

Міністерство освіти і науки України
Вінницький національний технічний університет

ПРИЙМАК НАТАЛІЯ ВАСИЛІВНА

УДК 004.8:004.415

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПОШУКУ АСОЦІАТИВНИХ ПРАВИЛ
ЩОДО ТРИВАЛОСТІ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**

05.13.06 – інформаційні технології

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня
кандидата технічних наук

Вінниця – 2020

Дисертацією є кваліфікаційна наукова праця на правах рукопису.

Робота виконана у Вінницькому національному технічному університеті Міністерства освіти і науки України.

Науковий керівник: кандидат технічних наук, професор
Савчук Тамара Олександрівна,
Вінницький національний технічний університет,
професор кафедри комп'ютерних наук

Офіційні опоненти: доктор технічних наук, професор
Гребеннік Ігор Валерійович,
Харківський національний університет
радіоелектроніки, завідувач кафедри системотехніки

доктор технічних наук, професор
Фісун Микола Тихонович,
Чорноморський державний університет імені Петра
Могили, завідувач кафедри інтелектуальних
інформаційних систем

Захист відбудеться «02» квітня 2020 р. о 12³⁰ годині на засіданні спеціалізованої вченої ради Д 05.052.01 у Вінницькому національному технічному університеті за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК, ауд. 210.

З дисертацією можна ознайомитися в бібліотеці Вінницького національного технічного університету за адресою: 21021, м. Вінниця, вул. Хмельницьке шосе, 95, ГНК, а також на сайті університету.

Автореферат розісланий «28» лютого 2020 р.

Вчений секретар
спеціалізованої вченої ради

С. М. Захарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Обґрунтування вибору теми дослідження. У сучасному світі програмне забезпечення (ПЗ) використовується для роботи, навчання, спілкування; для здійснення покупок, платежів, подорожей тощо. Програмне забезпечення відіграє важливу роль для різних галузей бізнесу та виробництва: використовується банками та страховими компаніями; в енергетиці та сільському господарстві, торгівлі, медицині, навчанні тощо.

Оскільки частка програмного забезпечення стає все більш вагомим в усіх галузях промисловості та послуг, то процес його розробки та інженерні засоби, необхідні для цього, потребують постійного розвитку та вдосконалення. Так як програмні продукти різноманітні та мають різний контекст використання: застосування прикладного програмного забезпечення в інформаційній системі чи програмного забезпечення, вбудованого в інші продукти, або ігрового чи навчального програмного забезпечення тощо, то виникає необхідність оцінки відповідності системи, компонента чи процесу визначеним вимогам: функціональним чи нефункціональним.

Однією з нефункціональних вимог до застосування програмного забезпечення, згідно зі стандартом IEEE Computer 1985, є вимога до його життєвого циклу: визначення обмежень щодо задіяних людських ресурсів та тривалості процесу розробки такого ПЗ. Точна оцінка графіка закінчення розробки програмного забезпечення необхідна менеджерам для планування роботи команди розробників та ефективного управління даним процесом. Окрім вказаної вимоги важливим показником є організація процесу розробки та якість отриманого програмного забезпечення.

Одним із способів перевірки, що програмне забезпечення надійне – є контроль усіх етапів його розробки. Традиційна модель розробки програмного забезпечення передбачає перевірку кожного етапу розробки (вимоги до документації, проектування, аналіз, кодування, перевірка та тестування) перед переходом до наступного. Сучасні програмні проекти розробляються більш динамічно, із застосуванням більш гнучких технологій та методологій розробки; але етапи, перераховані вище, присутні в будь-якому процесі, оскільки усе програмне забезпечення повинно бути розроблене, інтегроване, перевірене, випущене та впроваджене.

У зв'язку з тим, що складність програм зростає, то їх якість та доступність має важливе значення для споживача. В процесі написання програмного забезпечення його код часто змінюється з метою поліпшення функціональності або для виправлення помилок, які викликають збій у його роботі. Використання систем управління версіями програмного коду та баз даних (БД) завдань, що ставляться під час розробки ПЗ, дозволяють зібрати інформацію про усі зміни програмного коду та використати її для подальшого

аналізу. Наявні бази даних містять значну кількість інформації, яку неможливо проаналізувати вручну, а, отже, виникає необхідність у автоматизованому аналізі. Для вирішення даної задачі використовують різноманітні методи інтелектуального аналізу даних до яких належить і пошук асоціативних правил (АП). Такий метод дозволяє знаходити асоціативні залежності які можна використати для визначення часу, необхідного для реалізації конкретного завдання розробником. Отримана інформація може бути застосована менеджерами проектів для планування розробки ПЗ.

Сучасні методи пошуку асоціативних правил щодо тривалості розробки програмного забезпечення розглянуті в роботах Г. П'ятецький-Шапіро, В. Жоу, Ф. Жанг, Х. Жанг, П. Бхатачар'я, А. Ламканхфі, Е. Гігер. Питання моделювання систем представлено у роботах А. А. Молчанова, В.І. Скуріхіна, Н. Бусленко, В. М. Трояновського, Р. Н. Кветного, В. М. Дубового, Б. І. Мокіна, В. Б. Мокіна, О. В. Бісікала та інших дослідників. Один із підходів до вирішення задачі прогнозування трудомісткості розробки програмних систем запропоновано в роботі С. Д. Штовби.

Отже, наукова задача створення інформаційної технології пошуку асоціативних правил щодо тривалості розробки програмного забезпечення є актуальною, оскільки використання такої інформаційної технології дозволить підвищити швидкість процесу розробки програмного забезпечення.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій».

Результати дисертаційного дослідження увійшли у звіт з держбюджетної науково-дослідної роботи на тему «Розробка інформаційної системи пошуку асоціативних правил щодо тривалості розробки програмного забезпечення» № 47/5В (номер договору про творчу співпрацю № 47/5 від 15 січня 2018 року), де здобувач був у складі авторів проміжних та заключного звітів. Також впроваджено інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного забезпечення в ТОВ «Бізнес-Програми».

Мета і завдання дослідження. Метою дисертаційного дослідження є підвищення точності визначення тривалості виконання завдання розробником, а також скорочення тривалості пошуку асоціативних правил, які для цього застосовуються, за рахунок використання відповідної інформаційної технології.

Для досягнення поставленої мети необхідно розв'язати такі *завдання*:

1. Дослідити сучасний стан розвитку процесу розробки програмного забезпечення та проаналізувати існуючі методи визначення частих предметних наборів для пошуку асоціативних правил щодо тривалості даного процесу.

Розробити інформаційну модель (ІМ) процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

2. Удосконалити метод визначення частих предметних наборів FP-Growth для пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

3. Розробити інформаційну технологію, яка проводитиме пошук асоціативних правил щодо тривалості розробки програмного забезпечення з метою їхнього подальшого використання при розробці програмного забезпечення.

4. Провести моделювання процесу пошуку асоціативних правил з використанням удосконаленого методу FP-Growth.

5. Розробити інформаційну систему пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, що базується на запропонованій інформаційній технології.

Об'єктом дослідження є процес пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Предметом дослідження є інформаційна технологія пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Методи дослідження. Використано теоретичні та змішані методи, серед них: методи системного аналізу (порівняльний та структурний аналіз) для визначення властивостей процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення; методи інтелектуального аналізу даних (пошук асоціативних правил та класифікацію даних) для розробки інформаційної моделі вищевказаного процесу та удосконаленого методу визначення частих предметних наборів для пошуку асоціативних правил, що базується на цій моделі; методи об'єктного моделювання процесу пошуку асоціативних правил, а також розробки програмного забезпечення; комп'ютерне моделювання для розробки програмного забезпечення пошуку асоціативних правил.

Наукова новизна отриманих результатів полягає у тому, що:

1. Уперше запропоновано інформаційну модель процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, яка, на відміну від існуючих, враховує такі характеристики завдань (завдань, які потрібно вирішити при розробці програмного забезпечення), як: складність, пріоритет, критичність та тривалість їх виконання. Це дозволило підвищити точність визначення часу, необхідного на виконання завдання розробником певної кваліфікації.

2. Удосконалено метод визначення частих предметних наборів для пошуку асоціативних правил FP-Growth щодо тривалості розробки програмного забезпечення за рахунок уведення етапу класифікації завдань відповідно до таких характеристик завдань: складність, пріоритет, критичність

та тривалість їх виконання. Це дозволило скоротити час, необхідний для пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, які в, подальшому, використовуються для визначення часових характеристик реалізації завдання розробником певної кваліфікації.

3. Уперше розроблено інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного забезпечення в основі якої, на відміну від існуючих, використовується інформаційна модель та удосконалений метод визначення частих предметних наборів для пошуку асоціативних правил FP-Growth, що дозволило збільшити точність визначення тривалості виконання завдання розробником певної кваліфікації та скоротити тривалість пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Практичне значення отриманих результатів. Практична цінність результатів дисертаційного дослідження полягає в тому, що:

– розроблено алгоритм, що реалізує удосконалений метод визначення частих предметних наборів для пошуку асоціативних правил щодо тривалості розробки програмного забезпечення за рахунок уведення етапу класифікації завдань відповідно до характеристик завдань (складність, пріоритет, критичність та тривалість їх виконання), що дозволило скоротити тривалість пошуку таких асоціативних правил;

– розроблено алгоритм використання знайдених асоціативних правил щодо тривалості розробки програмного забезпечення, що дозволило автоматизувати процес розробки програмного забезпечення та підвищити точність визначення тривалості виконання завдання розробником певної кваліфікації;

– розроблено інформаційну систему для пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, що реалізує розроблену інформаційну технологію.

Отримані на основі наукових досліджень практичні результати впроваджено в компанії «Бізнес-Програми» у вигляді інформаційної технології пошуку асоціативних правил щодо тривалості розробки програмного забезпечення (акт впровадження від 27.07.18) та інформаційної системи пошуку асоціативних правил при розробці програмного забезпечення (акт впровадження № 47/5В від 17.09.18). Це забезпечило ефективність управлінських рішень щодо планування та управління розробкою програмного забезпечення.

Особистий внесок здобувача. Основні теоретичні та експериментальні дослідження та висновки дисертаційної роботи були отримані автором особисто.

У друкованих працях, опублікованих у співавторстві, автору дисертації належать: аналіз доцільності використання методу Frequent Pattern Growth щодо тривалості розробки програмного забезпечення [1], [8], розробка інформаційної моделі процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення [2], [9], розробка інформаційної технології процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення [3], [11], представлення процесу розробки програмного забезпечення з використанням марковських процесів [4], [13], аналіз доцільності використання методів ІАД при тестуванні програмного забезпечення [5] – [7], аналіз доцільності використання асоціативних правил щодо тривалості розробки програмного забезпечення [10], [12], удосконалення методу пошуку асоціативних правил FP-Growth, за рахунок класифікації даних [14], комп'ютерне моделювання [15].

Апробація матеріалів дисертації. Результати проведеного дослідження були апробовані на 10-ій та 11-ій міжнародних науково-практичних конференціях «Інтернет – Освіта – Наука, Вінниця (2016, 2018), XLV, XLVI, XLVIII науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів ВНТУ, Вінниця (2016, 2017, 2019), IX міжнародній науково-практичній конференції Грузинської математичної асоціації, Тбілісі, Грузія (2018), Всеукраїнській науково-практичній конференції здобувачів вищої освіти й молодих учених «Комп'ютерна інженерія і кібербезпека: досягнення та інновації», Кропивницький (2018).

Публікації. Матеріали досліджень опубліковані в 15 наукових працях: 1 стаття, що опублікована в іноземному фаховому журналі, який входить до наукометричної бази SCOPUS, 4 статті, що опубліковані в наукових виданнях з переліку фахових видань України, з них, в тому числі, 1 стаття у журналі, який входить до наукометричної бази SCOPUS; результати дослідження апробовано та опубліковано у вигляді 9 тез на шістьох науково-технічних конференціях (4 публікації у збірниках праць, 1 публікація у збірнику матеріалів конференції, 3 – у вигляді тез доповідей) та одній науково-практичній конференції; отримано 1 свідоцтво на реєстрацію авторського права на комп'ютерну програму.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Основний зміст роботи викладено на 108 сторінках друкованого тексту, що містить 34 рисунки та 16 таблиць. Список використаних джерел містить 161 найменування. Загальний обсяг роботи становить 164 сторінки.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертації, зазначено її зв'язок з науковими програмами, планами та темами, сформульовано мету та напрямок

досліджень, відзначено наукову новизну та практичне значення отриманих результатів, надано інформацію про їх впровадження, апробацію та публікації.

У першому розділі проведено аналіз засобів, що використовуються для управління процесом розробки програмного забезпечення та зроблено висновок про те, що вони не дозволяють спрогнозувати тривалість виконання окремого завдання і проекту в цілому, використовуючи дані про вже реалізовані завдання. Здійснено аналіз основних моделей процесу розробки програмного забезпечення. Він встановив, що при розробці програмного забезпечення враховується лише пріоритет завдання, в той час як такі характеристики завдання як складність, критичність та тривалість його виконання не беруться до уваги. Проведений аналіз також показав, що основні етапи (аналіз, проектування, програмування, тестування та виправлення дефектів, документування) процесу розробки програмного забезпечення є однаковими для розглянутих моделей і тривалість їх реалізації може бути прогнозованою за рахунок використання методів інтелектуального аналізу даних.

Проведений аналіз методів інтелектуального аналізу даних щодо розробки програмного забезпечення довів доцільність пошуку асоціативних правил щодо визначення тривалості виконання процесу розробки програмного забезпечення. Це дозволить автоматизувати процес розробки програмного забезпечення за рахунок визначення тривалості виконання як певного завдання, так і проекту в цілому.

Аналіз доцільності пошуку і використання асоціативних правил щодо тривалості розробки програмного забезпечення продемонстрував необхідність використання знайдених залежностей під час розробки програмного забезпечення. Також проведено аналіз методів визначення частих предметних наборів для пошуку таких асоціативних правил, що підтвердив доцільність використання методу визначення частих предметних наборів FP-Growth. Проте він потребує удосконалення з метою виправлення існуючих недоліків та його адаптації до предметної області, що розглядається. Удосконалення полягає у класифікації завдань перед пошуком асоціативних правил, що дозволить підвищити точність визначення тривалості виконання завдання розробником, а також скоротити тривалість пошуку асоціативних правил, які для цього використовуються. Аналіз методів класифікації даних виявив необхідність застосування методу дерев рішень. На основі проведеного аналізу визначено основні задачі дисертаційного дослідження.

У другому розділі запропоновано інформаційну модель процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Множина завдань, що виконані під час розробки програмного забезпечення T :

$$T = \{T_1, T_2, \dots, T_i, \dots, T_n\}, \quad (1)$$

де T_i – це i -е виконане завдання, $i = \overline{1, n}$, n – потужність множини T .

i -е завдання T_i представлено виразом:

$$T_i = \langle Tp, Pr, Sv, Cm, Cx, t, Dl, Sx, Rl \rangle, \quad (2)$$

де Tp – це тип i -го завдання; Pr – це пріоритет (ступінь важливості) i -го завдання; Sv – це критичність i -го завдання; Cm – це компонент розроблюваного ПЗ; Cx – це складність завдання, яке потрібно виконати під час розробки ПЗ; t – це час, необхідний для виконання завдання; Dl – рівень кваліфікації розробника, що виконав i -е завдання; Sx – стать розробника, що виконав i -е завдання; Rl – номер реліз, під час якого було виконано i -е завдання.

Множина розробників D , які виконують завдання:

$$D = \{D_1, D_2, \dots, D_j, \dots, D_m\}, \quad (3)$$

де D_j – це j -й розробник, $j = \overline{1, m}$, m – потужність множини D .

Інформаційна модель AR процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, відображена у виразі 4:

$$AR = \langle T, D, minsupp, minconf, \mu, \Omega, NT, PT \rangle \quad (4)$$

де T – множина завдань, що були виконані при розробці програмного забезпечення; D – множина розробників, які виконують завдання; $minsupp$ – мінімальне значення підтримки асоціативних правил; $minconf$ – мінімальне значення достовірності асоціативних правил; μ – метод, за допомогою якого буде здійснюватися пошук асоціативних правил; Ω – множина знайдених асоціативних правил; NT – множина завдань, для яких необхідно визначити тривалість їх реалізації розробником певного рівня кваліфікації; PT – множина завдань з визначеним часом, який необхідний на їх виконання розробником, що має певну кваліфікацію.

Оскільки не всі характеристики завдання однаково впливають на тривалість його розробки, було використано багатofакторний кореляційний аналіз (коефіцієнт Чупрова) для визначення ступеня зв'язку між

характеристиками завдань, що виконуються під час розробки програмного забезпечення, та тривалістю їх розробки. Визначено множину характеристик завдань, серед яких здійснюється пошук асоціативних правил щодо тривалості розробки програмного забезпечення: рівень кваліфікації розробника; тип, критичність, складність та пріоритет завдання; компонент ПЗ, що розробляється.

Запропонована інформаційна модель процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення розділена на два окремих блоки (рис. 1):

1. У блоці пошуку асоціативних правил щодо тривалості розробки програмного забезпечення представлено процес пошуку асоціативних правил у БД завдань, що виконувалися розробниками.

2. У блоці використання асоціативних правил представлено процес використання знайдених асоціативних правил щодо тривалості розробки ПЗ. Вони застосовуються для визначення часу, необхідного на виконання завдання конкретним розробником.

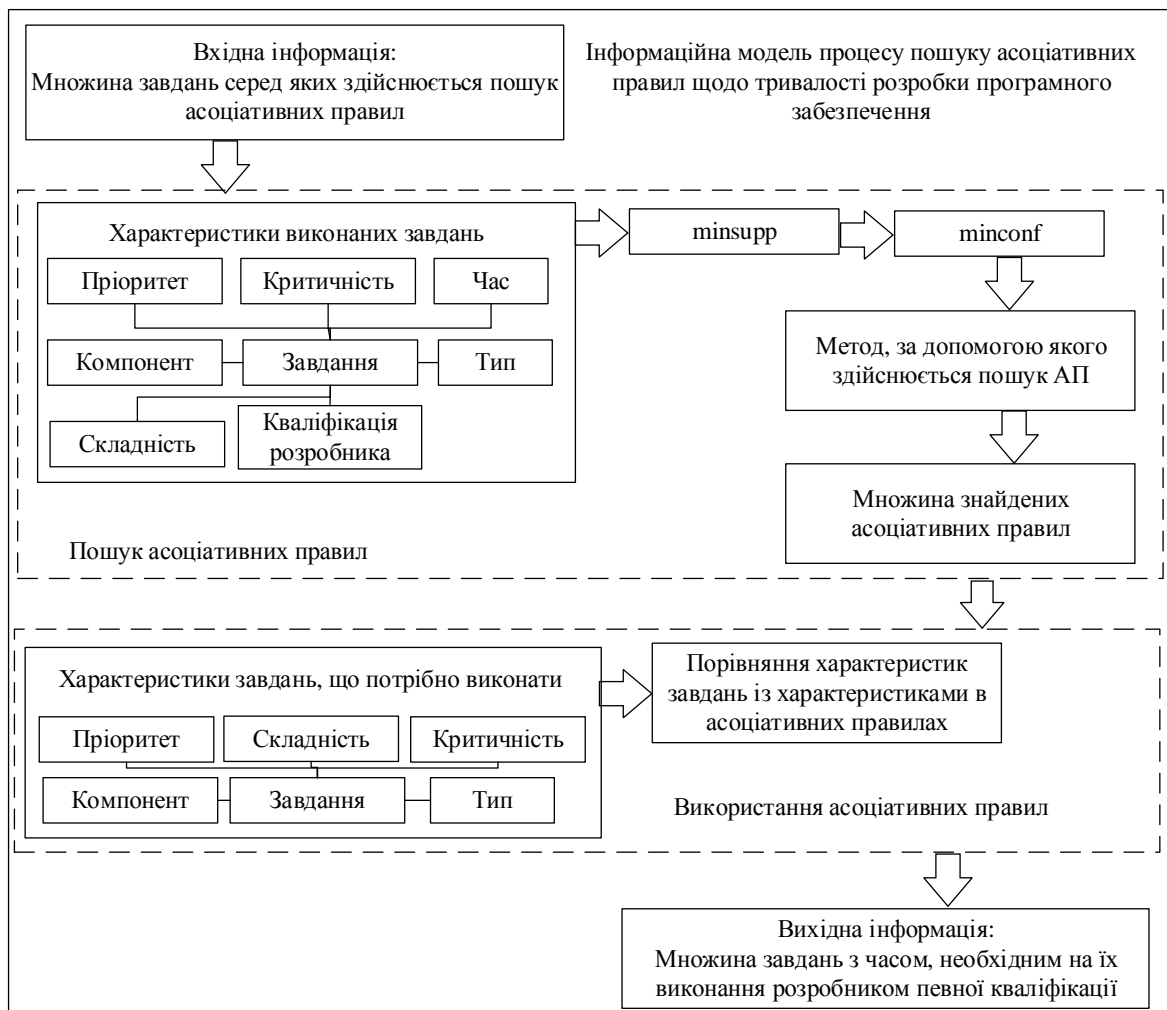


Рисунок 1 – Інформаційна модель процесу пошуку асоціативних правил щодо розробки програмного забезпечення

Удосконалено існуючий метод FP-Growth визначення частих предметних наборів для пошуку асоціативних правил за рахунок попередньої класифікації усієї множини завдань на три групи: завдання категорії 1 (завдання, які легко реалізувати), завдання категорії 3 (завдання, які важко реалізувати), завдання категорії 2 (завдання між легкими і важкими).

У формалізованому вигляді удосконалений метод FP-Growth містить такі основні кроки :

1. Задається значення достовірності *minconf*, відносно якого буде визначатися чи знайдене асоціативне правило корисне для експерта чи ні.

2. Відбувається сканування БД, а усі завдання, серед яких здійснюватиметься пошук АП, класифікуються на три класи.

4. Для кожного класу завдань визначається значення *minsupp*.

5. Здійснюється перетворення завдань у кожному класі у деревовидну структуру, що називається FP-дерево.

6. Відбувається визначення частих предметних наборів у FP-дереві з урахуванням обрахованого значення підтримки *supp*.

7. Здійснюється пошук асоціативних правил зі знайдених предметних наборів.

8. Обраховується значення достовірності *conf* для кожного асоціативного правила.

9. Відбувається запис усіх асоціативних правил, значення достовірності *conf* яких більша чи рівна значенню *minconf*, до множини асоціативних правил.

Оскільки обробка класів відбувається паралельно, то кроки 3-8 даного методу виконуються одночасно для трьох класів завдань.

Відповідно до кількості завдань, які виконані кожним розробником під час розробки ПЗ, характеристики, відносно яких здійснюється класифікація завдань, відрізняються:

– кількість виконаних завдань розробником до 50 – комплексна величина $CS (Sv + Cx)$, критичність завдання та тривалість його виконання;

– кількість виконаних завдань розробником після 50 – пріоритет завдання, його критичність та тривалість його виконання.

Це дозволило підвищити точність визначення тривалості виконання завдання під час розробки ПЗ, а також скоротити час пошуку асоціативних правил щодо тривалості розробки ПЗ.

Запропоновано алгоритм використання асоціативних правил для визначення часу, необхідного для виконання завдання розробником (рис. 2), де множина UT – це множина завдань, для яких тривалість розробки не визначено.

У **третьому розділі** розроблено інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, основними кроками якої є:

1. Завантаження вхідних даних, необхідних для навчання алгоритму

класифікації С 4.5.

2. Процес навчання алгоритму класифікації С 4.5 для подальшого його використання при класифікації завдань, серед яких необхідно знайти асоціативні правила.

3. Завантаження вхідних даних, що стосуються завдань, які було виконано під час розробки ПЗ і серед яких буде здійснюватися пошук асоціативних правил.

4. Процес класифікації вхідної множини завдань (серед яких необхідно знайти асоціативні правила) на три класи згідно їх характеристик.

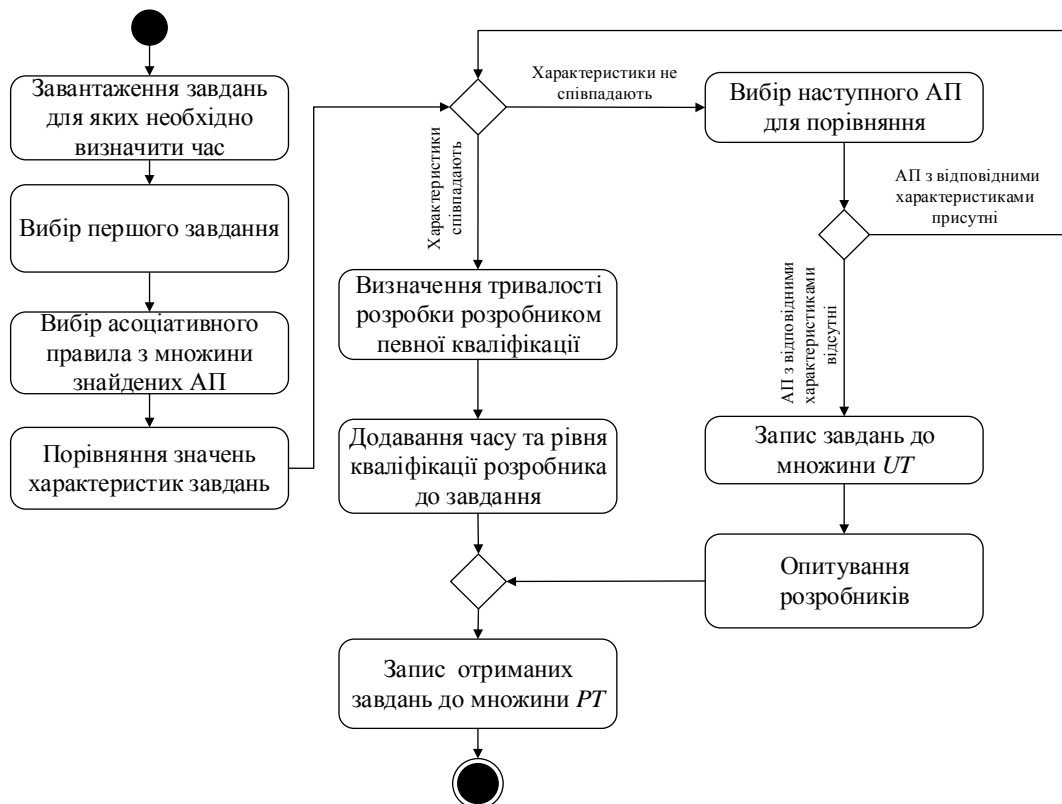


Рисунок 2 – UML-діаграма діяльності процесу використання асоціативних правил щодо тривалості розробки програмного забезпечення

5. Внесення значення мінімальної достовірності *minconf* експертом, в ролі якого виступає менеджер проекту, та обрахунок значення мінімальної підтримки для кожної групи завдань *minsupp*.

6. Для кожної групи завдань виконуються пункти 7 – 11.

7. Перетворення завдань у кожному із класів на деревовидну структуру даних FP-дерево.

8. Процес визначення частих предметних наборів у кожному із класів та підрахунок значення підтримки *supp* для кожного зі знайдених частих предметних наборів.

9. Відбір частих предметних наборів, у яких значення підтримки *supp* більше або дорівнює значенню мінімальної підтримки *minsupp*, заданого

експертом.

10. Процес пошуку асоціативних правил, серед відібраних частих предметних наборів та обрахунок значення достовірності *conf* для кожного із них.

11. Відбір асоціативних правил для яких значення підтримки *conf* більше або дорівнює значенню мінімальної підтримки *minconf*. Це необхідно зробити для визначення асоціативних правил, що є корисними для експерта, в ролі якого виступає менеджер проекту.

12. Збереження відібраних з кожного класу асоціативних правил для їх подальшого використання при визначенні тривалості виконання завдання розробником певної кваліфікації.

13. Процес використання знайдених асоціативних правил щодо тривалості розробки програмного забезпечення з метою визначення часу, необхідного на виконання завдання розробником певної кваліфікації.

Схему розробленої інформаційної технології пошуку асоціативних правил щодо тривалості розробки програмного забезпечення наведено на рис. 3:

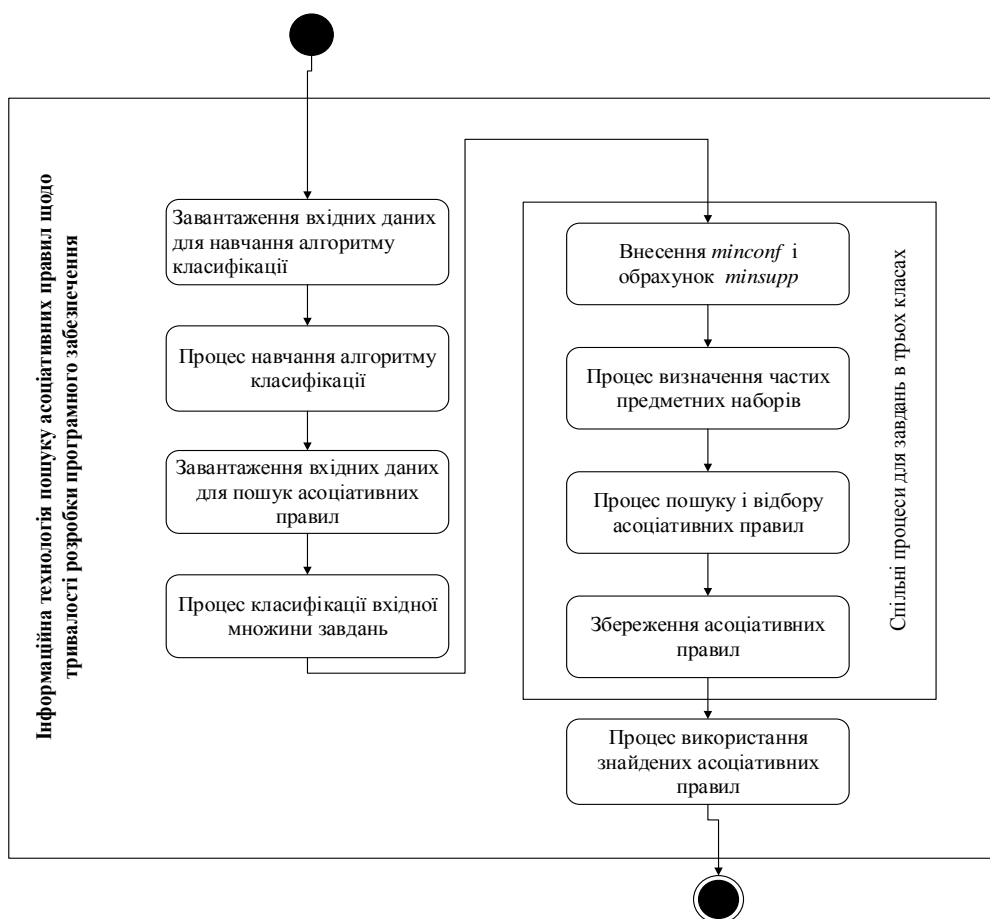


Рисунок 3 – Інформаційна технологія пошуку асоціативних правил щодо тривалості розробки програмного забезпечення

Здійснено представлення процесу розробки програмного забезпечення на

трьох рівнях деталізації (рис. 4) кожен з яких можна відобразити за допомогою єдиного математичного апарату – марковських ланцюгів з дискретними станами та дискретним часом.

Представлення процесу розробки програмного забезпечення *PCounter* з використанням марковських ланцюгів було здійснено за таким алгоритмом:

1. Визначено основні компоненти з яких складається дане програмне забезпечення, а саме: Reports, Counter, Routs, Admin, Settings, GPS, Archive, Users.

2. Для кожного компонента визначено основні етапи його розробки.

3. Для кожного етапу визначено основні кроки з яких він складається.

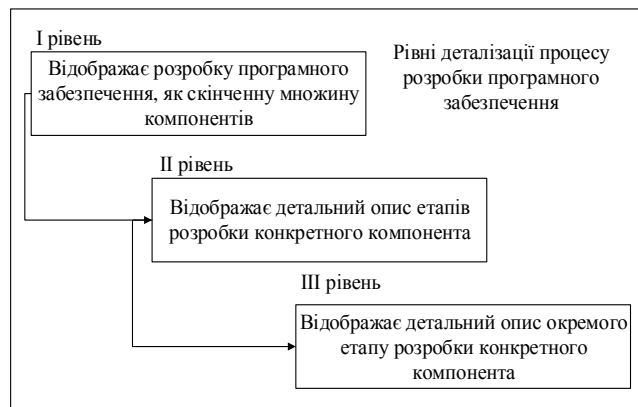


Рисунок 4 – Рівні деталізації розробки компонентів програмного забезпечення

На кожному із кроків алгоритму відбувався підрахунок компонентів, етапів та кроків відповідно. В результаті проведених досліджень виявлено такі закономірності:

1. При аналізі процесу розробки програмного забезпечення на першому рівні деталізації було виявлено 8 основних компонентів із яких складається комплексне програмне забезпечення *PCounter*. Кількість етапів та кроків процесу розробки вказаного програмного забезпечення на даному рівні не визначена.

2. Аналіз кожного окремого із 8 виявлених компонентів на другому рівні деталізації виявив 4 основних необхідних етапи для його реалізації. Кількість кроків процесу розробки вказаного програмного забезпечення на даному рівні не визначена.

3. При аналізі кожного етапу на третьому рівні деталізації було встановлено, що етап аналізу, проектування та документування виконується за 1 крок. Етап програмування та тестування передбачає такі основні 3 кроки: програмування, розробка і виконання тестів, оцінка результатів тестування. Таким чином, процес розробки одного компоненту програмного продукту *PCounter* на третьому рівні деталізації визначається в 6 кроків.

Представлення процесу розробки ПЗ показало, що із підвищенням рівня деталізації виникає необхідність пошуку асоціативних правил щодо тривалості процесу розробки ПЗ. Асоціативні залежності між розробником, що виконує завдання, та часом, необхідним для вирішення поставлених завдань, доцільно шукати лише на рівнях, де процес розробки ПЗ розглядається для окремого компонента а не системи в цілому, тому їх доцільно шукати на другому та третьому рівнях деталізації.

Здійснено моделювання процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення з використанням удосконаленого методу FP-Growth за допомогою пакету прикладних програм MATLAB. У табл. 1 наведено результати експериментальних досліджень з використанням відомого методу пошуку АП FP-Growth та удосконаленого методу.

Таблиця 1 – Результати моделювання процесу пошуку асоціативних правил з використанням існуючого методу FP-Growth та удосконаленого

Потужність вибірки (елементів)	Метод	
	FP-Growth	Удосконалений FP-Growth
	Тривалість пошуку (сек)	
100	1.988	1.981
250	3.058	3.021
500	6.258	6.184
750	9.654	9.526
1000	12.754	12.502

Визначено, що швидкість пошуку АП з використанням удосконаленого методу FP-Growth при кількості елементів: до 100 – зросла на 0,001 сек (0,35%), до 250 – зросла на 0,037 сек (1,2%), до 500 – зросла на 0,074 сек (1,2%), до 750 – зросла на 0,128 сек (1,3%), до 1000 – зросла на 0,252 сек (2%).

Проведено дослідження щодо залежності між кількістю реалізованих завдань під час розробки ПЗ та кількістю знайдених серед них АП щодо тривалості розробки ПЗ та точністю визначення тривалості виконання завдань з використанням таких правил.

Як видно з рис. 5 зі збільшенням кількості реалізованих завдань, що зберігаються в БД, збільшується кількість знайдених асоціативних правил. Так серед 50 завдань було знайдено 14 АП, а серед 1000 завдань – 153 правила, а це означає, що при збільшенні кількості завдань в 20 разів кількість знайдених АП зросла в 10,9 разів відповідно. Пошук асоціативних правил серед 1500 та 2000 збережених завдань показав, що кількість знайдених АП збільшується (158 та 161 знайдених правила відповідно). Але, оскільки, різниця у 8 знайдених правил при збільшенні кількості завдань на 1000 одиниць не є суттєвою, то

можна зробити висновок про недоцільність пошуку АП при кількості виконаних завдань більше 1000.

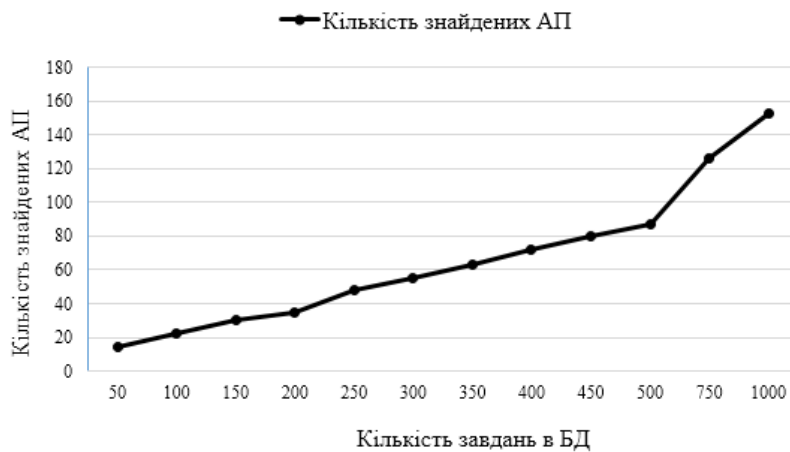


Рисунок 5 – Графік зміни кількості знайдених асоціативних правил щодо тривалості процесу розробки ПЗ від кількості завдань у базі даних

В свою чергу зі збільшенням кількості асоціативних правил збільшується точність визначення тривалості виконання завдань (характеризується похибкою, що розраховується відносно фактичного часу, витраченого розробником на виконання завдання): при знайдених 14 асоціативних правилах похибка визначення тривалості виконання завдання розробником становила 12%, а при збільшенні кількості правил до 153 точність зростає в 2,3 рази (значення похибки – 5,2%), що дозволило менеджерам здійснювати ефективне планування процесу розробки програмного забезпечення. Оскільки пошук асоціативних правил не доцільно здійснювати при кількості завдань більше 1000, то мінімальне значення похибки, яку можна досягнути для визначення точності визначення тривалості виконання завдань, – 5,2%. Результати наведено на рис. 6.

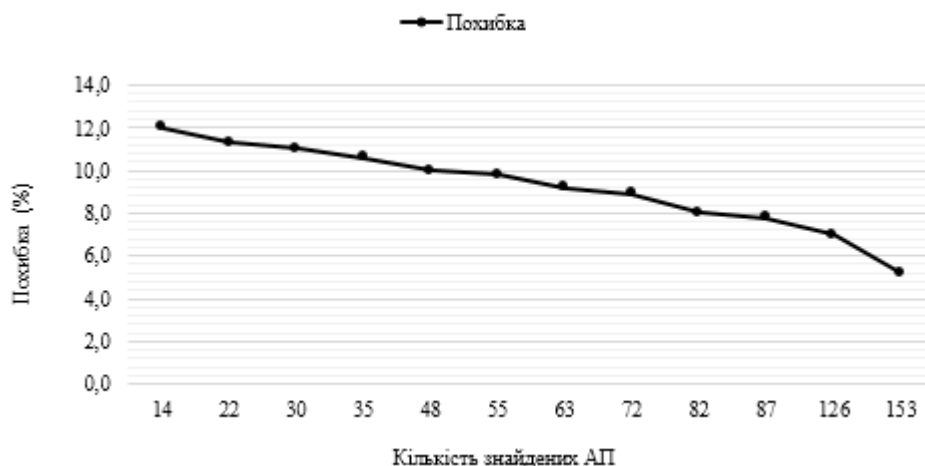


Рисунок 6 – Графік зміни похибки визначення тривалості виконання завдань від кількості знайдених асоціативних правил щодо тривалості процесу розробки ПЗ

Промодельовано процес використання АП щодо тривалості розробки ПЗ і, як видно із графіка, наведеного на рис. 7, тривалість визначення необхідного часу на виконання завдання розробником певної кваліфікації скоротилася в 5 разів (із 200 хв при стандартному підході до 40 хв – при запропонованому) для випадку, коли кількість завдань, для яких необхідно визначити час, дорівнює 20.

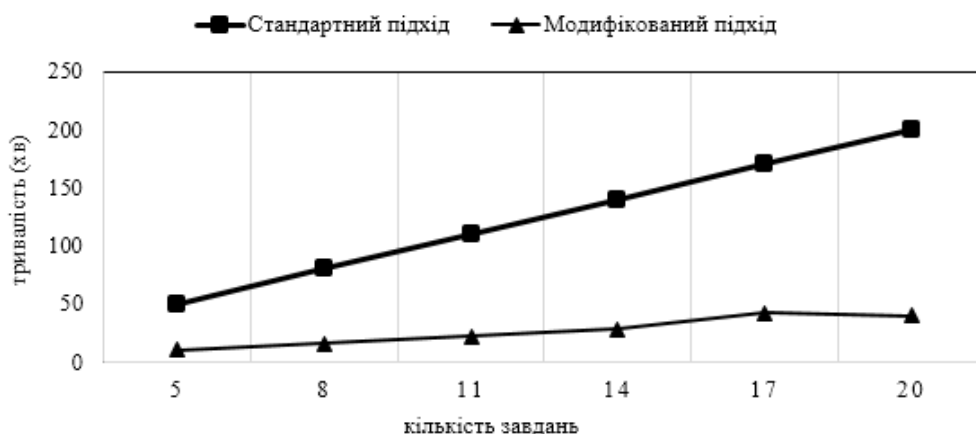


Рисунок 7 – Графік зміни тривалості визначення часу, необхідного на виконання завдань, із використанням знайдених асоціативних правил щодо тривалості процесу розробки ПЗ

У четвертому розділі розроблено структуру інформаційної системи пошуку АП щодо тривалості розробки ПЗ. При її проектуванні необхідно застосувати архітектурний шаблон Model-View-Controller для розмежування алгоритмічної частини від інтерфейсу та забезпечення спільного використання компонентів. Клієнтська частина побудована з використанням набору інструментів JavaFX. Дані, що передаються з клієнта на сервер, захищені від зовнішнього втручання та сторонньої модифікації за рахунок зберігання у закодованому вигляді.

Для розробки інформаційної системи пошуку асоціативних правил щодо тривалості розробки ПЗ використано мову програмування Java та середовище програмування NetBeans IDE 8.2. Встановлено, що реляційна модель даних є оптимальною для проектування бази даних інформаційної системи пошуку асоціативних правил щодо тривалості розробки ПЗ, а СУБД MS SQL SERVER доцільно застосовувати для маніпулювання даними з метою пошуку асоціативних правил щодо тривалості розробки ПЗ.

Запропоновану інформаційну систему пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, в основі якої лежить розроблена інформаційна технологія відповідного процесу, було впроваджено для роботи в компанії ТОВ «Бізнес-Програми» у вигляді програмного забезпечення для пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Результати оцінювання точності визначення тривалості виконання завдання з використанням розробленої інформаційної технології, для програм BAS та FREDO наведено у табл. 2.

Як видно з табл. 2 зі збільшенням кількості завдань в БД, серед яких здійснюється пошук АП, збільшується кількість асоціативних залежностей щодо тривалості розробки ПЗ.

Таблиця 2 – Результати оцінювання точності визначення тривалості виконання завдання з використанням розробленої інформаційної технології

Програмне забезпечення	Кількість завдань в БД	Кількість знайдених АП	Точність визначення (%)
BAS	100	18	9,3
	200	40	8,6
	300	51	7,8
	400	67	7
	500	90	6,4
	600	105	5,5
FREDO	100	21	8,9
	200	39	8,6
	300	52	7,7
	400	67	6,5
	500	91	6,1
	600	101	5,4

Таке зростання дозволило зменшити похибку визначення тривалості виконання завдань:

а) для програмного забезпечення BAS в 1,7 рази (з 9,3% при 18 знайдених АП до 5,5% при знайдених 105 АП);

б) для програмного забезпечення FREDO в 1,6 рази (з 8,93% при 21 знайдених АП до 5,4% при знайдених 101 АП).

Результати використання розробленої інформаційної технології пошуку асоціативних правил щодо тривалості розробки програмного забезпечення з метою визначення часу, необхідного на виконання завдання розробником певної кваліфікації, наведено у табл. 3. Для цього було розраховано відношення *Dur* тривалості визначення часу, необхідного для розробки ПЗ, з використанням стандартного підходу, до тривалості визначення часу, необхідного для розробки ПЗ, із використанням запропонованого підходу.

Визначення часу, необхідного на виконання завдання розробником певної кваліфікації, скоротилось від 2 до 6 разів у випадку застосування запропонованого підходу використання знайдених асоціативних правил щодо розробки програмного забезпечення: для програмного забезпечення BAS

тривалість визначення часу скоротилася від 2,75 до 6 разів, а для ПЗ FREDO – від 2 до 6 разів.

На рис. 8 наведено графік зміни швидкості визначення тривалості виконання завдання розробником певної кваліфікації для ПЗ BAS (а) та ПЗ FREDO (б).

Таблиця 3 – Результати використання розробленої інформаційної технології пошуку асоціативних правил щодо тривалості розробки програмного забезпечення

Програмне забезпечення	Кількість завдань	Тривалість визначення часу, необхідного для розробки ПЗ (хв)		Dur
		Стандартний підхід	Запропонований підхід	
BAS	10	60	10	6
	12	110	40	2,75
	8	60	15	4
	11	110	25	4,4
	15	140	40	3,5
	10	30	10	3
FREDO	5	60	10	6
	7	60	10	6
	10	120	40	3
	8	90	15	6
	6	60	10	6
	7	40	20	2

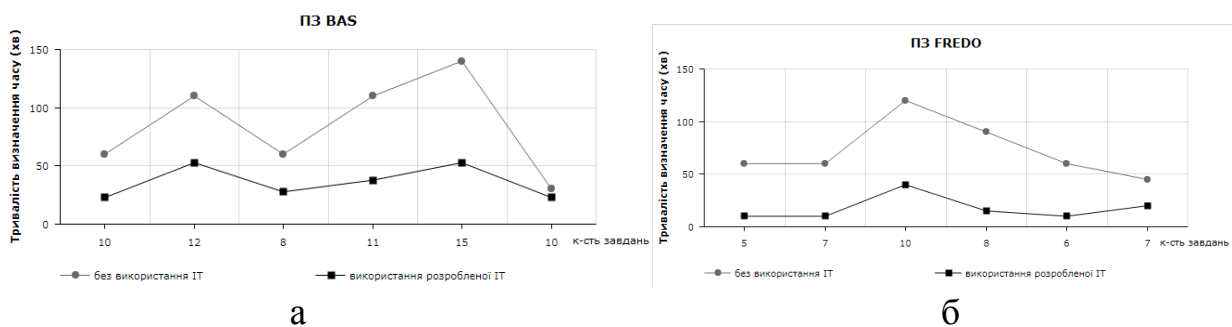


Рисунок 8 – Графік залежності тривалості визначення часу, необхідного на виконання завдання розробником певної кваліфікації для ПЗ BAS та ПЗ FREDO

ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ

1. Проаналізувавши сучасні засоби, що використовуються для планування та управління процесом розробки програмного забезпечення, було встановлено, що за допомогою них не можна прогнозувати тривалість виконання завдання чи проекту в цілому і, тому необхідно розробити інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного

забезпечення, яка б дозволила збільшити точність визначення тривалості виконання завдання розробником певної кваліфікації та скоротити тривалість пошуку асоціативних правил щодо тривалості розробки програмного забезпечення.

Аналіз сучасних моделей процесу розробки програмного забезпечення показав, що у них враховується лише така характеристика завдань як пріоритет. А тому існує необхідність у розробці інформаційної моделі, яка, на відміну від існуючих, враховувала б такі характеристики завдань (завдань, які потрібно вирішити при розробці програмного забезпечення) як: складність, пріоритет, критичність та тривалість їх виконання. Кожна із розглянутих моделей складається із основних етапів (вимоги до документації, проектування, аналіз, кодування, перевірка та тестування), тривалість яких можна визначити з використанням відповідних методів інтелектуального аналізу даних.

Розглянувши стан розвитку процесу розробки програмного забезпечення, було встановлено, що сучасні методи інтелектуального аналізу даних широко використовуються під час даного процесу. Тому варто застосовувати метод визначення частих предметних наборів для пошуку асоціативних правил FP-Growth щодо тривалості розробки програмного забезпечення. Даний метод потрібно удосконалити з метою виправлення існуючих недоліків та адаптації до відповідної предметної області.

2. Для визначення характеристик завдань, що впливають на тривалість розробки завдань розраховано коефіцієнт Чупрова і на основі отриманих даних розроблено інформаційну модель процесу пошуку асоціативних правил щодо тривалості розробки програмного забезпечення. Встановлено, що тип, складність, пріоритет та критичність завдання, рівень кваліфікації розробника, а також компонент програмного забезпечення, що розробляється впливають на тривалість розробки завдання. Запропонована інформаційна модель, на відміну від існуючих, враховує наступні характеристики завдань: складність, пріоритет, критичність та тривалість їх виконання. Для розробленої інформаційної моделі вхідними даними є множина завдань, серед таких завдань буде здійснюватися пошук асоціативних правил з використанням удосконаленого методу FP-Growth; вихідними даними є множина завдань з визначеним часом, необхідним на їх реалізацію розробником певної кваліфікації. Час реалізації визначається за рахунок застосування знайдених асоціативних правил.

3. Удосконалено метод визначення частих предметних наборів для пошуку асоціативних правил FP-Growth, який, на відміну від існуючих, передбачає етап класифікації завдань на три класи, залежно від характеристик завдань. Розроблено алгоритм використання таких асоціативних правил для визначення часу, необхідного на виконання завдання розробником з певним рівнем кваліфікації, що допоможе менеджерам проектів при плануванні та управлінні процесом розробки програмного забезпечення.

4. Розроблено інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного забезпечення в основі якої, на відміну від існуючих, лежить удосконалений метод визначення частих предметних наборів для пошуку асоціативних правил FP-Growth та проведено моделювання такого процесу. Результати моделювання показали, що точність визначення тривалості виконання завдань зросла у 2,3 рази при збільшенні кількості знайдених асоціативних правил у 11 разів; тривалість пошуку асоціативних правил скоротилася на 2%, якщо кількість завдань понад 750; кількість знайдених асоціативних правил зростає в 10,9 разів, у випадку збільшення кількості завдань у базі даних в 20 разів.

5. На основі запропонованої інформаційної технології розроблено інформаційну систему пошуку асоціативних правил щодо тривалості розробки програмного забезпечення. Аналіз результатів впровадження запропонованої інформаційної технології та інформаційної системи у ТОВ «Бізнес-Програми», показав що: точність визначення тривалості виконання завдання розробником певної кваліфікації зросла в 1,7 рази за рахунок використання знайдених асоціативних правил; швидкість визначення часу, необхідного на виконання завдання розробником певної кваліфікації, підвищилась до 6 разів в залежності від кількості завдань, для яких цей час визначався. Це дало можливість приймати ефективні управлінські рішення при плануванні та керуванні процесом розробки програмного забезпечення.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

[1] Т. О. Савчук, та Н. В. Приймак, “Обґрунтування доцільності використання методу fp-growth (fpg) при розробці програмного забезпечення”, *Вчені записки ТНУ імені В.І. Вернадського. Серія Технічні науки*, № 5, с. 41 – 45, 2018.

[2] Т. О. Савчук, та Н. В. Приймак, “Розробка інформаційної моделі процесу пошуку асоціативних правил при розробці програмного забезпечення”, *Інформаційні технології та комп’ютерна інженерія*, vol 42, № 2, с. 43 – 48, 2018.

[3] Т. О. Савчук, та Н. В. Приймак, “Розробка інформаційної технології процесу пошуку асоціативних правил при розробці програмного забезпечення”, *Наукові праці : наук. журн. / Чорном. нац. ун-т ім. П. Могили. Комп’ютерні технології*, вип. 305, т. 317, с. 6 – 11, 2018.

[4] Т. О. Savchuk, and N. V. Pryimak, “Modeling of software development process with the Markov processes”, *Eastern-European journal of Enterprise Technologies . Information technology*, vol. 3, no. 2, pp. 33 – 38, 2017.

[5] Т. О. Savchuk, N. V. Pryimak, A. Assembay, T. Zyska, M. Junisbekov, and A. Annabaev “The technology of searching the associative rules while

developing the software”, *Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments*, 2017, doi: 10.1117/12.2280900.

[6] Т. О. Савчук, Н. В. Приймак, “Використання методів інтелектуального аналізу даних при тестуванні програмного забезпечення”, на *X міжнародній конференції «Інтернет-освіта-наука-2016»*, Вінниця, 2016, с. 28 – 29.

[7] Т. О. Савчук, та Н. В. Приймак, “Використання методів штучного інтелекту для прогнозування часу для виправлення дефектів розробки програмного забезпечення”, на *X міжнародній конференції «Інтернет-освіта-наука-2016»*, Вінниця, 2016, с. 30 – 31.

[8] Т. О. Савчук, та Н. В. Приймак, “Обґрунтування вибору методу визначення частих предметних наборів для пошуку асоціативних правил при розробці програмного забезпечення”, на *XI міжнародній конференції «Інтернет-освіта-наука-2018»*, Вінниця, 2018, с. 45 – 46.

[9] Т. О. Савчук, та Н. В. Приймак, “Інформаційна модель процесу пошук і використання асоціативних правил при розробці програмного забезпечення”, на *XI міжнародній конференції «Інтернет-освіта-наука-2018»*, Вінниця, 2018, с. 43 – 44.

[10] Т. О. Savchuk, and N. V. Pryimak, “Feasibility study of using associative rules during the software development”, on *IX international Conference of the Georgian Mathematical Union*, Batumi, 2018, pp. 203 – 204.

[11] Т. О. Савчук, та Н. В. Приймак, “Інформаційна технологія пошуку асоціативних правил при розробці програмного забезпечення”, на *Всеукраїнській наук.-практ. конф. здобувачів вищої освіти й молодих учених*, Кропивницький, 2018, с. 100 – 102.

[12] Т. О. Савчук, та Н. В. Приймак, “Використання fpg-алгоритму для пошуку асоціативних правил при прийнятті рішень в управлінні процесами”, на *XLV НТК проф.-викл. складу, співробітників та студентів університету*, Вінниця, 2016, с. 929 – 930.

[13] Т. О. Савчук, та Н. В. Приймак, “Використання Марковських ланцюгів при моделюванні процесу розробки програмного забезпечення”, на *XLVI НТК проф.-викл. складу, співробітників та студентів університету*, Вінниця, 2017, с. 1019 – 1020.

[14] Т. О. Савчук, та Н. В. Приймак, “Метод пошуку асоціативних правил при розробці програмного забезпечення, що базується на попередньому розбитті множини завдань” на *XLVIII НТК проф.-викл. складу, співробітників та студентів університету*, Вінниця, 2019, с. 655 – 656.

[15] Т. О. Савчук, Н. В. Приймак, “Інформаційна система пошуку асоціативних правил при розробці програмного забезпечення”, *Свідоцтво про реєстрацію авторського права на твір №82173*; Жов. 11, 2018.

АНОТАЦІЯ

***Приймак Н.В.* Інформаційна технологія пошуку асоціативних правил щодо тривалості розробки програмного забезпечення. – Кваліфікаційна наукова праця на правах рукопису.**

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 «Інформаційні технології» – Вінницький національний технічний університет, Вінниця, 2020.

Дисертаційна робота присвячена підвищенню точності визначення тривалості виконання завдання розробником, а також скороченню тривалості пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, які для цього використовуються. Це досягається за рахунок використання відповідної інформаційної технології.

Удосконалено метод визначення частих предметних наборів FP-Growth, що використовуються для пошуку асоціативних правил, за рахунок класифікації вхідної множини завдань на три класи, відповідно до їх характеристик: завдання категорії 1 (завдання, які легко реалізувати), завдання категорії 3 (завдання, які важко реалізувати), завдання категорії 2 (завдання, по складності між легкими і важкими). Таке удосконалення підвищило точність визначення часу, необхідного на виконання завдання з використанням знайдених асоціативних правил, а також скоротило тривалість пошуку таких залежностей. Розроблено інформаційну технологію пошуку асоціативних правил щодо тривалості розробки програмного забезпечення, що базується на запропонованій інформаційній моделі.

Ключові слова: програмне забезпечення, процес розробки програмного забезпечення, інформаційна технологія, інформаційна модель, асоціативні правила, метод FP-growth, класифікація, дерева рішень, інтелектуальний аналіз даних.

АННОТАЦИЯ

***Приймак Н.В.* Информационная технология поиска ассоциативных правил относительно продолжительности разработки программного обеспечения – Квалификационный научный труд на правах рукописи.**

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 «Информационные технологии». – Винницкий национальный технический университет, Винница, 2020.

Диссертационная работа посвящена повышению точности определения продолжительности выполнения задания разработчиком, а также сокращению

продолжительности поиска ассоциативных правил относительно продолжительности разработки программного обеспечения, которые для этого используются. Это достигается за счет использования соответствующей информационной технологии.

На основе анализа состояния развития процесса разработки программного обеспечения, было установлено, что современные методы интеллектуального анализа данных широко используются при выполнении данного процесса. Поэтому следует применять метод определения частых предметных наборов для поиска ассоциативных правил FP-Growth относительно продолжительности разработки программного обеспечения. Данный метод необходимо усовершенствовать с целью исправления существующих недостатков и адаптации к соответствующей предметной области.

Разработана информационная модель процесса поиска ассоциативных правил относительно продолжительности разработки программного обеспечения после расчёта коэффициент Чупрова для определения характеристик задач, влияющих на продолжительность разработки задач. Такая информационная модель, в отличие от существующих, учитывает следующие характеристики: тип, сложность, приоритет и критичность задачи, уровень квалификации разработчика, а также компонент программного обеспечения, который разрабатывается.

Усовершенствован метод определения частых предметных наборов для поиска ассоциативных правил FP-Growth, который, в отличие от существующих, предусматривает этап классификации задач на три класса, в зависимости от их характеристик.

Разработан алгоритм использования таких ассоциативных правил для определения времени, необходимого на выполнение задания разработчиком с определенным уровнем квалификации, что будет учтено менеджерами проектов при планировании и управлении процессом разработки программного обеспечения.

Разработана информационная технология поиска ассоциативных правил продолжительности разработки программного обеспечения в основе которой, в отличие от существующих, лежит усовершенствованный метод определения частых предметных наборов для поиска ассоциативных правил FP-Growth и проведено моделирование такого процесса. Результаты моделирования показали, что точность определения продолжительности выполнения заданий выросла в 2,3 раза при увеличении количества найденных ассоциативных правил в 11 раз; продолжительность поиска ассоциативных правил сократилась на 2%, если количество задач более 750; количество найденных ассоциативных правил растет в 10,9 раз, в случае увеличения количества задач в базе данных в 20 раз.

На основе предложенной информационной технологии разработана информационная система поиска ассоциативных правил продолжительности разработки программного обеспечения. Анализ результатов внедрения данной информационной системы в компании ООО «Бизнес-программы» показал, что: точность определения продолжительности выполнения задания разработчиком определенной квалификации выросла в 1,7 раза за счет использования найденных ассоциативных правил; скорость определения времени, необходимого на выполнение задания разработчиком определенной квалификации, повысилась до 6 раз в зависимости от количества задач, для которых это время определялось.

Ключевые слова: программное обеспечение, процесс разработки программного обеспечения, информационная технология, информационная модель, ассоциативные правила, метод FP-growth, классификация, деревья решений, интеллектуальный анализ данных.

ABSTRACT

Pryimak N.V. The information technology of the associative rules search for the duration of software development – Qualification research paper, manuscript copyright.

Thesis for the degree of a candidate of technical science in specialty 05.13.06 “Information technology” – Vinnytsia National Technical University, Vinnytsia, 2020.

The thesis is dedicated to the study of increasing the accuracy of determining the duration of the task development, as well as of reducing the duration of the associative rules search which are used for this. This is achieved through the usage of appropriate information technology.

An information model of the associative rules search for the duration of software development has been suggested. It is based on improved FP-Growth method, which is used to find associative rules by classifying the input set of tasks into three classes according to their characteristics: category 1 (tasks easy to accomplish), category 3 (tasks difficult to accomplish), category 2 (tasks that are between easy and difficult). This improvement has increased the accuracy of determining the time required to complete a task using the found associative rules, as well as to reduce the duration of the search for such dependencies. Information technology of the associative rules search for the duration of software development has been developed, based on the proposed information model of this process.

Keywords: software, the process of software development, information technology, information model, associative rules, method FP-growth, classification, decision trees, data mining.

Підписано до друку 26.02.2020 р. Формат 29.7×42 ¼
Наклад 130 прим. Зам. № 2020-040
Віддруковано в інформаційному редакційно-видавничому центрі
Вінницького національного технічного університету
м. Вінниця, Хмельницьке шосе, 95. Тел.: 65-18-06
Свідоцтво суб'єкта видавничої справи
серія ДК №3516 від 01.07.2009 р