

データの活用 informaticsI-032

教科書 pp.119-120, p.124, pp.140-143

データの収集

- 問題解決のために、データの収集が役に立つ。インターネットの利用、図書館・書籍の利用、フィールドワークなどによりデータの収集が可能である。
 - インターネットの利用
 - Web ページの検索
 - 検索エンジン (Google や Yahoo, DuckDuckGo など) を用いて、誰かの作成した Web サイトから情報が得られる
 - AND 検索や OR 検索、NOT 検索を利用すると情報を見つけやすい
 - オープンデータ
 - インターネットなどを通じて配布される、二次利用可能で無償で利用できるプログラムで扱いやすいデータのこと
 - 図書館・書籍の利用
 - 書籍や新聞、学術雑誌などから、信頼性の高い情報を得ることができる
 - 司書に相談することで目的に合った資料を見つけてもらえる
 - 出版年や著者等を確認することで情報が古すぎないか、信頼できる情報かを確認できる
 - フィールドワーク
 - 実際の現場に行って観察や調査、アンケートを行い、直接データを集められる
 - 自分で収集したデータであるため、目的に合ったデータを得やすい
 - 調査対象によってはすべてのデータを集めるのは難しいため、一部のデータからデータ全体を推測することがある
 - 一部のデータを取り出すことを標本調査 (サンプリング) という

ファイル・フォルダ (ディレクトリ) によるデータの整理

- コンピュータ内でデータの整理を行う際は、フォルダ (ディレクトリ) を活用するとよい
 - 同じ分類のデータを同じフォルダにいれ、分類名をつけておくとファイル名を見つけやすい (趣味・勉強など)
 - フォルダの中にフォルダを作るなどして階層構造を作ることも役に立つ (勉強→高校→1年生→3学期→情報Iなど)

データベースとその必要性

- ある目的のためにデータを蓄積し、データをまとめ、複数のプログラムで共有できるようにして

利用価値を高めたものを(①

)という

- 自治体の住民情報、図書館の蔵書管理、Web ページの検索エンジンなど
- クラス出席簿・学年全体の生徒情報が別々で管理されているなどファイル・フォルダ管理の場合は、転校・転学が発生したときに両方のデータを書き換える必要があるが、データベース管理であれば、一つのデータを書き換えるだけでよい

データベース管理システム (DBMS (Data Base Management System))

- データベースは共有利用を前提としているが、機密情報や個人情報へのアクセスは制限して機密性を高める必要がある。また、可用性を高めるためにバックアップなども重要である。データベースの効率性を維持して安全に利用できるようにするのが DBMS である。

機能の例

- データの一貫性: 複数のユーザが同時にデータ操作をしても矛盾が生じないようにする
- データの整合性: データの重複や不正データの登録・更新を防ぎ、データの品質を維持する
- データの独立性: データベースとそれを利用するプログラムを独立して管理する
- データの機密性: 操作できるユーザのアクセス権を設定したりユーザ認証したりする
- データの可用性 (障害対策): 障害時のためのバックアップを作成する

リレーションナルデータベース (関係データベース)

- 表形式でデータを格納しているデータベース
 - 行 (レコード) と列 (フィールド) の二次元で構成され、行と列を指定することによってデータを取得する
 - 行によって、誰の情報かや何の情報かが区別され、列によって項目が決まる (行は英語基準、列は行じゃない方向)

- 行を特定する項目を(②) と呼ぶ (ある表においては他の行との重複は許されない)

- 複数の項目を組み合わせて行を特定するとき、

- 複数の項目をまとめて(③) と呼ぶ

- ほかの表と関連付けを行うための項目を(④) と呼ぶ

書籍表

主キー	列	項目	外部キー
行	書籍番号	書籍名	著者名
	913-36xx	津軽	太宰治
	913-75xx	孤高の人	新田次郎
	923-13xx	阿Q正伝	魯迅
	934-27xx	森の生活	ソロー

分類表

主キー	分類コード	分類
910	日本文学	
920	中国文学	
930	英米文学	
940	ドイツ文学	

表に対して行える操作は結合 (表を合わせる)・選択 (行を絞り込む)・射影 (列を絞り込む) の三つである

- 結合 (書籍表と分類表の二つの表を合わせて表示)

書籍番号	書籍名	著者名	分類コード	分類
913-36xx	津軽	太宰治	910	日本文学
913-75xx	孤高の人	新田次郎	910	日本文学
923-13xx	阿Q正伝	魯迅	920	中国文学
934-27xx	森の生活	ソロー	930	英米文学

- 選択 (書籍表の一部を行で絞り込みして表示 (分類コード910のみを表示))

書籍番号	書籍名	著者名	分類コード
913-36xx	津軽	太宰治	910
913-75xx	孤高の人	新田次郎	910

- 射影 (書籍表の一部を列で絞り込みして表示 (書籍名と著者名のみを表示))

書籍名	著者名
津軽	太宰治
孤高の人	新田次郎
阿Q正伝	魯迅
森の生活	ソロー

その他のデータベース

- リレーションナルデータベースでは SQL で操作するが、その他のデータベースは SQL を用いない。したがって、その他のデータベースを NoSQL と呼ぶことがある。ビッグデータに対応したデータベースを作る目的で使用されることがある。

- 例: キー・バリュー型、カラム指向型、グラフ指向型など