

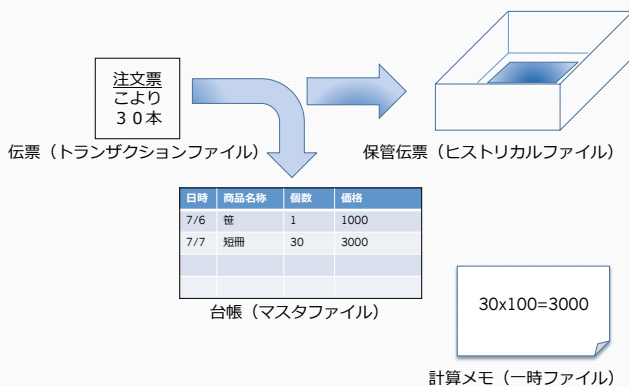
コンピュータ基礎(11)

10章 ファイルとデータベース

ファイルの種類(1)

- 使い道による分類
 - **基本ファイル** (基本ファイル)
 - 業務の根幹となるファイル。台帳的な性格を持つ。
 - **発生ファイル** (発生ファイル)
 - マスタファイルに対する更新情報を格納している。
 - これから処理する伝票のような性格のファイル。

ファイル種類のイメージ



ファイルの種類(2)

- ファイルの利用期間による分類
 - **パーマネントファイル** (パーマネントファイル)
 - そのシステムが使われている間は保存され、繰り返し利用される。
 - マスタファイルは一般に永久ファイルである。
 - **テンポラリーファイル** (テンポラリーファイル)
 - 処理中に、一時的に必要なファイル。
 - 必要なときに作成され、必要がなくなった時点で消去される。
 - 作業中のメモ書きのようなもの。

ファイルの種類(3)

- 利用者による分類
 - **システムファイル** (オペレーティングシステム) が使用するファイル。
 - OSのプログラムそのもの
 - OSの制御用・管理用の情報 など。
 - **ユーザーファイル** (計算機の利用者のファイル)。
 - 利用者が使用する情報を格納したファイル
 - 利用者が作ったり、インストールしたプログラム

ファイルの入力(1)

- 入力変換・媒体変換
 - 入力装置から、補助記憶装置にデータを入力する。
 - 誤りがないかどうかチェックする必要がある。
- **入力装置**の方法(1)
 - **目視チェック**: 目視検査。人手で読みあわせをするなど。
 - **数字チェック**: 数字検査。数字が入るべきところに数字以外の文字が入っているかどうかなど。
 - **限界チェック**: 限界検査。数値の範囲を調べる。たとえば、月なら1~12でなければおかしいなど。
 - **順番チェック**: 順番検査。たとえば、カードの学籍番号が順番になっているかどうかなどを調べる。



ファイルの入力(2)

- データチェックの方法(2)
 - 合計チェック**: 合計検査. 合計の数値が合うかどうかを検査する.
 - 平衡チェック**: 平衡検査. 一致するはずの項目同士を比べ, 合っているかチェックする.
 - 付加チェック**: 検査用の数字を付加しておき, それが正しいかどうかを調べる.
 - たとえば, センター試験の受験番号 1234X では, Xがチェックディジットになっている.
 - 全ての図書につけられている図書コード (ISBNコード) も最後の数字がチェックディジットになっている.

ISBNコードの例

- 右の書籍の例
 - ISBN-13 978-4-274-21014-3
- 頭の数字から順に, 1, 3, 1, 3, ... をかけた和を計算する
 - $9 \times 1 + 7 \times 3 + 8 \times 1 + \dots = 97$
- 合計をで割った余り (つまり, 1 の位) を求め, 10 から引く
 - この場合, $10 - 7 = 3$



センター試験の受験番号の例

- 試験場コードと受験番号から計算

には数字6桁の「試験場コード」と数字4桁+アルファベット1文字の「受験番号」が割り振られた。アルファベットはABCHKMRUXYZの11文字のどれかであり、受験番号はたとえば「1234X」となる。このアルファベットは**チェックディジット**である。試験場コード ($d_1 d_2 d_3 d_4 d_5 d_6$ としよう) と受験番号の数字4桁 ($d_7 d_8 d_9 d_{10}$) とから, それぞれに1~10をかけて, 和をとって

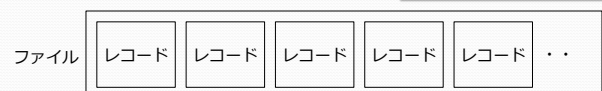
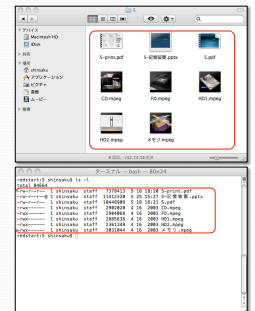
$$d_1 + 2d_2 + 3d_3 + 4d_4 + 5d_5 + 6d_6 + 7d_7 + 8d_8 + 9d_9 + 10d_{10} \equiv x \pmod{11} \quad (1)$$

と11で割った余りxについて, 下の表を引くことで一意に定まる。

表	11で割った余り	0	1	2	3	4	5	6	7	8	9	10
アルファベット		A	B	C	H	K	M	R	U	X	Y	Z

ファイルとレコード

- ファイルとは?
 - 補助記憶装置に置かれた, 1まとまりのデータ
 - ハードディスク
 - USBメモリやメモ리카ード
 - 主記憶 (メインメモリ) に読み込まれたデータではない
- レコード**とは?
 - ファイルの中に格納されるひとかたまりのデータ

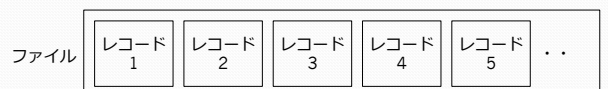


ファイルのアクセス方法(2)

- 順次アクセス
 - 順次呼び出し, シーケンシャルアクセス
 - 記録されている順番にアクセスする
 - 磁気テープは順アクセスしか出来ない
- 直接アクセス
 - ランダムアクセス
 - 記録順にかかわらず, 必要なレコードにアクセスする
 - 磁気ディスク装置などで可能

ファイル編成(1)

- ファイル編成: ファイル中へのレコードの配置方法
- 連続ファイル**
 - ファイルの先頭からレコードが順に記録されている



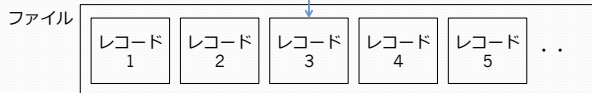
- 制御情報などがなく, 記録効率が良い (無駄がない)
- 磁気テープに使用可能
- ×ある特定のレコードだけをアクセスしたくても, 直接アクセス出来ないで, 先頭から特定のレコードまで読んでいかねばならない
- ×途中でレコードを挿入するには, 新しいファイルを作る必要がある

ファイル編成(2)

ファイル

- キー値からデータ格納場所を計算してアクセス

キー値 (例えば, 学生番号) → 計算 → レコードの位置

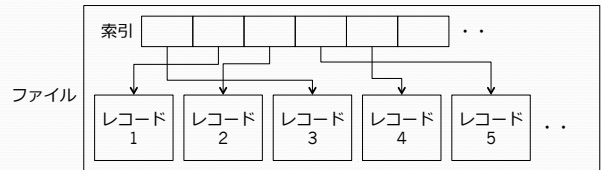


- 直接アクセスが最も速い
- ×ファイル内に空き領域ができてしまうことがある
- ×異なるキーが同じレコードを指す場合がある (工夫により回避する必要がある)

ファイル編成(3)

ファイル

- 順編成ファイルに索引をつけたもの

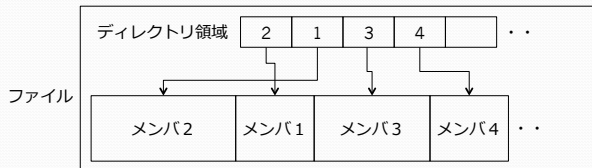


- ・索引 (インデックス域) を持つ
- 直接アクセス (ランダムアクセス) が出来る
- 順アクセスも出来る
- ・普通, レコードの大きさは一定 (固定長という)

ファイル編成(4)

ファイル

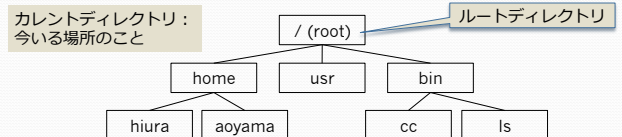
- 大きさの異なるデータを登録簿に基づいて格納



- メンバの大きさはまちまちで良い
- ディレクトリ領域をもとに直接アクセス可能
- ×メンバの削除が増えると様々な大きさの空きがあるので, 適宜メンバを移動して詰める必要がある

パソコンでは

- パソコン用のオペレーティングシステムは基本的に, 順編成ファイルの機能しか持っていない
- データへのアクセスの高度化は, それぞれのプログラムが自前で処理することで実現することも多い
- データを管理するソフトを別途, インストールして使うことも多い. **DBMS**という (後述).
- 数多くのファイルを整理するために [] が提供されている



ファイルの最適化

- ディスク上のファイルは, 記録・消去を繰り返しているうちに, 切れ切れになっていく
- これを [] または [] という.
- 最適化によりデータを並べ替え直して, 速くする.

●フラグメンテーションがない状態のハードディスク内部イメージ



※ファイルが1箇所に配置されているので読み書きがスムーズ.



※1つのファイルがあちこちに断片化されており, 読み書きに手間取る

データベース(1)

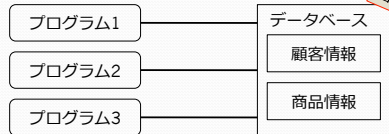
- 業務に用いるデータを管理するシステム
- データベースを使わないと



- それぞれのファイルへのデータ格納方法を, プログラムごとに決める・・・データの共有が難しい
- ファイルの形式を変えるにはプログラムを変えなくてはならない・・・保守コストの増大
- セキュリティやデータの安全性が低い・・・プログラムのミスなどでデータを失いやすい

データベース(2)

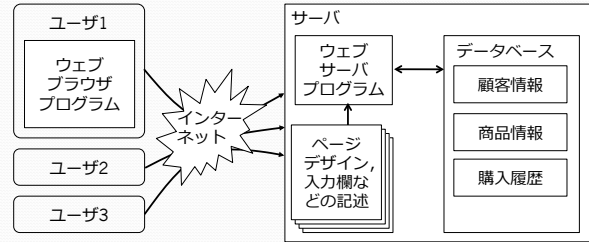
データベースを用いたシステム



- データベースが必要なデータを一元管理する
 - データの安全性が高い。プログラムのミスによりデータが壊れたりしにくい。
 - セキュリティ機能。権限のないユーザからのデータ削除などを禁じることが出来る
 - データをプログラムとは独立に管理できる。バックアップを取ったり、整理をしたり。
- データベースには、専用の言語（データベース言語）を用いてアクセスする。SQLが広く使われている

データベースの利用例

ウェブでのデータの入出力



- インターネットショッピング
- ブログ, SNS(mixi, twitter等)
- 検索エンジン (google等) などなど

データベースについて

- データベースの構造（教科書参照）
 - 階層型データベース・・木構造になっている。
 - ネットワーク型データベース・・網目状になっている。
 - データベース・・複数の表を組み合わせた処理ができるようなデータベース。
- データベースの管理
 - データベース定義・・表の項目を決めるような作業。名簿なら、氏名、住所、電話番号。．．など。
 - データベース操作・・データベースにデータを入れたり、1列分のデータを削除したりする。
 - データベース制御・・データの正しさを検証したり、障害回復などを行う。

データベースの利用

- 「台帳」としてのデータの管理
 - 社員名簿, 顧客名簿, 商品リスト, 売上, . . .
- 商売だけでなく、いろいろなものに使われている
 - ウェブサイトの構築には広く用いられている
 - ブログの各項目。記事内容, 更新日, コメント. . .
 - ニュースサイトの各ニュース記事
 - Wikipedia の各項目 などなど。
- その他の利用方法
 - データベースの内容から法則を探し出し、不正の防止や、よりよいサービス提供などに役立てる。
 - 教科書の例