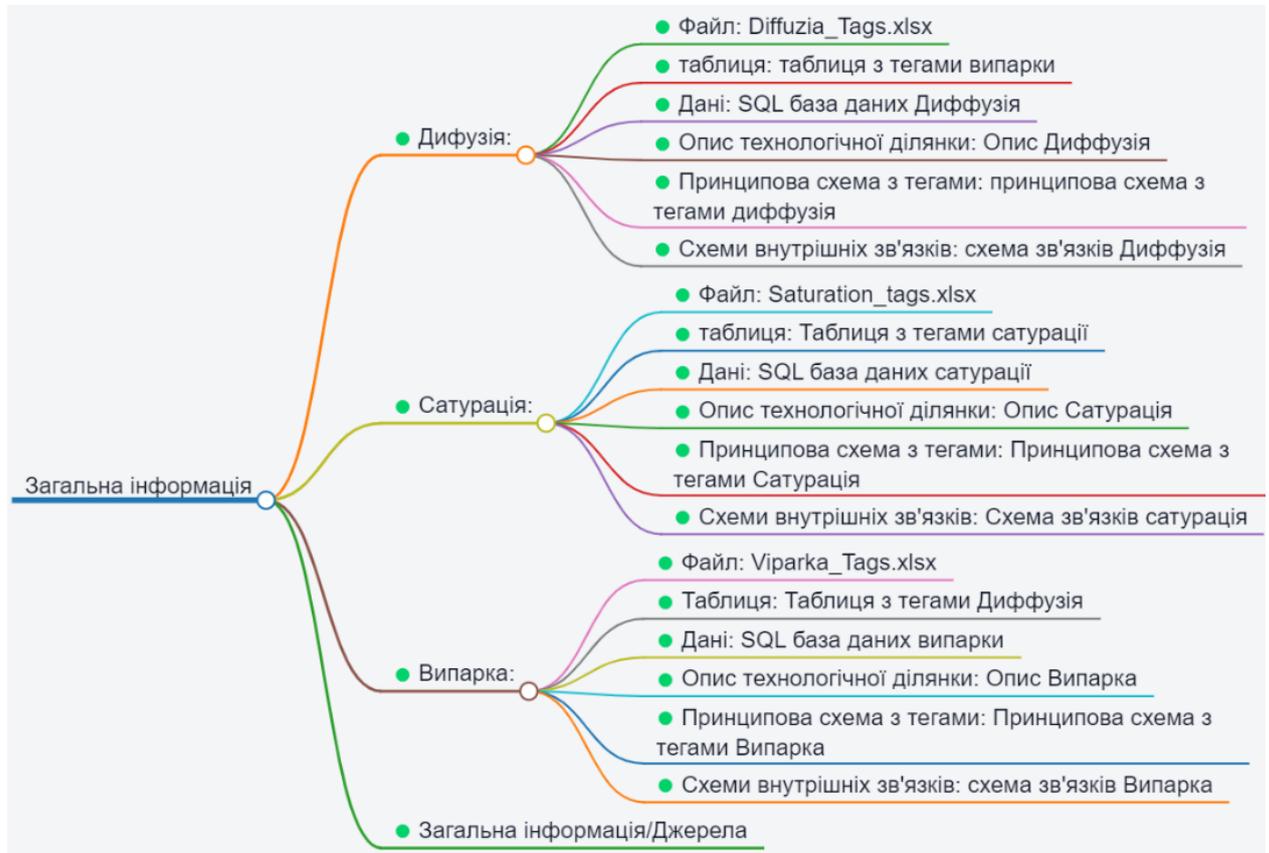


Рис 4. Загальна асоціативна карта зв'язків технологічної інформації про виробництво

У ході дослідження, генеруючи зв'язки, теги та метазнання, можна отримати семантичну модель бази знань. Відповідно до рис. 5, можна виділити 13 окремих скупчення, відповідно до технологічних даних систем автоматизації, баз даних, візуальних даних, зовнішніх джерел, та центральне скупчення №14, яке складає початкову вхідну точку до бази знань, та до якої належать загальні зв'язки про систему в цілому, та яка пов'язує усі джерела інформації, об'єкти та властивості між собою. Таким чином, кожна ділянка має свій рівень впливу на загальну базу.

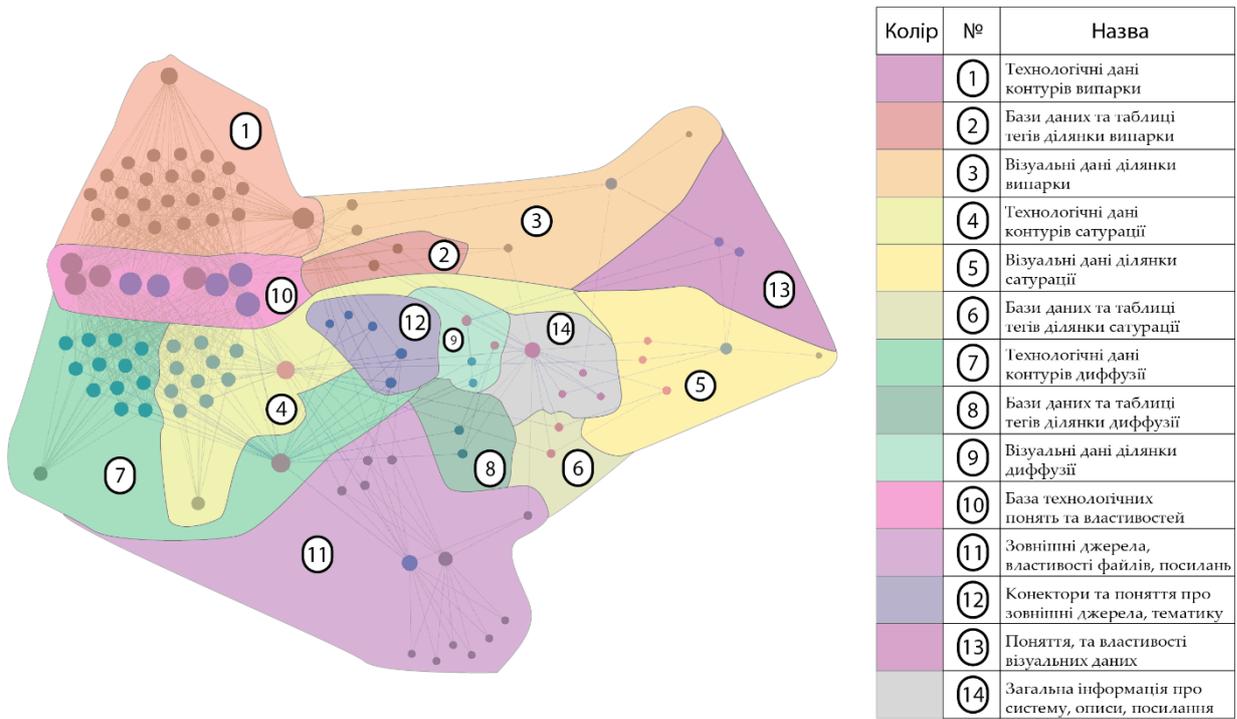


Рис 5. Семантична модель зв'язків бази знань

Описати зони можна так:

1, 4, 7 - Технологічні дані контурів – включають в себе таблиці, CSV файли з даними про кожен контур, тегами з SCADA систем та дані про ділянки відповідно дифузії, дефекосатурації та випарної станції, зони впливу, номери контурів, технологічний параметр керування зі збагаченням властивостями.

2, 6, 8 - Бази даних та таблиці тегів – містять повний збір таблиць тегів з SCADA-систем ділянок дифузії, дефекосатурації та випарної установки, а також бази даних історичних архівів.

3, 5, 9 - Візуальні дані ділянок – містять об'єкти, що пов'язують технологічну інформацію з зображеннями, скріншотами екранів операторів ділянок дифузії, дефекосатурації та випарної станції.

10 – База технологічних понять та властивостей, що пов'язують контури регулювання, насичують їх контекстом, зв'язують їх з іншими об'єктами та надають можливість посилатись і зв'язувати технологічні дані з різнотипними джерелами даних

11 – Зовнішні джерела, поняття про файлову систему, властивості, зовнішні джерела. Дана зона містить об'єкти, файли, поняття про файлову систему та зовнішні джерела, посилання та їх зв'язки

12 – Коннектори та поняття. Містять об'єкти та поняття, які надають можливість зв'язувати інформацію з зовнішніми джерелами, збагачувати зовнішні джерела властивостями для зв'язку з іншими зонами.

13 – Поняття та властивості візуальних даних. Містять об'єкти та сутності, що надають можливість зв'язувати зони технологічних даних з візуальними та зовнішніми джерелами

14 – Загальна інформація, описи ділянок, точка входу до бази знань.

Рис. 6 детальніше відображає зони семантичної моделі технологічних даних відділення випарювання. Зона 1 складається з контурів та об'єктів SCADA системи, інформації та властивостей окремих контурів, які об'єднані моделлю опису, яка об'єднує їх властивості та надає їм структури, окрім цього, кожен об'єкт окремо пов'язаний з моделлю даних відділення. Зона 2 відповідає за загальний зріз технологічних даних, в цю зону входить база даних SCADA системи, а також загальний експорт таблиці тегів. Зона 3 відповідає за візуальні поняття моделі, об'єднуючи експортовані таблиці з базами даних та окремими контурами регулювання.

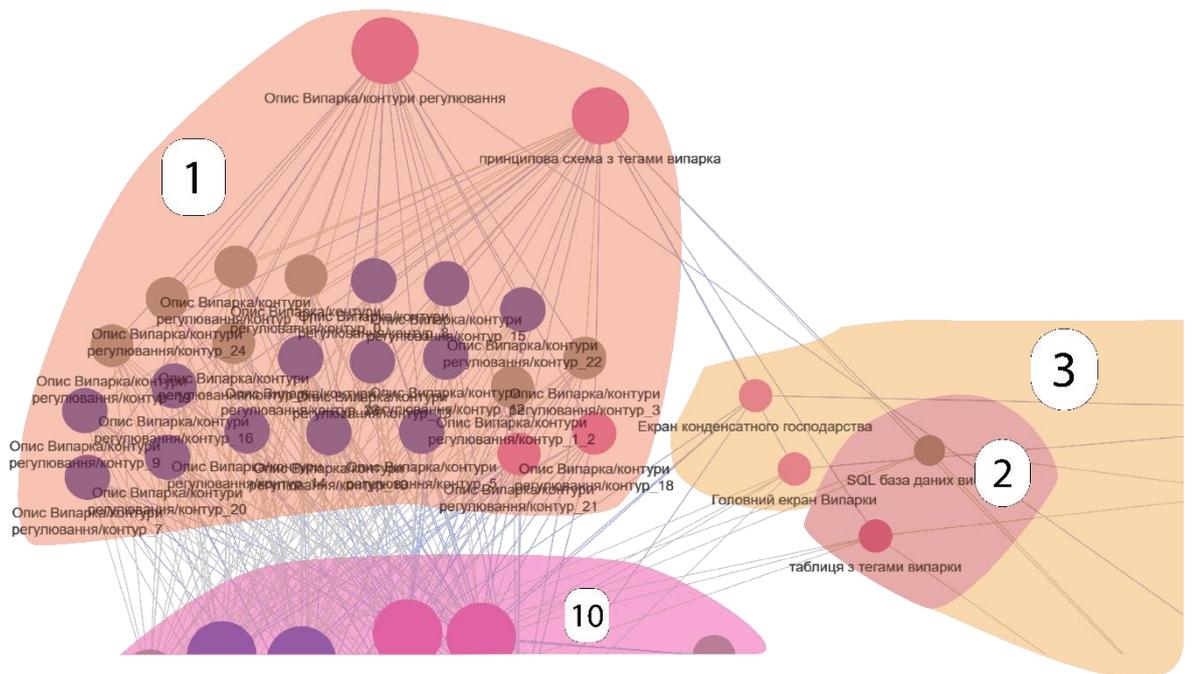


Рис. 6. Фрагмент семантичної моделі зв'язків, що відповідає технологічним даним відділення випарювання

Рис. 7 детальніше відображає зони семантичної моделі технологічних даних відділення дефекосатурації, а також зону понять 10. За аналогією з рис. 5, зона 4 складається з контурів та об'єктів SCADA системи, інформації та властивостей окремих контурів, які об'єднані моделлю опису, яка об'єднує їх властивості та надає їм структури. Зона 10 відображає властивості об'єктів, що надає їм контексту та

дозволяє проводити пошуки, генерувати посилання та пов'язувати об'єкти між собою, ця зона визначає та зберігає контекст даних, місцезнаходження інформації, зону її впливу, тип інформації, який контролюючий параметр використовується в регулюванні, яким середовищем даний контур керує, номер контуру та тип даних, що зберігається в об'єкті, що дозволяє більш детально визначати математичну модель впливу контуру системи автоматизації.

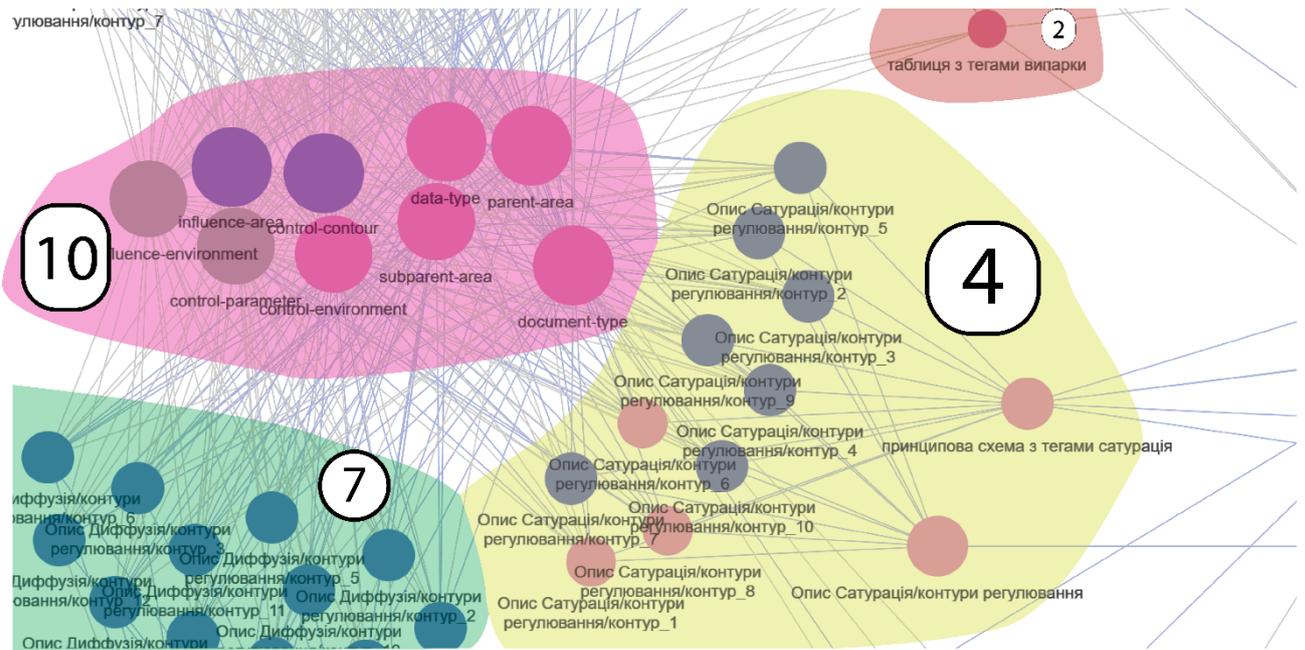


Рис. 7. Фрагмент семантичної моделі зв'язків технологічних даних відділення дефкосатурації

На рис. 8. зображена деталізовано семантична модель зв'язків технологічних даних відділення дифузії. За аналогією з рис. 5, зона 7 складається з контурів та об'єктів SCADA системи, інформації та властивостей окремих контурів, які об'єднані моделлю опису, яка об'єднує їх властивості та надає їм структури.

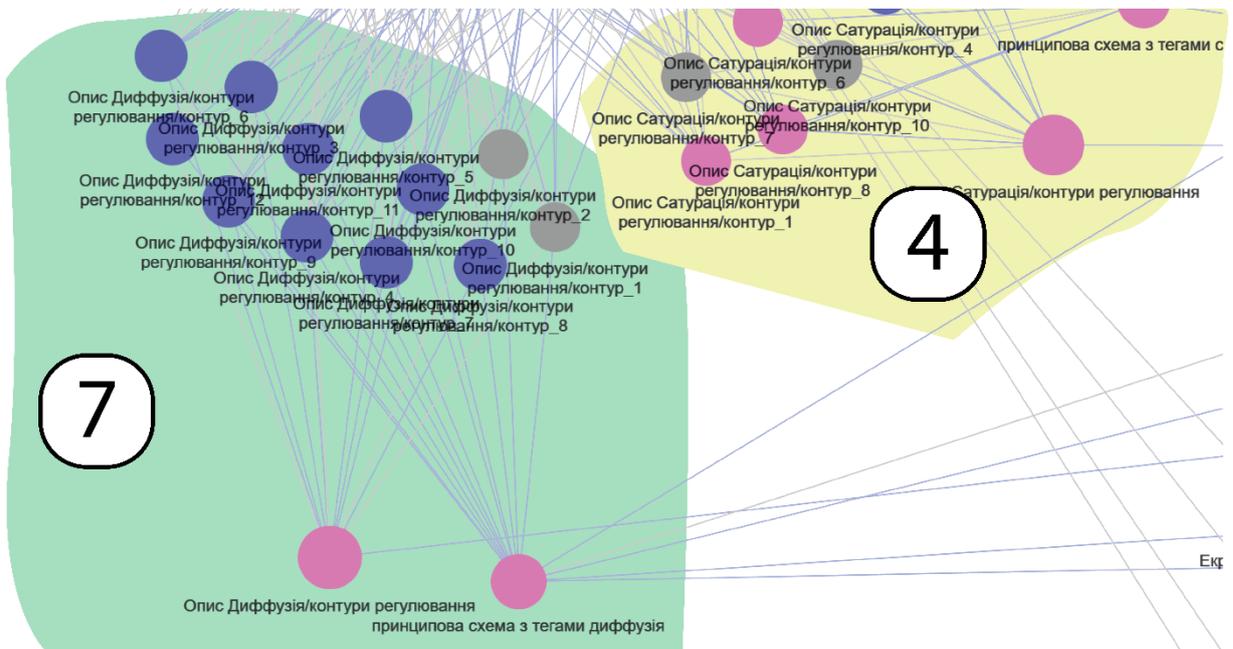


Рис 8. Фрагмент семантичної моделі зв'язків технологічних даних відділення дифузії

Рис. 9. відображає деталізовану частину семантичної моделі, що пов'язує гетерогенні дані. Зона 11 визначає файлову систему, властивості файлів, посилання, а також референсну модель, яка дозволяє іншим об'єктам створювати посилання на окремі частини документів, тобто локально розташовані зовнішні джерела. Зона 12 визначає об'єкти, що містять загальноприйняті поняття та властивості інформації, що міститься між зовнішніми джерелами та внутрішніми об'єктами. В даній зоні вказані властивості навчальних матеріалів, статей, веб-сторінки, способи впливу об'єктів з такою властивістю на зв'язки. Окрім того, на рисунку також відображені інші зони, окрім згаданих раніше.

Зони 6 та 8 визначають загальний зріз технологічних даних, в цю зону входять бази даних SCADA систем відділень дефекосатурації та дифузії, а також загальний експорт таблиць тегів. Зони 5 та 9 відповідають за візуальні поняття моделі, об'єднуючи експортовані таблиці з базами даних та окремими контурами регулювання відділень дифузії та дефекосатурації. Частково відображена зона 14 містить об'єкти верхнього рівня ієрархії, що представляють зріз та загальний опис моделі.

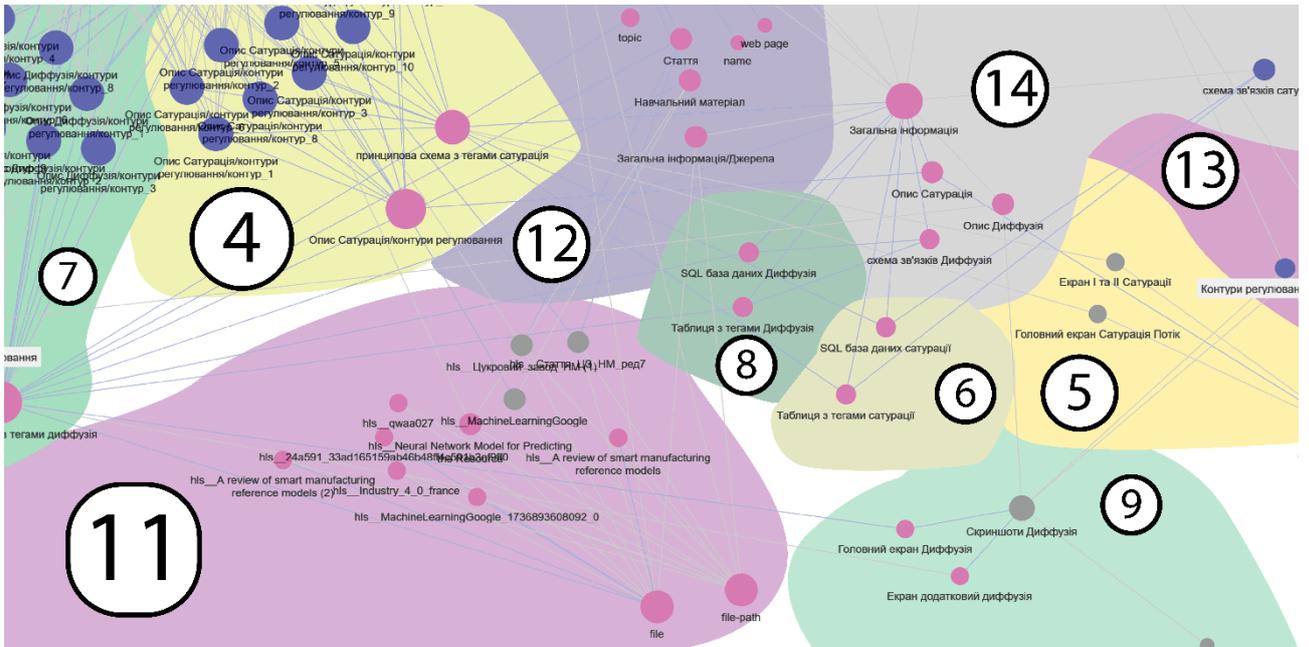


Рис 9. Фрагмент семантичної моделі зв'язків файлової бази даних (11) з поняттями (12)

Рис. 10. відображає деталізацію зв'язків Загальної інформації. На рисунку виділена основна зона 14, яка містить точку входу в базу знань, об'єкти з загальними властивостями та їх описи. Окрім того, також на рисунку відображені сусідні зони 12, 8, 3 та 5, згадані раніше

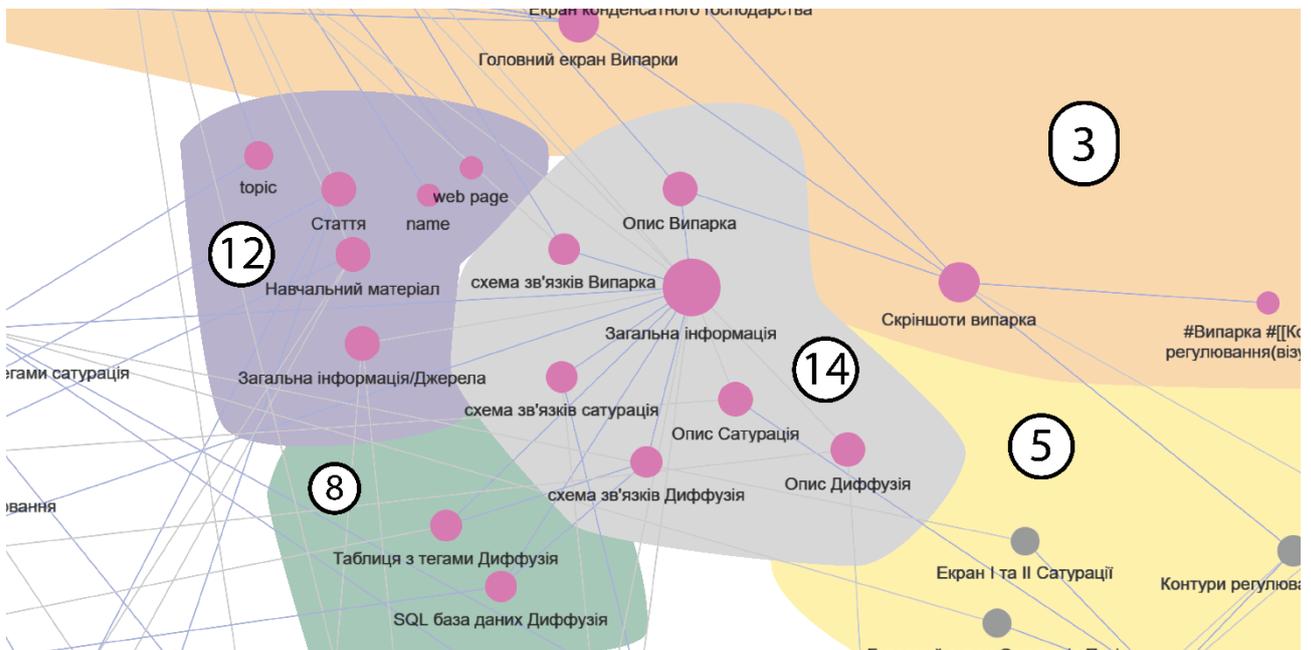


Рис 10. Фрагмент семантичної моделі зв'язків загальної інформації

Рис. 11 відображає більш детально Зону 3, яка відповідає за об'єкти з візуальною інформацією ділянки випарної станції.

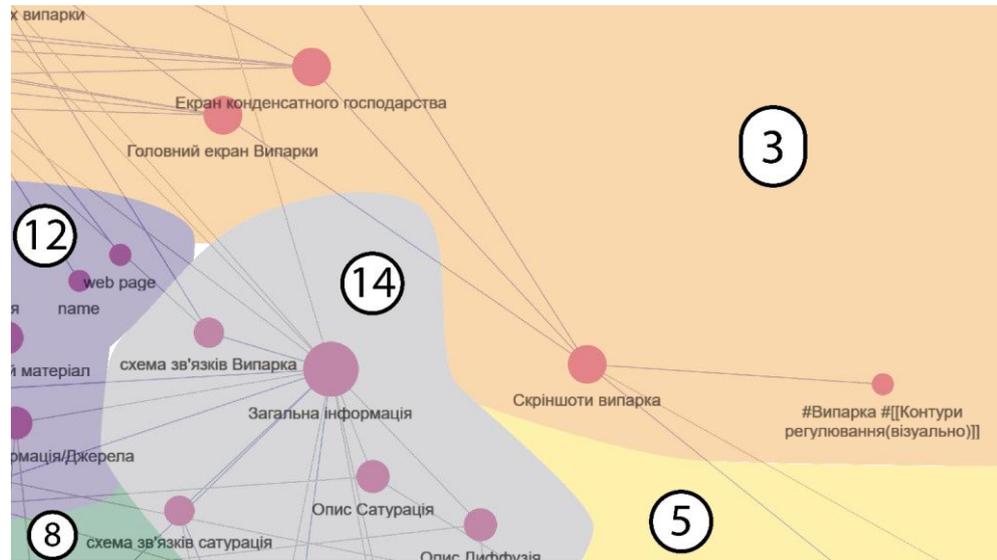


Рис. 11. Фрагмент семантичної моделі зв'язків для візуальної інформації ділянки випарної установки

Рис. 12 відображає деталізовану частину семантичної моделі зв'язків об'єктів візуальної інформації ділянки дифузії. Основною виділена зона 9, яка відповідає за об'єкти з візуальною інформацією ділянки дифузії, окрім того, на рисунку відображені зони:

5 – зона об'єктів з візуальною інформацією ділянки сатурації;

6 – зона об'єктів бази даних та експортованої таблиці тегів технологічної відділення дефекосатурації;

8 – зона об'єктів бази даних та експортованої таблиці тегів технологічної відділення дифузії;

11 – зона об'єктів файлової системи;

13 – зона загальних об'єктів, з властивостями про візуальну інформацію;

14 – зона об'єктів загальної інформації про систему, вхідна точка бази знань;

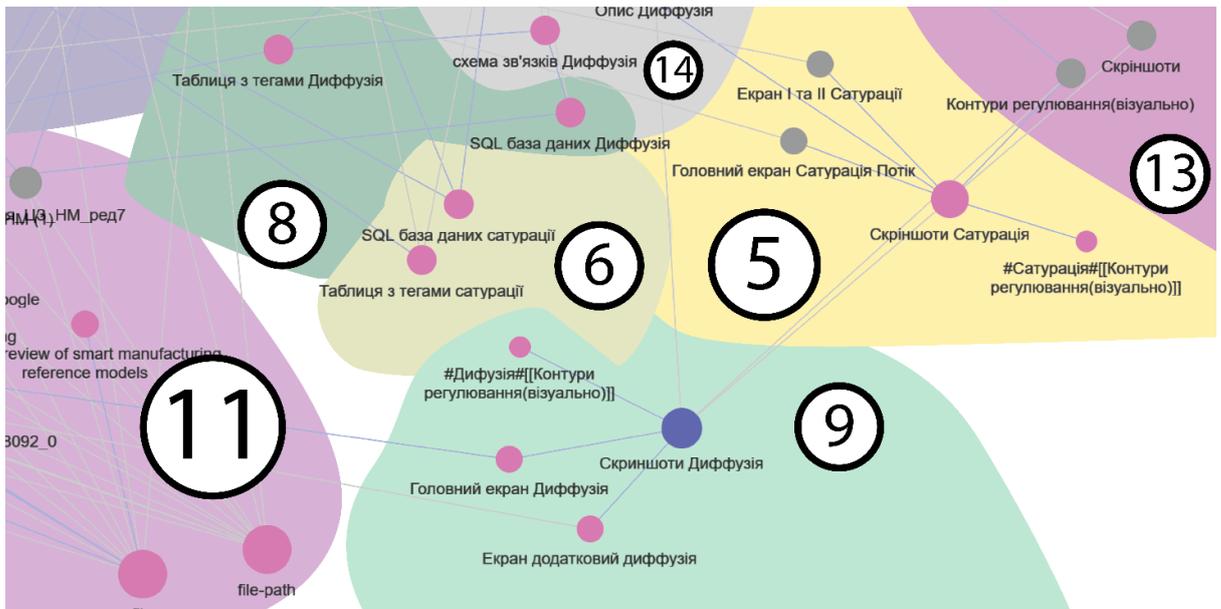


Рис 12. Фрагмент семантичної моделі зв'язків для візуальної інформації ділянки дифузії

Рис. 13 відображає деталізовану частину семантичної моделі зв'язків об'єктів, що описують візуальну інформацію ділянки дефекосатурації в зоні 5, а також об'єктів зони 13, що містять поняття і властивості зв'язків візуальних даних загалом.

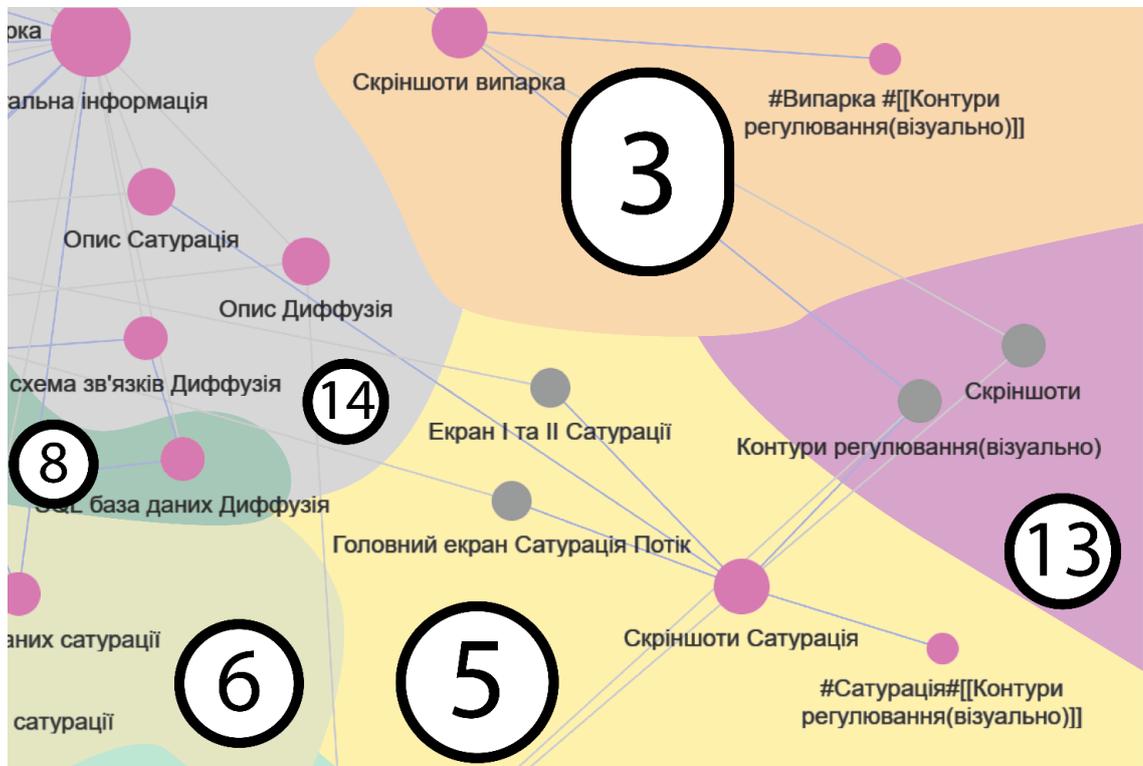


Рис 13. Фрагмент семантичної моделі зв'язків для візуальної інформації ділянки дефекосатурації а також понять про візуальну інформацію

Дослідивши зв'язки, генеровано аналітичний екран семантичної моделі, де кожен об'єкт відсортовано за тегами та посиланнями до відповідних ділянок, що зображено на рис. 14. На рисунку відображено кольорами об'єкти з тегами та прямими посиланнями, синім - до відділення дефекосатурації, зеленим – до відділення дифузії а червоним – до відділення випарної установки.

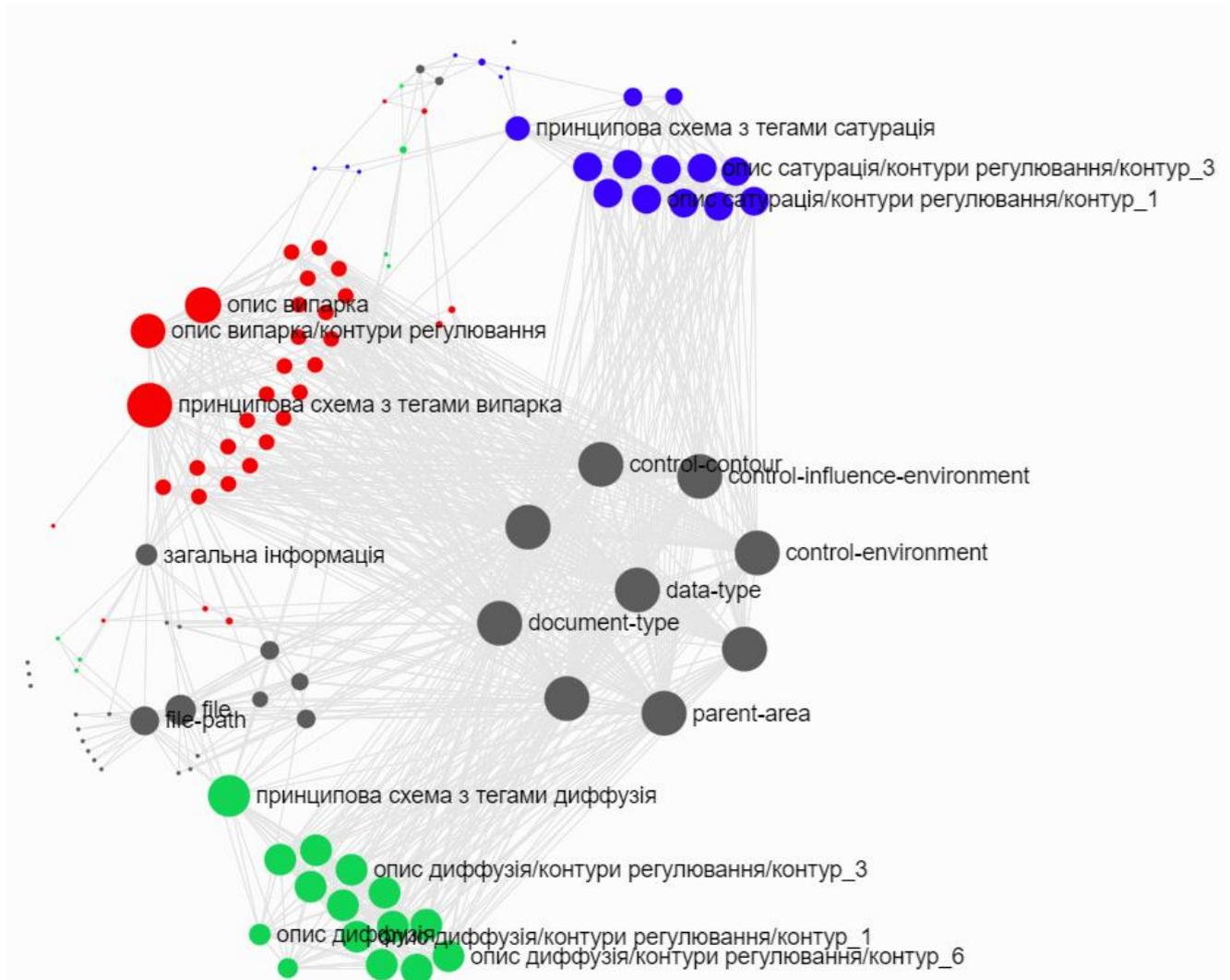


Рис 14. Семантична моделі зв'язків, утворених технологічною інформацією, розподілених за тегами та посиланнями до ділянок

Взявши дані дільниці випарювання, було створено ієрархію технологічних даних, як показано на рис 15.

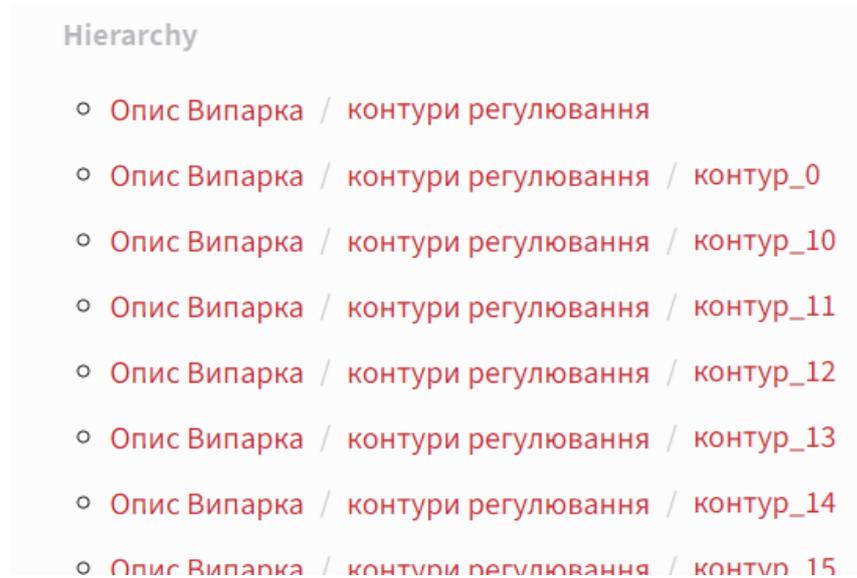


Рис 15. Ієрархія технологічних даних ділянки випарювання

Цю ієрархію об'єктів насичили контекстними властивостями, що притаманні системам автоматизації. Це дозволяє зв'язувати, створювати запити та обробляти дані. Таким чином, дані, що не мають строгої типізації можна піддавати паттернам та створювати структуру, що надає гнучкий підхід, залишаючись незалежною від типу даних, але з жорсткою структурою, що ефективно обробляється.

На рис. 16 наведено приклад простого запиту усіх контурів регулювання випарної установки після насичення їх контекстними властивостями за шаблоном.

parent-area	subparent-area	influence-area	control-con
Випарна станція	Випарна станція	ЗАК-4К	7
Випарна станція	Випарна станція	ЗБ-4К	8
Випарна станція	Випарна станція	4К-Конц	9
Випарна станція	Випарна станція	Уся випарна станція	0
Випарна станція	Випарна станція	2БК-ЗБК	6
Випарна станція	Випарна станція	Концентратор	18
Випарна станція	Випарна станція	1К-2А	3
Випарна станція	Випарна станція	Збірник сиропу	24
Випарна станція	Випарна станція	1 корпус	1_2
Випарна станція	Випарна станція	2БК-ЗАК	5
Випарна станція	Випарна станція	Конц-Сбірник сиропу	10
Випарна станція	Випарна станція	1 корпус	23
Конденсатне господарство	Конденсатне господарство	Збірник конденсату 4К	16
Конденсатне господарство	Конденсатне господарство	Збірник-випарювач конденсату 1К	12
Конденсатне господарство	Конденсатне господарство	Збірник конденсату 2К	13
Конденсатне господарство	Конденсатне господарство	Збірник конденсату 1К	11

Name	Description
FIX_LM110_F_CV	УВ на тиристорний привід (Ф_СВ)
FIX_LM120_F_CV	УВ на частотний привід (Ф_СВ)
FIX_LT102_F_CV	Рівень в 1К дублюючий (Ф_СВ)
FIX_LT110_F_CV	Рівень в 1К основний (Ф_СВ)
FIX_LZ110_F_CV	Рівень в 1К заданий (Ф_СВ)
FIX_LT101_F_CV	Рівень у збірнику соку (Ф_СВ)
FIX_CB_ZAD_F_CV	Зміст СР у сиропі задане (Ф_СВ)

Рис 16. Приклад запити усіх контурів регулювання випарної установки з відображенням властивостей об'єктів

Надалі, дані технологічного процесу об'єднали з іншими джерелами, отримавши модель, яка об'єднує файли з бази даних SCADA, у вигляді таблиць формату CSV, окремі вузли та контури регулювання, побудовані всередині середовища, базу даних історичних архівів, та скріншоти екранів оператора SCADA. Побудувавши між ними зв'язки зі своїми власними властивостями, представлено приклад єдиного середовища, що об'єднує різнотипові об'єкти. Завдяки такому підходу, можна об'єднувати технологічні дані, обробляти як дані систем реального часу, так і технологічну документацію з посиланнями до об'єктів, що мають власні властивості та зв'язки. Це дозволяє проводити швидкий аналіз, пошук та синтезувати моделі впливу систем один на одну.

Насичені дані візуалізовані та зв'язані між собою, утворюючи візуальні моделі даних. На рис. 17 зображено частину візуальної модель, яка відображає об'єкти

системи автоматизації ділянки випарювання: її початку, контури з тегами, експортованими з АРМ оператора, а також їх поєднання з документацією про моделювання нейронних мереж, утворюючи зв'язок, який об'єднує інтернет-джерело, документ про залежність вмісту сухих речовин від температури, об'єкт візуальної моделі та теги, на які дана залежність може впливати. Таким чином, при побудові запиту та аналізу інформації складається повна картина з урахуванням усіх змінних про вплив на технологічні параметри

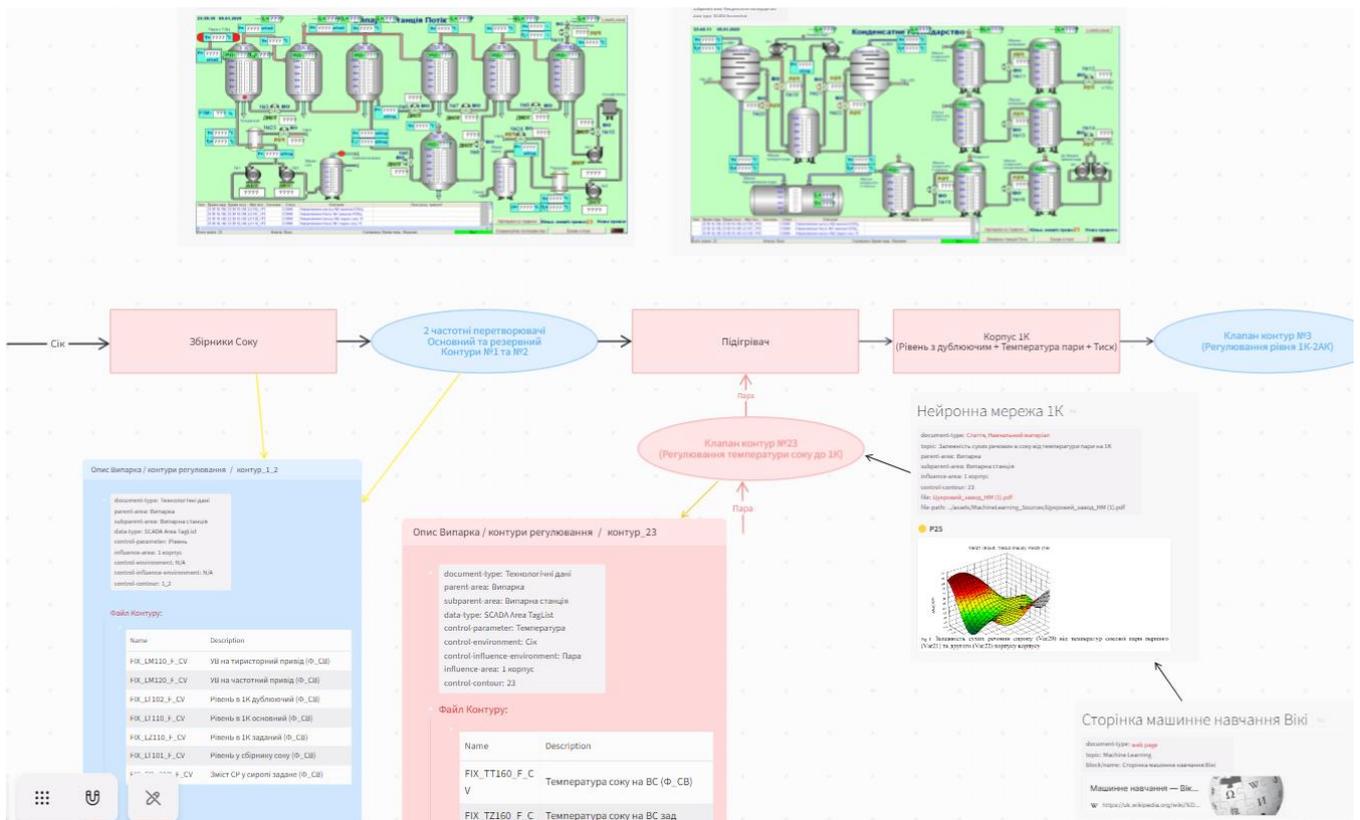


Рис 17. Візуальна моделі знань відділення випарки

У систему зв'язків семантичної моделі було впроваджено виділення через показник Adamic-Adar, що дозволяє визначати модель впливу та схожості об'єктів. Це дає можливість не лише аналізувати безпосередньо задані зв'язки між елементами, а й враховувати потенційні збурення та непрямі впливи, які не завжди явно позначені в структурі системи. Крім того, адаптація цього підходу в математичній моделі дає змогу шукати схожі аномальні зв'язки в інформації, що особливо важливо для аналізу даних SCADA-систем. Семантична модель схожості відображена на рис. 18. Такий підхід сприяє ефективному виявленню схожих

об'єктів на основі спільних посилань, визначенню зон впливу та аналізу їхньої чутливості до змін у знаннях про систему автоматизації, що розширює можливості побудови математичних моделей і робить їх адаптивнішими до складних взаємозалежностей:

$$S_{AA}(x, y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log(|N(u)|)}$$

де x, y – вузли, що порівнюються; $N(x), N(y)$ – множини сусідів вузлів x та y ; $N(u)$ – множини спільних сусідів, які одночасно зв'язані з x та y .

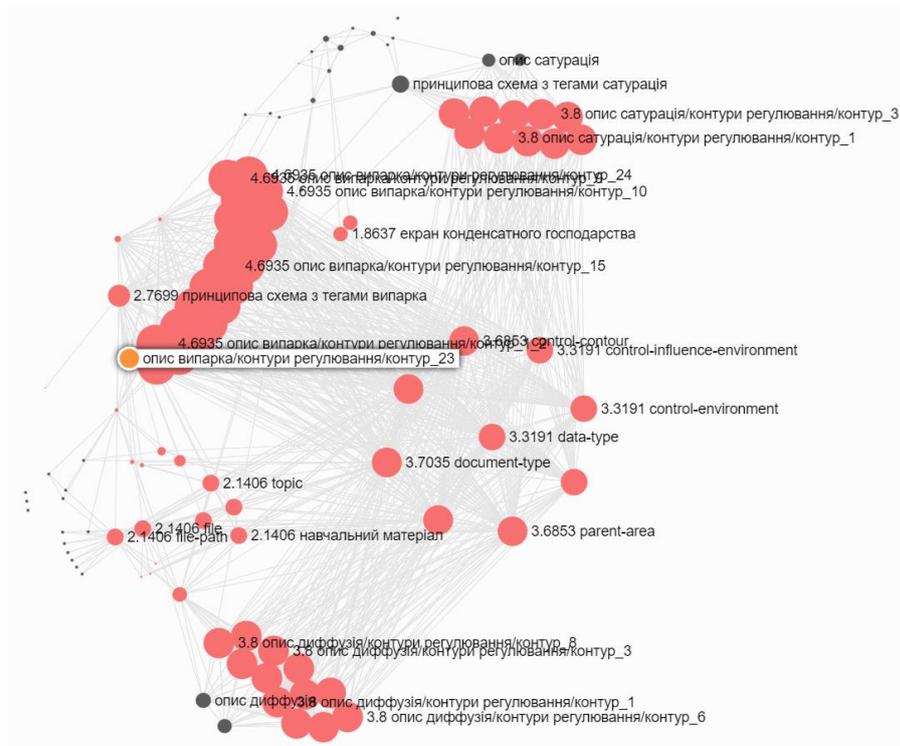


Рис 18. Семантична модель схожості та впливу за алгоритмом Adamic-Adar

Результати досліджень та їх обговорення. Дослідивши цей підхід, було застосовано принцип Semantic WEB для властивостей, що надає можливість поєднувати також інформацію з інтернет баз знань. На рис. 19 зліва відображено два запити – перший запит відображає усі елементи, що мають властивість «Навчальний матеріал», в той час як другий запит відображає усі

посилання на певний PDF-файл з місцем посилання, в правій частині екрану результат пошуку джерела «Machine Learning» в онлайн-базі знань Wikipedia.

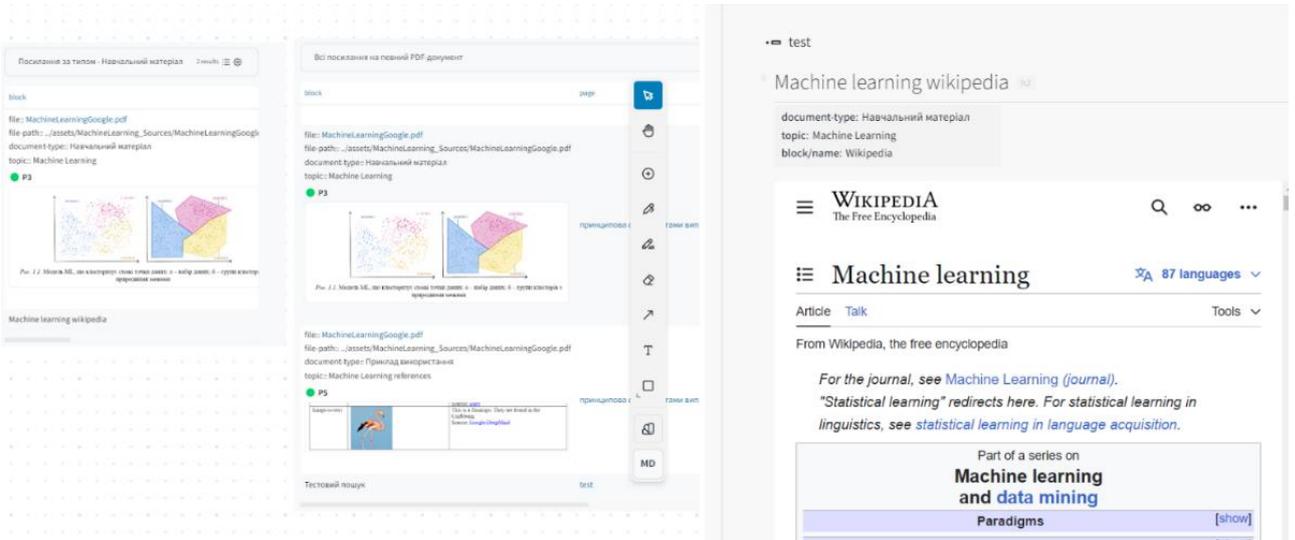


Рис 19. Властивості та метазнання гетерогенних даних

На рис. 20 зображено запит та його результат з пошуку контурів регулювання 1 корпусу випарної станції, та усієї інформації і документації, яка має посилання на 1 корпус за означеним запитом з використанням синтаксису Clojure.

```
clojure
1 #+BEGIN_QUERY
2 {:title "Документи та згадки про 1 корпус за властивістю впливу"
3  :query [:find (pull ?b [*])
4         :where
5         [?b :block/properties ?props]
6         [(get ?props :influence-area) ?area]
7         [(= ?area "1 корпус")]
8         [(get ?props :subparent-area) ?subparentarea]
9         [(= ?subparentarea "Випарна станція")]
10        [?b :block/page ?p]
11        [?p :block/name ?page-name]
12        [(not= ?page-name "принципова схема з тегами випарка")]]}
13 #+END_QUERY
```

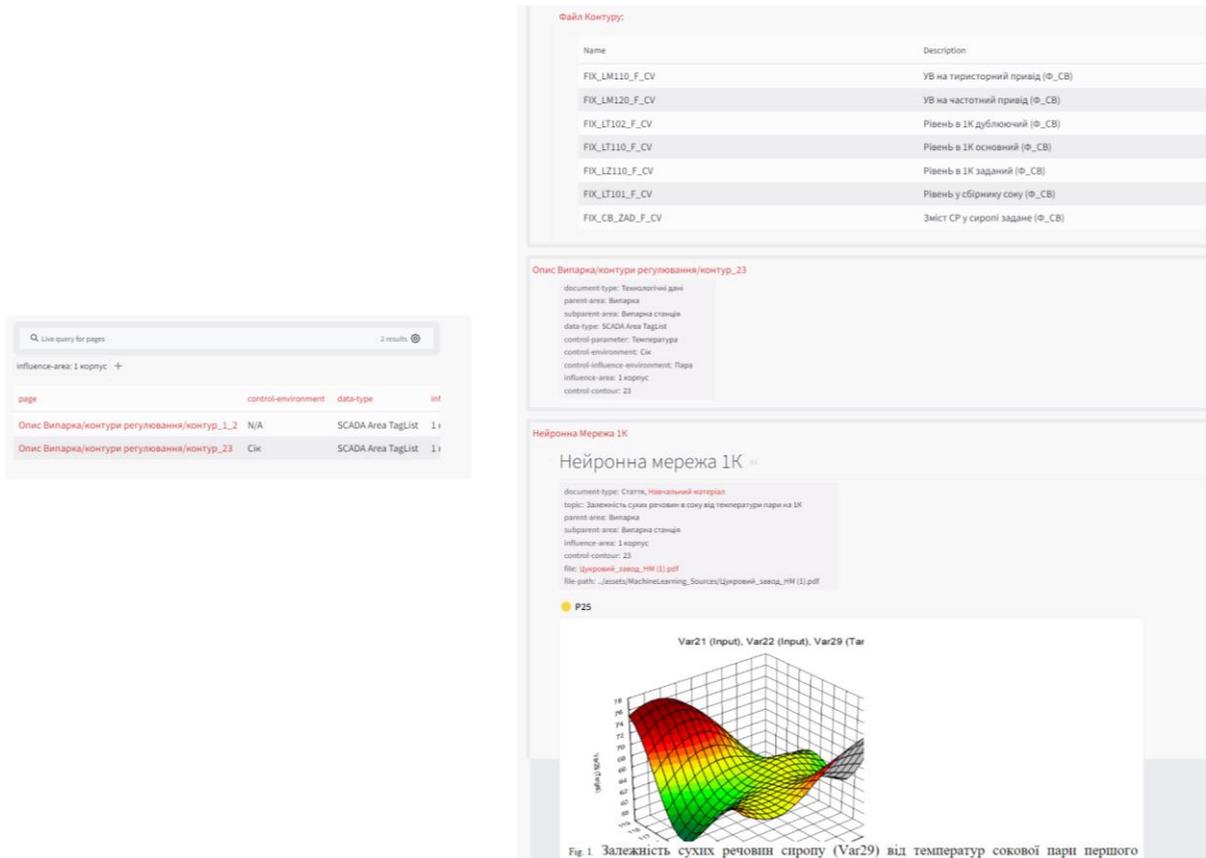


Рис 20. Приклад запити по пошуку контурів регулювання (справа) та усієї інформації і посилань, що впливають на 1 корпус випарної станції

Також чітко видно, що в результати попереднього запити потрапили як таблиці CSV, тексти, PDF-документи, так і що важливо, зовнішні-інтернет джерела знань. На рис. 21 зображено продовження результату запити, де видно окрім посилання на PDF - документ також посилання і на сторінку вебсайту, який вже раніше в моделі був прив'язаний до 1 корпусу випарної установки.



Рис 21. Результат запиту з пошуку гетерогенних даних

Така побудова надає переваги:

1. Швидкий аналіз впливу об'єктів, сторінок та блоків на технологічну ділянку та на один одного. Важливо враховувати взаємодію між окремими об'єктами, які можуть впливати на загальну ефективність технологічного процесу. Аналіз впливу допомагає оптимізувати процеси та зменшити ризики помилок або збоїв.
2. Враховувати як внутрішню, так і зовнішню інформацію на технологічному обладнанні при побудові математичних моделей процесу.
3. Швидкий доступ до динамічних навчальних та дослідницьких матеріалів, незалежно від їх зміни. Для успішного адаптування до нових умов важливо мати можливість оперативно змінювати моделі та алгоритми в реальному часі, використовуючи актуальні дослідження та дані.

4. Опис структури цифрового двійника, яка повинна чітко визначати, які дані в нього надходять, як вони обробляються та зберігаються, а також, як ці дані взаємодіють з іншими системами в межах технологічної ділянки.

5. Можливість поєднання цифрових двійників об'єктів в єдиний багатошаровий цифровий двійник, використовуючи не лише ієрархічні зв'язки для забезпечення комплексної моделі, що відображає всі аспекти технологічного процесу. Це дозволить отримати більш точну картину роботи системи та приймати ефективніші рішення на основі даних.

6. Інтеграція системи прогнозування та адаптації, що дозволяє коригувати моделі в залежності від умов для збереження стабільності і ефективності.

7. Гнучкість та масштабування з іншими технологіями та з перспективними розробками. Технології та процеси можуть змінюватися в майбутньому, тому система повинна бути гнучкою і мати можливість масштабування для підтримки нових об'єктів та технологій, що дозволяє адаптуватися до розвитку виробництва.

8. Керування версіями концептуальної моделі. Даний підхід дає змогу мати резервні копії моделі, відслідковувати зміни її компонентів та за необхідності повернення до попередніх версій.

9. Генерувати веб-сторінки з вмістом концептуальної моделі та керувати доступом до окремих її елементів

Висновки і перспективи. Розроблена семантична модель трьох відділень цукрового заводу описує об'єкти, знання, зв'язки між ними та їхній вплив на технологічний процес. Вона дозволяє швидко аналізувати взаємодію елементів, враховувати внутрішню й зовнішню інформацію для побудови математичних моделей, а також забезпечує доступ до актуальних навчальних і дослідницьких матеріалів. Важливо, що структура моделі, а в подальшому і цифрових двійників повинна чітко визначати, які дані в нього надходять, як вони обробляються та зберігаються, а також їх взаємодію та вплив з іншими системами. Передбачена адаптивність системи, її масштабування, керування версіями та створення інтерфейсів для доступу й контролю. Семантична мережа містить в собі об'єкти, блоки, концепти та зв'язки між ними, кожен об'єкт може мати властивості, які при

цьому можуть виступати як окремі джерела, або об'єкти інших систем автоматизації. У роботі побудовані запити, які дозволяють визначати зони впливу зовнішніх джерел, блоків та інформації на конкретні технологічні ділянки виробництва.

У перспективі планується розробка математичних моделей технологічних ділянок, які стануть основою цифрових двійників основних відділень цукрового виробництва, а також цифрового двійника виробничої лінії. Система обробки інформації цілком відповідає сучасним системам знань, заснованим на онтологічних базах, і надає можливості для майбутнього синтезу математичної моделі.

References

1. Gao, T., Chai, Y. and Liu, Y. (2018). A review of knowledge management about theoretical conception and designing approaches. *International Journal of Crowd Science*, 2 (1), 42–51. Available at: <https://doi.org/10.1108/ijcs-08-2017-0023>
2. Chavez, V., Wollert, J. (2024). Development of an Industry 4.0 Ontology to Enable Semantic Interoperability at the Field Level, *Sensors and Transducers*, 265 (2), 139–147. Available at: <https://www.researchgate.net/publication/381120487>
3. Singh, S., Shehab, E., Higgins, N., Fowler, K., Reynolds, D., Erkoyuncu, J.A., Gadd, P. (2020). Data management for developing digital twin ontology model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 235 (14), 2323–2337. Available at: <https://doi.org/10.1177/0954405420978117>
4. Kafeza, E., Drakpopoulos, G., Mylonas, P. (2024). Graph Neural Networks in PyTorch for Link Prediction in Industry 4.0 Process Graphs. *IFIP Advances in Information and Communication Technology*, 220–234. Available at: https://doi.org/10.1007/978-3-031-63219-8_17
5. Cimino, A., Longo, F., Mirabelli, G., Solina, V., Verteramo, S. (2024). An ontology-based, general-purpose and Industry 4.0-ready architecture for supporting the smart operator (Part II – Virtual Reality case). *Journal of Manufacturing Systems*, 73, 52–64. Available at: <https://doi.org/10.1016/j.jmsy.2024.01.001>
6. D'Amico, R.D., Sarkar, A., Karray, M.H., Addepalli, S., Erkoyuncu, J.A. (2024). Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins // *CIRP Journal of Manufacturing Science and Technology*, 52, 366–385. Available at: <https://doi.org/10.1016/j.cirpj.2024.06.007>
7. Zheng, X., Petrali, P., Lu, J., Turrin, C., Kiritsis, D. (2022). RMPFQ: A Quality-Oriented Knowledge Modelling Method for Manufacturing Systems Towards Cognitive Digital Twins // *Frontiers in Manufacturing Technology*, 2. Available at: <https://doi.org/10.3389/fmtec.2022.901364>
8. Rongen, S., Nikolova, N., Mark (2023). Modelling with AAS and RDF in Industry 4.0. *Computers in Industry*, 148, 103910: <https://doi.org/10.1016/j.compind.2023.103910>

9. Valdez-Valenzuela, A., Gómez-Adorno, H. (2024). text2graphAPI: A library to transform text documents into different graph representations. *SoftwareX*, 101888. Available at: <https://doi.org/10.1016/j.softx.2024.101888>
10. Blankenberg, C., Gebel-Sauer, B., Schubert, P. (2022). Using a graph database for the ontology-based information integration of business objects from heterogenous Business Information Systems. *Procedia Computer Science*, 314–323. Available at: <https://doi.org/10.1016/j.procs.2021.12.019>
11. Sun, Q., Luo, Y., Zhang, W., Li, S., Li, J., Niu, K., Kong, X., Liu, W. (2024). Docs2KG: Unified Knowledge Graph Construction from Heterogeneous Documents Assisted by Large Language Models. *arXiv*. Available at: <https://arxiv.org/abs/2406.02962>
12. Abonyi, J., Nagy, L., Tamás (2024). Ontology-Based Development of Industry 4.0 and 5.0 Solutions for Smart Manufacturing and Production. Available at: https://books.google.com.ua/books?hl=ru&lr=&id=bfrEAAAQBAJ&oi=fnd&pg=PR5&dq=ontology+knowledge+management+in+SCADA&ots=UDeFsHuOX&sig=b_9W_QTb-NfAHVicu-CQkJ6aeIE&redir_esc=y#v=onepage&q=ontology%20knowledge%20management%20in%20SCADA&f=false
13. Gebreab, S., Musamih, A., Salah, K., Jayaraman, R., Boscovic, D. (2024). Accelerating Digital Twin Development with Generative AI: A Framework for 3D Modeling and Data Integration. *IEEE Access*, 1–1. Available at: <https://doi.org/10.1109/access.2024.3514175>

MODERN APPROACHES TO PROCESSING TECHNOLOGICAL INFORMATION IN INDUSTRIAL AUTOMATION SYSTEMS BASED ON KNOWLEDGE BASES

R. Shevchenko, N. Lutska

Abstract. *This article is dedicated to the analysis of existing approaches and the synthesis of novel methods for processing technological information in industrial automation systems in the context of modeling digital twins of various interconnected sections and systems. The aim of the research is to analyze the challenges of knowledge base modeling and improve the efficiency of information processing by considering heterogeneous data. The study focuses on the automation systems of three sugar production sections: diffusion, defecosaturation, and evaporation. As part of the research, a general associative map was developed, enabling the structured organization of industrial automation systems, establishing relationships, and enriching data with context for further processing. A semantic knowledge base model was designed to define data influence zones and generate a data schema that accounts for data heterogeneity. The developed semantic model allows for rapid contextual and criteria-based searches, analysis of subsystem interactions, and the calculation of similarity levels in data and anomalies, which is crucial for mathematical modeling and the synthesis of multi-layered or cognitive digital twins of processes. Based on the semantic model, a production section visualization model was created, integrating technological documentation, SCADA system data, industrial data archives, external sources, and knowledge bases via the web using meta-knowledge. Using the Clojure programming language, search queries were*

generated to explore influence zones, context and mentions, as well as to calculate knowledge similarity. This approach enables the integration of multiple system models without significant additional costs, ultimately defining the structure of the digital twin while accounting for potential changes in knowledge systems. This ensures a systematic and efficient approach to data collection, processing, storage, and integration with other technological systems. An additional advantage is the ability to combine individual, already described digital twins into a multi-layered comprehensive model, providing a detailed representation of the technological process and enhancing decision-making efficiency.

Key words: *ontology, digital twin, knowledge base, data, intelligent knowledge management systems, automation*