

Тема: Організація машинної інформаційної бази

План.

1. Поняття машинного інформаційного забезпечення.
2. Передумови створення та основні переваги баз даних.
3. Поняття, класифікація і склад автоматизованого банку даних.
4. Характеристика інфологічної та деталогічної моделей баз даних.
5. Методи створення оптимальної моделі баз даних.

Література

1. Ананьєв О.М., Білик В.М. і ін.. Інформаційні системи і технології в комерційній діяльності: Підручник.-Львів: Новий Світ-2000, 2006.- 584 с.
2. Інформаційні системи і технології в економіці. / За ред .д.е.н., проф. В.С.Пономаренка. - К.:ВЦ "Академія", 2002. - 544 с.
3. Білик В.М., Костирко В.С. Інформаційні технології та системи: Навчальний посібник.- К.:Центр навчальної літератури, 2006. – 232 с.
4. Дятлова Н.В., Бубликова Н.Ф. Інформаційні системи і технології в комерційній діяльності. Навчально-методичний посібник. НМЦ, 2005- 162 с.

1. Поняття машинного інформаційного забезпечення

Інформаційне забезпечення — важливий елемент автоматизованих інформаційних систем обліку, призначений для відображення інформації, що характеризує стан керованого об'єкта і є основою для прийняття управлінських рішень.

Загальну структуру інформаційного забезпечення ілюструє рис. 1. Важливою складовою інформаційного забезпечення є інформаційна база, що складається з машинної та позамашинної інформаційної бази.

Машинна інформаційна база — частина інформаційної бази ІС, що являє собою сукупність інформаційних масивів, які зберігаються в пам'яті ЕОМ та на магнітних носіях.

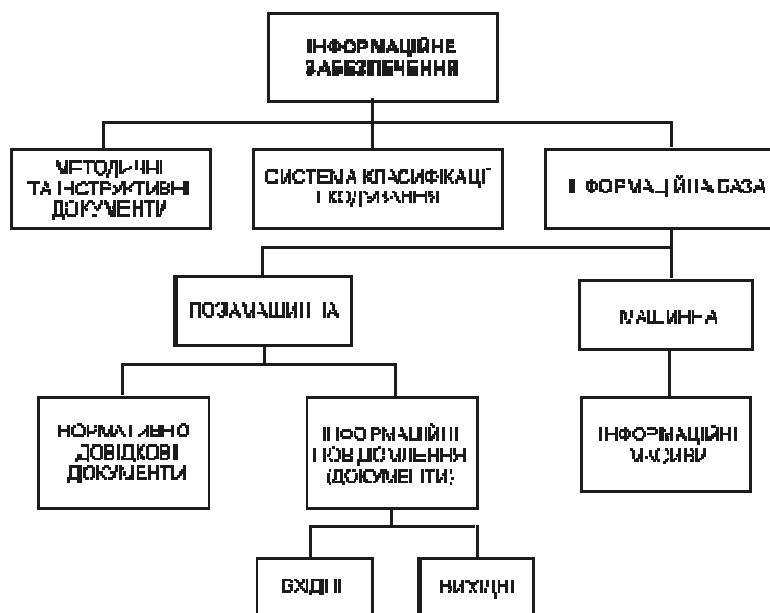


Рис. 1. Структура інформаційного забезпечення

Машинна інформаційна база складається з інформаційних масивів, які можуть бути організовані у вигляді окремих незалежних між собою, локальних інформаційних масивів, чи у вигляді бази даних, тобто інтегрованої сукупності пов'язаних між собою масивів, якими керує система управління базами даних (СУБД).

Масив — це ідентифікована сукупність примірників логічно пов'язаних між собою даних, які містяться поза програмою у зовнішній пам'яті і доступні програмі за допомогою спеціальних операцій.

У процесі еволюції розробки автоматизованих ІС машинна ІБ пройшла такі етапи розвитку:

- 1) підготовку інформаційних масивів (для кожної задачі окремо);
- 2) створення єдиної бази даних, яку можна використовувати для розв'язування певної кількості задач.

У сучасних ІС для організації інформаційного забезпечення використовується концепція баз даних (БД).

2. Передумови створення та основні переваги БД

У традиційних системах обробки інформації дані організовуються у вигляді не пов'язаних між собою локальних інформаційних масивів, які мають лінійну структуру. Сутність такого підходу до організації інформаційного забезпечення (ІЗ) полягає в тому, що інформаційні масиви проектуються окремо для кожної конкретної задачі чи для їх комплексів.

Такі системи називають іноді файловими. Попри відносну простоту організації файлові системи мають цілий ряд недоліків. Головними з них є такі.

1. Надмірність даних. Файлові системи характеризуються значною надмірністю, оскільки нерідко для розв'язування різних задач управління використовуються одні й ті самі дані.

Дублювання даних у різних файлах зумовлює неекономне використання пам'яті на зовнішніх запам'ятовуючих пристроях і призводить до того, що інформація одного й того самого об'єкта управління розподіляється між багатьма файлами.

2. Неузгодженість даних. Сутність цього недоліку полягає в тому, що, як уже зазначалося, при файловій організації ІЗ одна й та сама інформація може розміщуватись у різних масивах. При цьому технологічно складно простежити за внесенням змін одночасно в усі масиви. Через це може виникнути неузгодженість даних, коли одне й те саме поле в різних масивах може мати різні значення.

3. Залежність структур даних і прикладних програм. При файловій організації логічна та фізична структура файла має відповідати її опису в прикладній програмі.

Прикладна програма має бути модифікована при зміні логічної чи фізичної структури файла. Але з огляду на те, що зміни в одній програмі часто потребують внесення змін до інших інформаційно пов'язаних програм, а іноді простіше створити нову програму, ніж вносити зміни до старої, то цей недолік файлових систем спричинюється до значного збільшення вартості супроводу програмних засобів (зауважимо, що в ряді випадків вартість супроводу програмних засобів досягає 70 % вартості їх розробки).

Розвиток засобів обчислювальної техніки, створення запам'ятовуючих пристроїв прямого (безпосереднього) доступу створили передумови для вирішення проблем незалежності, неузгодженості і надмірності даних, а також сприяли створенню нової концепції організації ІЗ — концепції інтеграції даних, яка дістала назву автоматизованого банку даних (АБД).

Головні переваги організації ІЗ у вигляді АБД наступні:

1. Багаторазовість використання даних: одні й ті самі дані можуть використовуватися для розв'язування різних задач.

2. Економія витрат на створення й ведення ІЗ: організація ІЗ у вигляді БД характеризується нижчою вартістю на створення і меншими витратами на внесення змін в БД, оскільки зміни на фізичному рівні не потребують внесення змін до прикладних програм.

3. Зменшення надмірності даних. Необхідність розв'язування нових задач забезпечується здебільшого за рахунок існуючих файлів у БД, а не шляхом створення нових файлів. Дублювання даних у БД потрібне лише для забезпечення оперативності пошуку даних і організації зв'язку між файлами БД. Таке дублювання не є надмірним.

4. Швидкість обробки не передбачених запитів до системи. Для обробки таких запитів найчастіше не вимагається створення нової програми мовами програмування, оскільки ці процедури виконуються за допомогою спеціальних мовних засобів (мови запитів і мови генерації звітів), які входять до складу СУБД.

5. Простота і зручність внесення змін за рахунок єдиної системи ведення БД, яка підтримується засобами СУБД.

6. Логічна та фізична незалежність даних від прикладних програм.

3. Поняття і класифікація АБД

АБД — це система інформаційних, математичних, програмних, мовних, організаційних і технічних засобів, які необхідні для інтегрованого нагромадження, зберігання, ведення, актуалізації, пошуку і видачі даних.

АБД можна класифікувати за різними ознаками.

1. За призначенням АБД бувають:

- інформаційно-пошукові;
- спеціалізовані за окремими галузями науки та техніки;
- банки даних для автоматизації задач організаційно-економічного управління;
- банки даних для систем автоматизації наукових досліджень і виробничих випробувань;
- банки даних для систем автоматизованого проектування.

2. За архітектурою обчислювального середовища АБД бувають централізовані і розподілені.

3. За видом інформації, що зберігається, розрізняють банки даних, банки документів і банки знань.

4. За мовою спілкування користувача з БД розрізняють системи з базовою мовою (відкриті системи) та власною мовою (закриті системи).

У відкритих системах мовним засобом спілкування з БД є одна з мов програмування, наприклад SI, Паскаль тощо. В таких системах для спілкування з БД потрібний посередник, тобто програміст, який володіє вибраною мовою програмування.

Закриті системи мають власну мову спілкування, що, як правило, набагато простіша за мови програмування. Тому в таких системах не потрібний посередник-програміст для спілкування з БД. Самі користувачі за відповідної підготовки зможуть працювати з БД.

Основними складовими компонентами АБД є БД і система управління БД (СУБД).

База даних — це поійменована, структурована сукупність взаємопов'язаних даних, які характеризують окрему предметну область і перебувають під управлінням СУБД. БД являє собою інтегроване сховище даних, яке призначене для використання багатьма споживачами і забезпечення незалежності даних від прикладних програм. Зв'язок кінцевих користувачів та прикладних програм з БД відбувається через СУБД, яка слугує інтерфейсом між користувачами і БД.

Під предметною областю в даному разі розуміють один чи кілька об'єктів управління, інформація яких моделюється за допомогою БД і використовується для розв'язування різних функціональних задач.

Усі дані, які зберігаються в БД, поділяються на фонд і архів даних. Такий поділ пов'язаний з різницею в технологічних режимах використання даних. Фонд даних — це дані, які зберігаються на мета даних (МД) (при використанні машин типу ЕС ЕОМ)

чи вінчестері (при використанні ПЕОМ) і перебувають безпосередньо під управлінням СУБД. Архіви — це копії файлів БД, які зберігаються на магнітних стрічках чи гнучких магнітних дисках для відтворення БД на випадок її зруйнування при різних збійних ситуаціях.

Особливістю БД є те, що вона складається з даних і їх опису. Опис даних називають метаданими. Метадані дають змогу реалізувати незалежність даних від прикладних програм. При файловій організації даних потрібно в кожній прикладній програмі повністю описати структури відповідних інформаційних масивів, незалежних від того, скільки полів обробляється в тій чи іншій програмі. При використанні БД в програмі потрібно описувати лише поля, потрібні для обробки. Отже, метадані є незалежними від прикладних програм і являють собою самостійний об'єкт для зберігання.

В АБД існує ще таке поняття, як словник даних (СД). СД дуже важливий, особливо в умовах колективного використання даних, оскільки забезпечує вирішення проблеми вірогідності, збитковості і контролю за раціональним зберіганням та використанням даних.

Словник даних може містити відомості про джерело інформації, формати та взаємозв'язок між даними, відомості про частоту виникнення і характер використання даних, терміни коригування і осіб, відповідальних за це, і т.ін. Отже, СД являє собою базу даних про дані як особливий вид ресурсу.

Одне з основних призначень СД — документування даних. Нагромадження інформації про дані потрібно починати вже на ранніх стадіях проектування. Йому відводиться роль засобу централізованого ведення та управління даними на всіх етапах проектування системи, а також забезпечення ефективної взаємодії між всіма користувачами при розподіленій БД. В СД може бути занесена інформація про місце фізичного зберігання даних, а також відомості про обмеження секретного характеру, безпеки, доступу та інші питання, що характеризують фізичні параметри БД.

Усі питання, пов'язані з веденням СД, розглядатимемо далі в контексті із СУБД та автоматизованою формою його ведення. Зауважимо, проте, що всі СУБД в своєму складі мають цей засіб. Наприклад, широко відомі СУБД сім'ї DBASE його не мають. В такому разі потрібно вести СД вручну чи писати відповідні програми ведення такого словника.

Пакет програм ведення СД може інтегруватися із СУБД чи бути незалежним. На сучасному ринку програмних засобів є дуже мало СУБД, які мають у своєму складі засоби автоматизованого ведення СД.

Як приклад СУБД, що підтримує СД, може бути система ADABAS чи її русифікована версія ДИСОД.

До складу АБД обов'язково входить такий компонент, як СУБД, що є комплексом програмних і мовних засобів загального та спеціального призначення, необхідних для створення БД, підтримки її в актуальному стані, маніпулювання даними й організації доступу до них різних користувачів чи прикладних програм в умовах чинної технології обробки даних.

Усі численні функції СУБД можна згрупувати так:

1. *Управління даними.* Завданнями управління даними є підготовка даних та їх контроль, занесення даних до бази, структуризація даних, забезпечення їх цілісності, секретності.

2. *Доступ до даних.* Пошук і селекція даних, перетворення даних на форму, зручну для подальшого використання.

3. *Організація і ведення зв'язку з користувачем:* ведення діалогу, видача діагностичних повідомлень про помилки в роботі з БД і т.д.

Крім БД і СУБД до складу АБД входять мовні, технічні та організаційні засоби. Розглянемо кожний із них.

Мовні засоби потрібні для опису даних, організації спілкування та виконання процедур пошуку і різних перетворень з даними.

Ця класифікація, розроблена американським комітетом CODASYL з проектування та створення БД, має загальний характер і орієнтована на різні СУБД. Проте не кожна СУБД, які тепер використовуються на практиці і поширені на ринку програмних продуктів, має весь набір зазначених мовних засобів.

Мова опису даних (МОД) застосовується на різних рівнях абстракції: зовнішньому, логічному і внутрішньому. Згідно з пропозиціями CODASYL мови опису даних на логічному (концептуальному) і внутрішньому рівнях незалежні і різні. Проте в більшості промислових СУБД немає поділу на дві окремі мови опису логічної і фізичної організації даних, а існує єдина мова, яка ще називається мовою опису схем. Так, наприклад, у відомих і широко використовуваних на практиці СУБД сім'ї DBASE використовується єдина мова опису даних для подання їх на логічному і фізичному рівнях. Ця мова має свій синтаксис. Так, наприклад, ім'я файлу має не перевищувати восьми символів, а ім'я поля — десяти, при цьому кожне ім'я має починатися з букви, поля календарної дати позначаються символом D (DATA), символні поля — С (CHARACTER), числові — N (NUMERIC), логічні — L (LOGICAL), приміток — M

(МЕМО). Опис всіх імен, типів і розмірів полів зберігається в пам'яті разом з даними, ці структури при потребі можна проглянути й виправити

Якщо логічний і фізичний рівні розділені, то до складу СУБД має входити мова зберігання даних.

У деяких СУБД використовується ще мова опису підсхем (МОД — ПС), яка потрібна для опису частини БД, що відбиває інформаційні потреби окремого користувача чи прикладної програми. У складі СУБД сім'ї DBASE така мова не використовується.

Мова опису даних на зовнішньому рівні — це мова, яка використовується для опису потреб користувачів та прикладних програм, при створенні інфологічної моделі БД. Ця мова не має нічого спільного з мовами програмування. Так, наприклад, мовними засобами для інфологічного моделювання є звичайна мова чи її підмножина, а також мова графів і матриць.

Мова маніпулювання даними (ММД) — це мова, яка використовується для обробки даних, їх перетворень і написання програми. ММД може бути базовою чи автономною.

Базова мова (відкриті системи) — це одна з традиційних мов програмування — БЕЙСІК, СИ, ФОРТРАН тощо. Використання базових мов як ММД звужує коло осіб, які можуть безпосередньо звертатися до БД, оскільки для цього потрібне знання мови програмування. У такому разі для спрощення спілкування кінцевих користувачів з БД в деяких СУБД передбачена мова ведення діалогу і мова запитів, які простіші для опанування порівняно з мовою програмування.

Автономна ММД (закрита система) — це власна мова СУБД, яка дає змогу виконувати різні операції з даними. У деяких СУБД з допомогою операторів ММД можна виконувати і опис даних, тобто функції МОД, наприклад, у СУБД типу DBASE немає чіткої межі між операторами МОД і ММД.

До інших мовних засобів можна віднести мову ведення словника даних.

До технічних засобів АБД належать процесори, пристрої вводу і виводу даних, запам'ятовуючі пристрої, модеми, канали зв'язку. У кожному конкретному разі залежно від особливостей СУБД та особливостей об'єкта управління проектується і різна конфігурація технічних засобів. У технічній документації на СУБД зазначається мінімальна конфігурація технічних засобів, яка необхідна для організації БД, а також подаються різні обмеження на склад і кількість технічних засобів.

Тепер поряд з універсальними технічними засобами запроваджуються спеціальні машини баз даних, які безпосередньо призначені лише для зберігання та ведення баз даних. Потреба створення спеціальних машин БД, які реалізують на апаратному рівні

функції СУБД, пов'язана з необхідністю звільнення обсягів пам'яті, що мають відводитися для зберігання даних на тих ЕОМ, які виконуватимуть операції обробки даних.

Організаційні засоби АБД охоплюють персонал, який пов'язаний зі створенням і веденням БД, а також систему нормативно-технологічної і інструктивно-методичної документації з організації та експлуатації БД.

Певна посадова особа чи група таких осіб, які забезпечують створення, ведення і підтримку БД в актуальному стані, називається адміністрацією АБД. Основні функції адміністратора: спільна робота з проєктувальниками задач для визначення умов використання БД; розробка опису БД і початкове завантаження її; підтримка цілісності БД, організація захисту зберігання даних; відновлення БД при виникненні помилок програмного забезпечення чи збоїв пристроїв, які призводять до руйнування БД; нагромадження статистики щодо роботи з БД, реорганізація та реструктуризація БД згідно зі зміною потреб.

В умовах використання ПЕОМ функції адміністратора виконуються користувачем і частково програмістом, відповідальним за супровід тієї чи іншої системи. Користувач відповідає за завантаження БД та її підтримку в актуальному стані, програміст — за функції відновлення БД у випадках її зруйнування.

4. Характеристика інфологічної та даталогічної моделі баз даних

Проєктування даних пов'язане з багаторівневим їх поданням: зовнішнім, інфологічним, даталогічним, внутрішнім.

Зовнішній рівень являє собою вимоги до даних з боку користувачів і прикладних програм.

Вимоги користувачів до зовнішнього подання охоплюють сукупність даних, які потрібні для виконання запитів користувачів.

Вимоги з боку прикладних програм до зовнішнього рівня подання даних — це перелік даних з описом їх взаємозв'язків, які необхідні для реалізації певних функціональних задач.

Зовнішній рівень являє собою, як правило, словесний опис даних та їх взаємозв'язків і відбиває інформаційні потреби користувачів і прикладних програм. Іноді для опису зовнішнього рівня використовуються матричні або інші формалізовані методи. Опис зовнішнього рівня не виключає наявності дублювання, надлишковості, неузгодженості тощо.

Для того щоб спроектувати зовнішню модель БД, необхідно виконати обстеження ПО, вивчити систему вхідної і вихідної документації, дослідити й вивчити всі функціональні обов'язки майбутніх користувачів БД.

Американський комітет CODASYL пропонує три рівні: зовнішній, концептуальний, внутрішній. Іноді для зручності проектування вводять допоміжний рівень (проміжний), який називають інфологічним. Він може бути й самостійним або функціонувати як складова зовнішнього рівня.

Інтеграція всіх зовнішніх зображень виконується на інфологічному рівні. На цьому рівні формується інфологічна (канонічна) модель даних, яка не є простою сумою зовнішніх зображень даних.

Інфологічний рівень являє собою інформаційно-логічну модель (ІЛМ) предметної області, в якій виключена надмірність даних і відображені інформаційні особливості об'єкта управління, без урахування особливостей і специфіки конкретної СУБД.

Мета інфологічного проектування — створити структуровану інформаційну модель ПО, для якої розроблятиметься БД. При проектуванні на інфологічному рівні створюється інформаційно-логічна модель, яка має відповідати таким вимогам:

- коректність схеми БД, тобто адекватне відображення модельованої ПО;
- простота і зручність використання на наступних етапах проектування, тобто ІЛМ має легко відображатися в моделі БД, що підтримується відомими СУБД (сіткові, ієрархічні, реляційні);
- ІЛМ має бути описана мовою, зрозумілою проектувальникам БД, програмістам, адміністратору і майбутнім користувачам АБ.

Основною складовою інфологічної моделі є атрибути, які потрібно проаналізувати і деяким чином згрупувати для подальшого зберігання в БД.

Сутність інфологічного моделювання полягає у виділенні інформаційних об'єктів ПО (файлів), які підлягають зберігання в БД, а також визначенні характеристик об'єктів і зв'язків між ними. Характеристиками об'єктів є атрибути.

Даталогічний (логічний, концептуальний) рівень формується з урахуванням специфіки і особливостей конкретної СУБД. На цьому рівні будується концептуальна модель даних, тобто спеціальним способом структурована модель ПО, яка відповідає особливостям і обмеженням вибраної СУБД. Модель логічного рівня, яка підтримується засобами конкретної СУБД, іноді називають даталогічною. Залежно від типів моделей, які підтримуються засобами СУБД, є ієрархічні, сіткові і реляційні

моделі баз даних. Найпоширенішими на сучасному ринку програмних продуктів є реляційні СУБД (DBASE 111, FOXBASE, FOXPRO, CLIPPER і т.ін.).

Внутрішній рівень пов'язаний з фізичним розміщенням даних у пам'яті ЕОМ. На цьому рівні формується фізична модель БД, яка містить структури зберігання даних в пам'яті ЕОМ, включаючи опис форматів записів, їхнє логічне або фізичне упорядкування, розміщення за типами пристроїв, а також характеристики і шляхи доступу до даних.

Від параметрів фізичної моделі залежать такі характеристики функціонування БД: обсяг пам'яті і час реакції системи. Фізичні параметри БД можна змінювати в процесі її експлуатації (не змінюючи при цьому опису інших рівнів) з метою підвищення ефективності функціонування системи.

Визначення структури масивів БД відбувається на етапах інфологічного і логічного проектування, а формування структури — на етапі фізичного проектування БД.

Структура файлу — це поименована сукупність логічно взаємопов'язаних атрибутів.

5. Методи створення оптимальної моделі баз даних

Під оптимальною логічною моделлю баз даних розуміють модель, яка не має аномалій, пов'язаних з модифікацією БД, тобто проблем, що можуть виникнути у зв'язку із замінами, вставками і вилученнями даних із БД.

Для створення такої моделі баз даних назалежно від того, яка СУБД використовується — ієрархічна, сіткова чи реляційна — застосовується теорія нормалізації реляційних баз даних. Використання реляційного підходу дає змогу спроектувати оптимальну логічну модель БД, яка потім досить просто трансформується в ієрархічну чи сіткову модель.

В основу реляційних моделей покладено поняття відношення, яке подають у вигляді двовимірної таблиці.

Реляційна БД — це набір взаємопов'язаних відношень. Кожне відношення (таблиця) в ЕОМ подається як файл. Відношення можна поділити на два класи: об'єктні і зв'язкові.

Об'єктні відношення зберігають дані про інформаційні об'єкти предметної області.

Наприклад:

клієнт (код клієнта, назва клієнта, адреса, телефон) є об'єктним відношенням.

В об'єктному відношенні один з атрибутів однозначно ідентифікує окремий об'єкт. Такий атрибут називається *первинним* ключем відношення. В наведеному відношенні роль ключа виконує атрибут «код клієнта».

Ключ може вмикати кілька атрибутів, тобто бути складеним. В об'єктному відношенні не повинно бути рядків з однаковим ключем, тобто не допускається дублювання об'єктів. Це основне обмеження реляційної моделі для забезпечення цілісності даних.

Зв'язкове відношення зберігає ключі двох або більше об'єктних відношень. Ключі зв'язкового відношення мають на меті встановлення зв'язків між об'єктними відношеннями.

Наприклад, розглянемо ще одне об'єктне відношення БАНК(код банку, назва банку, адреса банку).

Тоді зв'язкове відношення БАНК-КЛІЄНТ (код банку, код клієнта) буде сполучним між двома об'єктними відношеннями БАНК і КЛІЄНТ. У зв'язковому відношенні можуть дублюватися ключові атрибути. Крім ключів, за якими встановлюють зв'язок у зв'язковому відношенні, можуть бути ще й інші атрибути, які функціонально залежать від цього складового ключа.

Ключі в зв'язкових відношеннях називаються зовнішніми ключами, оскільки вони є первинними ключами інших відношень. Реляційна модель накладає на зовнішні ключі обмеження, яке називають посилковою цілісністю. Воно необхідне для забезпечення цілісності даних. Це означає, що кожному зовнішньому ключеві має відповідати рядок якогось об'єктного відношення. Без такого обмеження може статися так, що зовнішній ключ посилається на об'єкт, про який нічого не відомо.

У реляційній БД накладається ще одне обмеження — відношення мають бути нормалізовані.

Контрольні запитання

1. Розкрийте поняття машинного інформаційного забезпечення.
2. Дайте означення машинній інформаційній базі.
3. Що таке масив даних машинної ІБ?
4. Які передумови виникнення баз даних?
5. Назвіть основні переваги організації ІЗ в у вигляді автоматизованих банків даних.
6. Дайте означення автоматизованому банку даних.
7. Як класифікуються АБД?
8. Що таке база даних? Система керування базами даних (СКБД)?
9. Які функції виконує СКБД?
10. Мета інфологічного та деталогічного проектування баз даних?
11. Які методи створення оптимальної моделі баз даних.