

コンピュータ概論B - ソフトウェアを中心に -

#14 データベース (後編)

Yutaka Yasuda

データベース

- 外見

データを集合・蓄積したもの
一定のフォーマット、レコードの存在

- 目的

入力・更新 / 高速な検索、再利用 / 共有

- 内部構造

DBMS の仲介によってデータの一貫性保持と保護を実現
大量データ処理のためのさまざまな工夫の集合体

種類

- 幾つかのタイプ：データモデルに適したものを選ぶ
- カード型
 図書館蔵書カードのような一件一枚のもの
- ネットワーク(型)データベース
 データの親子関係に注目
- リレーショナル(型)データベース (Relational Database)
 データの関係 (relation) に注目
 現在もっとも良く使われている
- 学生情報データベースを考える

カード型による学生情報データベース

- 一人一件

- 利点

全ての情報がカードの中にあるのでカードを見つけられればあとの処理が簡単

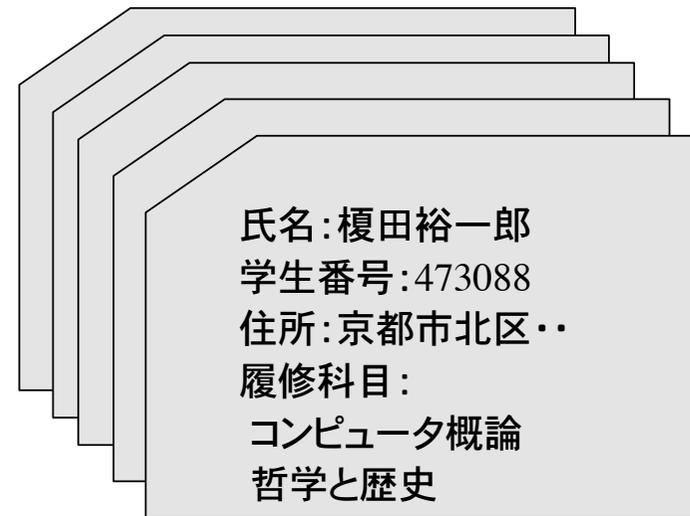
- 欠点

柔軟な検索が出来ない

キー以外の検索は一枚一枚繰ることには？

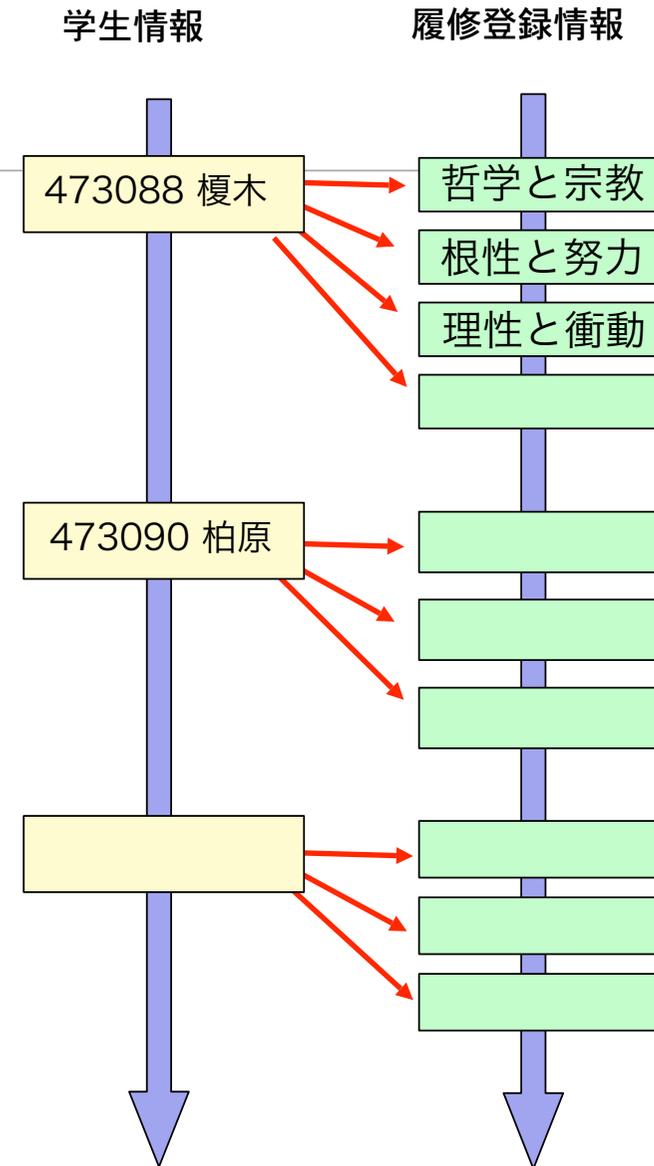
通常はキーでソートして検索を容易にする

インデックス（複数）の利用



ネットワーク型データベース

- データの親子関係に注目
- 利点
 - よく適合する用途には非常に高速
(検索処理が実質不要)
- 欠点
 - 柔軟なデータ構成がとれない (設計時に完全に決定しておかなければならない)
 - 適合しない用途が後から現れると非常に効率が下がる
 - 例：三回生の多い履修科目は？
- 銀行、業務システムなど変化の少ない用途には向いている



関係データベース

- Codd (1970, IBM) が理論的モデルを提唱
- データを表組みで表現
- 表と表の関係処理を集合演算モデルで定義
- 数学的に完成したモデルと言える

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1

関係データベース

- 歴史

1973 の SystemR (IBM), Ingres (UCB バークレー校)

1979 Oracle

SQL の発明 (1986, ANSI 標準となる)

現在もっとも市場で多く使われているタイプ

- Ingres

UCB で開発され、商用化

後の PostgreSQL (Post Ingresから)

RDB における表

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1

- データは表形式

行と列による表現

多様なデータを表と項目
の関係で記述

GNO	GAKUHI	SIHARAI
473088	1223000	643000
859674	1200000	1200000

- 学生情報一人分は：

学生レコード一行

学費レコード一行

履修登録レコード複数行

GNO	KAMOKU	UNIT
473088	科学と哲学	4
473088	基礎演習	2
473088	人生航路	4
859674	科学と哲学	4

RDBにおける演算

- 集合と見なして演算
- 部分集合

GNOが473088の行
を抜く

GNOとGAKUBUだ
けを取り出す

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1

GNO	GAKUBU
473088	E
859674	B

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2

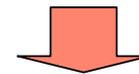
RDBにおける演算

- 足す (集合和)
- 同じ項目名の行をそのまま加える (追加する)

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1



GNO	NAME	GAKUBU	GAKUNEN
785412	暁三四郎	E	1
325698	空手一大	J	3



GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1
785412	暁三四郎	E	1
325698	空手一大	J	3

RDBにおける演算

- 表どうしを結ぶ

共通の項目(key)で
突き合わせ

JOIN

GNO	NAME	GAKUBU	GAKUNEN
473088	榎田裕一郎	E	2
859674	明日田勇作	B	1

GNO	GAKUHI	SIHARAI
473088	1223000	643000
859674	1200000	1200000

GNO	NAME	GAKUBU	GAKUNEN	GAKUHI	SIHARAI
473088	榎田裕一郎	E	2	1223000	643000
859674	明日田勇作	B	1	1200000	1200000

RDBMS

- RDB の DBMS

 - テーブル、項目の管理

 - ログ管理

 - アクセス制限

- 特徴

 - SQL 問い合わせ言語によるアクセス

 - 簡易問い合わせシステムがついている

 - プログラムからもSQLでアクセス

SQL

- Full Spec 無し(略語ではない)
元はあったが今は SQL として仕様化
- 集合演算をプログラミング言語風に簡略化
- 選択

```
SELECT * FROM GAKUSEI  
WHERE GAKUBU="E"
```

```
SELECT * FROM GAKUHI  
WHERE SIHARAI > 600000
```

SQL

- 選択（項目抜きだし）

```
SELECT GNO, GAKUBU FROM GAKUSEI
```

- 突き合わせ

```
SELECT * FROM GAKUSEI, GAKUHI  
WHERE GAKUSEI.GNO = GAKUHI.GNO
```

- カウント他

```
SELECT COUNT(*) FROM GAKUSEI  
WHERE GAKUBU="E"
```

```
SELECT GNO, GAKUHI-SIHARAI  
FROM GAKUSEI, GAKUHI  
WHERE GAKUSEI.GNO = GAKUHI.GNO
```

関係データベース

- 利点

 - 柔軟、プログラムとデータが独立

 - SQL という問い合わせ言語の便利さ

 - 数学的完全性

- 欠点

 - 概して低速

 - データ格納効率が高くない

- 動かしながら開発したり将来変更があるシステムに向く

- 現在もっとも多く市場で使われているタイプ

Open Source Software Project

- LAMP or LAPP

Linux / Apache / MySQL (or PostgreSQL) / Perl

- 業務利用に耐えうるDBMS

過去には商用のみ

現在はオープンソースのものが多い

OpenSource Software 普及の一つの原動力

Oracle / DB2 も Open になった