

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
“ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”

Кукленко Дмитро Вікторович

УДК 004.65:004.825

**МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
ІНТЕЛЕКТУАЛЬНОГО ІНЖИНІРИНГУ ДАНИХ
В АВТОМАТИЗОВАНИХ СИСТЕМАХ УПРАВЛІННЯ ТЕХНОЛОГІЧНИМИ ПРОЦЕСАМИ**

Спеціальність 05.13.06 – автоматизовані системи управління та прогресивні
інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків - 2005

Дисертацією є рукопис.

Робота виконана в Національному технічному університеті “Харківський політехнічний інститут” Міністерства освіти і науки України.

- Науковий керівник: кандидат технічних наук, доцент
Ткачук Микола В'ячеславович,
Національний технічний університет
“Харківський політехнічний інститут”,
доцент кафедри автоматизованих систем управління.
- Офіційні опоненти: доктор технічних наук, професор
Федорович Олег Євгенович,
Національний аерокосмічний університет ім. М.Є. Жуковського
”ХАІ”,
завідувач кафедри інформаційно-управляючих систем,

кандидат технічних наук
Євланов Максим Вікторович,
Харківський національний університет радіоелектроніки, доцент
кафедри інформаційних управляючих систем.
- Провідна установа: Одеський національний політехнічний університет, Міністерство
освіти і науки України,
кафедра системного програмного забезпечення,
м. Одеса.

Захист відбудеться 16 червня 2005 р. о 14³⁰ на засіданні спеціалізованої вченої ради Д 64.050.07 у Національному технічному університеті “Харківський політехнічний інститут”, за адресою: 61002, Харків – 2, вул. Фрунзе, 21.

З дисертацією можна ознайомитись у бібліотеці Національного технічного університету “Харківський політехнічний інститут”.

Автореферат розісланий 14 травня 2005 р.

Вчений секретар
спеціалізованої вченої ради

Голоскоков О.Є.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Однією з галузей сучасних комп'ютерних технологій, що розвиваються найбільш швидко, є технологія створення автоматизованих систем управління технологічними процесами (АСУ ТП). Значні обсяги даних, що обробляються в таких системах, висока інтенсивність інформаційного обміну, що обумовлена швидкістю протікання ТП, разом з розподіленим характером цих систем приводять до необхідності розробки інтелектуальних моделей і процедур для обробки інформації в таких системах.

Поточний етап розвитку АСУ ТП характеризується впровадженням в процес управління ТП обчислювальної техніки. Сучасні інформаційні системи використовуються для вирішення таких задач, як збір даних про хід ТП та відображення його стану. Одним з підходів до розробки програмного забезпечення АСУ ТП є використання систем контрольованого управління доступом та збору даних (Supervisory Control Access and Data Acquisition, або SCADA). Застосування SCADA-систем дозволяє здійснювати автоматизовану розробку АСУ ТП, забезпечує можливість контролю стану ТП у режимі реального часу із використанням зручних для операторів ТП засобів відображення інформації. Однак на сьогоднішній час моделі та інформаційні технології, що забезпечують інтелектуальну обробку даних в складних багаторівневих АСУ ТП, не були достатньо розвинуті. Існуючі комерційні системи, що відносяться здебільшого до класу SCADA-систем, вирішують лише такі традиційні задачі обробки даних, як збір даних, їх накопичування та візуалізація. Все це говорить про актуальність наукових досліджень щодо підвищення ефективності обробки даних в АСУ ТП, які виконувались у дисертаційній роботі.

Зв'язок роботи з науковими планами, програмами, темами. Дисертаційна робота виконувалася в межах держбюджетної теми № 0103U001543 „Розробка інформаційних моделей для реалізації процедур структурного синтезу в комп'ютерно-інтегрованих системах”, (наказ МОН України № 633 від 05.11.2002 р.), а також згідно госпрозрахункового договору №80133 між кафедрою автоматизованих систем управління НТУ “ХПІ” та КБ „Промавтоматика” (м. Харків). Здобувач брав участь у виконанні зазначених робіт як виконавець.

Мета і задачі дослідження. Мета дослідження полягає в підвищенні ефективності обробки даних в АСУ ТП за рахунок застосування сучасних технологій баз даних (БД) та моделей і процедур інтелектуальної обробки даних. Для досягнення цієї мети в роботі вирішуються такі задачі:

- аналіз сучасних тенденцій і підходів до розробки систем обробки даних для АСУ ТП;
- визначення основних задач, що мають бути вирішені для забезпечення ефективної обробки даних в АСУ ТП;
- розробка математичних та інформаційних моделей, що забезпечують вирішення основних задач ефективної обробки даних в АСУ ТП;

- реалізація інформаційних технологій, що реалізують розроблені математичні та інформаційні моделі;
- перевірка ефективності та продуктивності розроблених моделей та інформаційних технологій.

Об'єкт дослідження - процес розробки систем обробки даних, що використовуються в АСУ ТП.

Предмет дослідження - методи інтелектуальної обробки інформації, що накопичена в БД АСУ ТП, та моделі й інформаційні технології, що дозволяють реалізувати інтелектуальну обробку даних в АСУ ТП.

Методи дослідження. Для досягнення поставленої мети використовувалися: методи об'єктно-орієнтованого аналізу і проектування інформаційно-управляючих систем - для розробки архітектури інтелектуального інжинірингу даних (ІД) та визначення його процесів і компонент; методи теорії реляційних і пост-реляційних БД - для розробки моделей компонент, що реалізують процеси ІД; методи штучного інтелекту - для виявлення закономірностей про хід ТП у накопичених даних.

Наукова новизна отриманих результатів. Наукова новизна роботи полягає в постановці та вирішенні актуальної проблеми забезпечення ефективною обробки даних в АСУ ТП. У результаті вирішення цієї проблеми:

1) вперше дано визначення ІД в АСУ ТП як модельно-технологічного комплексу (МТК), впровадження якого спрямовано на підвищення ефективності обробки даних;

2) удосконалено модель активного правила в системах БД за рахунок інтерпретації умови активного правила як специфікації особливої ситуації, яка виникає в ТП, що дозволило використовувати існуючу концепцію активних БД для специфікації особливих ситуацій в ТП та визначення реакції АСУ ТП в разі виникнення таких ситуацій;

3) вперше запропоновано метод адаптації системи активних правил, які описують структуру проблемних ситуацій в ТП. Адаптація активних правил здійснюється на основі результатів аналізу ретроспективної інформації про хід ТП та дозволяє зменшувати складність існуючих активних правил, запобігаючи порушенням обмежень реального часу під час їх обробки;

4) одержали подальший розвиток моделі та процедури синхронізації даних в багаторівневій інформаційно-управляючій системі, результатом чого стало спрощення процесу відображення реляційних даних до формату XML;

5) одержали подальший розвиток підходи до моделювання продуктивності програмних компонент, що дозволило оцінити умови застосування процесів і компонент ІД, які безпосередньо пов'язані з обмеженнями реального часу.

Практичне значення одержаних результатів. Розроблено моделі даних (МД) та відповідну інформаційну технологію інтегрованої бази даних вузлу (ІБДВ), яка в архітектурі ПД служить засобом зберігання та обробки даних в АСУ ТП. Запропоновано моделі, алгоритми та інформаційну технологію механізму активних правил (МАП), який дозволяє операторові АСУ ТП задавати специфікації особливих ситуацій в ТП та реакцію системи в разі виникнення таких ситуацій. Розроблено алгоритм декомпозиції реляційної схеми даних, використання якого у процесі синхронізації даних дозволяє відобразити дані, що зберігаються в реляційній БД, до набору документів у форматі XML.

Одержані в роботі результати використані при розробці ІУС АСУ ТП Ул'янівської установки комплексної переробки газу (УКПГ) Перещепінського нафтогазопромислу, а також Сахалінської УКПГ в Краснокутському районі Харківської області, що підтверджено відповідними актами впровадження.

Апробація результатів дисертації. Результати дисертаційної роботи були представлені на наступних наукових конференціях: 1-й та 3-й міжнародних конференціях “Information Systems Technology and its Applications (ISTA)” (м. Харків, 2001 р., м. Солт-Лейк-Сити, США, 2004 р.), IX-й та XI-й міжнародних наукових конференціях “Інформаційні технології: наука, техніка, технологія, освіта, здоров'я” (м. Харків, 2001 р. та 2003 р.), 3-й міжнародній науково-практичній конференції з програмування “УкрПРОГ’2002” (м. Київ, 2002 р.), “First EurAsian Conference on Advances in Information and Communication Technology (EurAsia)” (м. Шираз, Іран, 2002 р.), “International Association of Science and Technology for Development (IASTED) Conference of Intelligent Systems and Control” (м. Зальцбург, Австрія, 2003 р.), “Remote Engineering Workshop (REV)” (м. Філах, Австрія, 2004 р.) та на наукових семінарах кафедри автоматизованих систем управління НТУ “ХПІ”.

Особистий внесок здобувача. Здобувачеві належать: постановка задачі розробки МТК для ефективною обробки даних в багаторівневих розподілених АСУ ТП; визначення інтелектуального інжинірингу даних в АСУ ТП та архітектура процесів ПД; розширена модель активного правила для специфікації особливої ситуації в ТП та визначення процедури обробки даних у системі в разі виявлення такої ситуації; метод адаптації існуючого набору активних правил на основі результатів проблемно-орієнтованого аналізу даних із використанням операторів модифікації; процедура декомпозиції реляційної схеми даних із подальшим її перетворенням до набору XML-документів; метричний підхід до вивчення функціонування програмних компонент; інформаційна технологія, що реалізує моделі та алгоритми ПД в АСУ ТП .

Публікації. Основні результати дисертаційного дослідження опубліковані в 11 наукових публікаціях, серед яких 6 - у фахових виданнях, що відповідають переліку ВАК України, та 5 – у тезах доповідей міжнародних конференцій.

Структура та обсяг дисертації. Дисертація складається зі вступу, 4-х розділів, висновків та додатків. Повний обсяг дисертації складає 162 сторінки. Робота містить 59 ілюстрацій; 16 таблиць; 2 додатки на 15 сторінках; 143 найменування використаних джерел на 8 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовується актуальність проведеного дослідження, визначені об'єкт і предмет дослідження, сформульовані мета і задачі дисертаційної роботи, показані наукова новизна і практична цінність отриманих результатів, подано інформацію щодо впровадження результатів дисертаційного дослідження, їх апробації та наукових публікацій за темою дослідження.

У **першому розділі** описується актуальність дослідження проблем ефективної обробки даних у сучасних АСУ ТП, наведено аналіз основних тенденцій в дослідженнях щодо БД, систем та технологій обробки накопичених даних. Особливу увагу приділено проблемам обробки даних в АСУ ТП та системах SCADA.

У першому параграфі обґрунтовується актуальність дослідження. Показано важливість газотранспортної та газо-видобувної галузей для економіки України та основні фактори, що мають призвести до підвищення ефективності роботи цієї галузі. Одними з основних таких факторів є запровадження комп'ютеризованих систем диспетчерського контролю стану ТП та попередження виникнення в ньому особливих (тобто позаштатних та аварійних) ситуацій. Як можливий засіб вирішення останньої задачі пропонується підвищення якості інформації, що може бути надано комп'ютеризованою системою операторам ТП, та за допомогою якої стане можливим зменшити кількість випадків виникнення особливих ситуацій. Наведено концепцію SCADA-системи, яка дозволяє здійснювати оперативне управління ТП та збір даних щодо його ходу. Показано, що SCADA-система за своєю суттю є одним з рівнів в архітектурі сучасних АСУ ТП, головними функціями якого є виконання операцій, які пов'язані з обмеженнями реального часу, на відміну від рівню підтримки прийняття рішень щодо управління ТП.

Другий параграф розглядає основні тенденції сучасних наукових досліджень у галузі проектування БД та обробки інформації, що накопичується в них. Переглянуто основні пост-реляційні МД; зроблено висновок про те, що темпоральні (або часові) МД можуть стати основою для проектування БД АСУ ТП (це стосується у першу чергу проектування архіву значень параметрів ТП). Також підкреслюється значення XML- стандарту представлення даних та технологій обробки даних, що базуються на цьому стандарті, які можуть бути використані як засіб вирішення задачі синхронізації даних, що виникає в розподіленій АСУ ТП. Розглянуті основні задачі та моделі технології інтелектуальної обробки даних (Data Mining).

Третій параграф містить постановку задачі, яка полягає в необхідності розробки комплексу інформаційних та математичних моделей та інформаційних технологій, що реалізують ці моделі, для забезпечення ефективної обробки даних в багаторівневих розподілених АСУ ТП. При цьому МТК, що має бути розроблений, повинен враховувати наступні фактори: 1) структуру ТП, що контролюється; 2) топологію розподіленої АСУ ТП; 3) характер особливих ситуацій, що виникають в ТП, що контролюється.

В **другому розділі** розробляється МТК ІД в АСУ ТП та визначається архітектура його процесів і компонент.

Перший параграф містить методологічну базу ІД в АСУ ТП. Вона включає в себе наступні моделі та технології обробки даних: 1) темпоральні (часові) МД. Наводяться мотиви щодо використання лінійної часової моделі із дискретним часом; 2) розширену модель активного правила для специфікації особливих ситуацій в ТП та архітектуру активної системи управління базою даних (СУБД), що ґрунтується на цій розширеній моделі та дозволяє визначати особливі ситуації, що виникають під час ходу ТП; 3) моделі асоціативних правил, використання яких дозволить відображати закономірності, що можуть бути виявлені під час аналізу значних за обсягом масивів ретроспективних даних, що накопичуються в АСУ ТП; 4) стандарт XML та технології, що базуються на ньому, як засіб вирішення задачі синхронізації даних.

Другий параграф містить опис природи даних, які мають оброблятися в АСУ ТП. В цьому параграфі надаються визначення наступних понять, які в подальшому будуть використані під час побудови МД:

- пристрій - агрегат, що входить до складу ТП;
- параметр - кількісна характеристика пристрою або ТП в цілому; при цьому виділяються чисельні параметри, бульові параметри (або повідомлення) та уставки (граничні значення, що визначають режими функціонування пристроїв);
- протокол - визначає набір правил, згідно з якими здійснюється обмін даними між SCADA-системою та програмними контролерами;
- мнемосхема - графічне відображення фрагменту ТП в системі, яке включає до себе зображення пристроїв ТП та надає можливість бачити значення параметрів, що характеризують ці пристрої;
- особлива ситуація - це аварійна або позаштатна ситуація в ТП, або така ситуація, що вимагає особливого контролю з боку персоналу;
- вузол - це елемент розподіленої АСУ ТП, в якому відбувається обробка даних щодо ТП.

Також представлено концепцію ІБДВ, яка виступає в архітектурі АСУ ТП в якості компонента зберігання та обробки даних. ІБДВ включає наступні схеми даних: 1) конфігураційну схему, яка призначена для зберігання структури ТП у вузлі; 2) ретроспективну схему, яка призначена для

накопичування даних про хід ТП (архіву ТП); 3) схему протоколів, яка призначена для зберігання інформації щодо протоколів обміну даних, які використовуються в системі; 4) схему облікових записів користувачів системи.

Третій параграф містить визначення ПД в АСУ ТП та архітектуру його процесів і компонент. ПД в АСУ ТП визначено як МТК, який є орієнтованим на використання знань та який призначено для здійснення ефективної обробки даних в АСУ ТП. Архітектуру процесів та компонент ПД показано на рис.1.

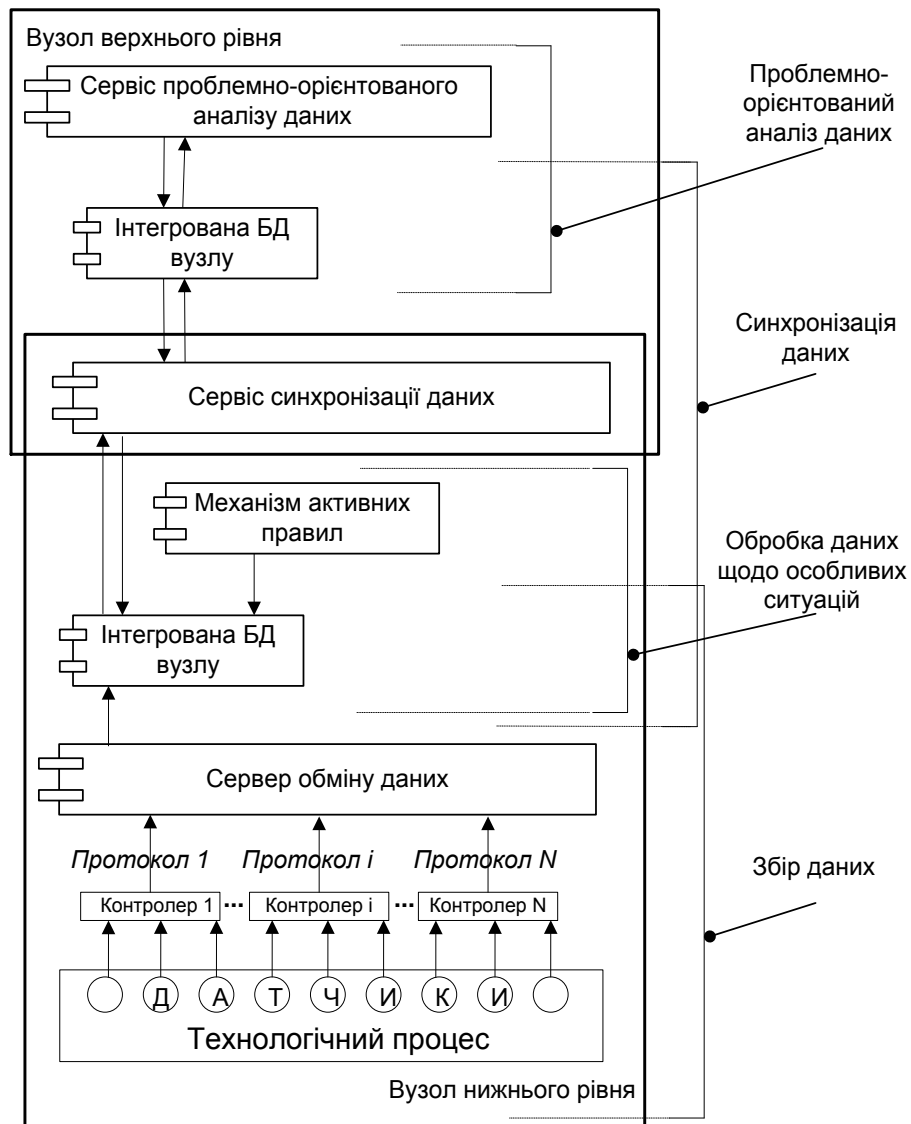


Рис. 1. Архітектура процесів та компонент ПД

ПД містить чотири процеси:

1) процес збору даних. Головною метою цього процесу є здійснення збору даних про хід ТП та забезпечення їх накопичення в системі. Сервер обміну даних (СОД) у відповідності до протоколу обміну даних надає значення параметрів ТП, які зберігаються в ІБДВ. Ці дані згодом опрацьовуються у відповідності до часових розбивок (секунда, хвилина, година, доба) за допомогою процедур агрегації значень параметрів;

2) процес обробки даних щодо особливих ситуацій. Метою цього процесу є виявлення особливих ситуацій в ТП. Ця задача вирішується за допомогою МАП;

3) процес синхронізації даних. Метою цього процесу є забезпечення обміну даними між ІБДВ різних вузлів. Компонент, що забезпечує вирішення цієї задачі, - сервіс синхронізації даних (ССД);

4) процес проблемно-орієнтованого аналізу даних (ПОАД). В межах цього процесу вирішуються задачі знаходження залежностей між параметрами ТП у вигляді асоціативних правил та спрощення структури існуючих активних правил на основі таких залежностей. Вирішення цих задач забезпечує сервіс проблемно-орієнтованого аналізу даних.

В четвертому параграфі запропоновано метод адаптації існуючих активних правил на базі результатів процесу проблемно-орієнтованого аналізу даних. Загальну схему, яка ілюструє застосування цього методу, показано на рис. 2.

Рис. 2. Загальна схема застосування методу адаптації активних правил

Ідея методу полягає в тому, що після знаходження асоціативних правил сервіс ПОАД намагається спростити структуру існуючих активних правил, при цьому використовується колекція операторів модифікації. Структура оператора модифікації P має наступний вигляд: $R' = P(R, A, U)$, де R' - модифіковане активне правило (AcR), R – вихідне активне правило, A – знайдене асоціативне правило (AsR), U – сукупність умов застосування оператора модифікації.

В п'ятому параграфі розглядається роль та місце ПД в еталонній архітектурі багаторівневої розподіленої інформаційно-управляючої системи (БРІУС). Моделі та інформаційні технології, що використовуються для вирішення задач процесів ПД, забезпечують функціональність компонент еталонної архітектури, які відносяться безпосередньо до ПД. Крім цього, дані, що зберігаються та обробляються в ІБДВ, використовує сервіс візуалізації даних, який здійснює відображення поточного стану ТП, та менеджер облікових записів, який забезпечує виконання в системі обмежень щодо доступу користувачів до даних. Дворівнева архітектура процесів ІБДВ забезпечує обробку даних для вузлів базового та кушового рівня еталонної архітектури БРІУС.

У **третьому розділі** наведені моделі та процедури для розробки інформаційної технології, що реалізують основні процеси інтелектуального інжинірингу даних.

В першому параграфі розглядаються моделі і процедури, які використовуються на етапі збору даних та обробки даних щодо особливих ситуацій. Наводиться загальна структура ІБДВ, розглядаються особливості розробки схем даних ІБДВ.

Розглянемо підхід до розробки ретроспективної схеми ІБДВ. Оскільки дані про хід ТП надходять у систему у виді послідовності телеметричних кадрів, що містять актуальні значення чисельних та бульових параметрів (такі, що відповідають останній відповіді, що була одержана від контролера), то їх збереження в БД здійснюється з прив'язкою до значення системного часу, у який

були одержані ці дані. Через те, що для бульового параметру можливими значеннями є лише 0 (повідомлення неактивне) і 1 (повідомлення активне), доцільним представляється зберігати для кожного з бульових параметрів лише моменти часу, у яких він змінював свій стан. Таким чином, у момент часу фіксуються значення чисельних параметрів, що надходять, й ідентифікатори бульових параметрів, які змінили в цей момент часу свій стан. Крім цього, обчислюються агреговані значення параметрів ТП для визначених часових розбивок як результат застосування операторів агрегації. Періоди обчислення агрегованих значень задаються окремо для кожної i -ї часової розбивки (da_i та db_i відповідно для чисельних та бульових параметрів, $i = \overline{1, T}$, T – загальна кількість часових розбивок). Після обчислення агрегованих значень для більш великої часової розбивки відповідні значення з більш дрібною часовою розбивкою можуть бути вилучені, при цьому періоди видалення значень (dra_i та drb_i відповідно для чисельних та бульових параметрів, $i = \overline{1, T}$) не може бути більшими за відповідні періоди обчислення агрегованих значень.

Розглянемо формальне представлення операторів агрегації для чисельних та бульових параметрів. Введемо наступні позначення: $np_1, \dots, np_j, \dots, np_N$ - чисельні параметри, які утворюють множину чисельних параметрів NP ; $m_1^{(i)}, \dots, m_k^{(i)}, \dots, m_Q^{(i)}$ - моменти часу, що відносяться до i -ї часової розбивки; ці моменти часу утворюють множину $M^{(i)}$. Кожному з елементів декартового добутку $NP \times M_i$ поставимо у відповідність число v – значення відповідного чисельного параметру у відповідний момент часу, при цьому $v \in D_j$ (D_j - область можливих значень j -го чисельного параметру). Таким чином ми одержали множину NPV_i , яка містить кортежі виду $\langle np_j, m_k^{(i)}, v \rangle$. Елементи множини NPV_1 - це поточні значення чисельних параметрів, а елементи множин $NPV_i, i > 1$ являють собою середні значення за відповідний часовий інтервал агрегації.

Позначимо як $v_{jk}^{(i)}$ значення v , яке належить кортежеві $\langle np_j, m_k^{(i)}, v \rangle \in NPV_i$. Значення параметрів, що відповідають $i+1$ -й часовій розбивці, можуть бути одержані на основі значень параметрів i -ї часової розбивки наступним чином:
$$v_{jk}^{(i+1)} = \frac{1}{a} \sum_{z \in M'} v_{jz}^{(i)},$$
 де $M' = \{m_z \in M_i : m_k - da_i \leq m_z \leq m_k\}$, a – потужність множини M' .

Для бульових параметрів $bp_1, \dots, bp_j, \dots, bp_B$, які утворюють множину BP , множина BPV_i може бути одержана аналогічним шляхом. Ця множина містить кортежі виду $\langle bp_j, m_k^{(i)}, u \rangle$, де u – значення відповідного бульового параметру ($u \in \{0, 1\}$). Позначимо як $u_{jk}^{(i)}$ значення u з кортежу $\langle bp_j, m_k^{(i)}, u \rangle \in BPV_i$. Тоді елементи множини BPV_2 можуть бути одержані на основі значень

елементів множини BPV_1 за допомогою співвідношення: $u_{jk}^{(2)} = \sum_{z \in M''} h_z$, $h_z = \begin{cases} 1, u_{jz} < u_{jz+1} \\ 0, u_{jz} \geq u_{jz+1} \end{cases}$, де

$M'' = m_z \in M_1 : m_k^{(1)} - de_1 \leq m_z \leq m_k^{(1)}$. Для часових розбивок з більшим порядковим номером співвідношення матиме вигляд: $u_{jk}^{(i+1)} = \sum_{z \in M'''} u_{jz}^{(i)}$, де $M''' = m_z \in M_i : m_k^{(i)} - db_i \leq m_z \leq m_k^{(i)}$.

Введемо сигнатуру імплікативних операцій, що використовуються в подальшому розгляді: $\Sigma = \mapsto, \Rightarrow$. Операція $X \mapsto Y$ використовується для запису активних правил і означає, що в разі справедливості умови X активного правила виконується його дія Y . Операція $X \Rightarrow Y$ використовується для запису асоціативних правил і означає наслідок правої частини правила (імплікації Y) з його лівої частини (умови X).

У другому параграфі розглянуті моделі для процесу обробки даних щодо особливих ситуацій. Для цього процесу використовується розширення класичної моделі активного правила, яке задає реакцію активної СУБД на виникнення специфічних для БД ситуацій:

$$\langle Rule \rangle = \langle Event \rangle \langle Condition \rangle \langle Action \rangle, \text{ або } R \subseteq E \times C \times A.$$

В класичній моделі активного правила розділ *Event* (подія) дозволяє задавати зовнішні по відношенню до БД явища, при виникненні яких перевіряється умова або набір умов щодо поточного стану БД, що задаються в розділі *Condition* (умова), і якщо умови виконані, СУБД виконує набір операцій, зазначених в розділі *Action* (дія).

В комплексі ПД активні правила виступають як засіб специфікації особливих ситуацій, що виникають в ТП, та реакції АСУ ТП на виникнення таких ситуацій. Під реакцією системи мається на увазі здійснення додаткового опрацювання значень параметрів ТП, що характеризують такі проблемні ситуації. При цьому як зовнішнє по відношенню до БД явище, що описується в розділі *Event*, виступає виклик з боку СОД відповідних процедур ІБДВ, що здійснюють занесення значень параметрів в таблиці ретроспективної схеми. У розділі *Condition* за допомогою набору умов до значень параметрів ТП, які можуть з'єднуватися за допомогою логічних операторів, формується предикат, який задає специфікацію проблемної ситуації в ТП. Розділ *Action* містить перелік параметрів ТП, значення яких є суттєво важливими для аналізу стану ТП протягом особливої ситуації, розгляду причин її виникнення (перед-історії) та наслідків (пост-історії). Значення параметрів, що містяться в розділі *Action*, зберігаються в ІБДВ і не можуть бути видаленими внаслідок виконання операторів агрегації.

Таким чином, розширена модель активного правила для специфікації особливих ситуацій в ТП та поведінки системи в разі виникнення таких ситуацій має вигляд:

$$R' \subseteq E \times C' \times A' \times HA \times HA,$$

де C' – це розширена умова активного правила, яка має узагальнений вигляд $C' = \tilde{\Phi}(S, L)$, S – множина елементарних умов, L – множина логічних операторів, $\tilde{\Phi}$ – узагальнена суперпозиція. Як елементарні умови розглядаються умови наступних типів: 1) порівняння значення чисельного параметра із константою: $np_i \omega const$, де $np_i \in NP$, $\omega \in \Omega$, $\Omega = \{<, \leq, >, \geq, =, \neq\}$; 2) порівняння значень двох чисельних параметрів: $np_j \omega np_k$, $np_j, np_k \in NP$, $\omega \in \Omega$; 3) умова на значення бульового параметру: $bp_s^{\beta_s}$, $bp_s \in BP$, $bp_s^{\beta} = \begin{cases} bp, & \beta = 1 \\ \overline{bp}, & \beta = 0 \end{cases}$. Дія правила A' задається як множина

$NP' \subseteq NP$, яка містить параметри ТП, значення яких мають зберігатися в разі виникнення особливої ситуації, що описана даним активним правилом. HA – це множина можливих довжин інтервалів перед-історії та пост-історії активного правила.

Введемо наступну форму запису для активного правила, яке задає специфікацію особливої ситуації в ТП (при цьому будемо вважати, що подією для нього виступає факт отримання нового пакету зі значеннями параметрів ТП): $\Phi(S, L) \mapsto NP'$. Тут $\Phi(S, L)$ – функція, що задає умову активного правила. Значення цієї функції в момент часу t позначимо як $\Phi(t)$. Це значення визначає, чи виконується умова активного правила в момент часу t , та може дорівнювати 0 чи 1. Запис $\Phi(S, L) \mapsto NP'$ означає, що при виконанні умови $\Phi(S, L)$ мають бути збережені значення для параметрів, що належать до множини NP' . Позначимо тривалість перед-історії активного правила як \underline{t} , тривалість його пост-історії як \bar{t} . Також позначимо як t_1 момент часу, для якого вірні наступні співвідношення: $\Phi(t_1) = 1; \Phi(t_1 - 1) = 0$, та як t_2 – момент часу, для якого вірні наступні співвідношення: $\Phi(t_2) = 1; \Phi(t_2 + 1) = 0$. Моменти часу t_1 та t_2 є відповідно моментами початку та закінчення виконання умови активного правила. Тоді значення параметрів з множини NP' зберігатимуться за допомогою розглянутого активного правила протягом часового інтервалу $[t_1 - \underline{t}, t_2 + \bar{t}]$.

У третьому параграфі розглядаються розроблені процедури й алгоритми етапу проблемно-орієнтованого аналізу даних. Спочатку розглядається запропонована інтерпретація класичної концепції асоціативних правил в контексті АСУ ТП. Для запису закономірностей, що були знайдені на основі аналізу значень чисельних параметрів ТП, використовуються кількісні асоціативні правила. В роботі запропоновано наступну модель таких правил:

$$\int \cup \quad \cup \int \quad , \quad U \cap V = \emptyset, \quad U \subseteq NP, \quad V \subseteq NP.$$

Тут $\underline{c}'_j, \overline{c}'_j, \underline{c}'_k, \overline{c}'_k$ – це відповідно нижні та верхні границі інтервалів можливих значень чисельних параметрів p_j, p_k ($p_j, p_k \in NP$). Для закономірностей, що були знайдені на основі

аналізу значень булевих параметрів ТП, використовуються булеві асоціативні правила у вигляді:
 $E_L \Rightarrow E_R, E_L \subset BP, E_R \subset BP, E_L \cap E_R = \emptyset$.

Для кожного асоціативного правила $A \Rightarrow B$, яке було побудовано на відповідній множині транзакцій БД, визначаються його рівень довіри c та рівень підтримки s . Оскільки в процесі ПОАД генерація асоціативного правила A_j здійснюється на основі накопиченої історії виконання активного правила R_i , то ми можемо ввести поняття рівня довіри та підтримки асоціативного правила відповідно для зони перед-історії активного правила (зона 1), зони розвитку особливої ситуації (зона 2) та зоні пост-історії (зона 3). Такі параметри позначаються відповідно $s_i^{(1)}, s_i^{(2)}, s_i^{(3)}, c_i^{(1)}, c_i^{(2)}, c_i^{(3)}$.

Визначення. Асоціативне правило A_j є ефективним для зони особливої ситуації активного правила R_i , якщо рівні підтримки та довіри правила A_j , обчислені для цієї зони, є вищими за поперед задані мінімальні рівні s_{\min}, c_{\min} . Асоціативне правило A_j є ефективним для активного правила R_i в цілому, якщо наведені вище умови виконуються для всієї множини транзакцій БД, пов'язаних із активним правилом R_i .

Положення. Якщо асоціативне правило A_j є ефективним для активного правила R_i в цілому, то воно є ефективним хоча б для однієї з його зон.

В роботі визначається кілька предикатів, які використовуються під час розробки колекції операторів модифікації активних правил. Позначимо умову, що асоціативне правило A є ефективним для зони n активного правила R , як предикат $Eff(A, R, n)$, а умову, що асоціативне правило A є ефективним для активного правила R в цілому, - як предикат $Eff(A, R)$.

Позначимо як $r(p_i)$ інтервальну умову на значення параметру p_i наступного вигляду:
 $\underline{p}_i \leq p_i \leq \overline{p}_i$, де $\underline{p}_i, \overline{p}_i$ - відповідно нижня та верхня границі інтервалу значень параметру p_i .
 Предикат $Rng(R, A, p_i)$ являє собою умову перетинання інтервалів для активного правила R , умова якого містить $r(p_i)$ як одну з елементарних умов, та асоціативного правила A , заданого у формі $\tilde{r}(p_j) \rightarrow \tilde{r}(p_i)$. Цей предикат має наступний вигляд: $(\tilde{r}(p_1) \cap \tilde{r}(p_2)) \neq \emptyset$, де

Раніше на рис. 2 було наведено загальну схему методу адаптації існуючих активних правил на основі результатів ПОАД. В цьому методі можливі типи модифікацій визначаються за допомогою операторів модифікації. Запропоновано колекцію операторів модифікації, яка складається з наступних п'яти операторів: 1) оператор виключення залежного чисельного параметру з умови активного правила; 2) оператор з'єднання активних правил; 3) оператор

скорочення перед-історії активного правила; 4) оператор скорочення пост-історії активного правила; 5) оператор виключення залежного бульового параметру з умови активного правила.

Розглянемо оператор виключення залежного чисельного параметру. Цей оператор передбачає, що якщо асоціативне правило A свідчить про наявність залежності між двома параметрами, що входять до умови активного правила R , то в разі виконання умови U ми можемо замінити активне правило R на активне правило R' , умова якого вже не міститиме залежного параметру. Складові компоненти оператору мають наступний вигляд:

$$R: r(p_1) \wedge r(p_2) \wedge \dots \wedge r(p_n) \mapsto p_{k_1}, \dots, p_{k_m} ; A: \tilde{r}(p_1) \Rightarrow \tilde{r}(p_2) ;$$

$$U: (Eff(A, R, 2)) \wedge (Rng(A, R, p_2)) \wedge (Eff(\tilde{r}(p_1) \Rightarrow \tilde{r}(p_2), R, 2)) = I ;$$

$$R': r(p_1) \wedge r(p_3) \wedge \dots \wedge r(p_n) \mapsto p_{k_1}, \dots, p_{k_m} .$$

Нарешті, у четвертому параграфі викладені процедури й алгоритми етапу синхронізації даних. Запропонована процедура синхронізації складається з підготовчого та основного етапів. На підготовчому етапі виконуються наступні операції: 1) визначення шаблонів синхронізації; 2) проведення декомпозиції вихідної реляційної схеми; 3) побудова SQL-запитів до БД, які б надавали дані у форматі XML. Протягом основного етапу виконуються наступні дії: 1) виконання запитів, одержання XML-документів; 2) передача даних по каналах зв'язку; 3) аналіз одержаних даних, їх занесення до БД. Всі шаги підготовчого етапу та шаг 1 основного етапу виконуються у вузлу - сервері публікацій, шаг 3 основного етапу виконується у вузлі – клієнті.

Шаблон синхронізації має наступне формальне визначення: $X = \langle S_i, \lambda_i, \tau \rangle$, де S_i - таблиці бази даних, які було віднесено до даного шаблону, λ_i - ознака зміни даних у таблиці (по доданню даних або по їх зміні), τ - період синхронізації для таблиць шаблону.

Розглянемо алгоритм декомпозиції реляційної схеми. Нехай $\tilde{S} = s_i, i = \overline{1, \tilde{n}}$ - вихідна множина таблиць. Введемо на цій множині відношення імпорту Im первинних ключів:

$$Im(s_1, s_2) = \begin{cases} I, & \text{таблиця } s_1 \text{ містить первинний ключ таблиці } s_2; \\ 0, & \text{інакше.} \end{cases}$$

Введемо також множини $I(s) = s_i, Im(s, s_i) = I$ та $E(s) = s_i, Im(s_i, s) = I$. В алгоритмі для кожної таблиці встановлюються значення трьох параметрів: *schema* - номер схеми, у яку будуть поміщені дані з таблиці; *tag* – номер XML-елементу, що відповідає даній таблиці; *parent* – номер предка даної таблиці. Алгоритм складається з наступних кроків:

Крок 1. Визначити множину $S = \tilde{S} / S_I$, де $S_I = s_i : I(s) = \emptyset, E(s) = \emptyset$. Для $s_i \in S_I$ прийняти *schema*=*tag*=*parent*=0. Подальші дії виконуються для таблиць, що належать до множини S .

Крок 2. Сформувати матрицю D . $d_{ij} = \begin{cases} 1, \text{Im}(s_j, s_i) = 1 \\ 0, \text{інакше} \end{cases}, i, j = \overline{1, n}$.

Крок 3. Сформувати матрицю C : $c_{ij} = \begin{cases} 1, d_{ij}^{\Sigma} > 0 \\ 0, d_{ij}^{\Sigma} = 0 \end{cases}, D^{\Sigma} = D^1 + \dots + D^n$.

Крок 4. Прийняти $f_j^0 = 0, j = \overline{1, n}, k=1$.

Крок 5. Визначити $f_j^k = \begin{cases} k, c_{kj} \neq 0 \\ f_j^{k-1}, c_{kj} = 0 \end{cases}, j = \overline{1, n}$. Якщо $k = n$, йти до кроку 8.

Крок 6. Знайти z^k - кількість різних елементів з $F^k = f_j^k$. Якщо $z^k \leq z^{k-1}$, то прийняти $k = k + 1$ і перейти до кроку 5.

Крок 7. Прийняти $f_j^{k-1} = k$ для $f_j^{k-1} = 0$ і $c_{kj} \neq 0 (j = \overline{1, n})$, $f_j^k = f_j^{k-1}$, $k = k + 1$.

Перейти до кроку 5.

Крок 8. Сформувати множину Z різних елементів з F^k . Для всіх таблиць $s_j, j \in Z$ встановити $schema=i, tag=j, parent=0 (i = \overline{1, z})$. Побудувати множину $E(s_j)$. Якщо $E(s_j) \neq \emptyset$, то для усіх $e_i \in E(s_j)$ встановити: $schema=j, tag=i+1, parent=e_i$.

В результаті роботи розробленого алгоритму формується ієрархічна структура XML-документів, до яких будуть перетворені дані з таблиць вихідної схеми БД.

У **четвертому розділі** розглядаються питання реалізації інформаційної технології ПД в АСУ ТП, а також оцінювання їхньої ефективності та продуктивності.

В першому параграфі досліджується продуктивність інформаційної технології, що реалізує моделі та процедуру ПД, на основі метричного моделювання. Оскільки продуктивність ПД має першорядне значення, коли йдеться про виконання обмежень реального часу, розглядається метрика, що пов'язана із часом виконання окремих процедур інформаційної технології. В роботі запропоновано процедуру метричного моделювання продуктивності, що складається з наступних кроків: 1) розробка діаграм діяльності для програмних компонент, що моделюються; 2) визначення вектору \vec{P} вхідних параметрів метричної моделі; 3) побудова графів виконання для програмних компонент на основі діаграми діяльності; 4) побудова метричної моделі у вигляді $EC = F(\vec{P})$, де EC – ресурсоемність (Execution Cost), що вимірюється в системних одиницях. $F(\vec{P})$ - аналітичний вираз для знаходження ресурсоемності за значеннями параметрів метричної моделі; 5) знаходження ресурсоемності програмних компонент, що вимірюється в системних одиницях, за допомогою програм-аналізаторів; 6) перехід від системних до часових одиниць, які визначаються окремо для

кожної апаратно-програмної конфігурації за допомогою програм-профайлерів; 7) аналіз результатів та знаходження критичних значень параметрів моделі.

В роботі було досліджено продуктивність програмних компонент, що реалізують моделі та процедури процесу збору даних та обробки даних щодо особливих ситуацій, оскільки саме ці процеси є найбільш хитливими з погляду виконання обмежень реального часу. Такими компонентами є ІБДВ та МАП. Під час дослідження був сформований наступний вектор параметрів метричної моделі: $\vec{P} = \langle C, M, P, A, D, Pn, p_1, p_2, p_3, p_4 \rangle$. Тут C – кількість контролерів, M – кількість булевих параметрів, P – кількість чисельних параметрів, A – кількість активних правил, D – середня складність активного правила, Pn – середня кількість параметрів в умові активного правила, p_1 – ймовірність зміни пакету булевих параметрів, p_2 – ймовірність зміни хоча б одного з булевих параметрів, p_3 – середня ймовірність зміни значення булевих параметрів, p_4 – ймовірність того, що поточний чисельний параметр хоча б у одному активному правилі входить до умови, що виконується в даний момент часу.

Розроблена модель для обчислення ресурсоемності компонент ІБДВ та МАП має вигляд:

$$EC = 0,29 + C \cdot (0,03 + p_1 \cdot (M \cdot (0,04 + p_2 \cdot 0,03) + 0,07) + P \cdot (0,04 + p_3 \cdot 0,02)) + \\ + A \cdot (0,06 + 0,01D + p_4 \cdot (0,03 + 0,04P_n)).$$

Щоб дослідити, які параметри мають найбільший вплив на ресурсоемність системи, було проведено експеримент, результати якого наведені на рис.3. Фіксувалися значення всіх параметрів моделі, окрім одного, значення якого збільшувалося із коефіцієнтом k (вісь абсцис на рис. 3). По вісі ординат відкладалися значення ресурсоемності, що були одержані (для параметрів p_1, p_2, p_3 та p_4 максимальне значення в експерименті було прийнято рівним 1).

Рис. 3. Результати дослідження метричної моделі

Найбільший вплив на продуктивність системи мають значення параметрів A (кількість активних правил) та D (середня складність активного правила).

В другому параграфі розглядається дослідження ефективності запропонованого МТК ПД. Для цього проводиться серія експериментів з метою спрощення існуючих активних правил із застосуванням операторів модифікації. На рис. 4 показаний результат спрощення активних правил внаслідок застосування операторів модифікації (по осях відкладені номер активного правила та складність активного правила, світлим кольором показана складність активного правила після модифікації, темним – „виграш” в результаті спрощення активного правила).

Було одержано сумарне зменшення складності активних правил, що розглядалися під час експерименту, на 12,7%. При цьому було досягнуто спрощення 8 активних правил з 10.

Неможливість спрощення решти активних правил пояснюється відсутністю відповідних асоціативних правил, які могли б спричинити застосування операторів модифікації.

Рис. 4. Результати експерименту щодо дослідження ефективності МТК ПД

В третьому параграфі розглядається процес розробки інформаційної технології, що реалізує моделі, процедуру та алгоритми ПД в АСУ ТП. Розглянуто логічні моделі даних окремих схем ІБДВ, особливості програмної реалізації операторів агрегації даних (процес збору даних), процедур визначення стану активного правила та виконання його дії (процес попереднього аналізу даних), етапів та кроків процесу синхронізації даних та методу адаптації набору активних правил (процес ПОАД).

ВИСНОВКИ

У роботі вирішено важливу науково-практичну задачу підвищення ефективності обробки даних в складних розподілених АСУ ТП. В результаті виконання даної роботи були одержані наступні результати.

1) Проведено аналіз проблеми обробки даних в сучасних АСУ ТП. Обґрунтовано актуальність вирішення цієї проблеми на основі розподілу функцій обробки даних між вузлами АСУ ТП. Критерієм такого розподілу є критичність функції обробки даних з погляду обмежень реального часу.

2) Наведено визначення МТК ПД в АСУ ТП. Запропоновано архітектуру процесів і компонент ПД, у межах якої виділено чотири процеси: збору даних, обробки даних щодо особливих ситуацій у ТП, проблемно-орієнтованого аналізу даних та синхронізації даних.

3) Одержано моделі ІБДВ, яка забезпечує зберігання та накопичування даних щодо ТП. Виділено та запропоновано моделі схем даних ІБДВ, зокрема, конфігураційної схеми, схеми протоколів, ретроспективної схеми.

4) Запропоновано розширену модель активного правила для специфікації особливих ситуацій в ТП. Розроблено механізм активних правил, який здійснює моніторинг особливих ситуацій та забезпечує збереження значень технологічних параметрів, які характеризують такі ситуації.

5) Запропоновано моделі асоціативних правил для представлення закономірностей у параметрах ТП, значення яких накопичуються в системі. Генерація асоціативних правил здійснюється в процесі ПОАД.

6) Розроблено метод адаптації набору активних правил на основі результатів ПОАД. Його використання дозволяє спростувати структуру активних правил, що призводить до підвищення продуктивності їхньої обробки та запобігає порушенням обмежень реального часу.

7) Одержано процедуру синхронізації даних у розподіленій АСУ ТП. Запропоновано алгоритм декомпозиції реляційних схем даних, який спрощує їх відображення до формату XML.

8) Розроблено інформаційну технологію, що реалізує моделі та процедури ПД в АСУ ТП. Проведено дослідження продуктивності та ефективності розробленого МТК ПД.

9) Результати дисертаційної роботи знайшли практичне застосування під час розробки ІУС АСУ ТП Ул'янівської установки комплексної переробки газу (УКПГ) Перещепінського нафтогазопромислу та Сахалінській УКПГ (Краснокутський район Харківської області), що підтверджено актами впровадження.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Кукленко Д.В., Ткачук Н.В. Ретроспективная база данных в АСУ ТП: концепция и опыт реализации // Вісник Національного технічного університету „Харківський політехнічний інститут”. – Харків: НТУ „ХПІ”. – 2001. - Вип..8. - С. 62-68.

Здобувачеві належать моделі ретроспективної бази даних та описані технології відображення накопиченої в ній інформації.

2. Tkachuk N.,Kuklenko D., Ovasapov S., Schekotichin K. Knowledge-based Maintenance Environment for Large Information Handling Systems // Lecture Notes in Computer Science. – Germany, Bonn. -2001. - P. 139-154.

Здобувачеві належать концептуальна модель успадкованої системи управління станційним обладнанням газокompресорної станції та концептуальна модель бази даних цільової системи.

3. Kuklenko D. Design and Development Issues of Retrospective Database in Complex Technological Control Systems //Lecture Notes in Computer Science.– Germany, Bonn.- 2001. -P.21-31.

4. D.V. Kuklenko, H.C. Mayr, M. V. Tkachuk, K. M. Shchekotykhin. Web-Based Information Systems For Technological Process Control: Architectural Framework And Software Solutions // Проблемы программирования. - К.:2002. - № 1-2. - С. 317 – 325.

Здобувачеві належить аналіз можливих СУБД й мотивація вибору MS SQL Server.

5. Ткачук Н.В., Кукленко Д.В. Применение концепции SCADA-систем для интеллектуального реинжиниринга данных в АСУ ТП // Автоматизированные системы управления и приборы автоматики. – Харьков: ХНУРЭ. – 2002. - Вып. 121. - С.129-136.

Здобувачеві належить аналіз вимог до підсистеми накопичування даних у складі SCADA-системи, а також моделі даних та підсистеми відображення для кількох її реалізацій.

6. Mykola V. Tkachuk, Heinrich C. Mayr, Dmytro V. Kuklenko, Michail D. Godlevsky. Web-based Process Control Systems: Architectural patterns, Data Models, and Services. // Lecture Notes in Computer Science. – Germany, Berlin. – 2002. - P. 721-729.

Здобувачеві належить концепція інтегрованої бази даних вузлу (Integrated Node Database) та механізму правил (Supervisory Rules Engine) для обробки даних у проблемних ситуаціях.

7. Кукленко Д.В. Подход к интеллектуальной обработке данных в АСУ ТП, основанный на применении концепции активных баз данных // Високі технології в машинобудуванні. – Харків: НТУ „ХПІ”. – 2003. -Вип..2. - С.51-56.

8. Кукленко Д.В. Концепция и модели интегрированной базы данных узла как средства хранения и обработки данных в АСУ ТП // Вісник Національного технічного університету „Харківський політехнічний інститут”. – Харків: НТУ „ХПІ”. – 2003. - Вип. 7, т. 1. - С. 67-78.

9. Ткачук М.В., Кукленко Д.В. Про один підхід до побудови архітектури розподіленої обробки даних в Intranet-системах управління технологічними процесами // Системні дослідження та інформаційні технології. – Київ. – 2003. - №4. – С.35-43.

Здобувачеві належить концепція інтелектуального інжинірингу даних та його архітектура.

10. N.V. Tkachuk, V.A. Shekhovtsov, D.V. Kuklenko, and H.C. Mayr. An Approach to Efficient Data Handling in Web-based Process Control Systems // Proc.IASTED International Conf. on Intelligent Systems and Control. – Austria, Salzburg. – 2003. - P. 242-247.

Здобувачеві належить модель активного правила для опису проблемної ситуації в технологічному процесі та інформаційна технологія механізму активних правил.

11. Kuklenko D., Gamzayev R., Goloborodko V., Tkachuk M. An Approach to Advanced Data Synchronization in Complex Process Control System // Lecture Notes in Computer Science. – Germany, Bonn. – 2004. - P. 221-227.

Здобувачеві належать процедура синхронізації даних в багаторівневій розподіленій АСУ ТП та процедури оцінювання ефективності функціонування програмних компонентів бізнес-логіки бази даних ретроспективної інформації з застосуванням метричного підходу.

АНОТАЦІЇ

Кукленко Д.В. Моделі та інформаційні технології інтелектуального інжинірингу даних в автоматизованих системах управління технологічними процесами. – Рукопис.

Дисертаційна робота на здобуття вченого ступеню кандидата технічних наук за спеціальністю 05.13.06.–Автоматизовані системи управління та прогресивні інформаційні технології. – Національний технічний університет “Харківський політехнічний інститут”. Харків. – 2005.

У дисертаційній роботі вирішується актуальна науково-практична задача розробки модельно-технологічного комплексу (МТК), що забезпечує ефективну обробку даних у розподілених ІУС АСУ ТП. Цей МТК отримав в роботі назву інтелектуальний інжиніринг даних (ІІД) в АСУ ТП. ІІД включає до себе наступні процеси обробки даних у складних АСУ ТП: збір

даних; попередня обробка даних щодо особливих ситуацій у технологічному процесі; проблемно-орієнтований аналіз даних; синхронізація даних між вузлами розподіленої багаторівневої АСУ ТП.

Розроблено архітектуру компонент і процесів ІД в АСУ ТП. Наведено розширену модель активного правила для опису особливих ситуацій в технологічному процесі та реакції АСУ ТП у випадку виникнення таких ситуацій; запропоновано метод адаптації набору активних правил на основі одержаних в результаті проблемно-орієнтованого аналізу даних асоціативних правил. Розроблений метод базується на використанні колекції операторів модифікації активних правил. Запропоновано алгоритм декомпозиції реляційної схеми даних, який використовується на етапі синхронізації. Було досліджено ефективність та продуктивність розробленого комплексу ІД, при цьому було використано реальну вихідну інформацію.

Ключові слова: інформаційна технологія, АСУ ТП, інтелектуальний інжиніринг даних, активне правило, асоціативне правило, бази даних.

Кукленко Д.В. Модели и информационные технологии интеллектуального инжиниринга данных в автоматизированных системах управления технологическими процессами. – Рукопись.

Диссертационная работа на соискание ученой степени кандидата технических наук по специальности 05.13.06 – Автоматизированные системы управления и прогрессивные информационные технологии. – Национальный технический университет “Харьковский политехнический институт”. Харьков. – 2005.

В диссертационной работе решается актуальная научно-практическая задача разработки модельно-технологического комплекса (МТК), обеспечивающего эффективную обработку данных в распределенных ИУС АСУ ТП. Данный МТК получил в работе название интеллектуальный инжиниринг данных (ИИД) в АСУ ТП. ИИД включает следующие процессы обработки данных в сложных АСУ ТП: сбор данных, предварительная обработка данных об особых ситуациях в технологическом процессе, проблемно-ориентированный анализ данных, синхронизация данных между узлами распределенной многоуровневой АСУ ТП.

Разработана архитектура компонент и процессов ИИД в АСУ ТП. Предлагается расширенная модель активного правила для описания особых ситуаций в технологическом процессе и реакции АСУ ТП в случае возникновения такой ситуации; разработан метод адаптации набора активных правил на основе полученных в результате проблемно-ориентированного анализа данных асоціативных правил. В основе метода лежит использование коллекции операторов модификации активных правил. Также в работе предложен алгоритм декомпозиции реляционной схемы данных, используемый на этапе синхронизации данных. Были исследованы эффективность и

производительность разработанного комплекса ИИД, при этом были использованы реальные исходные данные.

Ключевые слова: информационная технология, АСУ ТП, интеллектуальный инжиниринг данных, активное правило, ассоциативное правило, базы данных.

Kuklenko D.V. – Models and information technologies of intelligent data engineering in Process Control Systems. – Manuscript.

Thesis for a candidate's degree by specialty 05.13.06 – computer-aided management systems and progressive information technologies. – National Technical University “Kharkiv Polytechnic Institute”. Kharkiv – 2005.

The important scientific problem of designing the framework of models and information technologies (CMIT) that provides effective data processing in distributed Process Control Systems (PCS) has been solved in the work. CMIT proposed has been named as Intelligent Data Engineering (IDE) in PCS. IDE contains the following processes of data handling in complex PCS: collecting data; pre-processing data with respect to specific situations in technical process; problem-oriented data analysis; synchronizing data between nodes of distributed multi-level PCS.

The architecture of processes and components of IDE has been designed. The enhanced Active Rule model that describes the specific situations in technical process as well as the system's reaction on such situation has been elaborated; the framework for active rule adaptation based on the results of problem-oriented data analysis in the form of association rules is presented (this framework is grounded on the concept of Active Rule modification pattern). An algorithm that allows to decompose the relational scheme (process of synchronizing data) is also described in the thesis. The effectiveness and performance of elaborated IDE CMIT have been investigated based on the real-life case-study.

Keywords: information technology, Process Control System, Intelligent Data Engineering, Active Rule, Association Rule, database.