

В. А. Андруник¹, Л. Б. Чирун¹, Л. В. Чирун²
Національний університет “Львівська політехніка”,
¹кафедра інформаційних систем та мереж,
²кафедра програмного забезпечення

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ МАТЕРІАЛЬНО-ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ СТРУКТУРНОЇ ОДИНИЦІ НАВЧАЛЬНОГО ЗАКЛАДУ

© Андруник В. А., Чирун Л. Б., Чирун Л. В., 2015

Нові інформаційні, телекомунікаційні технології сприяють оптимізації управління навчальним процесом. Запропоновано структуру побудови ІС-аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу.

Ключові слова: інформаційні технології, інтелектуальний аналіз даних, матеріально-технічне забезпечення.

New information, telecommunication technologies contribute to the optimization in the management of studies. In the article the author suggests the structure of IS analysis of support's guide.

Key words: information technology, data mining, logistical support.

Вступ. Загальна постановка проблеми

Процес переходу до інформаційного суспільства, а також соціально-економічні зміни, що відбуваються в Україні, вимагають суттєвих змін у багатьох сферах діяльності держави [1]. Насамперед це стосується реформування галузі освіти та інновацій в забезпеченні навчального процесу. Тенденція використання інформаційних та телекомунікаційних технологій полягає в створенні систем, які здатні не тільки виконувати інформаційно-аналітичні функції, але й створювати умови для оперативного керівництва розподіленим дистанційним навчальним процесом, бути ефективним середовищем організації й управління процесом навчання та укладання ефективного забезпечення навчального процесу [1]. Це дає можливість зробити універсальним процес планування навчального процесу та підготовки навчально-методичних матеріалів для різних форм навчання та різних структур навчальних програм. Використання інформаційних комунікаційних технологій повинно оптимізувати процеси управління навчальним процесом і створити ефективну організаційну структуру. Забезпечення навчального процесу за допомогою технічних, інформаційних, інтелектуальних та програмних інновацій надзвичайно важливо в контексті управління сучасним розподіленим навчальним процесом [1, 6, 9–11].

Складність потоків даних, обсяги інформації, що передається, в інформаційній інтелектуальній системі аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу (структурної одиниці навчального закладу) вимагає введення в її структуру допоміжних засобів інтелектуальної навігації, опрацювання інформації, каталогізації та планування [6, 9–11]. Більшість задач адміністрування і управління структурною одиницею покладаються на агента-секретаря, якого можна вважати інтелектуальним ядром структурної одиниці навчального закладу. Це не просто утиліта, що спрощує процес інтерактивного діалогу із системою, але й активний учасник її діяльності: інтерфейсний агент, що має індивідуальність і виконує вказівки користувача з питань планування, комунікації і управління [1–4].

Аналіз останніх досліджень та публікацій

Сьогодні на ринку України є небагато програмних продуктів, які можуть підтримувати прийняття рішень у процесі обліку та аналізу матеріально-технічного забезпечення структурної одиниці ВНЗ. Крім того, більшість із них розроблені за кордоном і тому мало застосовні до

українських реалій (слід врахувати особливості законодавчої бази та економічної системи України). Популярні програмні продукти управління навчальним процесом вимагають попередньої підготовки для роботи з ними, і тому рідко застосовуються на Україні. Тому виникла необхідність у написанні такого програмного продукту [1–5, 9–11].

Інтелектуальна інформаційна система аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу є основною інтелектуальною інформаційною компонентою освітньої комплексної інформаційно-комп'ютерної мережі забезпечення навчального процесу. Вона містить систему баз даних, бази знань і комплекс аналітичного опрацювання інформації з метою відпрацювання оптимальних процедур підготовки, підтримки та прийняття управлінських рішень у навчальному закладі або його складовій (інститут, кафедра) [1-4]. У системі виділимо шість функціональних підсистем, зображених на рис.1. Кожна підсистема призначена для розв'язування окремих задач.

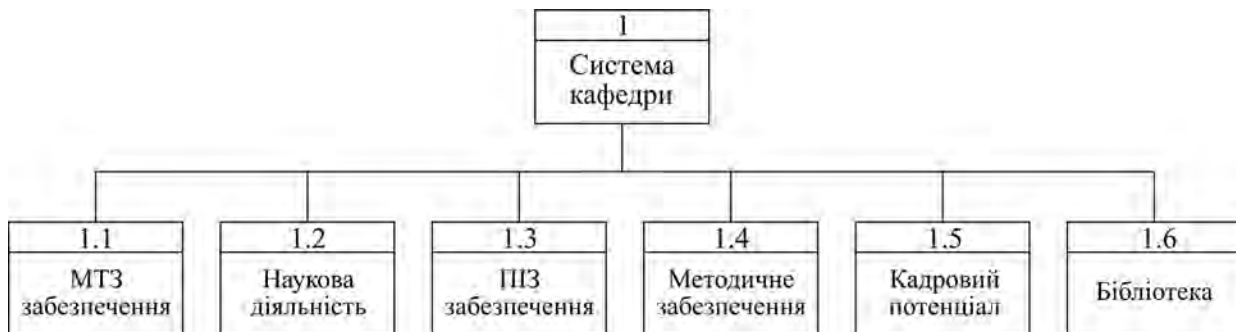


Рис. 1. Діаграма підсистем інформаційної системи структурної одиниці навчального закладу

Підсистема обліку техніки та матеріального забезпечення структурної одиниці навчального закладу призначена для обліку місця розташування техніки, призначення, інвентарних номерів, ремонту, гарантійного обслуговування тощо. Спрощує облік техніки в навчальному закладі та полегшує роботу матеріально відповідальних осіб та інженерів. Предметною областю (ПО) цієї інтелектуальної інформаційної системи є навчально-методичне забезпечення структурної одиниці навчального закладу в складі вищого навчального закладу (ВНЗ) [1, 2].

Інтелектуальний аналіз даних (DM, data mining) є складовою процесу видобування знань з баз даних (KDD, knowledge discovery in data bases), який дає змогу розкрити суть прихованих залежностей в даних, виявити взаємні впливи між властивостями об'єктів, інформація про які зберігається в базах даних, виділити закономірності, властиві певному набору даних [7, 8, 12–21].

Актуальність проблеми дослідження та опрацювання даних підтверджується широким практичним та комерційним використанням систем інтелектуального аналізу. Найчастіше їх застосовують у науковій сфері та бізнесі [1, 9–11]. У загальному випадку процес видобування знань може бути ітеративним і складатися з чотирьох основних кроків.

1. Відбирання даних.
2. Попереднє опрацювання даних.
3. Інтелектуальний аналіз даних.
4. Оцінювання та інтерпретація побудованих моделей та знайдених залежностей.

Послідовність етапів [7, 8] видобування знань зображено на рис. 2.

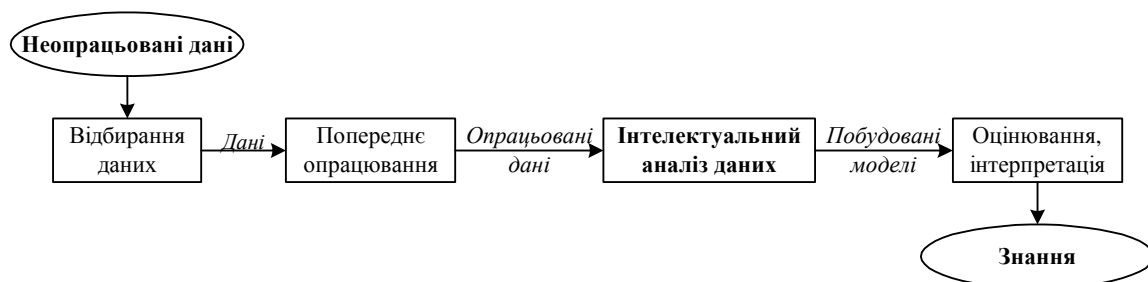


Рис. 2. Загальна схема процесу видобування знань для інтелектуальної інформаційної системи

Об'єкти предметної області описують множинами їхніх властивостей. Найзручніше подавати інформацію про властивості об'єктів таблицями, стовпці яких позначено іменами властивостей, а елементи рядків містять значення властивостей. Рядок таблиці в термінах машинного навчання є *прикладом*, стовпці таблиці називають *атрибутами*. Таблиця є *таблицею прийняття рішень*, якщо у таблиці визначено *атрибути прийняття рішення*, значення якого вказує на належність прикладу до певного класу. Тоді всі атрибути, окрім атрибутів прийняття рішень, називають *умовними атрибутами*. Проблема відсутніх даних у таблицях та невідомих значень атрибутів у прикладах з'являється тоді, коли існують приклади, визначені не на всій множині атрибутів, тобто хоча б одне значення атрибута невідоме. Під невідомим розуміємо таке значення атрибута, визначити яке вже немає можливості, оскільки неможливо повторити умови, у яких було отримано всі інші дані у таблиці. Це значення можна довизначити на основі певних міркувань, формуванню яких присвячені подальші дослідження. Приклади в таблицях прийняття рішень можуть мати невідомі значення як умовних атрибутів, так і атрибута прийняття рішення. Надалі розглядатимемо приклади, у яких можуть бути невідомими лише значення умовних атрибутів. Таблицю з даними називатимемо заповненою, якщо відомі значення всіх атрибутів усіх прикладів. Незаповнена таблиця матиме невідомим принаймні одне значення атрибута принаймні одного прикладу. Таблиця є порожньою, якщо вона не містить даних. Для інтелектуального аналізу таблиць даних, атрибути в яких мають невідомі значення, планується застосувати порівняно новий підхід, який ґрунтується на понятті *наближеної множини* (rough set) [7, 8, 12–21]. Наближені множини – це символічна індуктивна методологія, яка поряд з нейронними мережами, розмитими множинами, генетичними алгоритмами належить до методологій *м'яких обчислень* (soft computing), які успішно застосовують в інтелектуальному аналізі даних та машинному навчанні. Причинами появи [7, 8, 12–21] відсутніх даних у таблиці прийняття рішень можна вважати, по-перше, недбалість осіб, що заповнюють форму замовлення, по-друге, зміну множини атрибутів у процесі збирання даних, а також небажання розголошувати індивідуальні дані.

Приклади, які є описом об'єктів предметних областей та на основі яких доводиться приймати рішення, у своїй більшості містять невідомі значення атрибутів. У разі побудови систем прийняття рішень доводиться враховувати такі дані. Це пов'язано з тим, що у більшості випадків здійснення додаткових досліджень з метою покращення даних неможливе або вартісне. Виділяють [7, 8, 12–21] такі основні групи методів опрацювання таблиць із невідомими значеннями атрибутів:

- ігнорування відсутніх даних;
- видалення прикладів із невідомими значеннями атрибутів;
- доповнення відсутніх даних;
- безпосереднє опрацювання таблиць з відсутніми даними.

Ігнорування відсутніх даних полягає у доповненні домена атрибута значенням, яким позначене невідоме значення цього атрибута.

Видалення прикладів або атрибутів. Використовують два підходи до видалення даних. Перший з них здійснює спеціаліст, який, з огляду на свої знання та досвід, приймає рішення про видалення прикладу чи атрибута. Такий підхід не є алгоритмічним, оскільки прийняття рішення про видалення залежить від досвіду людини-експерта. Другий підхід до видалення прикладів допустимий тоді, коли немає залежності між невідомими значеннями та причиною їх появи [7, 8].

Проте, разом із прикладами з невідомими значеннями атрибутів можна вилучити й такі, що мають суттєві властивості такої таблиці прийняття рішень. Видалення даних може бути повним або частковим. У результаті повного видалення у таблиці залишаться лише заповнені рядки.

Пропорції та обсяги видалення прикладів із невідомими значеннями атрибутів залежать від конкретних даних та задач. У [7, 8, 12–21] описано випадок видалення з таблиці понад 50% прикладів, у яких понад 30% значень атрибутів були невідомими. Вилучення прикладів чи атрибутів, що містять навіть одне невідоме значення, може суттєво зменшити розмірність таблиці.

Доповнення таблиць даних [7, 8, 12–21]. Невідомі значення атрибута заповнюють за певним критерієм, який формують на основі відомих його значень. У разі доповнення таблиці необхідно

розрізняти дані, що об'єктивно існують, та такі, що не існують. До перших належать такі дані, які можна отримати, але вони з певних причин не були внесені в таблицю. Наприклад, якщо невідомі дані про вік працівника, то їх можна доповнити на основі інших відомих даних.

Доповнення таблиць із відсутніми даними не змінює розмірності таблиці, але вносить інформаційний шум до даних. Універсальні методи доповнення таблиць з відсутніми даними дають змогу застосовувати відомі методи опрацювання заповнених таблиць прийняття рішень. Теоретичні міркування, що стосуються заповнених таблиць прийняття рішень, можна застосувати для дослідження і незаповнених таблиць без аналізу невідомих значень.

Існують такі основні методи доповнення таблиць з відсутніми даними:

- комбінаторне доповнення;
- глобальне доповнення;
- локальне доповнення з огляду на атрибут та на рішення;
- доповнення методом k найближчих сусідів;
- перепрофілювання змінних і використання системи прийняття рішень.

Доповнення невідомих значень є універсальним способом розв'язування задачі про неповний опис об'єктів. Водночас доповнення невідомих даних має небезпеку внесення істотних змін у дані, що ускладнить пошук зв'язків між умовними атрибутами та рішенням.

Безпосереднє прийняття рішень з допомогою неповних даних. Одним із способів безпосереднього опрацювання даних з невідомими значеннями є методи поділу, за допомогою яких таблицю прийняття рішень з відсутніми даними поділяють так, щоб утворити заповнені таблиці. Такі таблиці опрацюють методами для заповнених таблиць. Існують також методи опрацювання прикладів з невідомими значеннями атрибутів, які полягають у прийнятті рішень безпосередньо на основі неповних даних. Такий підхід реалізовано в алгоритмах MLEM2 [19], URG-2 [20]. Негативний аспект такого підходу полягає у індивідуальному налаштуванні алгоритму до заданого набору даних.

Порівняно нова методологія наближених множин природно пристосована до роботи з недосконалими даними. Теорію наближених множин створив Ж. Павлак (Z. Pawlak) [18] як математичний інструмент для подолання суперечностей у даних і виявлення в них прихованих закономірностей, або шаблонів. Фундаментальний принцип алгоритму навчання на прикладах з використанням наближених множин полягає у виявленні надлишковості серед наявних ознак, що описують приклад. Так виявляють сильні ознаки, які впливають на класифікацію прикладу. Далі вилучають стовпці, атрибути яких не впливають на класифікацію. Для аналізу залишають лише такі атрибути, від яких залежить класифікація прикладів таблиці. Множину атрибутів, що залишилися, називають *редуктом*. Іншими словами, редукт – це підмножина усіх атрибутів таблиці, які забезпечують той самий результат класифікації всіх прикладів таблиці, як і вся множина прикладів. Знаходження редукта є NP-складною задачею. Однак існують достатньо ефективні методи, які дають змогу розв'язувати цю задачу за прийнятний час. До таких методів належать, зокрема, *логічне виведення* (boolean reasoning), алгоритм Джонсона [7, 8, 12–21].

Актуальність проблеми

Необхідно визначити основні пріоритети та найважливіші завдання реального реформування вищої освіти, особливо в контексті прийняття нового Закону України “Про вищу освіту”. В цьому законі питання забезпечення якості вищої освіти розглянуто відповідно до європейської практики. Нова парадигма вищої освіти та необхідність перевірки ліцензійних вимог вимагають постійної достовірної та об'єктивної інформації про матеріально-технічне забезпечення кафедр як структурних одиниць вищих навчальних закладів. З цього погляду проведення інтелектуального аналізу та створення інтелектуальної системи аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу є актуальними.

Для цього дослідження предметною областю є клієнтська база інтелектуальної інформаційної системи аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу, де атрибутами таблиці прийняття рішень є інформація про працівника, техніку, лабораторії, корпуси, програмне забезпечення тощо (загалом 64 атрибутів).

Атрибутом прийняття рішень є висновок про збільшення або зменшення матеріально-технічного забезпечення (6 атрибутів). У таблиці станом на початок 2015 року є 230 прикладів. У результаті проведеного дослідження розв'язано такі основні задачі:

- визначення потреби аналізу існуючої клієнтської бази для визначення залежностей, що істотно впливають на процеси списання або ремонту технічного або програмного забезпечення структурної одиниці навчального закладу;

- опис даних інтелектуальної інформаційної системи аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу;

Наявні дані описано з метою уточнення подальших завдань аналізу та для попереднього опрацювання даних. З усього набору даних клієнтської бази сформовано таблицю прийняття рішень. Дані необхідно трансформувати до формату, якого вимагає алгоритм аналізу.

Постановка задачі

За об'єктно-орієнтованим підходом можна повно описати модель систем аналізу матеріально-технічного забезпечення навчального процесу структурної одиниці навчального закладу з індуктивною компонентою та її поведінку. Об'єктно-орієнтована модель подання знань – відносно новий метод представлення знань. На думку багатьох спеціалістів, ця модель представлення знань разом із об'єктно-орієнтованими мовами програмування дозволяє легко переходити від концептуального до формального опису моделі, а потім і процесу реалізації. Однак, формальна математична модель представлення знань за допомогою об'єктів недостатньо досліджена. Запропоновано математичну модель представлення знань за допомогою об'єктів, яка ґрунтується на понятті схеми об'єктів. Розглянемо основні поняття, що відображають структурну суть об'єктно-орієнтованої моделі. З цією метою введемо такі позначення: C – клас; C^* – множина класів, яка наслідуються із C . Модель ПО інтерпретує класи. Інтерфейс класу визначають за методами, що є шаблоном для всіх об'єктів, які є екземплярами класу. Методи мають ім'я функцій, параметрами і результатом якої є об'єкти. Метод – це відображення деякої підмножини аргументів об'єктів в новий об'єкт. Кожен метод має форму

$$C_1 \times C_2 \times \dots \times C_n \rightarrow C_{result},$$

де C_1, C_2, \dots, C_n – класи, до яких належать об'єкти аргументу; C_{result} – клас об'єкта результату.

Усі класи формують дерево, де вершина кореня – найзагальніший клас об'єктів і будь-який інший клас може мати батьків. Підкласи наслідують поведінку від своїх батьків, а також можуть визначати власні додаткові методи. Отже, дерево класів забезпечує поліморфізм включення [12–21], який дозволяє об'єкту класу C використовуватись в будь-якому контексті, що визначається суперкласом C . Об'єкти інкапсулюють стан і поведінку. Стан об'єкта зафіксовано його значенням. Методи, що визначені на класі, об'єкт якого є екземпляром, визначають поведінку об'єкта. Поведінка проявляється при застосуванні методу до об'єкта. Результат прикладної програми методу – деякий інший об'єкт. Дерево цілей системи аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу має такий вигляд (рис. 3):



Рис. 3. Дерево цілей системи аналізу матеріально-технічного забезпечення навчального процесу структурної одиниці навчального закладу

Також важливе місце у майбутній проєктованій системі займатиме БД. Наведемо декілька діаграм проєктованої системи з використанням методології ERD (рис. 4, 5).

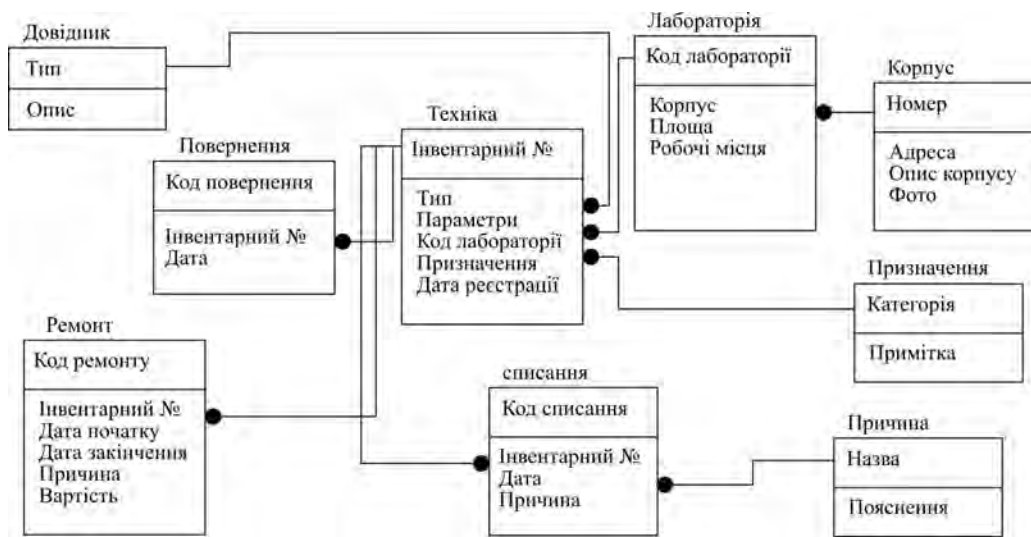


Рис. 4. ERD-діаграма системи аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу



Рис. 5. ERD-діаграма системи аналізу програмного забезпечення навчального процесу структурної одиниці навчального закладу

Також важливим етапом розроблення інтелектуальної системи є функціональне моделювання проєктованої системи. З цією метою доцільно використати методологію моделювання потоків даних (DFD). Результати моделювання наведено на рис. 6–10.



Рис. 6. DFD-діаграма системи аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу першого рівня

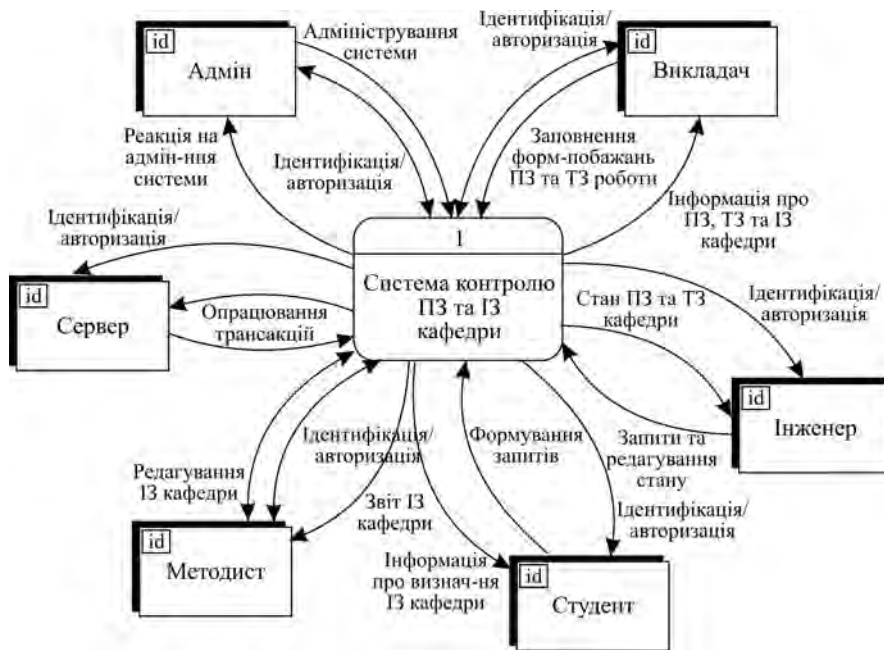


Рис. 7. DFD-діаграма системи аналізу забезпечення структурної одиниці навчального закладу першого рівня

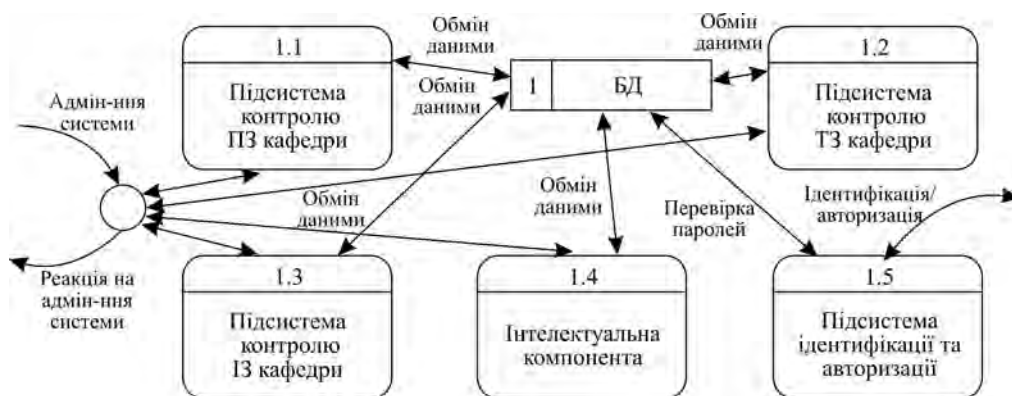


Рис. 8. DFD-діаграма системи аналізу забезпечення структурної одиниці навчального закладу другого рівня

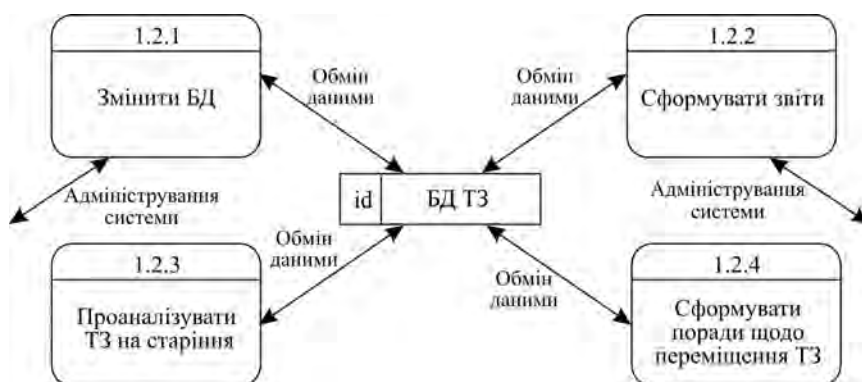


Рис. 9. DFD-діаграма системи аналізу забезпечення структурної одиниці навчального закладу третього рівня

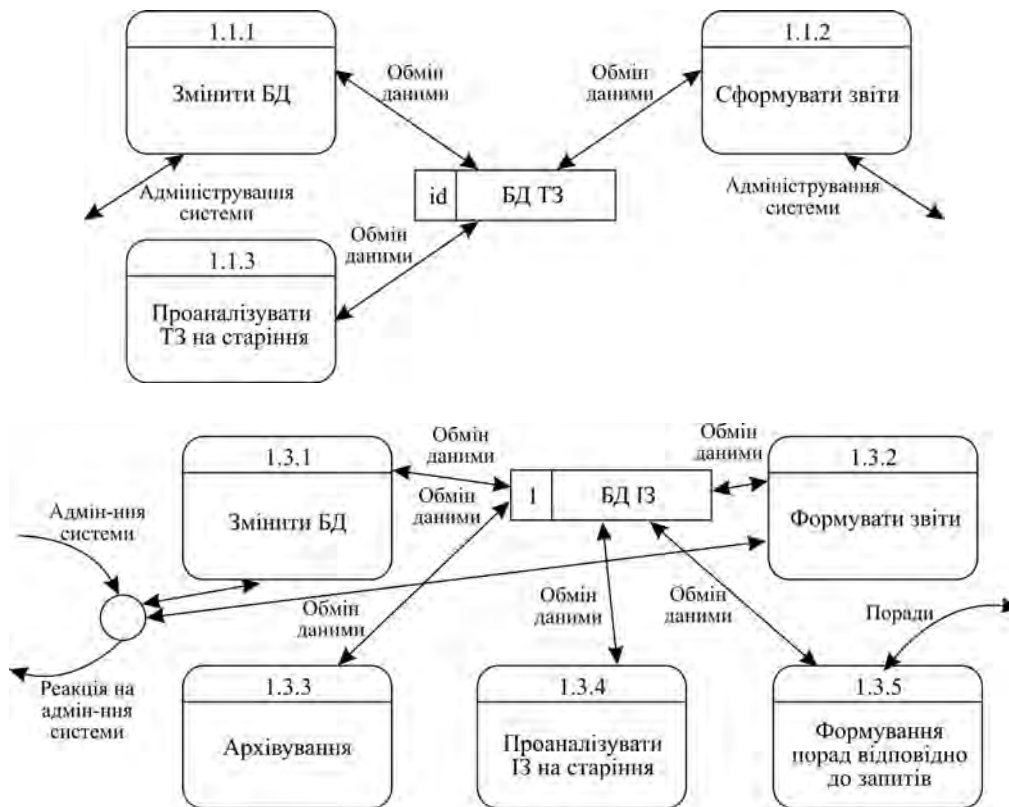


Рис. 10. DFD-діаграма системи аналізу забезпечення структурної одиниці навчального закладу третього рівня

Призначення системи аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу

Завданням є опис предметної області, на основі якого буде спроектовано реляційну базу даних. Обрано таку предметну область: є множина пристроїв з визначеним інвентарним номером, які знаходяться на визначеній кафедрі та у визначених лабораторіях і володіють певними характеристиками, програмне забезпечення. Потрібно побудувати базу даних для роботи з інформаційною системою аналізу матеріально-технічного забезпечення навчального процесу, тобто вести облік наявних пристроїв на кафедрі та історію для формування відповідних звітів.

Концептуальна модель об'єкта матеріально-технічного забезпечення структурної одиниці навчального закладу

Характеристики предметної області, що підлягають відображенню у базі даних, описує така множина атрибутів: назва пристрою, інвентарний номер, тип пристрою, призначення пристрою, характеристики пристрою, місцезнаходження пристрою, інформація, яка вказує на наявність чи відсутність пристрою в даний момент часу на кафедрі (в ремонті або списаний). Для зручності роботи з пристроями та процесом ремонту/списання пристрою можна ввести такі характеристики, як дата ремонту/списання та причина ремонту/списання, і дата повернення з ремонту та результат ремонту.

Логічна схема бази даних матеріально-технічного забезпечення структурної одиниці навчального закладу

Всі перелічені раніше характеристики пристроїв, лабораторій та ремонту пристроїв можна представити такими інформаційними відношеннями:

TechDevice – інформація про пристрої (назва пристрою, його розташування, призначення та інвентарний номер пристрою);

DovAssingment – довідник, призначення пристрою на кафедрі;

DovDeviceType – довідник, тип пристрою;

FromRepair – інформація про повернення пристрою з ремонту (дата, інвентарний номер);

Laboratory – інформація про лабораторії (адреса розташування лабораторії);

Param – таблиця, що фіксує основні параметри для реалізації багатокритерійного пошуку за базою;

RemoveDevice – інформація дати та причину списання пристрою;

ToRepair – інформація про дату, причину та ціну ремонту пристрою.

Особливості системного аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу вимагають під час системного аналізу та планування розглянути призначення системи.

Інтелектуальність розроблюваної системи полягає у визначенні таких показників, що враховуються при визначенні відповідності структурної одиниці навчального закладу ліцензійним вимогам щодо надання освітніх послуг:

- *Частка площі навчально-лабораторних приміщень структурної одиниці навчального закладу від загальної кількості приміщень Інституту з розрахунку на 1-го студента стаціонару*

- *Частка кафедральних робочих місць від загальної кількості комп'ютерних місць Інституту з розрахунку на 1 студента 4–6-х курсів*

- *Частка вартості придбаного кафедрою за рік обладнання відносно Інституту для навчального процесу з розрахунку на 1-го студента стаціонару*

- *Частка вартості придбаного кафедрою за рік обладнання відносно Інституту для наукових досліджень з розрахунку на 1-го науковця*

Визначення та розрахунок перелічених показників прямо пропорційно залежать від правильності, точності, повноти та своєчасності заповнення бази даних інтелектуальної інформаційної системи аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу.

Інтелектуальний аналіз даних інтелектуальної ІС аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу

Необхідно проаналізувати базу даних для інтелектуальної інформаційної системи аналізу матеріально-технічного забезпечення навчального процесу структурної одиниці навчального закладу. Працівник (матеріально відповідальний) структурної одиниці заповнює базу даних про характеристики пристроїв, лабораторій і ремонт пристроїв, в якій зазначає назву пристрою, його розташування, призначення та інвентарний номер пристрою, вказує його призначення та тип, інформацію про повернення пристрою з ремонту (дата, інвентарний номер), про лабораторії (адреса розташування лабораторії), фіксує основні параметри для реалізації багатокритерійного пошуку за базою, інформацію про дату та причину списання пристрою, інформація про дату, причину та ціну ремонту пристрою. За своєю структурою множину умовних атрибутів можна розбити на три підмножини за принципом заповнення значень таблиці.

1. До першої підмножини належать 39 атрибутів, значення яких заповнюється автоматично, навіть якщо працівник не зробив вибору. Ці атрибути описують довідник призначення пристрою (2), інформація про лабораторії (4) та корпус (4), довідник тип пристрою (2), основні параметри для реалізації багатокритерійного пошуку (18), причини списання пристрою (2), застосування програмного забезпечення (2), розповсюдження програмного забезпечення (2), дата введення даних, логін та пароль працівника – по одному атрибуту (3).
2. До другої підмножини умовних атрибутів належать 17 атрибутів, що можуть містити відсутні дані, якщо працівник не зробив свого вибору. Це атрибути, що описують програмне забезпечення (6), техніку (5), лабораторію (3) та корпус (3).
3. Третя підмножина умовних атрибутів містить 2 атрибути, значення яких можуть бути або порожніми, або довільними. Це адреса працівника та його посада.

Атрибут прийняття рішення містить висновок про те, чи доцільно збільшувати або зменшувати технічне і/або програмне забезпечення, тобто матеріально-технічне забезпечення навчального процесу для структурної одиниці навчального закладу .

Структуру набору даних зображено на рис. 11.

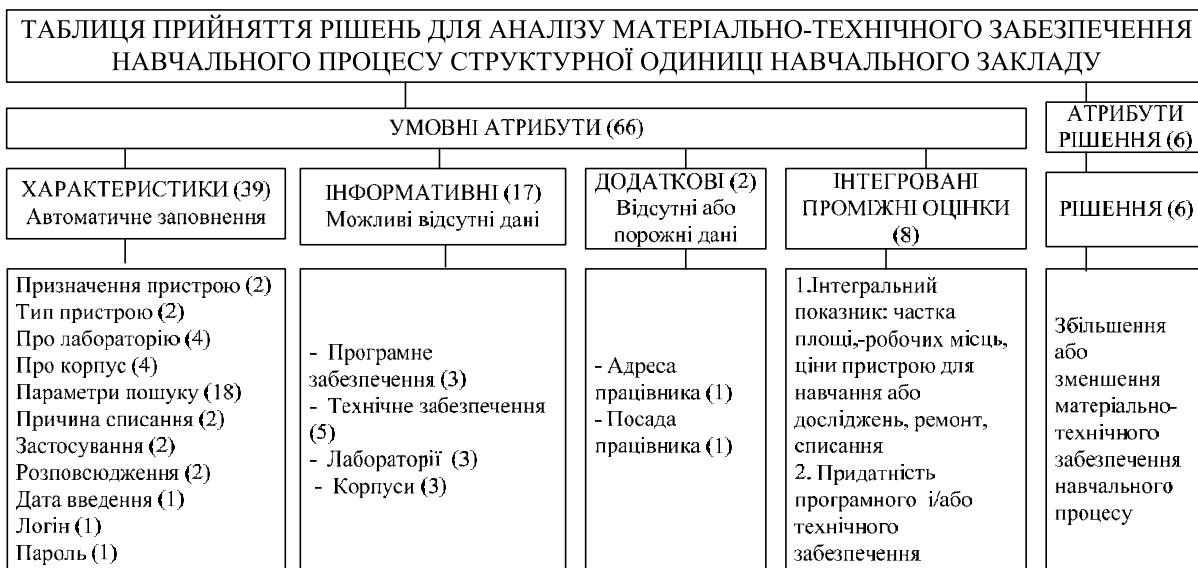


Рис. 11. Структура таблиці даних інтелектуальної інформаційної системи аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу

Атрибут „придатність” програмного або технічного забезпечення оцінює працівник на основі інтегрального показника та власних характеристик забезпечення. Цей атрибут може набувати таких значень: бп (безумовно придатний), п (придатний), уп (умовно придатний), р (група ризику).

Недоліки даних. Описані дані мають кілька недоліків.

1. По-перше, існує певна надлишковість даних. Деякі атрибути містять неістотну для аналізу інформацію, наприклад, адресу працівника чи логін. На етапі відбирання даних необхідно вилучити атрибути, що містять надлишкову інформацію.

2. По-друге, у таблиці даних невідомі деякі значення атрибутів, через що незастосовні методи інтелектуального аналізу даних, які працюють лише із заповненими таблицями.

3. Нарешті, загальним недоліком прийнятих рішень є їх суб’єктивність. Власне, суб’єктивність прийнятих рішень вимагає розв’язування задачі знаходження залежностей у даних та атрибутів, які реально впливають на прийняття рішень. Знаходження таких атрибутів дасть змогу вдосконалити форму замовлення та саму структуру інтелектуальної ІС аналізу матеріально-технічного забезпечення навчального процесу для структурної одиниці навчального закладу у майбутньому.

Розглянемо детально етап відбирання даних, що відповідає загальному процесу видобування знань, зображеному на рис. 2.

Етап відбирання даних. Відбирання даних полягає у видаленні атрибутів таблиці, не важливих для аналізу. Під час відбирання даних було вилучено атрибути “адреса працівника”, “дата введення даних”, “пароль” та “логін”. З вилученням вказаних атрибутів змінилася і структура таблиці (рис. 12.)

Після відбирання даних загальна кількість атрибутів таблиці зменшена, і разом з атрибутом прийняття рішення становить 68 замість 72. Отже, отримана таблиця даних має 230 прикладів та 68 атрибутів. Наступним є етап попереднього опрацювання даних, у процесі якого вилучають зайві дані, замінюють типи даних, кодують окремі значення, опрацьовують відсутні дані тощо. Для подальшого дослідження необхідно здійснити інтелектуальний аналіз даних з метою пошуку прихованих залежностей та оцінити ці залежності. Для аналізу даних та виявлення атрибутів, які

найбільше впливають на прийняте рішення, пропонується використати теорію наближених множин та алгоритм Джонсона. Серед даних, які залишилися після етапу відбирання даних з початкової таблиці, будуть приклади із невідомими значеннями атрибутів. Необхідність опрацювання таких неповних таблиць даних зумовлює здійснення кількох серій експериментів над таблицями з різними способами опрацювання невідомих значень атрибутів для порівняння якості підходів опрацювання відсутніх даних. Таблиця прийняття рішень з оцінкою якостей матеріально-технічного забезпечення навчального процесу містить атрибути, які поділено на чотири групи.

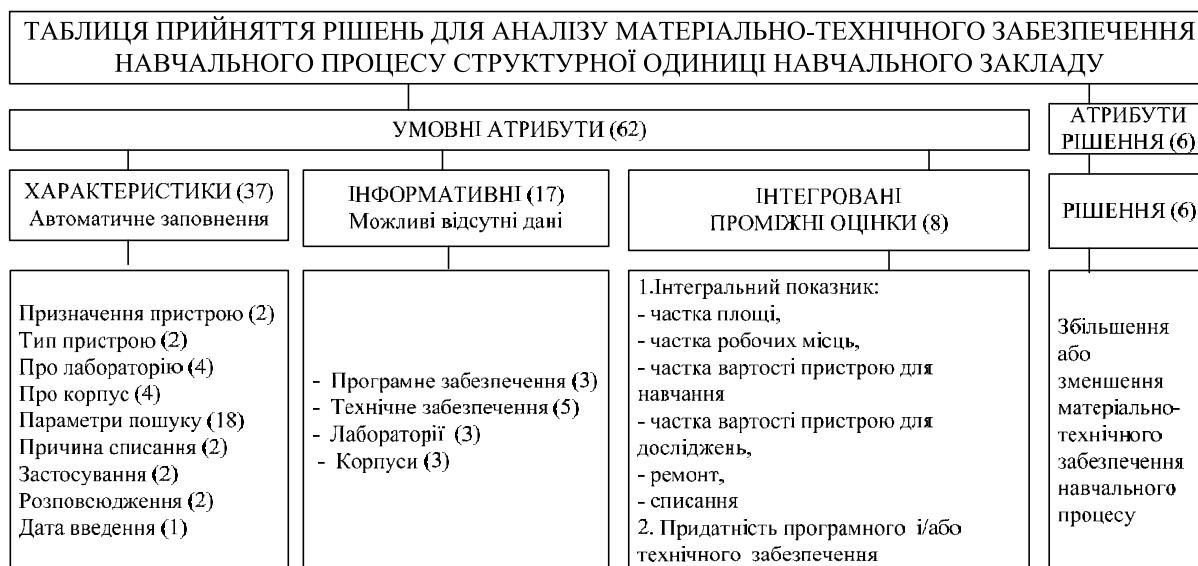


Рис. 12. Структура таблиці після етапу відбирання даних інтелектуальної інформаційної системи аналізу матеріально-технологічного забезпечення навчального процесу для структурної одиниці навчального закладу

1. *Атрибути-характеристики* (37). Ці атрибути містять загальну інформацію матеріально-технічного забезпечення. Всі атрибути-характеристики поділено на такі підгрупи: призначення пристрою, тип пристрою, характеристики лабораторії та корпусів, основні параметри пошуку, причини списання, застосування пристрою, розповсюдження програмного забезпечення та дата введення даних. У таблиці даних для кожного з атрибутів-характеристик, крім дати введення даних (оцінювання рівня оновлення даних залежно від теперішньої дати), існує відповідний атрибут з оцінкою результату тесту за чотирибальною шкалою. Ці атрибути можуть набувати значень *v* (відмінно), *д* (добре), *з* (задовільно), *н* (незадовільно) [*v*, *д*, *з*, *н*].
2. *Інформативні атрибути* (17). До інформативних належать атрибути: програмне забезпечення, технічне забезпечення, лабораторії, корпуси. Атрибут “лабораторії” описує адресу розташування лабораторії. Атрибут “корпуси” описує адресу розташування корпусів.
3. *Інтегровані проміжні оцінки* (8). До проміжних оцінок належать “інтегральний показник” (ІП) та “придатність” матеріально-технічного забезпечення, які набувають значення в межах [0,1] залежно від якості забезпечення цих показників навчального процесу. Оцінювання інтерпретується для *v* (відмінно) в межах [0,75;1], *д* (добре) – [0,5;0,74], *з* (задовільно) – [0,25;0,49], *н* (незадовільно) – [0;0,24].
 - 3.1. Частка площі навчально-лабораторних приміщень структурної одиниці навчального закладу від загальної кількості приміщень Інституту з розрахунку на 1-го студента стаціонару.
 - 3.2. Частка кафедральних робочих місць від загальної кількості комп'ютерних місць Інституту з розрахунку на 1 студента 4–6-х курсів.
 - 3.3. Частка вартості придбаного кафедрою за рік обладнання відносно Інституту для навчального процесу з розрахунку на 1-го студента стаціонару.
 - 3.4. Частка вартості придбаного кафедрою за рік обладнання відносно Інституту для наукових досліджень з розрахунку на 1-го науковця.

- 3.5. Частка ремонту технічного забезпечення.
- 3.6. Частка списання матеріально-технічного забезпечення.
- 3.7. Атрибут “інтегральний показник”. Атрибут “інтегральний показник” обчислено як середнє арифметичне значень атрибутів 3.1–3.6 з результатами тестування якостей матеріально-технічного забезпечення навчального процесу. Значення інтегрального показника узагальнено значенням атрибута “оцінка МТЗ” за чотирибальною шкалою. Атрибут “оцінка МТЗ” набуває значення з домену [в, д, з, н].
- 3.8. Атрибут “придатність” матеріально-технічного забезпечення. Атрибут “придатність” матеріально-технічного забезпечення оцінює матеріально відповідальний працівник структурної одиниці навчального закладу на основі інтегрального показника та власних характеристик цього матеріально-технічного забезпечення. Цей атрибут може набувати таких значень: *бн* (безумовно придатний), *п* (придатний), *уп* (умовно придатний), *р* (група ризику).
4. *Атрибути прийняття рішення* (б). Атрибутами прийняття рішень є якість, ефективність, успішність та надійність конкретного матеріально-технічного забезпечення. Оцінював успішність та надійність роботи працівника матеріально відповідальний працівник за п’ятибальною шкалою в процесі заповнення окремої анкети. Атрибути прийняття рішень набувають значень з домена [1–5]. Атрибут “характеристика” обчислено як середнє арифметичне значень атрибутів “якість”, “ефективність”, “успішність” та “надійність” конкретного матеріально-технічного забезпечення. Атрибут “рекомендація” щодо збільшення або зменшення матеріально-технічного забезпечення вираховують у результаті отриманого значення атрибута “характеристика” – для значення в межах [1, 2] приймається рекомендація списати це матеріально-технічне забезпечення (якщо це пристрій) та оновити його новим еквівалентом. Для лабораторій та корпусів при оцінці в межах [1, 2] рекомендується провести ремонт. Для програмного забезпечення – оновити його. У разі оцінки [3, 4] – косметичний ремонт для лабораторій та корпусів, оновлення версій програмного забезпечення та ремонт технічного забезпечення. При оцінці 5 – нічого не рекомендувати.

Недоліки даних. Описані дані мають кілька недоліків.

По-перше, існує певна надлишковість даних. Більшість характеристик працівника подано в таблиці двома атрибутами. Один атрибут містить кількість балів, які є результатом відповідного тестування, інший – оцінку цього результату за певною шкалою. Тому загальна кількість умовних атрибутів складається з 62 атрибутів. На стадії відбирання даних необхідно вилучити атрибути, що містять отримані бали, та залишити для аналізу лише атрибути зі значеннями оцінок.

По-друге, у таблиці даних невідомі деякі значення атрибутів, через що незастосовні методи інтелектуального аналізу даних, які працюють лише із заповненими таблицями.

По-третє, оскільки тестування здійснювалось тільки раз у чотири роки, то це ускладнює відслідкування змін параметрів у часі. З огляду на вимоги конфіденційності відсутність частини даних, а також необхідність уніфікації кодування значень атрибутів потрібно попередньо опрацювати дані.

Нарешті загальним недоліком прийнятих рішень є їхня суб’єктивність. Власне суб’єктивність прийнятих рішень вимагає розв’язування задачі знаходження залежностей у даних та атрибутів, які реально впливають на прийняття рішень. Знаходження таких атрибутів дасть змогу вдосконалити здійснення психологічних досліджень у майбутньому.

Етап попереднього опрацювання даних. На цьому етапі було здійснено опрацювання атрибутів з оцінкою загальних власних характеристик матеріально-технічного забезпечення. Ці характеристики під час тестувань оцінювали за різними шкалами. Домени цих атрибутів перетворені з [1, 2, 3] на [в (відмінно), д (добре), з (задовільно)]. Такою заміною уніфіковано подання оцінок в цих атрибутах, оскільки були різними шкали, за якими ставили ці оцінки. Серед даних, які залишились після етапу відбирання даних з початкової таблиці, є приклади із невідомими значеннями атрибутів. Необхідність опрацювання таких неповних таблиць даних зумовила здійснення двох серій експериментів.

У першій серії експериментів на етапі попереднього опрацювання даних застосовано метод видалення прикладів з невідомими значеннями. Разом було видалено 18 таких прикладів, і розміри таблиці становили 188×38. У другій серії при опрацюванні даних було використано глобальне доповнення, за яким невідомі значення доповнені вибраним параметром. Параметр, яким доповнено відсутні дані, для числових атрибутів обчислено як середнє арифметичне значень відповідного домена, а для символічних – значення їхніх доменів, які зустрічаються найчастіше.

Перша серія експериментів. Таблицю, яку буде використано для аналізу у першій серії експериментів була опрацьована з використанням методу видалення прикладів з невідомими значеннями атрибутів. Розміри таблиці 188×38. Її структуру наведено у табл. 1.

Таблиця 1

Фрагмент таблиці для першої серії експериментів

Група	Атрибути-характеристики						Інформативні атрибути			Інтегровані проміжні оцінки		Атрибути рішення		
	Назва атрибута	Призначення	Тип	...	Застосування	Розповсюдження	Дата	ПЗ	...	Корпуси	Інтегрований показник	Придатність МТЗ	Характеристика	Рекомендація
	2	1	...	1	1	1	[в, д, з, н]		[в, д, з, н]	[0,1]	п	4	+/-	
	3	2	...	1	2	1				[0,1]	р	4	+/	
	4	2	...	1	1	1				[0,1]	уп	3	-	
	4	3	...	1	1	1				[0,1]	бп	4	+	
	4	1	...	1	3	1				[0,1]	п	4	+/	
	2	2	...	1	1	2				[0,1]	п	4	+/	
	3	3	...	1	4	4				[0,1]	п	4	+/	
	4	4	...	1	1	1				[0,1]	п	4	+	
	5	3	...	2	1	1				[0,1]	п	5	+	
	3	2	...	1	1	3				[0,1]	п	4	-	
...														
	5	4	...	1	4	4				[0,1]	п	4	+/-	

Над даними з табл. 1 виконано низку експериментів. Для кожного експерименту було відібрано свій набір атрибутів.

1. Для експерименту E1.1 відібрано 6 інформативних атрибутів, атрибут „оцінка МТЗ”, 4 атрибути з оцінками власних якостей МТЗ, 7 атрибутів з даними про мотиви ремонту/списання, атрибут „придатність”, „успішність” та атрибут прийняття рішення – „надійність”. Разом – 21 атрибут. Розміри таблиці прийняття рішень – 188×21.

2. Для експерименту E1.2 використано всі 38 атрибутів з початкової табл. 1. На відміну від експерименту E1.1, використано 17 атрибутів з результатами тестів, на основі яких розраховується „оцінка МТЗ”. Атрибут прийняття рішення – „надійність”. Розміри таблиці даних – 188×38.

3. Для експерименту E1.3 відібрано 6 інформативних атрибутів, атрибут „оцінка МТЗ”, 4 атрибути з оцінками особистих якостей, 7 атрибутів з даними про мотиви ремонту/списання, атрибут „придатність” та „успішність”. Атрибут прийняття рішення – „успішність”. Загальний розмір таблиці – 188×20.

4. Для експерименту E1.4 використано всі атрибути з початкової табл. 3 за винятком атрибута „надійність”. У цьому експерименті таблиця прийняття рішень містить атрибути з результатами тестів, на основі яких розраховано „оцінку МТЗ”. Атрибут прийняття рішення – „успішність”. Розміри таблиці даних – 188×37.

5. Для експерименту E1.5 відібрано 6 інформативних атрибутів, атрибут „оцінка МТЗ”, 4 атрибути з оцінками особистих якостей МТЗ, 7 атрибутів з даними про мотиви, атрибут „придатність” та „надійність”. Атрибут „успішність” не відібрано з метою перевірки того, чи пов’язана ця суб’єктивна оцінка з іншими атрибутами таблиці. Атрибут прийняття рішення – „надійність”. Загальний розмір таблиці – 188×20.

Побудова редуктів. У табл. 2 наведено результати п’яти експериментів (E1.1), метою яких було знаходження редукта кожної таблиці – набору атрибутів, які найбільше впливають на прийняття рішення. Символом “+” у таблицях позначено атрибути, які утворюють редукти за різних параметрів алгоритму.

Таблиця 2

Результати експерименту E1.1 за методом Джонсона для атрибуту прийняття рішення „надійність” (розмір таблиці – 188×21)

Ступінь підтримки, %	Призначення	Тип	Застосування	Оцінка МТЗ	Розповсюдження	Лабораторія	Частка вартості	Частка робочих місць	Успішність	Ремонт	Списання
100	+	+	+	+	+	+	+	+		+	+
98	+	+	+				+	+	+		
95		+	+				+	+	+		
90		+	+				+		+		
81		+	+				+				

Аналіз таблиці прийняття рішень за методом Джонсона дав змогу з 21 атрибуту, що входили до неї, відібрати до редукта зі ступенем підтримки у 100% десять атрибутів, які наведено у табл. 2. Зі зменшенням ступеня підтримки виявилось, що серед атрибутів, які впливають на прийняте рішення, з’являється „успішність” (у редуктах зі ступенем підтримки 98, 95 та 90%). Найбільше на „надійність” впливають „тип”, „застосування” та „частка вартості”, оскільки ці атрибути входять до редукта зі ступенем підтримки 81%.

Друга серія експериментів. Для другої серії експериментів було здійснено глобальне доповнення відсутніх значень вибраним параметром. Як параметр для числових значень атрибута обрано середнє, для символічних – значення, що зустрічається найчастіше.

Перед доповненням з табл. 1 було вилучено 12 прикладів, для яких значення атрибута прийняття рішення було відсутнім. Таблиця прийняття рішень містила 37 умовних атрибутів, тобто 37 атрибутів×194 прикладів = 7178 значень атрибутів. Серед них було шість прикладів з відсутніми значеннями атрибутів, що становить 3.1% від загальної кількості прикладів; відсутніх значень атрибутів – 57, або 0.79%. Як вказано раніше, такий рівень відсутніх значень цілком прийнятний для аналізу. Було заповнено відсутні значення для 6-ти прикладів. Для 5-ти з них було заповнено відсутні оцінки особистісних характеристик та мотивів. Для одного прикладу заповнено значення двох атрибутів з оцінками результатів психологічних і психофізіологічних тестів. Остаточні розміри готової для аналізу таблиці прийняття рішень 194×38.

На основі цієї таблиці вибрано дані для серії з 5 експериментів та побудовано таблиці з їх результатами. Ці таблиці мають номери 9–13. Вони відповідають експериментам, які позначені як E2.1 – E2.5. Сформовані таблиці даних складаються з тих самих атрибутів, що й відповідні їм таблиці у експериментах E1.1 – E1.5, тобто мають однакову кількість стовпців. Відмінність полягає в тому, що всі нові таблиці даних мають 194 рядки, на відміну від таблиць з першої серії експериментів, які містили 188 рядків.

Побудова редуктів. У табл. 3 наведено результати експериментів з побудови редуктів для таблиць з доповненими даними. Результати такого аналізу незначно відрізняються від аналогічних результатів першої серії: видно, що найвпливовіші „тип”, „застосування” та „частка вартості”. Редукт таблиці прийняття рішень з експерименту E2.1 демонструє, що порівняно з результатами

експерименту E1.1, видаленням прикладів з неповним описом, видалено ті з них, які визначали важливість атрибутів „причина списання”, „дата” та „корпус”. Заповнення відсутніх даних підтвердило важливість особистісних характеристик, але показало, що мотиви списання не впливають на атрибут прийняття рішення.

Таблиця 3

Результати експерименту E2.1 за методом Джонсона для атрибута прийняття рішення „надійність” (розмір таблиці – 194×21)

Ступінь підтримки, %	Призначення	Тип	Причина списання	Застосування	Дата	Оцінка МТЗ	Корпус	Частка вартості	Частка робочих місць	Успішність
100	+	+	+	+	+	+	+	+	+	+
98	+	+		+				+	+	+
95		+		+				+	+	+
91		+		+				+		+
82		+		+				+		

Висновки та перспективи подальших наукових розвідок

Основною метою і основним результатом дослідження стало виявлення набору атрибутів, які найбільше впливають на прийняте рішення щодо успішності та надійності впровадження матеріально-технічного забезпечення в навчальний заклад. На основі цих результатів надалі можна буде здійснювати більш обґрунтований вибір окремих, лише вагомих, атрибутів для експериментів над даними, чи для здійснення майбутніх тестувань МТЗ. Виявлено, що із 20-38 (в різних випадках) атрибутів тільки 6-10 атрибутів впливають на прийняте рішення.

У процесі здійснення експериментів в процесі опрацювання таблиць даних потрібно було врахувати відсутність частини даних. Тому було здійснено дві серії експериментів, у кожній з яких застосовано окремий МТЗ врахування відсутніх даних.

Заповнено відсутні значення і згенеровано правила, проте для повного оцінювання результатів цих дій необхідні подальші дослідження. Надалі важливо проаналізувати, який спосіб опрацювання невідомих значень для такого типу даних дає кращі результати. Необхідно порівняти результати, отримані після доповнення даних різними методами, та обрати оптимальний.

У результаті проведеного системного аналізу і функціонального моделювання спроектовано інтелектуальну систему аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу. Запропонована інтелектуальна система є складовою комплексної інтелектуальної інформаційної системи аналізу матеріально-технологічного забезпечення навчального процесу для структурної одиниці навчального закладу. Предметною областю (ПО) цієї інтелектуальної інформаційної системи є матеріально-технічне забезпечення структурної одиниці навчального закладу. Отримано готовий продукт – ІС, що здатна приймати рішення із застосуванням методів та засобів штучного інтелекту щодо відповідності матеріально-технічного забезпечення структурної одиниці навчального закладу ліцензійним вимогам. Частину поставлених задач вже реалізовано, інші розробляються. Побудована інтелектуальна система аналізу матеріально-технічного забезпечення структурної одиниці навчального закладу дає можливість полегшити та прискорити роботу працівників структурної одиниці навчального закладу, забезпечить швидкий та легкий доступ до необхідної інформації, підвищить якість та ефективність функціонування забезпечення структурної одиниці навчального закладу.

1. Рашкевич Ю. М. Болонський процес та нова парадигма вищої освіти. – Львів: Видавництво Львівської політехніки, 2014. – 166 с. 2. Катренко А. Ю. Системний аналіз об'єктів та процесів комп'ютеризації. – Львів, 2003. 3. Берко А. Ю. Застосування баз даних. – Львів, 2007. 4. Ліцензійні

умови надання освітніх послуг в сфері вищої освіти. Затверджено наказом Міністерства освіти і науки України від 24.12.2003 N9847. 5. Брукс Ф. Как создаются программные системы. – СПб.: Символ, 2001. – 495 с. 6. Голощук Р. О. Математичне моделювання процесів дистанційного навчання / Р. О. Голощук, В. В. Литвин, Л. В. Чирун, В. А. Висоцька // Інформаційні системи та мережі: Вісник Нац. ун-ту “Львівська політехніка”. – 2003. – № 489. – С. 100–109. 7. Завалій Т. І. Інтелектуальний аналіз результатів психологічного тестування / Т. І. Завалій, Ю. В. Нікольський, Т. В. Шестакевич // Вісник Нац. ун-ту “Львівська політехніка”. – 2008. – № 631: Інформаційні системи та мережі. – С. 113–138. – Бібліографія: 15 назв. 8. Чирун Л. В. Інтелектуальний аналіз таблиць прийняття рішень у системах електронної комерції / Л. В. Чирун, Т. В. Шестакевич // Вісник Нац. ун-ту “Львівська політехніка”. – 2008. – № 621: Інформаційні системи та мережі. – С. 230–239. – Бібліографія: 11 назв. 9. Берко А. Ю. Intranet архітектура інтелектуальних систем електронного навчання / А. Ю. Берко, В. А. Висоцька // Інформаційні системи та мережі. Вісник Нац. ун-ту “Львівська політехніка”. – Львів 2001. – № 438. – С. 3–10. 10. Інтерактивна взаємодія та зворотній зв'язок в системі дистанційного навчання / Р. О. Голощук, В. А. Висоцька // Інформаційні системи та мережі: Вісник Нац. ун-ту “Львівська політехніка”. – 2002. – № 464. – С. 44–53. 11. Висоцька В. А. Система опрацювання структури електронного підручника / В. А. Висоцька // Інформаційні системи та мережі: Вісник Нац. ун-ту “Львівська політехніка”. – 2003. – № 489. – С. 49–63. 12. Mitra Sushmita, Pal Sankar K., Mitra Pabitra. Data mining in soft computing framework: a survey, IEEE Transactions on Neural Networks, Vol. 13, Issue 1, 2002. 13. Acuna E., Rodriguez C. The Treatment of Missing Values and its Effect in the Classifier Accuracy // In D. Banks, L. House, F. R. McMorris, P. Arabie, W. Gaul (Eds), Classification, Clustering and Data Mining Applications, Springer-Verlag Berlin-Heidelberg, 2004, c.639–648. 14. Jan Komorowski, Lech Polkowski, Andrzej Skowron, Rough Sets: A Tutorial. // Eds. S. K. Pal and A. Skowron, Rough Fuzzy Hybridization: A New Trend in Decision-Making, Spriner-Verlag, Singapore, 1998. 15. Grzymala-Busse J. W. Rough Set Strategies to Data with Missing Attribute Values. // Proceedings of the Workshop on Foundation and New Directions in Data Mining, associated with the third IEEE International Conference on Data Mining, November 19-22, 2003, Melbourne, FL, USA, c.56-63. 16. Latkowski R. Metody wnioskowania w oparciu o niekompletny opis obiektow, Warszawa, 2001. <http://logic.mimuw.edu.pl/Grant2003/prace/BMscLatkowski.pdf>. 17. Øhrn A. ROSETTA Technical Reference Manual, 2001. <http://www.idi.ntnu.no/~aleks/>. 18. Pawlak Z. Rough Sets – Theoretical Aspects of Reasoning about Data, volume 9 of Series D: System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, Dordrecht, 1991. 19. Grzymala-Busse J. W. MLEM2 – Discretization During Rule Induction. // Proceedings of the IIPWM'2003, International Conference on Intelligent Information Processing and WEB Mining Systems, Zakopane, Poland, June 2-5, 2003, Springer Verlag, c.499-508. 20. Konias S., Chouvarda I., Vlahavas I., Maglaveras N. A Novel Approach for Incremental Uncertainty Rule Generation from Databases with Missing Value Handling: Application to Dynamic Medical Databases. Medical Informatics & The Internet in Medicine, Taylor & Francis, Vol. 2005, Issue 5, 2005. 21. Komorowski J., Polkowski L., Skowron A. Rough Sets: A Tutorial. // Eds. Pal S. K. and Skowron A. Rough Fuzzy Hybridization: A New Trend in Decision-Making, Spriner-Verlag, Singapore, 1998.