

# 明德資訊及通訊科技

► 資訊處理

## 第二課 數據控制及數據組織





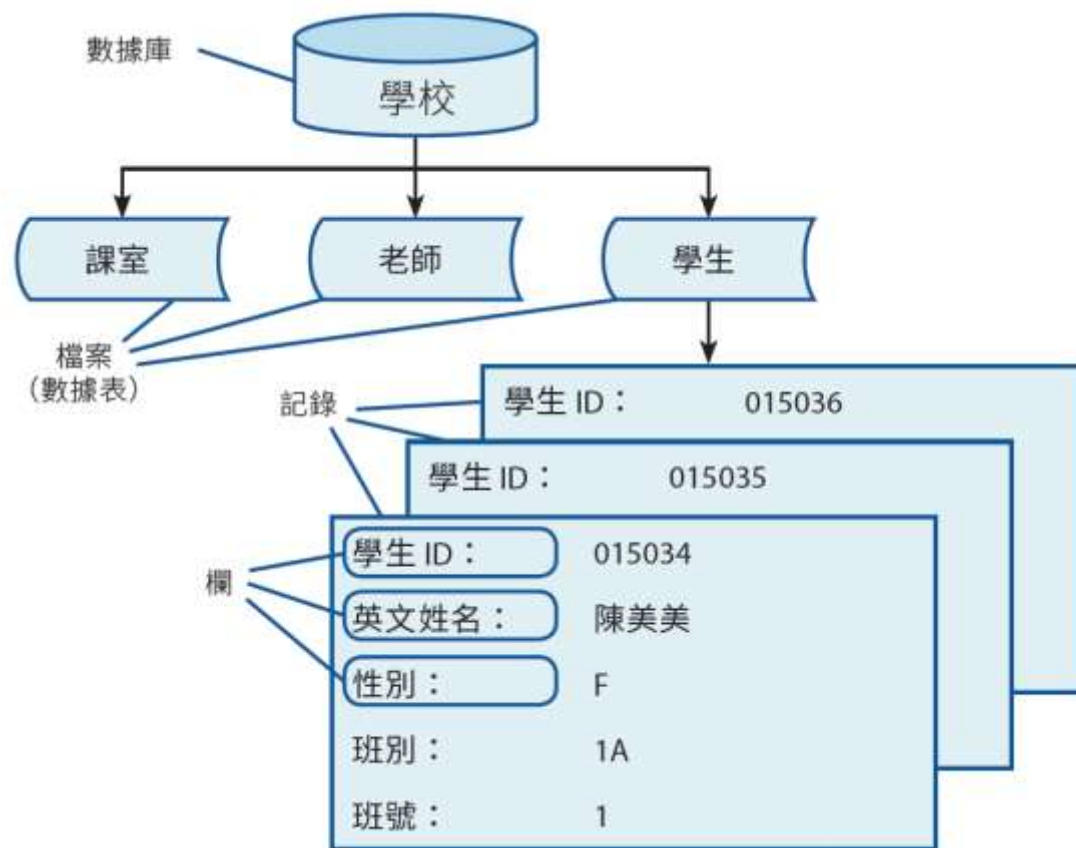
## 為什麼醫生能知道你的醫療記錄？

2008年，香港政府開發了一個電子平台「醫健通」。此平台會展示所有參與者的電子健康記錄，將參與者的個人資料數碼化，並儲存於數據庫中。

這就是組織和控制數據的例子。

## 2.1 數據分級

數據通常以**分級形式**儲存。在數據分級中，每個較高層的數據都是較低層項目的集合。



## 2.2

## 數據控制

數據控制指管理數據的過程。此過程包括：

- 數據有效性檢驗
- 數據驗證

## A

## 無用輸入，無用輸出

「無用輸入，無用輸出」指輸入不可靠或不協調的數據用作數據分析，導致無法得出有用的輸出結果。數據輸入錯誤包括：

| 錯誤     | 說明             | 例子                                     |
|--------|----------------|--|
| 數據來源錯誤 | 數據來源提供了不正確的數據。 | 用戶錯誤地提供了一個不存在的電子郵件地址。                  |
| 轉錄錯誤   | 數據被誤讀、誤聽或誤植。   | 打字員將「冶」打錯為「治」。                         |
| 换位錯誤   | 數據位置不正確。       | 某人於記錄一項交易時誤將兩個相鄰數字調轉，令「\$15」變成「\$51」。。 |

## B

# 數據有效性檢驗

數據有效性檢驗可確保該數據合理並符合特定規則。

數據有效性檢驗通常由系統或程式運行，並被廣泛應用於線上平台。

## 申請表

|                |   |                                  |
|----------------|---|----------------------------------|
| 學生 ID          | <input type="text" value="s034930"/>      | 擁有此 ID 的學生已經申請了。                 |
| 姓名             | <input type="text"/>                      | 必須輸入姓名。                          |
| 性別             | <input type="text" value="15"/>           | 性別必須為 M 或 F。                     |
| 年齡             | <input type="text" value="M"/>            | 年齡必須大於 12。                       |
| 電話號碼           | <input type="text" value="2590872"/>      | 電話號碼必須為 8 個位。                    |
| 電郵地址           | <input type="text" value="alexchan.com"/> | 電郵地址必須包含「@」。                     |
| 可出席日期          |   |                                  |
| 開始日期           | <input type="text" value="31/08/2025"/>   | <input type="button" value="📅"/> |
| 結束日期           | <input type="text" value="01/08/2025"/>   | <input type="button" value="📅"/> |
| 結束日期必須在開始日期之後。 |   |                                  |

**B**

## 數據有效性檢驗

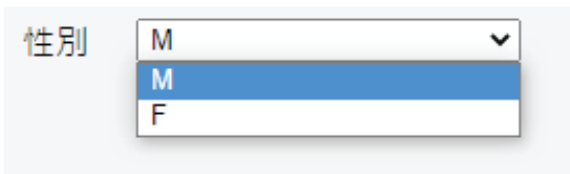
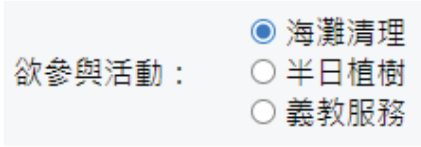


數據有效性檢驗有許多不同類型，包括：

| 類型    | 功能                | 例子                 |
|-------|-------------------|--------------------|
| 完整性檢查 | 確保必填欄位中有數據。       | 申請人必須提供姓名。         |
| 固定值檢查 | 確保數據只能來自一系列特定的值。  | 申請人的性別必須是「M」或「F」。  |
| 類型檢查  | 確保數據必須為特定類型。      | 申請人的年齡必須是數字。       |
| 範圍檢查  | 確保數據必須位於指定範圍內。    | 申請人的年齡必須大於 12 歲。   |
| 長度檢查  | 確保數據的字符數必須在指定範圍內。 | 申請人必須提供 8 個位的電話號碼。 |
| 格式檢查  | 確保數據必須採用特定格式。     | 申請人的電郵地址必須包含「@」符號。 |

## B

## 數據有效性檢驗

不同用戶界面有助使有效性檢驗更容易進行。

| 類型    | 有效性檢驗   | 例子  |
|-------|---|---|
| 下拉式選單 | 固定值檢查：只可輸入其中一個特定值。  |    |
| 單選按鈕  | 固定值檢查：只可輸入其中一個特定值。  |    |
| 複選框   | 固定值檢查：只可輸入特定值，可選取多個值。   |   |
| 日期選擇器 | <p>類型檢查：只可輸入日期。用戶不能選擇像「32/08/2025」這樣的無效日期。</p> <p>格式檢查：日期必須為特定格式，如「DD/MM/YYYY」。</p> |  |



## B

## 類型檢查

為方便組織和分析數據，同一欄位的數據應屬同一數據類型。

$$(\text{"八十五"} + \text{"Seventy two"} + 91 + 79.5) \div 4 = \text{ERROR}$$
$$(85.0 + 72.0 + 91.0 + 79.5) \div 4 = 81.875$$

## B

## 類型檢查

以下為常用於數據欄位的數據類型：

| 數據類型     | 說明                            | 數據值例子    |       | 用途例子           |
|----------|-------------------------------|----------|-------|----------------|
| 整數       | 可以是正數或負數。                     | 74       | -29   | 學生人數、某人的年齡     |
| 浮點數 / 實數 | 可以是正數或負數。<br>可以有小數點和小數位。      | 2.09     | 1.00  | 某學生的分數、某人的身高   |
| 字串 / 文字  | 由字符組成，其中包括英文字母、阿拉伯數字、中文字、符號等。 | "RA^1\$" | "日，月" | 學生姓名、某人的地址     |
| 布爾       | 只能是TRUE（真）或FALSE（假）。          | TRUE     | FALSE | 某學生是否合格、某人是否在世 |

## B

## 檢查數位

檢查數位是一種數據有效性檢驗方式，常用於數字數據，例如身份證號碼、書號、信用卡號碼等。

檢查數位由特定算法生成，並附加到數據的結尾。檢查數位可以有多於一個數位。

## B

## 奇偶檢測

數據以二進制代碼的形式儲存在電腦中，並僅由 0 和 1 組成。

在二進制代碼的末尾附加類似檢查數位的奇偶檢驗位，可以確保代碼中 1 的總數是偶數或奇數。這就稱為奇偶檢測。

奇偶檢測可分為兩類：

| 原始二進制代碼  | 1 的總數 | 根據奇偶檢測類型添加奇偶檢驗位後的二進制代碼 |           |
|----------|-------|------------------------|-----------|
|          |       | 偶數檢測                   | 奇數檢測      |
| 0000 000 | 0     | 0000 0000              | 0000 0001 |
| 1010 101 | 4     | 1010 1010              | 1010 1011 |



## 例題

### 2.1

1. 要以檢查數位來檢驗一組香港身份證號碼的有效性，我們應：

(1) 確保號碼開首的字符為一或兩個英文字母。

(2) 確保號碼的加權和可被 11 整除。

(3) 確保號碼不可包含多於 9 個字符，括號除外。

A. 只有(1)

B. 只有(2)

C. 只有 (1) 和 (3)

D. 只有 (2) 和 (3)



## 例題

2.1

### 分析

(1) 是格式檢查, (3) 是長度檢查, 兩者都是檢驗香港身份證號碼有效性的合理方法。然而, 只有 (2) 是利用檢查數位進行有效性檢驗。

### 答案

正確答案是 B。



## 例題

2.1

2. 某數據傳輸系統接收了四組二進制代碼。如只有一組代碼通過了奇偶檢測，它是以下哪一項呢？

- A. 1101 1000
- B. 1110 1011
- C. 0010 0101
- D. 0100 1000



## 例題

### 分析

問題沒有指明這裏使用的是哪一種奇偶檢測，但它明確表示只有一組代碼通過了奇偶檢測。因此，我們需要點算每組代碼中 1 的總數。

| 選項 | 二進制代碼     | 1 的總數 |
|----|-----------|-------|
| A  | 1101 1000 | 4     |
| B  | 1110 1011 | 6     |
| C  | 0010 0101 | 3     |
| D  | 0100 1000 | 2     |

只有 1 的總數為單數的代碼會通過檢測。

### 答案

正確答案是 B。





## 測試站

### 2.1

試完成課本第 40 頁至第 41 頁的測試站 2.1。

1. B
2. D
3. C

## C

## 數據驗證

數據驗證是數據控制的一部分，能夠協助驗證數據在媒體之間傳輸時是否保持準確和一致。

不同類型的數據驗證包括：

| 類型                   | 說明  |
|----------------------|---|
| Inputting data twice | It requires an operator to input the same data twice.                                 |
| Double data entry    | It requires two operators to input the same data independently.                       |
| Proofreading data    | It requires another operator to compare the input data against the original document. |



## 例題

### 2.2

1. 阿發有一個網上論壇的帳戶。他想更改帳戶的電郵地址時，論壇要求他填寫以下的網上表單。阿發嘗試提交表單時，發生了一些錯誤。這張網上表單應用了以下哪些數據控制方式？

(1) 類型檢查

(2) 完整性檢查

(3) 輸入數據兩次作驗證

- A. 只有 (1) 和 (2)
- B. 只有 (1) 和 (3)
- C. 只有 (2) 和 (3)
- D. (1)、(2) 和 (3)

|             |   |              |
|-------------|---|--------------|
| 舊電郵地址：      | <input type="text"/>                          | * 此欄位必須填寫    |
| 新電郵地址：      | <input type="text" value="fredchan.ict.gov"/> | * 無效的電郵地址    |
| 重新輸入新的電郵地址： | <input type="text" value="fredchan@ict.gov"/> | * 請輸入相同的電郵地址 |



## 例題

2.2

### 分析

「舊電郵地址」一欄不可留空。這意味著系統應用了完整性檢查，以確保該欄已被填妥。

「新電郵地址」一欄有錯誤訊息，指該電郵地址無效。原因是電郵地址並無包含「@」符號，所以格式不正確。這項檢查是格式檢查而非類型檢查。

每名用戶均須輸入其新電郵地址兩次，代表系統用到輸入數據兩次作驗證。

### 答案

正確答案是 C。



1. 操作員正將紙本檔案中的幾個 8 位二進制代碼輸入到資訊系統中，但其中一個 8 位二進制代碼被輸入錯誤，如下所示：

原始二進制代碼

0010 0000

錯誤輸入的二進制代碼

0100 000

以下哪項/些數據控制方式可以幫助操作員發現此錯誤？

- (1) 雙重數據輸入
  - (2) 奇偶檢測
  - (3) 長度檢查
- 
- A. 只有 (1)
  - B. 只有 (2)
  - C. 只有 (1) 和 (3)
  - D. 只有 (2) 和 (3)

## 2.3 數據庫的功能

你知道大量的數據（如圖書館中的圖書數據）是如何組織、儲存和提取的嗎？  
此過程須要運用數據庫。



# A

## 組織

數據庫中的數據是根據數據庫結構來組織的。

良好的數據庫結構可以使不同數據表之間的傳遞更流暢，並減少數據重複。

舉例來說：

| 欄位名稱  | 數據類型    | 說明  |
|-------|---------|---|
| SID   | String  | The unique student ID of a student. A number cannot have leading zeros, thus it has to be stored as text. |
| NAME  | String  | The name of a student.  |
| SEX   | String  | The sex of a student.   |
| CLASS | String  | The class that a student belongs to.  |
| CNO   | Integer | The class number of a student.  |

## B

## 儲存

我們可以在數據庫中插入新記錄、更改或刪除現有記錄。

結構化查詢語言 (SQL)

```
INSERT INTO STUDENT (SID, ENAME, SEX, CLASS, CNO)  
VALUES ("015044", "Peter Chan", "M", "1A", "11");
```

圖形用戶界面 (GUI)

| SID    | ENAME           | SEX | CLASS | CNO |
|--------|-----------------|-----|-------|-----|
| 015034 | Suzanne Rivera  | F   | 1A    | 1   |
| 015035 | Zack Ashton     | M   | 1A    | 2   |
| 015036 | Roberta Barrera | F   | 1A    | 3   |
| 015037 | Caspian Browne  | M   | 1A    | 4   |
| 015038 | Nolan Humphrey  | M   | 1A    | 5   |
| 015039 | Kalvin Kearney  | M   | 1A    | 6   |
| 015040 | Gracie Manning  | F   | 1A    | 7   |
| 015041 | Aaisha Proctor  | F   | 1A    | 8   |
| 015042 | Eilidh Rennie   | M   | 1A    | 9   |
| 015043 | Liliana Sykes   | F   | 1A    | 10  |
| 015044 | Peter Ch        |     |       | 0   |



## C

## 提取

我們可以使用 SQL 或通過 GUI 從數據庫中提取特定數據。

結構化查詢語言 (SQL)

```
SELECT * FROM STUDENT WHERE SEX = "M"
```

圖形用戶界面 (GUI)

| SID    | ENAME          | SEX | CLASS | CNO |
|--------|----------------|-----|-------|-----|
| 015035 | Zack Ashton    | M   | 1A    | 2   |
| 015037 | Caspian Browne | M   | 1A    | 4   |
| 015038 | Nolan Humphrey | M   | 1A    | 5   |
| 015039 | Kalvin Kearney | M   | 1A    | 6   |
| 015042 | Eilidh Rennie  | M   | 1A    | 9   |
| 015044 | Peter Chan     | M   | 1A    | 11  |

## C

## 提取

我們可以為頻繁搜尋的欄位建立索引，以加快提取數據的過程。

為 ENAME 加上索引會建立一個索引表，其中僅包含索引欄位和指向相關記錄的指標。

索引表

| ENAME           | 指標 |
|-----------------|----|
| Aaisha Proctor  | 8  |
| Caspian Browne  | 4  |
| Eilidh Rennie   | 9  |
| Grace Manning   | 7  |
| Kalvin Kearney  | 6  |
| Liliana Sykes   | 10 |
| Nolan Humphrey  | 5  |
| Peter Chan      | 11 |
| Roberta Barrera | 3  |
| Suzanne Rivera  | 1  |
| Zack Ashton     | 2  |

數據表

| SID    | ENAME           | SEX | CLASS | CNO |
|--------|-----------------|-----|-------|-----|
| 015034 | Suzanne Rivera  | F   | 1A    | 1   |
| 015035 | Zack Ashton     | M   | 1A    | 2   |
| 015036 | Roberta Barrera | F   | 1A    | 3   |
| 015037 | Caspian Browne  | M   | 1A    | 4   |
| 015038 | Nolan Humphrey  | M   | 1A    | 5   |
| 015039 | Kalvin Kearney  | M   | 1A    | 6   |
| 015040 | Grace Manning   | F   | 1A    | 7   |
| 015041 | Aaisha Proctor  | F   | 1A    | 8   |
| 015042 | Eilidh Rennie   | M   | 1A    | 9   |
| 015043 | Liliana Sykes   | F   | 1A    | 10  |
| 015044 | Peter Chan      | M   | 1A    | 11  |

## 2.4 檔案存取方式

數據可以儲存在不同的儲存設備中，它們決定了存取數據的方式。

常見的檔案存取方式有兩種：

順序存取：  
存取第五項元素前先處理前面所有元素。



直接存取：  
存取第五項元素前不須處理前面所有元素。

# A

## 順序存取

使用順序存取的儲存設備會按順序從頭開始寫入數據。順序存取的存取速度，平均而言會比直接存取慢。

只會使用順序存取的儲存設備：



磁帶

## B

## 直接存取

使用直接存取的儲存設備可以在任何儲存位置寫入數據。直接存取的存取速度一般更穩定，且速度比順序存取快。

能使用直接存取的儲存設備：



硬碟



隨身碟



## 例題

### 2.3

1. 一位老師想將1A班學生的資料儲存在一個數據庫表格中。數據表的結構如下：

| 欄位     | 數據類型 | 說明 | 例子  |
|--------|------|----|-----|
| CNO    | 數字   | 班號 | 1   |
| NAME   | 文本   | 姓名 | 陳大文 |
| GENDER | 文本   | 性別 | M   |

以下哪種 / 些類型的有效性檢驗適合運用到 CNO 欄位中？

- (1) 唯一性檢查
  - (2) 類型檢查
  - (3) 完整性檢查
- A. 只有 (2)
- B. 只有 (1) 和 (2)
- C. 只有 (2) 和 (3)
- D. 只有 (1)、(2) 和 (3)



## 例題

2.3

### 分析

就 (1) 而言，由於表中的所有資料都來自同一班的學生，即 1A 班，所以學生的班號不應重複。

就 (2) 而言，由於 CNO 欄位中的數據類型為數字，故必須進行類型檢查。

就 (3) 而言，每個學生都必須有班號。

### 答案

正確答案是 D。



## 測試站

2.3

試完成課本第 51 頁的測試站 2.3。

1. B

2. C