**Лекція 3. Алгоритми і структури даних**

Питання студентам для контролю знань попередньої лекції і матеріалу для самостійного вивчення.

1. Назвіть основні етапи розв’язування задач за допомогою комп’ютера.
2. Що таке «припустима елементарна операція»?
3. Як Ви можете визначити інформаційну систему, чим вона характеризується?
4. Які основні характеристики інформаційної системи?
5. В чому полягає місія та основні задачі інформаційної системи.?
6. Які компоненти ІС і як вони приймають участь в обробці інформації?
7. Які основні функції ІС?
8. Назвіть етапи розвитку інформаційних систем.
9. В чому полягає системний підхід?
10. Назвіть етапи розвитку інформаційних технологій?
11. Які типи ІС Ви знаєте?
12. На які групи можна поділити задачі обробки інформації?
13. Що таке програмне забезпечення?
14. За якими ознаками можна класифікувати програмне забезпечення?
15. Що забезпечують прикладні програми?
16. Які функції виконують системні програми?
17. Для чого використовуються сервісні програми?
18. Що таке операційна система?
19. Які сучасні операційні системи Ви знаєте?
20. Що входить до функцій операційної системи?

Сучасний рівень розроблення програмного забезпечення передбачає застосування структур даних, як необхідних атрибутів програмних конструкцій. Атрибут – в перекладі з латинської /присукупленний/, істотний признак, властивість будь чого, невід’ємна приналежність предмету тощо. Програмування передбачає формулу представлення: **структуризація предмета програмування + структури даних + алгоритми + програма + реалізація алгоритму**. Так як дані можуть бути різними (числа, символи, строки тощо), тому вони мають різну структуру представлення, що необхідно враховувати в програмуванні. Історично можна виділити три етапи розвитку структуризації (організації структур даних) в мовах програмування.

- Перша спроба організації структур – через адресність (пам’яті) машин.

- Друга спроба – через типи даних.

- Третя спроба – повна типізація на всіх рівнях обробки програм і конструювання типовості структур даних.

Існує понад 5 тис. середовищ програмування (інженерні, економічні, інтелектуальні, тощо) зі своїми технологіями організації структур даних. Сучасне системне програмування застосовує технологію третього етапу організації структур даних. Вирішення задач предметних областей потребують структуризації даних, розробки алгоритмів, представленням яких є алгоритмічна програма (програма) написана на тій чи іншій мові. Реалізація алгоритмів за їх представленням проводиться людиною, штучним виконавцем тощо. Ефективність представлення і виконання алгоритму рішення задачі залежить від вибраних структур даних.

**Інформація і дані**

Будь який предмет природи або штучний об’єкт «породжують» інформацію. Людина сприймає інформацію з інформаційного простору через свої сенсори зір, слух та інше. Інформація може бути визначена як відображення реального світу, яке має властивості: збереження, обробки, передачі. Інформація має носій, на який вона накладається. Носієм інформації є текстове, звукове, відео або інше повідомлення. Програма, як об’єкт ЕОМ є не тільки послідовність строк, операторів деякої штучної мови програмування, але і набір інформаційних об’єктів імен даних над якими виконуються ті чи інші дії операторів програми. Отже програма є носієм інформації. Програмний носій як правило представляється у вигляді текстів повідомлення, утворених на алфавітах природних або штучних мов, або конструктивних об’єктів, за допомогою яких будуються фрагменти мов. Символ – базовий об’єкт для побудови конструктивних об’єктів текстів повідомлення.

***Інформація*** – відображення реального світу, яке має властивості: здобування, передавання, збереження (інформації). Носієм інформації є повідомлення, яке формально складається з символів, позначок та іншого.

***Дані*** – інформація у вигляді повідомлення спеціальним чином представлена технічними пристроями, наприклад, ЕОМ або людиною

**Структура даних.**

**Структура даних (СД)** - загальна властивість інформаційного об'єкта, з яким взаємодіє та або інша програма. Ця загальна властивість характеризується:

* множиною допустимих значень цієї структури;
* набором допустимих операцій;
* характером організованості.

Найпростіші структури даних називаються також ***типами даних***.

Правильний підбір структур даних є надзвичайно важливим для ефективного функціонування відповідних алгоритмів їх опрацювання. Добре побудовані структури даних дозволяють оптимізувати використання машинного часу та пам'яті комп'ютера для виконання найбільш критичних операцій.

Відома формула **≪Програма = Алгоритми + Структури даних≫** дуже точно виражає необхідність відповідального ставлення до такого підбору. Тому іноді навіть не обраний алгоритм для опрацювання масиву даних визначає вибір тієї чи іншої структури даних для їх збереження, а навпаки.

У програмуванні та комп'ютерних науках структури даних—це способи організації даних у комп'ютерах. Часто разом зі структурою даних пов'язується і специфічний перелік операцій, які можуть бути виконаними над даними, організованими в таку структуру. Структури даних поділяються на вбудовані (реалізовані в мовах програмування) та похідні (утворюються користувачами). Класифікація СД у програмах користувача та пам'яті комп'ютера подана на рис. 1.



**Рисунок 1.** Класифікація СД

Важливою ознакою для класифікації є зміна структур даних під час виконання програми. Наприклад, якщо змінюється кількість елементів і/або відношення між ними, то такі структури даних називаються динамічними, інакше – статичними.

**Рівні описування структур даних**

Розрізняють наступні рівні описування/подання даних:

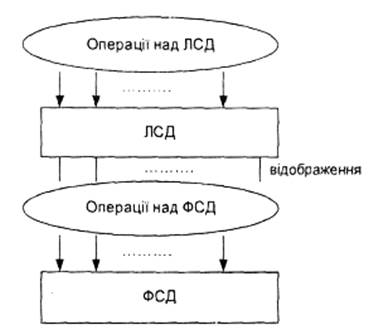
• абстрактний (математичний) рівень;

• логічний рівень;

• фізичний рівень.

**Логічний рівень (ЛСД)**– подання структури даного на тій чи іншій мові програмування.

**Фізичний рівень (ФСД)** — відображення у пам'ять комп'ютера інформаційного об'єкту відповідно до логічного описування.

Оскільки пам'ять комп'ютера обмежена, то виникають питання розподілу пам'яті й керування нею.

Логічний і фізичний рівні відрізняються один від одного, тому в обчислювальних системах здійснюється відображення фізичного рівня на логічний і навпаки (рис. 2).

**Рисунок. 2.** Зв'язок між логічним та фізичним рівнями подання СД.

Структурування інформації пов'язане з необхідністю її зберігання, обробки чи передачі. Структуру економічної інформації визначає її будова, відокремлення тих чи інших елементів. Ці елементи називають інформаційними одиницями. Вони можуть бути простими або складними. Прості елементи не піддаються подальшому поділу. З них утворюються складні, формуються різні рівні структурної побудови інформації.

З погляду логіки управління та розміщення інформації на носіях розрізняють логічну та фізичну структуру даних.

Логічне структурування інформації виділяє елементи в залежності від їх функціонального призначення та особливостей. Для економічних задач це такі як: символ, реквізит/атрибут, показник, інформаційне повідомлення, інформаційний масив, інформаційний потік, інформаційна підсистема, інформаційна система.

Символ - це найпростіший елемент даних, сигнал інформації (літера, цифра, знак), який окремо не має змісту.

**Реквізит - інформаційна одиниця найнижчого рівня, яка складається з цифр, літер, символів і має зміст.** В сучасній термінології найчастіше замість терміну «реквізит» використовують «атрибут», що пов'язано з використанням саме цього терміну в міжнародній економічній літературі.

Реквізит відображає окремі властивості об'єктів - кількісні або якісні. Тому реквізити бувають двох видів: реквізити-ознаки та реквізити-основи (реквізити-величини). Реквізит-ознака (область, матеріал, спеціальність) описує якісні властивості об'єкта чи обставини, за яких відбувався той чи інший процес. Реквізит-основа (сума, дата, ціна) розкриває абсолютне або відносне кількісне значення реквізиту-ознаки.

Реквізити можуть бути різного типу: числового, текстового, логічного, дата тощо.

У випадку машинного представлення інформації синонімами поняття "реквізит" можуть бути "поле", "елемент", "атрибут". У спеціальній літературі вживають також інші синоніми реквізиту - "терм", "ознака" і. т.п.

Розрізняють форму і значення реквізитів. Форма реквізиту включає найменування і структуру (формат).

Найменування служить для звернення до реквізиту. Наприклад: "Оклад", "Посада".

Структура реквізиту - це спосіб подання його значень. Вона включає довжину і тип. Довжина - це кількість символів, що утворюють значення реквізиту. Наприклад, "Код працівника" може містити три позиції, "Код підприємства" - 10 позицій; "Ідентифікаційний код фізичної особи" містить 10 позицій.

Значеннями реквізитів є послідовності символів (літер, цифр, різних знаків і спеціальних позначень). Під час обробки інформації над реквізитами-основами виконують арифметичні операції, а за допомогою реквізитів-ознак здійснюють пошук інформації, її сортування, вибірку, порівняння (логічні операції). Однорідні реквізити-ознаки об'єднуються у номенклатуру (наприклад, номенклатура товарів).

З реквізитів утворюється показник, що характеризує певний об'єкт з кількісного та якісного боків. Це найменша інформаційна одиниця, з якої утворюється самостійний документ. Сутність економічної інформації розкривається через економічний показник. Наприклад, показник "Об'єм продажу ноутбуків фірми Acer складає 33 тисячі одиниць" є носієм кількісної та якісної характеристики відповідної величини. Показники є основними одиницями інформації, за допомогою яких формуються бази даних.

Сукупність показників, достатня для характеристики певного процесу (явища, факту), утворює повідомлення. Наприклад: вхідні дані надходять в інформаційну систему у вигляді інформаційних повідомлень.

Однорідні повідомлення, об'єднані за певною ознакою, складають інформаційний масив даних. Прикладом масиву може бути сукупність даних про рух грошових коштів на підприємстві.

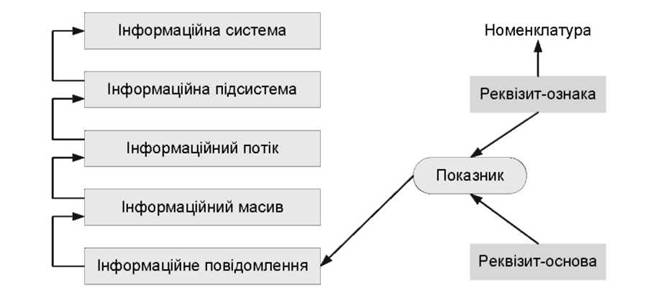
Масив є основною структурною одиницею при автоматизованій обробці інформації, зокрема при запису даних в пам'ять машини.

**Інформаційний потік - сукупність масивів, що відносяться до однієї з частин процесу управління об'єктом**

Для інформаційних технологій важливим є визначення інформаційних потоків від джерел інформації до користувача. Сукупність інформаційних потоків, які характеризують роботу, пов'язану з виконанням певної функції чи з діяльністю певної галузі, називають інформаційною підсистемою.

**Інформаційна система (ІС) - сукупність інформаційних підсистем, що характеризують управління об'єктом загалом**

ІС є структурною одиницею вищого рівня і охоплює всю інформацію об'єкта (цеху, підприємства, установи, організації, галузі):



**Рисунок 4**. Взаємозв'язок між елементами логічної структури інформації

**Одиниці фізичної структури даних**

З точки зору подання інформації на певних носіях (фізичне структурування) відповідні одиниці визначаються залежно від носія інформації та способу її фіксації. Це пов'язано з розміщенням масивів даних у пам'яті ПК. Як правило, виділяють такі одиниці фізичної структури даних (від найнижчої до найвищої): символ, поле, агрегат даних, запис, файл, база даних.

Поняття символу наведено вище.

**Поле** - множина символів, яка створює мінімальний семантичний елемент масиву.

**Агрегат даних** - це пойменована сукупність двох і більше елементів нижчого рівня, яка має окремий зміст . До агрегату даних можуть належати як елементи, так і інші агрегати даних. Прикладом агрегату даних можуть бути групи елементів, які утворюють "Адресу" або "Дату народження".

**Запис** - пойменована сукупність полів, об'єднаних за змістовним принципом.

Агрегати даних і записи реалізуються на практиці шляхом організації списків, черг, стеків, таблиць.

**Файл** - іменована сукупність записів про об'єкти одного типу. Як правило, записи, що входять у файл, мають однакову структуру. Прикладом файлу може бути сукупність записів про колір пікселів, що складають зображення.

**База даних** - це іменована сукупність взаємозв'язаних даних, що відображає стан об'єктів і їх відношення в даній предметній області. Наприклад, база "Студент" містить інформацію про вік, стать, домашню адресу, успішність та інші дані про студента.

Організація даних у базі характеризується певною структурою, тобто формою ї способом їх впорядкування. За характером взаємозв'язків елементів усі структури даних можна поділити на лінійні та нелінійні.

До лінійних структур належать послідовні структури, в яких елементи розміщуються в тому порядку, який необхідний під час їх обробки, наприклад, список, елементами якого є записи (рядкова структура).

До нелінійних структур даних належать складні списки, дерева, мережі, табличні та гібридні структури:

* складні списки містять вкладені елементи (підсписки) меншого обсягу;
* деревоподібні структури - елементи розміщуються на різних рівнях і сполучаються за допомогою адрес;
* сіткові структури являють собою розширення дерева за рахунок нових адрес зв'язку;
* табличні структури даних призначені для зберігання інформації про ключові ознаки даної інформаційної сукупності;
* гібридні структури даних містять фрагменти різних структур.

Зі структурою пов'язана **модель даних** - фіксована система понять і правил для представлення структури даних.

Існує декілька моделей даних, які становлять основу інформаційних систем. Найчастіше використовуються три: ієрархічна, мережева (сіткова) та реляційна. Останнім часом з'явився четвертий тип: об'єктно-орієнтовані моделі даних, які використовують поєднання традиційної та об'єктної моделей. В моделі сіткових баз використовують представлення даних у вигляді довільного графу. В ієрархічних - дані представляються у вигляді ієрархічної (деревоподібної) структури. Найпоширеніші моделі баз даних - реляційні. В них дані представляються у вигляді таблиць. Таке подання даних, що складається зі стовпців, які розташовуються у певному порядку зліва направо, називається відношенням. Тому ці таблиці називаються реляційними (від англ. relation - відношення).

Під час проектування інформаційних систем та розробленні алгоритмів оброблення даних в них необхідно враховувати такі властивості економічної інформації:

* вхідна інформація, в основному, фіксується в первинних документах, які не завжди придатні для автоматичного введення в комп'ютерну пам'ять;
* ті самі вхідні дані використовуються багаторазово для здобуття показників у різних економічних розрізах для всіх служб і видів господарської діяльності;
* основна частина економічної інформації підлягає періодичному, регулярному оновленню;
* здобута вихідна інформація часто використовується як вхідна при подальших розрахунках;
* економічна інформація характеризується тривалістю збереження.

**Базові структури даних**

Поняття структури даних можна визначити як особливу систему організації взаємозв’язаних елементів даних. Структура самих елементів залежить від конкретної задачі. Вони можуть бути як дуже простими (наприклад, цілі числа або строки символів) так і дуже складними (наприклад, для зображення матриці часто використовують одномірний масив, кожний елемент якого в свою чергу також є одномірним масивом). Розглянемо деякі структури даних, що найчастіше зустрічаються.

**Лінійні структури даних**

Серед простих структур даних можна виділити два найважливіших – одномірний масив і зв’язаний список.

**Одномірний масив** – це послідовність із n однотипних елементів, що розташовані підрід в оперативній пам’яті комп’ютера, доступ до яких виконується по значенню індексу масиву:

|  |  |  |  |
| --- | --- | --- | --- |
| Елемент [0] | Елемент [0] | … | Елемент [n-1] |

В більшості випадків індекс масиву, що складається із n елементів, є цілим числом, що має значення від 0 до n-1 або від 1 до n В деяких мовах програмування допускається, щоб індекс набував любих цілочисельних значень, включаючи від’ємні. Головне, щоб він не виходив за межі раніше встановленого діапазону, що встановлює верхню і нижню межу зміни індексу.

Іноді в мовах програмування для доступу до елементів масиву можуть використовуватись нечислові індекси. Такі масиви називаються асоціативними. Наприклад, можна створити масив із 12 елементів, що відповідає місяцям року, доступ до яких здійснюється по назві місяців.

Час звернення до любого елементу масиву однаковий і не залежить від його положення в масиві. Ця особливість відрізняє масиви від зв’язаних списків, мова про які піде пізніше. Проте кожен елемент масиву займає однакову кількість чарунок пам’яті в комп’ютері.

На основі одномірних масивів створюються різні структури даних, серед яких особливу увагу заслуговують рядки.

*Рядки* – послідовність символів раніше обумовленого алфавіту, що закінчується особливим символом, що служить ознакою кінця рядка. Рядки, елементами яких є лише 0 та 1, називаються двійковими (бінарними) або бітовими.

Рядки є основним елементом, що використовується при обробці текстових даних, визначення синтаксису мови програмування, компілювання створених на ній програм, а також при вивченні абстрактних обчислювальних моделей.

Операції, що виконуються над масивами, залежать від його типу – операції, що виконуються над рядками, відрізняються від операцій, що виконуються над масивом чисел (визначення довжини рядків, об’єднання двох рядків для створення однієї, порівняння двох рядків).

**Зв’язаний список** – послідовність (в т.ч. і порожня) кількох елементів даних, що називаються вузлами. В кожному вузлі зберігається інформація двох видів: самі дані та одна або більше посилань на інші вузли списку, що називаються вказівниками. Якщо поточний елемент є останнім, то в нього поміщується нульовий вказівник зі спеціальним значенням NULL – це вказує на те, що вказівник ні на що не вказує.

В *однонапрямленому* зв’язаному спискові кожний вузол вміщує лише один вказівник, в якому міститься посилання на наступний елемент списку, або NULL, якщо поточний елемент є останнім.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Елемент [0] |  |  | Елемент [1] |  | … | Елемент [n-1] | Null |

Для досягнення заданого елемента зв’язаного списку необхідно обрати перший вузол списку, потім пересуватися по ланцюгу елементів до необхідного вузла. Тому час доступу до конкретного елемента списку залежить від місця його положення в списку і може дуже суттєво відрізнятися від того, що затрачується в масиві для подібної операції. Це є недоліком зв’язаного списку. Перевага зв’язаного списку полягає в тому, що для списку немає необхідності резервувати оперативну пам’ять комп’ютера. Крім того, вставка та видалення елементів списку не завдають труднощів і виконуються достатньо швидко шляхом зміни вказівників.

Існує модифікація структури однонапрямленого зв’язаного - двонапрямлений зв’язаний список. Відміна між ними в тому, що кожен вузол двонапрямленого списку крім першого та останнього включають в себе вказівники на попередній та наступний вузли:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Null | Елемент [0] |  |  |  | Елемент [0] |  | … |  | Елемент [n-1] | Null |

Масив та зв'язаний список частіше всього використовують для зображення ще однієї абстрактної структури даних – лінійний список або просто список. Список являє собою скінчену послідовність елементів даних, що організовані в заданому лінійному порядку. До цієї структури даних можуть бути застосовані наступні операції: пошук, видалення та додавання елемента.

Серед всієї множини можна виділити два типи списків: стеки та черги.

***Стек*** – це такий список, в якому можна виділити тільки останній елемент, а новий елемент можна додати тільки в кінець. Цей кінець називають вершиною стеку, оскільки стеки зазвичай зображують на малюнках у вигляді вертикального прямокутника. Принцип роботи стека – LIFO (last-in-first-out), оскільки першим видаляється елемент, котрий був останнім у стеці. Цей стек нагадує купку тарілок, оскільки ми можемо взяти з неї тільки верхню тарілку, а нову покласти наверх купки. Стеки широко використовуються на практиці – без них не обійтися при реалізації рекурсивних алгоритмів.

***Черга*** – це такий список, елементи якого видаляються з однієї сторони структури (вона називається головою), а додаються з іншої (вона називається хвостом). Перша операція називається видаленням елемента з черги, а друга – постановкою в чергу. Принцип роботи черги – FIFO (first-in-first-out). Аналогія – черга в магазині. Область використання – алгоритми розв’язку задач теорії графів.

В багатьох випадках необхідно вибрати серед множини елементів, що динамічно змінюється, елемент з найвищим пріоритетом. Для розв’язку таких задач використовують структуру даних, що називається чергою з пріоритетами або просто пріоритетною чергою. До основних операцій над чергою з пріоритетами відносяться: пошук найбільшого елемента, видобування найбільшого елемента та додавання нового елемента. Дві останні операції повинні бути реалізовані так, щоб в результаті їх виконання утворилась нова черга с пріоритетами або перевпорядкована стара. Найпростіше реалізувати таку структуру на основі масиву або впорядкованого масиву, хоча так не досягається максимальна ефективність. Кращою є реалізація на основі структури даних, що називається пірамідою.

**Найпростіші стандартні та складені типи даних.**

**Перелік спискових типів даних, до яких належать: лінійні списки, деревоподібні структури та графи**

**Складеним типом даних** назвемо тип даних, що складається із скінченної та наперед заданої множини елементів певного типу, які не обов'язково є атомарними.

Перерахуємо складені типи даних та дамо їм коротку характеристику.

**Множина** – скінчена сукупність елементів, в якої R = Ø.

**Послідовність** — абстрактна структура, у якої множина R складається з одного відношення лінійного порядку (тобто для кожного елемента, крім першого і останнього, є попередній і наступний елементи).

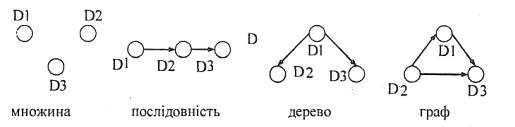
**Матриця** – структура, в якої множина R складається із двох відношень лінійного порядку.

**Дерево** – множина R складається з одного відношення ієрархічного порядку.

**Граф** – множина R складається з одного відношення бінарного порядку.

**Гіперграф** – множина R складається із двох і більше відношень різного порядку.

Приклади СД подано на рис. 5.



**Рисунок 5**. Приклади подання структур даних

***Для самостійного вивчення*:** Поглибити матеріал лекції за наданою літературою. Вивчення лекційного матеріалу та додаткових джерел. Розгляд запитань і виконання завдань для самостійної роботи, запропонованих на лекції.

***Література***

1. Ковалюк Т. В. Алгоритмізація та програмування: Підручник. — Львів: «Магнолія 2006», 2013. — 400 с., ил.
2. Джейс Либерти Освой самостоятельно С++ за 21 день: 3-е изд. пер. с англ.: Уч. пос. – М.: Издательский дом „Вильямс”, 2001. – 816 с.: ил..
3. Куприянова Л.М. Программирование, алгоритмические языки и вычислительная математика: Учеб. пособие. — М.: Финансы и статистика, 1985. —223 с.

**Контрольні питання**

1. Надайте визначення інформації та даних.
2. Як пов’язані інформація і дані?
3. Поняття структури даних. Як вона характеризується?
4. Класифікація структур даних у програмах користувача.
5. Визначте рівні подання структур даних.
6. Назвіть елементи логічної структури економічної інформації.
7. Надайте визначення інформаційного масиву даних та інформаційного потоку.
8. Окресліть взаємозв'язок між елементами логічної структури економічної інформації.
9. Визначте одиниці фізичної структури даних.
10. Наведіть приклади лінійних та нелінійних структур даних.
11. Чим відрізняється стек від черги?
12. Назвіть складені типи даних.