

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

データの収集

- 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

データの収集

- 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

データの収集

- 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

- ・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データベースという
 - 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

データベースとその必要性

- ・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)

- 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。

データベース管理システム(DBMS)

- 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

データベース管理システム(DBMS)

- 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):

障害時のためにバックアップを作成する

リレーショナルデータベース(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

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	

参照

リレーショナルデータベース(関係データベース)

- 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)

書籍表

主キー	列	項目	外部キー	分類表
行	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
	923-13xx	阿Q正伝	魯迅	920
	934-27xx	森の生活	ソロー	930

参照

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

リレーションナルデータベース(関係データベース)

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

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	

参照

リレーションナルデータベース(関係データベース)

- 複数の項目を組み合わせて行を特定するとき、複数の項

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	

参照

リレーションナルデータベース(関係データベース)

- ほかの表と関連付けを行うための項目を**外部キー**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	

参照

リレーショナルデータベース(関係データベース)

- 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

リレーショナルデータベース(関係データベース)

- 結合(書籍表と分類表の二つの表を合わせて表示)

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	

参照



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

リレーションナルデータベース(関係データベース)

- 選択(書籍表の一部を行で絞り込んで表示)

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	

参照

↓

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

リレーションナルデータベース(関係データベース)

- 射影(書籍表の一部を列で絞り込んで表示)

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	

参照



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