これで情報Iはバッチリ
共通テスト対策　情報I 図解ノート

共通テスト対策 情報I
生徒用学習支援サイト
で提供している教材の一つです。
https://beyondbb.jp/indexEntInfomatic1.html

この「情報I 図解ノート」は共通テスト対策として、基礎的な知識を学習、習得することを手助けすることを目的に作成しました。

各項目について必要な事項を記憶、理解しやすいように図式、表と簡単な文書で、コンパクト(全部で32ページ)にまとめたものです。

各ページは下半分がノートになっているので、学習を進める中で、補足が必要な内容など書き込んで自分のノートに仕上げてください。

補足: 本ノートは知識中心でプログラミンやシミュレーションの演習問題には対応していません。これらについてはWeb公開している「プログラミンやシミュレーションの問題対策」教材を利用してください。

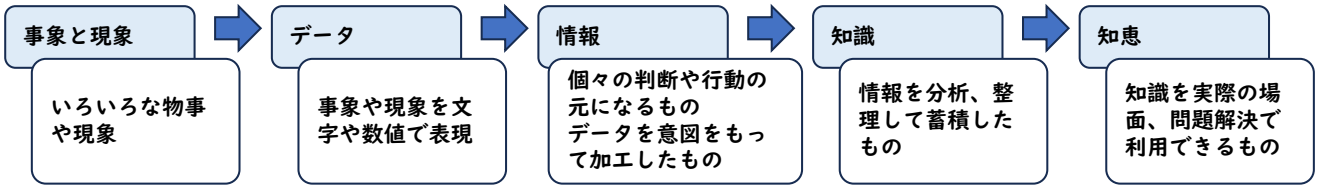
第1版　2024年8月　(太田 剛)

Index

情報の定義と分類

図解ノート

情報の定義



情報の特徴

形がない		他の文字や数値などのシンボルなどで表示や保存可能
簡単に複製できる	複製性	短時間に大量の複製ができる、複製しても劣化しない。
消えない	残存性	長く保存が可能である。複製した場合は、元のものが残る。
容易に伝播する	伝播性	手軽に素早く移動や引き渡しが可能である。

一次情報/二次情報

一次情報	自分で直接的に見聞きした物事や事象。自分で直接的に調査した内容
二次情報	他人が伝えた、調査したデータや情報(政府など権威があっても二次情報)

情報の分類(基礎情報学での分類)

生命情報: 個々の生物が持つ個別に意味付けられた講義の情報:感情、個体の反応等

社会情報: 人間を含む個々の生物が他者とのやりとりを行うための社会的な情報であり、文脈の中で交換、理解される物。 声、ジェスチャーなど

機械情報: 文字や数値など背景・文脈から切り離されたシンボル・記号が独立した情報

メディアとは

図解ノート

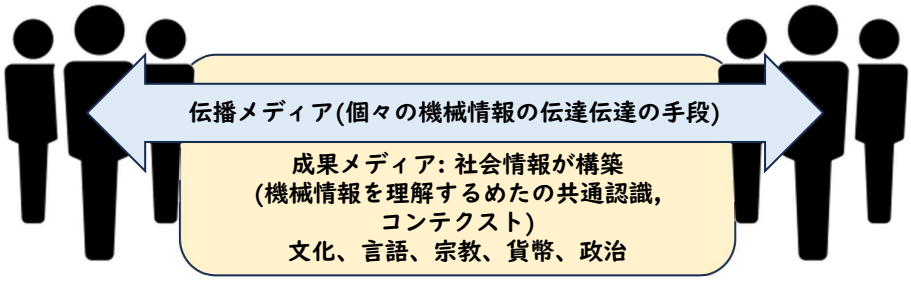
(伝播)メディアの定義: 機械情報を伝える仕組み

表現のためのメディア	情報自体を伝えるために表現する手法: 表現メディアにする段階で加わる情報、削除される情報がある。	文字、音声、音楽、動画、静止画、数値、表、図、グラフなど
伝達のためのメディア	表現されたメディアを伝えるための手段: 大きくテレビなどの方法と、電場などの媒介に分かれる	新聞、テレビ、インターネット、電話など 媒介: 紙、空気、電場、電線など
記録のためのメディア	表現されたメディアを記録しておく手段: 手書き文字などアナログやコンピュータ情報なとのデジタルがある。	ノート、本、ビデオ録画、ハードディスク、USBメモリ、DVDなど

いろいろなメディア: メディアに関連した用語

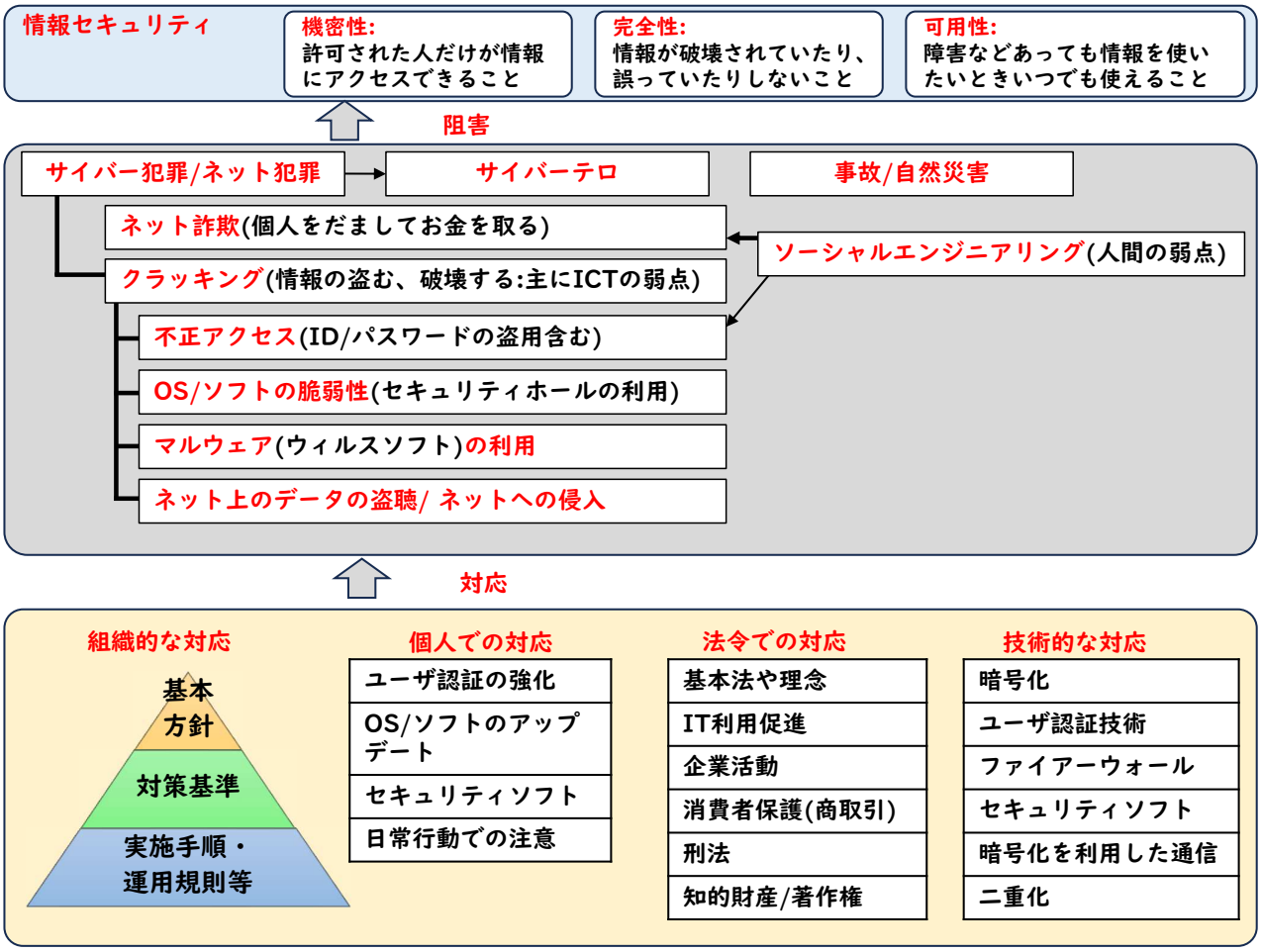
マスメディア	テレビ、新聞など多数の人を対象に一方通行で情報を提供するメディア
ネットワークメディア	インターネットを介したメディア SNS:インスタ,LINE, Xなど他の人とつながり情報を共有できる仕組み CGM(Consumer Generated Media) YouTubeなど利用者が作成したコンテンツを共有する仕組み

発展: 伝播メディアと成果メディア



情報セキュリティ

図解ノート



サイバー犯罪/ネット犯罪

図解ノート

ネットワーク利用詐欺など

- 架空請求
- ワンクリック詐欺(クリックしたとたん入会)
- フィッシング(偽サイト・メールで情報盗む)
- スキミング(暮れ実施
- バイティング(偽のアプリのダウンロード等)

ソーシャルエンジニアリング

- ショルダーハッキング(肩越しに覗き見)
- なりすまし
- ピギーバック(鍵のかかった部屋に一緒に入る)
- ゴミ箱あさり
- 廃棄データ修復

マルウェア(広義のコンピュータウィルス)の分類　(必ずしも明確な分類基準は無い)

- 自己感染機能: 自ら複製して伝染する
- 潜伏期脳: 発病まで潜伏している
- 発病機能: プログラムやデータを破壊したり、意図しない動作を引き起こす

感染方法や挙動による分類

- ウィルス型(ファイル等を経由して増殖感染)
- ワーム型(ネット上で自己増殖)
- トロイの木馬型(時限で発動)
- ボット型(外部からの命令で動作)
- バイティング(偽のアプリのダウンロード等)

被害内容による分類

- キーロガー(キー入力内容を盗む)
- スパイウェア(情報を盗む)
- ランサムウェア(データを暗号化し身代金要求)
- バックドア(外部からの侵入窓口の構築)
- アドウェア(偽の広告や警告の表示)

情報化社会/情報セキュリティに関する法令

図解ノート

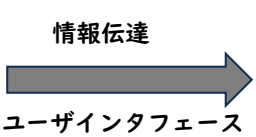
区分	略称(既存法令追加・改定)	概要
基本法	デジタル社会形成基本法	国の大きな方針(以前はIT基本法(略称))
	サイバーセキュリティ基本法	情報セキュリティに関する国の指針に関する法律
利用促進	官民データ活用推進基本法	官民のデータ活用の推進を行う法律
	デジタル手続法	行政のデジタル化、オンライン化について定めた法律
	情報公開法	国や行政機関に対して情報を公開要求する法律
	電子署名法	電子的な署名/認証を認め推進する法律
企業活動	特定電子メール法	迷惑メール等の防止。2008年の改定でオプトイン方式明記
	個人情報保護法	個人情報を扱う企業に対する活動指針に関する法律
	プロバイダ責任制限法	ネット上での誹謗中傷や著作権などの問題に対する責任に関する法律
	出会い系サイト規制法	出会い系サイトの運営方法について定めた法律
	青少年インターネット環境整備法	青少年が有害サイトにアクセスできないように通信業者の運用について定めた法律
	(特定商取引法)	ネット販売、オークションについての商取引の規則。クーリングオフ
消費者保護	(消費者契約法)	消費者と事業者間のルールに関する法律
	電子契約法	消費者のミスなどによる購入に対する法律
刑法等	不正アクセス禁止法	不正アクセスの手段の入手と実施の禁止に関する法律 (不正アクセスするまで、その後で実施したことは刑法等の対象)
	(刑法)	コンピュータ・電磁的記録対象犯罪(データの改ざん、不正使用、盗用、金品の盗用) 不正指令電磁的記録に関する犯罪 (マルウェアの配布と利用) ネットワーク利用犯罪(名誉棄損(誹謗中傷)、わいせつ物の配布、販売等、インターネット上などでの詐欺)

情報デザイン(1)　(いろいろな見方や考え方がある)

図解ノート

コミュニケーションと情報デザイン

情報伝達



ユーザインタフェース
(アフォーダンス)

受け手の特性
・人間としての情報処理能力の限界
・個人としての情報処理能力の特性
・既存知識
・興味/動機

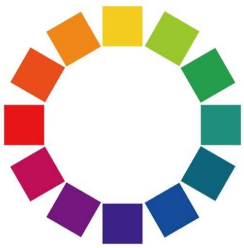
情報デザイン:
受け手の特性を考慮して情報を的確に伝えるための方法

受け手の特性を考慮した情報デザインのいろいろな考え方

- 適切な情報の固まり(チャンク)と構造化
- 5つの整理/分類基準(場所、アルファベット、時間、カテゴリー、階層)
- 視覚化(図表、グラフ化、インフォグラフ)
- 具体化/比喩、抽象化
- 要素の配置(近接、整列、反復、対比/強弱)

抽象化の例:
ピクトグラム

情報デザインの構成要素の一つ(配色の工夫)



色相環:
色の違い(赤、黄色、緑等)の関係を表した図
補色:
向かい合った色(派手な印象になる/目立つ)
類似色:
隣あった色(まとまりのある印象)

情報デザイン(2)　使いやすいデザイン

図解ノート

使いやすさの考え方

ユーザインターフェース

ユーザビリティ

使いやすさの度合い(Usability ＝ use ＋ able)

アクセシビリティ

高齢者・障害者を含むすべての人の使いやすさの度合い・考え方
(Accessibility ＝ Access ＋ able)

バリアフリー

物理的/制度的などの障壁を取り除く
(Barrier-free ＝ Barrier(障害))

ユニバーサルデザイン

具体的考え方

すべての人が使えるデザイン
(Universal Design 普遍的)

	対象/考え方	成果/普及
バリアフリー	高齢者や障碍者を対象に、生活しやすい街や家	主に行政指導型で、建物、道路、交通機関など
ユニバーサルデザイン	すべての人にとって使いやすい物	民間主導で、いろいろな使いやすい道具、製品など

例: Webアクセシビリティ

知覚可能	・背景色や文字色が適切な配色になっている。 ・文字サイズを変更できる。
操作可能	・キーボードだけで操作できる。 ・情報入力に十分な時間がとれる。 ・理解しやすいエラーメッセージを表示する。
理解可能	・できるだけ簡単に用語で説明する。 ・サイト全体で操作性が統一されている。
堅牢	・正しいHTMLで作成されている。 ・外部のテキスト読み上げ機能などと確実に連携できる。

コンピュータのユーザインターフェース(UI)

CUI: キャラクターユーザインタフェース

Character ＝ 文字
文字(テキスト)のUI

GUI: グラフィカルユーザインタフェース

Graphical ＝ 図形的な
図によりUI
デスクトップメタフォー: 机の上の比喩

NUI: ナチュラルユーザインタフェース

Natural ＝ 自然な
音声や身振りなど

問題解決の基礎

図解ノート

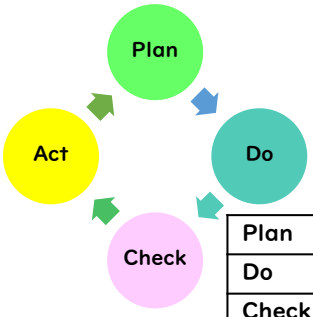
問題解決の一般的にプロセス
(教科書によって、似たようなプロセスや用語が使用されている)

計画
- 問題の発見/問題の明確化
  (現状と目的の整理と分析/定義)
- 情報の収集と整理・分析
- 解決方法の探索/立案
  (計画の立案)

計画・解決方法の実行

評価(反省・振り返り)

実社会でよく使用されているプロセス
PDCAサイクル



Plan	計画
Do	実行
Check	評価
Act	改善

問題解決の関連用語:
フィードバック:問題解決や目標達成に向けて前段階の作業軌道修正を促すための評価や結果の情報
トレードオフ:ある選択をすると、ひとつの課題が改善していても、ほかの課題が悪化し、両立ができないこと

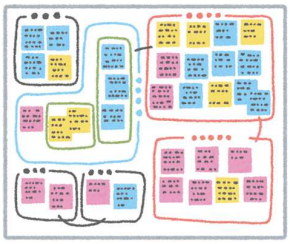
代表的な問題の発見/明確化のフレームワーク(手法)

問題解決を進める上で，情報を収集・分析し，原因や解決策を効果的に考えることは欠かせない。そのために，現状や目に見えないアイデアや考えなどの情報を可視化することは非常に重要である。

ブレーンストーミング:
数人で状況や改善方法を自由に話し合う
ブレーンストーミングのルール

質より量	アイデアの良い悪いは問わず，とにかくたくさんのアイデアを出す
制約を設けない	テーマに関係しそうなことならどんな奇抜な発想でも受け入れる
便乗の奨励	他人の意見を参考にさらに発想を広げるアイデアも歓迎する
批判の禁止	批判すると萎縮してしまい発想の範囲が狭まる恐れがあるため，アイデアを出すときは批判をしない

KJ法
1.アイデアをまとめてカードに記入する(1カード1アイディア)
2.カードをグループに分ける
3.グループ同士の関係を図解化
4.図解化した関係を文章化



その他の手法
・コンセプト(マインド)マップ
・ロジックツリー
・マトリクス表
・フィッシュボーン図
などいろいろ。

知的財産:著作権と産業財産権(1)

図解ノート

知的財産(無体財産)

- 知的財産
  - 著作権
    - 創作活動(文学、学術、美術、音楽等)
    - 無方式主義(創造と同時に自然に発生)
  - 産業財産権
    - 工業、農業等の分野に発明、発見等
    - 方式主義(届け出して発生)
  - その他
    - 回路配置、種苗、営業情報等

著作権の内容

- 著作権
  - 著作権(創造者の権利)
    - 著作人格権 → 譲渡・相続不可
      - 公表権
      - 氏名表示権
      - 同一性保持権
    - 著作権:財産権 → 譲渡・相続可
      - 複製権
      - 上映権・演奏権・上映権
      - 公衆送信権
      - 口述権
      - 展示権
      - 頒布権
      - 譲渡・貸与権
      - 翻訳権・翻案権
      - 二次的著作物の利用に関する権利
  - 著作隣接権(伝達者の権利)
    - 演奏家人格権
    - 財産権 → 譲渡・相続可

産業財産権の内容

- 産業財産権
  - 特許権
    - 自然法則を利用した高度な発明
    - (経済活動、精神活動等は含まない)
  - 実用新案権
    - 物品の形状、構造の考案
    - (プログラム、化合物等は含まない)
  - 意匠権
    - 見た目の新しいデザイン等
  - 商標権
    - サービスや商品を区別するための、
    - マーク、商標

権利の保護期間

著作権関係	著作者人格権	著作者死亡時消滅
	著作権:財産権 小説等の個人の著作物	著作者死亡後70年間保護
	団体名義の著作物 映画の著作物	公開後70年間保護
産業財産権関係	特許権	出願から20年保護 (新薬など特例とした最大5年延長)
	実用新案権	出願から10年保護
	意匠権	出願から25年保護
	商標権	登録から10年保護 (更新可能)

知的財産:著作権と産業財産権(2)

図解ノート

著作権の例外規定

身の回り	家庭	家族のための複製	○
	学校	授業の中での少量の複製	○
		授業のレポート内の引用	○
		高校のクラブ活動での使用	×
		運動会/文化祭での使用	○
		先生の研修での複製	×
		学校ホームページでの使用	×
	YouTube	楽曲の使用	×
		楽曲の演奏	○
その他	図書館での一部コピー		○
	教科書、入試問題内での一部使用		○
	障がい者のための複製		○
	行政情報/データ(特に禁止の明記無し)		○
	営利を目的としない上演・演奏		○

引用時の条件

・すでに公表されている著作物であること
・引用を行う「必然性」があること
・カギ括弧などにより「引用部分」が明確になっていること
・引用部分とそれ以外の部分の「主従関係」が明確であること
・引用される分量が必要最小限度の範囲内であること
・「出所の明示」が必要

解放特許例

QRコードは（株）デンソーが開発し、オープン化

CCライセンス

著作者がみずからの著作物の再利用を許可するという意思表示を手軽に行えるようしたもの

BY(表示)	NC(非営利)	ND(改変禁止)	SA(継承)
著作権者の表示を要求する。	非営利目的での利用に限定する。	いかなる改変も禁止する。	元になった作品と同じライセンスを継承させた上で頒布を認める。

知的財産のオープン化

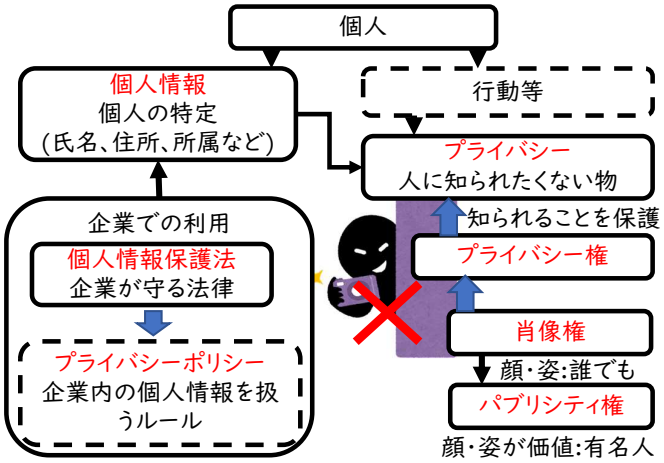
・知的財産の共有による、知的活動の活性化
・知的財産の(条件付き)の共有および自由利用

- オープンライセンス
- 解放特許
- オープンソース
- オープンデータ
- CCライセンス

個人情報とプライバシー

図解ノート

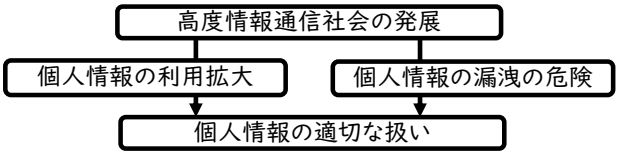
個人情報とプライバシーに関する用語と低具



個人情報の種類と分類

区分	情報	
公知: (基本4情報) 法令に基づき公開される場合有	氏名、性別、 生年月日、住所、	
非公知: 必要とされる場合に取得さくることがある	職業、趣味、所得、 学歴、資格、人種 健康状態、病歴	要配慮 個人情報
機微 (センシティブ) 本人の同意の元取得されなければならない	犯罪歴、思想信条、 宗教、思想信条	
	本籍地、性癖	

個人情報の利用に関して(個人情報保護法などで規定されている)



- 利用目的の制限
- 利用目的の取得時の説明
- 個人情報データの正確性の保持
- 個人情報データの安全管理
- 第三者提供の制限

いくつかの例

○オプトイン方式(オプトアウト方式)
関連する情報や広告を希望する人のみに提供する方式

○ 匿名加工情報
特定の個人が識別できないように加工されたデータ。
第三者への提供が可能。

アナログとデジタル/情報量と文字の表現

図解ノート

アナログとデジタル

アナログ
連続データ
(水銀体温計/
レコード等)

AD変換 →

← DA変換

デジタル
離散データ
(デジタル体温計
/CD等)

ビットとバイト

一つのON/OFFの状態 = ビット
2進法 ON＝1, OFF 0

8ビット ＝ 1バイト

デジタル情報の処理と特徴

処理	概要
複製	簡単に複製出来て、なくならない。複製してもノイズなど劣化しない。
編集	簡単に編集、加工できる。
統合	文字、数値、音、図、映像などを統合できる。
圧縮	その内容を壊さずにデータ量を減らすこと。必要に応じてもとに復元できる
検索	要な情報を効率よく正確にさがすことができる。

情報量の単位

単位	読み方	倍率	
bit(b)	ビット		
Byte(B)	バイト	x8	1B＝8b
KB	キロ	x1024	1KB＝1024B(1000B)
MB	メガ	x1024	1MB＝1024KB(1000KB)
GB	ギガ	x1024	1GB＝1024MB(1000MB)
TB	テラ	x1024	1TB＝1024GB(1000GB)

文字コードの例

コード	サイズ	論理文字数	概要
ASCII	7b/(拡張1B)	128(256)	英数と制御コードを表した文字コード
シフトJIS	2B	65536	日本語の漢字などを表すための文字コード
Unicode			世界中の文字に対して番号を割り当てて管理する方法
UTF-8	1～6B		Unicodeの実現方法のひとつで最も普及している。

ASCIIコード表(一部)

		0000	0001	0020		0111
		0	1	2		7
0000	0	NUL	DLE	(SP)		p
0001	1	SOH	DC1	!		q
0010	2	STX	DC2	“		r
1100	C	FF	FS	,		|
1101	D	CR	GS	-		}
1110	E	SO	RS	.		~
1111	F	SI	US	/		DEL

n進法とその変換

n進法の意味

10進法

105

10^2	10^1	10^0	
100	10	1	位の重み
1	0	5	
100 ＋	0 ＋	5	＝ 105

16進法

069

16^2	16^1	16^0	
256	16	1	位の重み
0	0	9	
0 ＋	96 ＋	9	＝ 105

2進法

110 1001

2^6	2^5	2^4	2^3	2^2	2^1	2^0	
64	32	16	8	4	2	1	位の重み
1	1	0	1	0	0	1	
64 ＋	32 ＋	0 ＋	8 ＋	0 ＋	0 ＋	1 ＝	105

2進/10進/16進対応表

10進	2進	16進	10進	2進	16進
0	0000	0	8	1000	8
1	0001	1	9	1001	9
2	0010	2	10	1010	A
3	0011	3	11	1011	B
4	0100	4	12	1100	C
5	0101	5	13	1101	D
6	0110	6	14	1110	E
7	0111	7	15	1111	F
			16	10000	10
			31	11111	1F
			32	100000	20

n進法の変換(いろいろ方法があるなかの例)

○2進法<>16進法

2進数の4桁が16進数の1桁に対応するため。

2進数 > 16進数：2進数を4桁に区切り対応する16進数にする。

16進数 > 2進数: 16数の1桁をそれぞれ2進数に直してくっつける。

2進数	0	1	1	0	1	0	0	1
16進数	6				9			

○2進法 > 10進法

n進法の意味で示した、2進数の各桁の重みを掛けた数値を合計して10進数を計算する。

○16進法 > 10進法

n進法の意味で示した、16進数の各桁の重みを掛けた数値を合計して10進数を計算する。

○10進法 > 16進法

n進法の意味で示した、16進数の各桁の重みで割った数を並べて、10進数とする。256の上の桁は4096

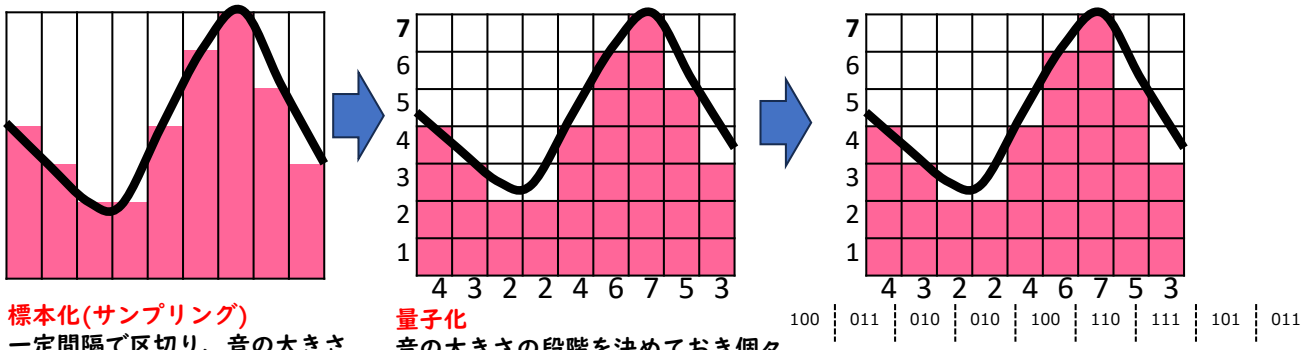
○10進法 > (16進法) > 2進法

10進法>16進法で示した方法でいったん16進数に変換したあと、2進数に変換する(教科などでは2で割っていく方法が一般的であるが計算ミスが多くなる)

音のデジタル化

図解ノート

標本化・量子化・符号化



標本化(サンプリング)
一定間隔で区切り、音の大きさ(振れの大きさ)を取り出す。
標本化周期: 1秒間に区切る回数
標本化周波数: 1秒間/表法周波数

量子化
音の大きさの段階を決めておき個々の取り出しだ最も近い値がいくつになるか変換する。
量子化ビット数: 大きさの段階を決めるもの。例:8ビット=256段階

符号化(コード化)
量子化で変換した数値をコンピュータで扱えるように2進数で表現する。

音のデジタル化のデータ量の計算

1秒間のデータ量 = (1回に発生するデータ量: 量子化ビット数) x (1秒間に発生数データ数: 標本周波数)
x (チャネル数: ステレオは左右2, モノラルは1)
n秒間のデータ量 = 1秒間のデータ量 x 秒数

計算例: 標本化周波数44000Hzで量子化ビット16ビットでデジタル化する3分のステレオ音楽の場合、何MBになるか。
(1B = 8bit, 1KB=1000B, 1MB=1000KBの場合)
1秒間のデータ量 = 16ビット x 44000Hz x 2チャンネル= 1408,000ビット = 176KB
n秒間のデータ量 = 176KB * 60秒x3分 = 31.68MB
ポイント 16ビット x 44000Hz の部分を2バイト x 44KHに変換しておくと計算が楽。

画像のデジタル化(1)

図解ノート

色の合成

加法混色(光)

赤 R	緑 G	青 B	何色?
		○	青
	○		緑
	○	○	水色(シアン)
○			赤
○		○	マゼンタ
○	○		黄色
○	○	○	白

減法混色(印刷など)

水色	黄色	マゼンタ	何色?
		○	マゼンタ
	○		黄色
	○	○	赤
○			水色
○		○	青
○	○		緑
○	○	○	黒 *

* 塗料自体に白が混じっているため完全な黒にはならない

イメージ画像のデジタル化

距離間隔で区切った物:画素
(ドット、ピクセル)
総素数 = 縦の画素数 x 横の画素数
=解像度
画素の単位(1インチあたりの数)
= ppi (pixel per inch)
or dpi(dot per inch) 印刷で使用

標本化:画像を一定の距離間隔で区切って分割し，各点(ドット)の各RGB(赤緑青)の明るさ（濃淡）を取り出す。

量子化: 各RGB(赤緑青)の濃淡を、それぞれ1,2,3などの段階で表現する。諧調=量子化レベル数. 例8ビット=256諧調
フルカラー=256階調(人間の目で自然に見える)

符号化: 量子化した大きさを2進数(デジタル)で表す

諧調と1ピクセル(ドット)データサイズ

256諧調 カラー			
R	G	B	1ピクセル= 8bit x 3色 = 24bit = 1B x 3色 = 3B
8b	8b	8b	
1B	1B	1B	
16諧調カラー			
White			1ピクセル= 8bit x 1色 = 8bit = 1B
8b/1B			

イメージ画像のデジタル化のデータ量の計算

1枚のイメージ画像 = (1ピクセルのデータ量) x (総画素数)
=1色のデータ量 x (カラーor) x 縦の画素数 x 横の画素数

計算例: 24ビットのフルカラーで、横1600画素 x 縦1200画素の写真は何MBか。
(1B = 8bit, 1KB=1000B, 1MB=1000KBの場合)
1枚のイメージ画像 = 8 bit x 3色 x 1600 x 1200
= 46080000ビット = 5.76MB
ポイント 単位を変換しておくと計算が楽。
8 bit x 3色 x 1600 x 1200
= 8 bit x 3色 x 1600 x 1200 / 8/ 1000/1000
= 1 × 3 × 1.6 × 1.2 = 5.76

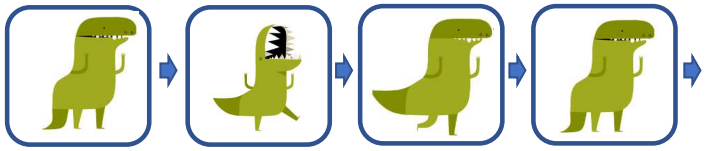
画像のデジタル化(2)

2種類の画像

形式	ベクター形式	ラスター形式 (ビットマップ形式)
例	Office系の図形 ポリゴンのゲーム	写真/詳細なイラスト ドット絵のゲーム
		
使用ソフト	ドロー系ソフト	ペイント系ソフト
仕組み	複数図形を組み合わせて作成する。 *1	ピクセル(ドット)の集まりで作成する。
特徴	・拡大、縮小が計算でできる、見た目が変わらない。 ・詳細な図形の表現にコンピュータパワーが必要 ・修正変更が楽。 ・データ量は少	・拡大するとギザギザ(ジャギー)が目立つ ・詳細な図形でもコンピュータパワーはそれほど必要ない。 ・修正変更が手間。 ・データ量は大

補足*1
　ベクター形式の画像でも組み合わせる図形の数が膨大になるとペイント同様に滑らかに見える。最近のゲームの一人のキャラクターは1万以上のポリゴンを組み合わせている。

動画のしくみ



1フレーム　1フレーム　1フレーム　1フレーム

一定時間内に高速で画像を切り替えていくことにより、
あたかも画像を動いているように見せる。
フレーム: 一枚一枚の画面
フレームレート: 一秒間に何フレーム切り替えるか
単はfps = frame per second
　映画の場合は24fps
　ビデオや地デジテレビの場合30fps
　(この程度のfpsだと人間には自然に見える)

動画のデジタル化のデータ量の計算(*2非圧縮時)

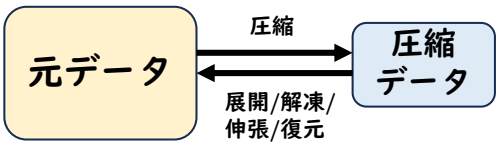
一連の動画のサイズ
　= (1フームのサイズ) x (fps数) x (動画の秒数)

計算例: 24ビットのフルカラーで、横1600画素 x 縦1200画素の24fpsで30秒の動画は何MBか。
　(1B = 8bit, 1KB=1000B, 1MB=1000KBの場合)
前述した写真の計算例を使用して
　1フレームのサイズ = 5.76MB
5.76 x 24 x 30 = 4147.2MB ≒ 4.1GB*2

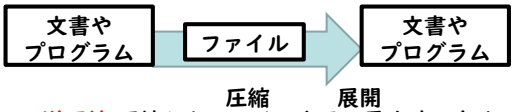
補足*2 圧縮/非圧縮:
　動画データは非常に大きくなるため通常は圧縮して使用する。同様にラスター形式(ビットマップ形式)の画像も圧縮して使用される。

データの圧縮

圧縮: ある方法にしたがってデータサイズを小さくする。
展開/解凍/伸張/復元: 圧縮されたデータを元のデータに戻す。



可逆圧縮と非可逆圧縮



圧縮　展開

可逆圧縮:圧縮したファイルを元に戻す時、全く同じじゃないと困る。

圧縮　展開

非可逆圧縮: 音、画像、動画など圧縮したファイルを元に戻す時、ほどほど似ていればいい。高圧縮できる

圧縮の具体的な方法例

ランレングス法:同じデータが続く場合、一つのデータとその繰り返し回数に変換する。
辞書式:同じデータのパターンが前にあったら、それと同じであることで示して圧縮する。
差分式: 映像などのシーケンスなデータの場合は、直前のデータの差分だけを使用する。
知覚符号化: 音など人間が知覚できない周波数をカットして圧縮する。

圧縮ファイル形式(非圧縮を含む)

種別	形式	概要
データ	ZIP	可逆圧縮で、複数のファイルをまとめて圧縮
音声	WAV	Windowsで使用。CD並みで知覚符号化のみ。
	MP3	高圧縮で、多きのプレーヤーでサポート
	AAC	MP3の改良版で最近多く利用されている。
	FLAC	知覚符号化をしない(ロスレス)で高品質
静止画	BMP	画像のピクセルのすべての情報。非圧縮
	GIF	256色のみ使用。初期のWebページで使う。
	PNG	フルカラーで可逆圧縮。透過が可能
	JPEG	フルカラーで写真などで広く利用されている
動画	AVI	Windowsで使用
	MPEG2	MPEGシリーズでDVD,地デジで使用
	MP4	MPEGシリーズで高圧縮。ネット配信も可

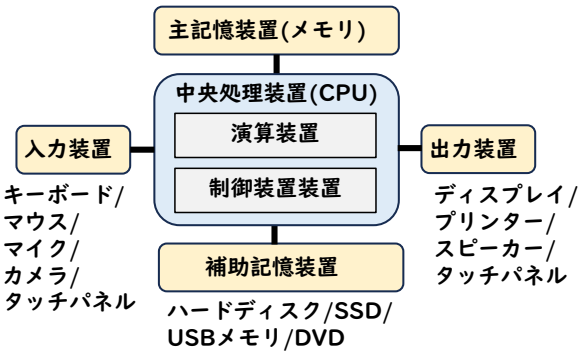
JPEG/MPEGファミリー

静止画	JPEG	カメラの写真等
動画	MPEG	JPEGの動画への応用
	(MPEG1)	(最初の基準)
	↳MP3	Layer 3 オーディオ圧縮
	MPEG2	デジタル放送/DVD
	(MPEG3)	(MPEG2に吸収)
	MPEG4	低速での使用を前提
	↳MP4	MPEG4 Layer14 MPEG4やMP3のファイルの集まり。主にインターネット上の動画で使用

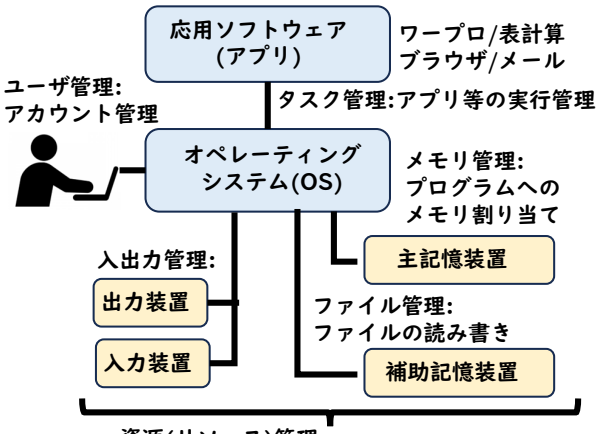
コンピュータの構成と性能

図解ノート

ハードウェア



ソフトウェア

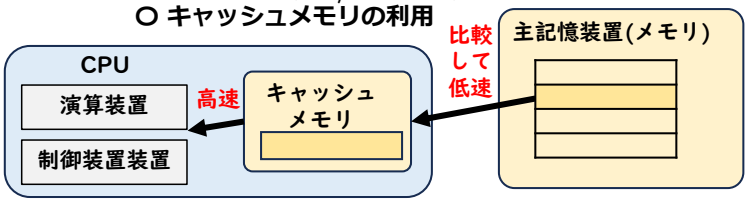


コンピュータ/CPUの処理速度

コンピュータ内部では、複数の器機が調和して動作するにはメトロノームのようなリズムがある
タイミングのための信号=クロック信号
　　クロック信号が早いほどCPUの動作は早い
1秒間信号数 = クロック周波数(Hz)

コンピュータ/CPUの処理速度の向上の工夫

- ○ クロック周波数を大きくする。
- ○ マルチコアCPU
  一つのCPUの中に複数のCPU機能を入れる。
- ○ GPU(Graphics Processing Unit)の利用
  画像処理専用のプロセッサ(最近ではAIの並行処理理計算でも活用されている)
- ○ 高速な記憶装置/補助記憶装置の利用
- ○ キャッシュメモリの利用



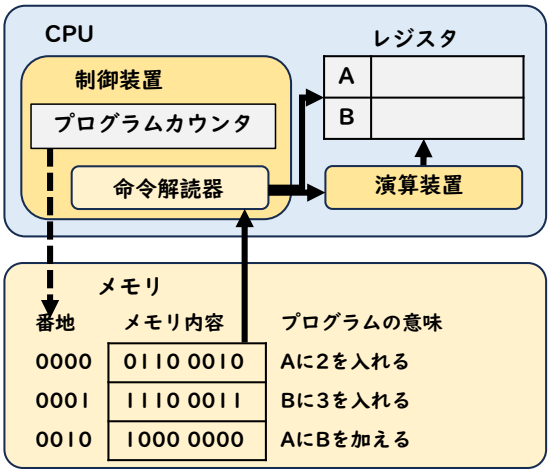
メモリと主記憶の間の転送より高速なキャッシュメモリを使用する。
予めCPUで使用すると想定されるメモリの内容をキャッシュメモリに転送して利用する。
(想定通りキャッシュメモリの内容が利用される率= ヒット率)

◎ CPUのメモリ内容の平均アクセス時間の計算
- ・キャッシュメモリのアクセス時間 20ナノ秒
- ・主記憶装置(メモリ)へのアクセス時間　80ナノ秒
- ・ヒット率　0.6　(60%は想定通りキャッシュに入っている)

平均アクセス時間 = 0.6 x 20 + (1-0.6) x 80 = 44 [ナノ秒]
(キャッシュに入っていない場合(1-0.6)は主記憶装置(メモリ)から取り出すため)

コンピュータの内部動作

CPUのプログラム実行の仕組み



CPUとメモリの機能

命令解読器	メモリの中に入っている、01のパターンのプログラムを読み取り実行する。
プログラムカウンタ	解読器が読み込む、メモリの番地を格納する。通常1命令読み込むと、この番地は1増加する。
演算装置	加算などの計算を実施する
レジスタ	一時的にデータを格納する。
メモリ	01のパターンのプログラムやデータを番地で区切って格納する。

プログラムの実行例(左図を例に)

① 初期状態: プログラムカウンタの内容は0000
② 命令解読器が0000の内容を読み込み実行。
　レジスタAに2を格納
　プログラムカウンタの内容は0010
③命令解読器が0010の内容を読み込み実行。
　レジスタBに3を格納
　プログラムカウンタの内容は0011
③命令解読器が0011の内容を読み込み実行。
　レジスタAにレジスタBの内容を加算装置が加えて
　レジスタAに5を格納

コンピュータによる数値や計算結果の誤差

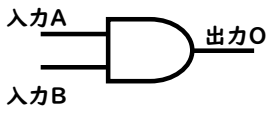
　コンピュータ内容では数値を領域の限られた2進数で表現するため、数値自体や計算結果に次のような誤差が発生する。
①桁あふれ誤差(オーバーフロー/アンダーフロー)
　計算結果が扱える数値の最大値又は最小値を超える場合。
② 丸め誤差
　数値が最小値より小さい部分について切り捨てや四捨五入などによる誤差。
③情報落ち
　大きい数と小さい数を計算した場合、小さい数の数値が反映されない誤差
④打ち切り誤差
　1/3や円周率など永遠に続く数値を途中で表現した誤差
⑤桁落ち
　丸め誤差がある数値同士で計算した場合、下位の方の桁に誤差が発生する。

論理回路

論理回路:

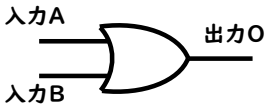
コンピュータ内部で0 1のビットを処理する仕組み。基本的にCPUの内部動作(命令の解読や演算など)も、この論理回路の組み合わせによって出来上がっている。論理回路はAND, OR, NOTが基本でこれを複雑に組み合わせる。(以下の図式はMIL記号)

AND回路:



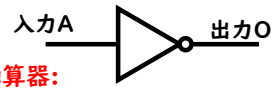
入力A	入力B	出力O
0	0	0
0	1	0
1	0	0
1	1	1

OR回路:



入力A	入力B	出力O
0	0	0
0	1	1
1	0	1
1	1	1

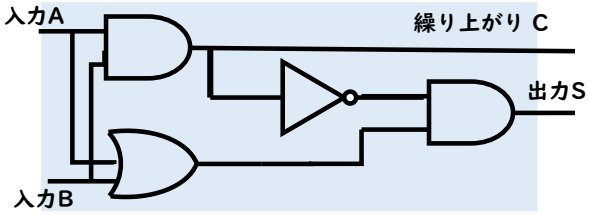
NOT回路:



入力A	出力O
1	0
0	1

半加算器:

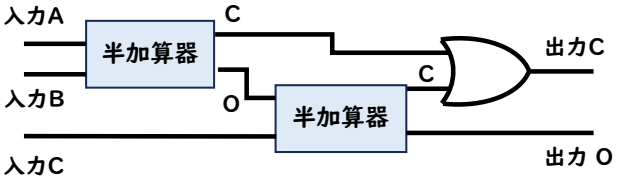
1桁の2進数同士の加算をする回路(繰り上がり C有)



入力A	入力B	出力S	繰り上がりC
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

全加算器:

下位の位からの繰り上がりを考慮した1桁の2進数同士の加算をする回路(これを連続して組み合わせることで複数桁の加算が可能)



入力A	入力B	入力C	出力S	出力C
0	0	0	0	0
0	1	0	1	0
1	0	0	1	0
1	1	0	0	1
0	0	1	1	0
0	1	1	0	1
1	0	1	0	1
1	1	1	1	1

アルゴリズムとプログラム

図解ノート

アルゴリズムとプログラム

アルゴリズム: 問題を解決するための方法や手順
プログラム: アルゴリズムをコンピュータで実行できるようにプログラム言語で表現したもの。

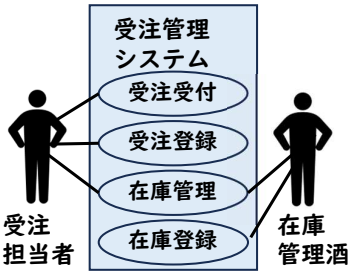
実行から見たプログラミング言語

CPUで実行可能なのは機械語のプログラミングだけ

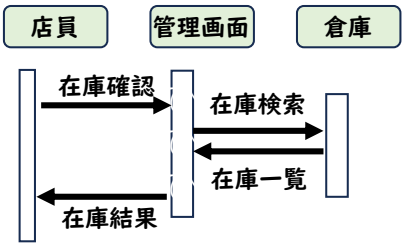
機械語	CPUごとに動作が決められた01のパターン
アセンブリ言語	機械語と一対一に対応しているが、やや人間が理解できるもの、プログラムを一括して翻訳して機械語に変換 ADD A, 10
コンパイラ	人間の言語に近い形で、プログラムを一括して機械語に翻訳して機械語に変換　A＝A＋10
インタープリター	人間の言語に近い形で、実行時に1行ずつ読み取り対応した動作をCPUに指示を機械語で与える

代表的なアルゴリズムの表現図式(設計技法)

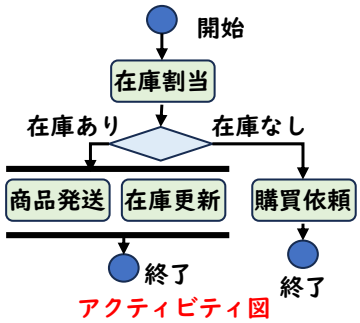
ユースケース図	ユーザがシステムをどのように使い、それに対してシステムがどう動作するか機能的に示す。
シーケンス図	システムの構成要素間(ユーザ含む)の相互作用を示すため、メッセージのやり取りを時間軸に示す。
アクティビティ図	実行の順序や条件分岐，並行処理など，制御の流れを記述する。
状態遷移図	システムの構成要素が、特定の条件やイベントによってどのように状態が変化するか示した図



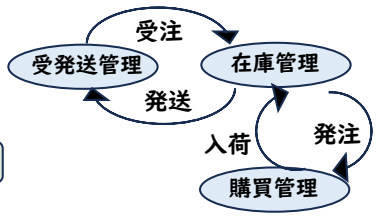
ユースケース図



シーケンス図



アクティビティ図



受注

状態遷移図

アルゴリズムによる実行効率の違い

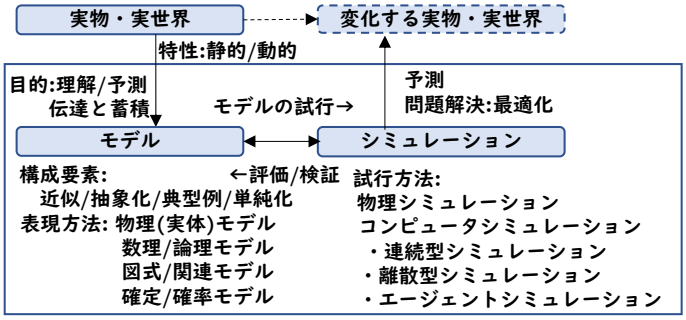
検索のアルゴリズムの違いによる**計算量**(実行回数など)の違い

方法	概要	平均検索回数(n=1000)
線形検索	最初(又は最後)の数から一個づつ比較して検索する。	(1+n)/2 (≒500回)
二分検索	予め並び替えか行われている数列を検索。位置を二分割しながら検索する。	Log_2n (≒10回

モデル化とシミュレーション

図解ノート

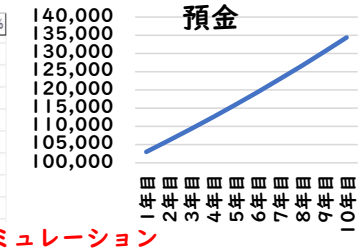
モデル化:複雑な実物や実社会を、理解、状態の予測、または知識として伝達・蓄積するため、それらを近似、抽象化、単純化または典型例として表現すること。
静的:時間経過しても変化しないものをモデル化
動的:対象の振舞い又は時間による変化をモデル化
確定モデル: 変動する要素がなくいつも結果が同じになる。
確率モデル: 変動する要素があり、結果が異なる。



シミュレーション:実物や実世界を表現したモデルを使って、試行・実験する。
シミュレーションの役割:
①モデルの評価し、その対象の本質を理解することができる。
②いろいろな実体、事象や現象の予測ができる。
③実体、事象や現象の最適状態を見つけることで問題解決できる。
コンピュータシミュレーション:
・連続型シミュレーション:自然科学や工学など時間などの経過によって連続的に変化する状態等を計算式によりシミュレートする
・離散型シミュレーション:確率的に発生する事象の状態や変化を、待ち行列型モデルなどのイベントの発生によりシミュレートする。
・エージェントシミュレーション:一定のルールに基づいて、自律的に行動する物の振舞いの集まりとして事象や世界をシミュレートする。

確定モデル/連続型シミュレーション
例: 複利計算
次期の金額＝現在の金額＋利息
利息＝現在の金額×利率×時間間隔

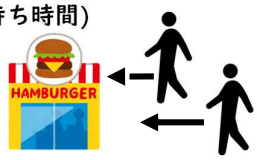
元金	100,000	利率(年)	3%
	預金	利息	
1年目	103,000	3,000	
2年目	106,090	3,090	
3年目	109,273	3,183	
4年目	112,551	3,278	
5年目	115,927	3,377	
6年目	119,405	3,478	
7年目	122,987	3,582	



確率モデル/離散型シミュレーション
例: 待ち行列(ハンバーガー屋の待ち時間)

時間間隔の平均		調理時間		30人平均待ち時間	
3.00	分	2.00	分	0.47	分

人目	時間間隔	到着時間	開始時刻	終了時刻	待ち時間
1		0.00	0.00	2.00	0.00
2	1.86	1.86	2.00	4.00	0.14
3	2.88	4.75	4.75	6.75	0.00
4	0.50	5.25	6.75	8.75	1.50
5	1.13	6.38	8.75	10.75	2.37
6	4.39	10.77	10.77	12.77	0.00
7	2.88	13.65	13.65	15.65	0.00
8	2.10	15.75	15.75	17.75	0.00
9	1.20	16.95	17.75	19.75	0.80
10	5.23	22.17	22.17	24.17	0.00
11	2.54	24.72	24.72	26.72	0.00
12	4.63	29.35	29.35	31.35	0.00



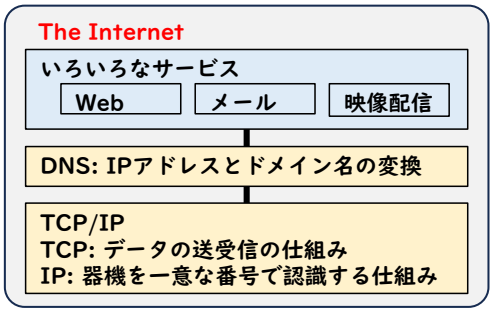
・店側はワンオペ
・客は平均3分間隔でランダムに来店
・注文を受けてから2分で調理し渡す。
・調理中は次の注文は受けないで客は待つ。

いろいろなシミュレーション

最適化/リソース割当	部屋割つけ，配車計画等
物理現象	振り子の動作，動体の移動等
グラフ/最短経路	最短経路: 移動距離等
予想/平均変化率	売り上げ予想，ローン計算等
待ち行列	ファーストフード店，交通渋滞
ライフゲーム	伝播状態: 感染状態等
モンテカロル法	ゲーム,不確実性を計算、円周率

IPアドレスとドメイン名

インターネットの仕組みの基本



IPアドレス

インターネットで通信を行う器機ごとにつけられた番号

○ IPv4とIPv6

IPv4　32ビット(8ビットx 4)　約43億個(不足している)

IPv6　128ビット(16ビットx 8)　約43億の4乗個

・IPV4

	8ビット	8ビット	8ビット	8ビット
10進表示	0～255	0～255	0～255	0～255
表示例	10.	56.	192.	16
	192.	168.	0.	1

○グローバルIPアドレスとプライベートIPアドレス

グローバルIP: 世界中につながるインターネットで通信するために、一意に割り当てられた番号

プライベートIP: 会社、学校、家庭内部で通信するために独自に割り当てられた番号

○ NAT (Network Address Translation)

通常、会社、学校、家庭には一つのグローバルIPアドレスが割り当てられ、内部では複数のプライベートIPアドレスが使用される。このプライベートIPアドレスからインターネットを利用できるように、切り替えてグローバルIPアドレスを使用する仕組み

ドメイン名

IPアドレスは人間が分からないので、IPアドレスに対応するようにつけられた名前

https://www.yahoo.co.jp/

プロトコル名　ホスト名　ドメイン名

FQDN(ホスト名/ドメイン名)

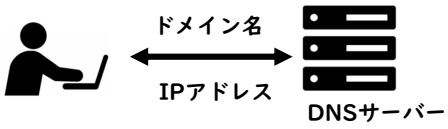
URL

yahoo.co.jp

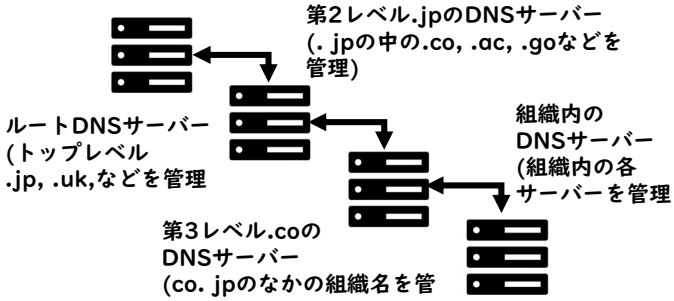
組織名　組織区分(種別)　国別コード

DNS

IPアドレスとドメイン名を変換する仕組み

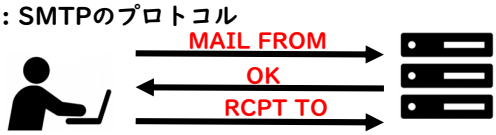


DNSの階層化による管理イメージ



プロトコル

プロトコル
通信を行う場合の約束、やりとの手順
例: SMTPのプロトコル

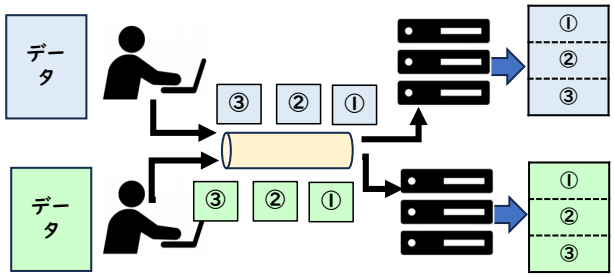


インターネットの通信のためのプロトコル= TCP/IP
全4層の全体を総称してTCP/IPと呼んでいる

層	名称	機能	プロトコル例
4層	アプリケーション層	Webやメールなどのアプリ	HTTP, SMTP, FTP等
3層	トランスポート層	データを効率よく通信	TCP/UDPなど
2層	インターネット層	IPアドレスをもとに経路確保	IPなど
1層	ネットワークインターネット層	物理的な電気や光信号に通信を変換	Wifi, 光ファイバー、イーサネット等

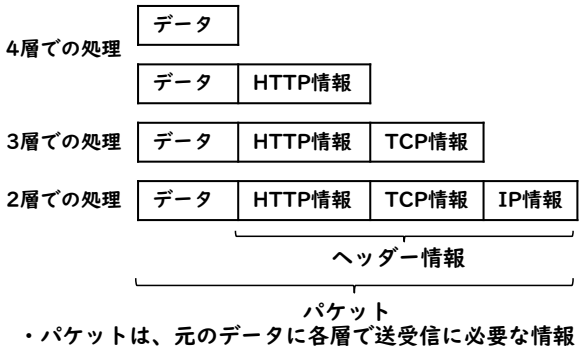
	名前	役割
4層	HTTP	Webページの送信
	HTTPS	HTTPのセキュリティ強化
	SMTP	メールの送信
	POP	メールの受信
	FTP	ファイルの転送
3層	TCP	送受信確認の有るデータ通信
	UDP	送受信確認の無いデータ通信

パケットとパケット交換方式



・データを送受信する時に、パケットという細かい単位に小さく分ける。
・パケットには送信元、受診先、順番などの情報があり、1つの通信経路を複数の通信が同時に使用できる。

TCP/IPの階層とパケット



・パケットは、元のデータに各層で送受信に必要な情報をヘッダー情報として付け加えていくことで出来上がる。

ルーティングとネットワーク

図解ノート

LANとWAN

LAN: 同一の建物や敷地内をつないだ、比較的小規模のネットワーク

WAN: LANの範囲を超えて、広域につないだ大規模ネットワーク

集中処理と分散処理

集中処理システム

・大型コンピュータで複数台の端末の処理を行う

分散処理システム

・複数のコンピュータが分散して処理を行う

ピアツーピアシステム

・2台のコンピュータが直接接続して処理を行う

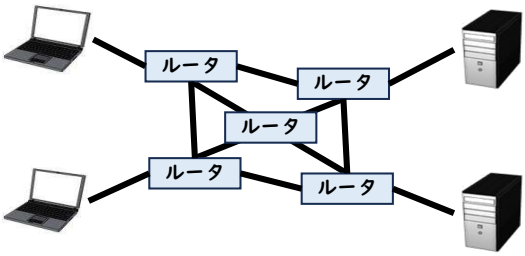
クライアントサーバシステム(分散処理システムの一種)

サーバ: サービスを提供する側

クライアント: サービスを要求利用する側

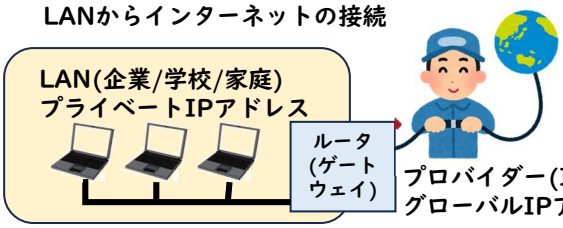
名前	役割
Webサーバ	Webページの閲覧
メイルサーバ	メールの送受信
プロキシサーバ	Webサイトへの代理アクセス Webページのキャッシング
DNSサーバ	DNS機能の提供
DBサーバー	データベース機能の提供

ルーティング: インターネットの経路



ルータ: パケットの経路を決定する装置: 経路情報を持ち、その時最適な経路でパケットを通信する。
実際各機材は、どの経路でパケットが流れているかはあまり気にしない。

LANからインターネットの接続



プロバイダー(ISP):企業/学校/家庭などにインターネットへの接続を提供する会社(グローバルIPアドレスを提供)

LANとインターネットの通信の区別

LANの中の器機はLAN内の他の器機のIPアドレスを知っているので(ネットマスクで分かる)、LAN内では直接通信。それ以外にIPアドレスと、とりあえずルータ(ゲートウェイ)を使って外と通信。

情報システム

図解ノート

情報システム:
社会・個人や企業に必要な情報の収集・蓄積・処理・伝達・利用にかかわる人間や機械の仕組みであり、狭い意味では、コンピュータや通信を利用したシステムである。
・社会情報システム
　情報システムは大きく、企業内部で利用される企業情報システムと、社会的な利益と利便性を目指した社会情報システムに分類される。

分野	システム例
商業	POS(店舗における商品、顧客管理) B2C(オンラインショップ等)
金融	インターネットバンキング 電子決済システム，電子マネー，ATM
交通	ITS(高度道路交通システム) ETC，座席予約システム
ビジネス・生活	GPS，生成AI，SNS CGM(ブログ，口コミサイト，掲示板)
行政	住民基本台帳関連業務，税関連業務 健康保険/介護関連業務
防災・災害対策	防災気象情報システム 安否確認システム
教育	LMS(学習管理システム，Google classroom等)

情報社会の影
デジタルデバイド: インターネットやコンピュータを使える人と使えない人の格差。年齢、教育、経済状態、国などの要因がある。
VDT障害:インターネットやコンピュータを長時間使用することによる身体的な症状。
テクノストレス: 情報社会でICTを使用したり、生活することによるストレスなどの精神的な症状

移動体情報処理システム
従来の集中・分散情報システムに加えて、近年ではスマホなどの通信技術を利用した移動体情報システムが構築されている。

オープンデータ
単なるITシステムだけでなく、行政などは積極的にオープンデータを公開して、多くの人が利用できるようにしている。

データベース

電子的に保存され、アクセスできる組織化(データモデル)されたデータの集合

タイプ	概要
階層型	上位から下位へ階層構造でデータを整理
ネットワーク型	個々のデータのリンク関係で整理
関係型(リレーショナル型)	データを複数の表で管理。表の間は共通列の中の値の関連で管理。現在主流
NoSQL	特にデータ間の関係を定義しないで、キーを元にしてデータの保存、検索に特化。ビックデータで利用されている。

リレーショナル型DB(RDB):

テーブル(表)　　フィールド(列)

顧客番号	顧客名	性別	会員種別	住所
2002002	浅野沙織	女	月額A	大分県
2003002	細川成美	女	月額B	岐阜県
2003004	佐々百花	女	月額B	京都府
2003009	金森瞳	女	月額B	埼玉県
2004004	里見節子	女	月額A	島根県
2004007	堀秀久美子	女	月額A	徳島県
2005001	武田彩花	女	月額B	長崎県
2007001	柴田陽子	女	月額B	大阪府
2009004	龍造光子	女	月額B	石川県

レコード(行)

表の中の各フィールドは主キーで一意に決まる。(主キーの重複は無い)

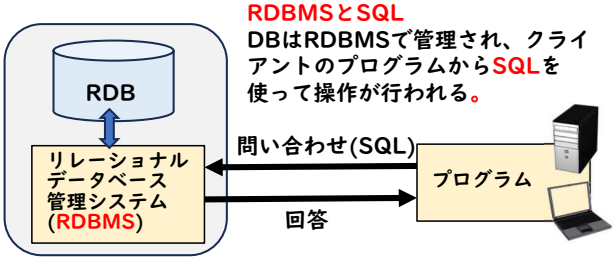
複数のテーブルは、特定のフィールで対応がつけられる(中の値が一致。)

主キー

貸出番号	顧客番号	貸出月	貸出日	商品番号
R0101543	2004001	1	1	E2040
R0102365	2009008	1	2	T2011
R0102365	2009008	1	2	C4967
R0103974	2007009	1	3	A3907
R0103990	2007003	1	3	T3499

商品番号	種別	種類	分野	タイトル
A1951	ＤＶＤ	アニメ	ゲーム	戦国乙女絵巻
A6849	ＤＶＤ	アニメ	ゲーム	ソルジャーキャ
A3011	ＤＶＤ	アニメ	ゲーム	僕らのサマーバ
A3313	ＤＶＤ	アニメ	恋愛	ようこそ花やし
A4966	ＤＶＤ	アニメ	恋愛	上級生と下級生
A2349	ＤＶＤ	アニメ	恋愛	ウェディングラ

RDBの操作

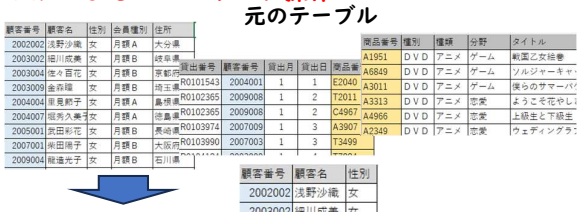


RDBMSとSQL

DBはRDBMSで管理され、クライアントのプログラムからSQLを使って操作が行われる。

SQLによるRDBのデータ操作

元のテーブル



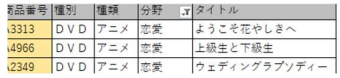
射影

一つの表から特定のフィールドを取り出す

顧客番号	顧客名	性別
2002002	浅野沙織	女
2003002	細川成美	女
2003004	佐々百花	女
2003009	金森瞳	女
2004004	里見節子	女
2004007	堀秀久美子	女
2005001	武田彩花	女
2007001	柴田陽子	女

選択

一つの表から条件に合レコードを取り出す



結合

複数の表を関連付けたせれたフィールドをもとに結合して、新しい表データを作成する。

情報セキュリティの技術

情報セキュリティ(機密性、完全性、可用性)を実現するための、いろいろな技術

○ 認証方式

認証タイプ	概要
知識認証	本人しか知らないパスワードや質問の答え
生体認証	顔、指紋、声などの身体的特徴
所持認証	スマホなど本人しか持っていない機材等

二要素(多要素)認証: 異なる認証タイプを組み合わせて使用
二段階認証: 複数の認証タイプ(同一タイプ可)を段階的に使用

○ コンテンツフィルタリング

利用者が不適切なWebサイトへのアクセスを制限する
ブラックリスト: アクセス不可のリストは接続不可
ホワイトリスト: アクセス可のリストのみ接続許可

○ ファイアーウォール

外部と内部のネットワークの間に置き、データの内容を判断し通信の許可又は拒否を行う。
ポート番号: 一つのIPアドレスは実際には複数のポートがあり、プロトコルにより使用するポート番号が決まっている。このポート番号ごとに許可/拒否を設定することにより、使用できるプロトコルを制限する。

○ 電子すかし

データやコンテンツの不正利用や複製を防ぐため、もとの情報に予め、電子なすかしを入れておくことで、不正に利用したものか判断する。

○ マルウェア対策(技術的なもの)

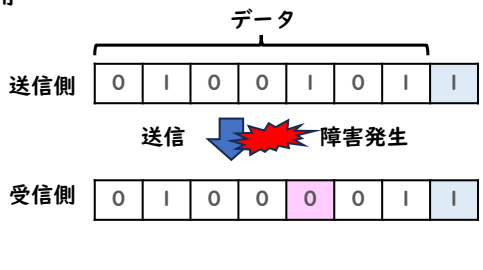
・セキュリティソフトの利用: マルウェアの検出と削除
・ソフトのアップデート: セキュリティホールを埋める
・ファイアーウォールやコンテンツフィルタリングの利用

○ 暗号化

記憶装置や通信におけるデータを暗号化し、他者に内容がわからないようにする[暗号化とその応用を参照]

○ 誤り検出/訂正

記憶装置や通信におけるデータのビット単位の誤りを検出したり訂正する機能。
誤り検出符号(パリティビット): 例えば、通信時に7or8ビットのデータごとに1ビットのパリティビットを付加する。パリティビットの値はデータを含めて含まれてる1の数が偶数になるように設定する(受信後偶数個でなければ、どっかの1ビットが誤っていると判断できる)



○ VPN（Virtual Private Network）

インターネット上での安全な通信(会社で社員が自宅から会社のシステムを利用する時等)を確保するため、仮想的な専用線(一本の線だけで接続)を設置する方法。

○ ブロックチェーン

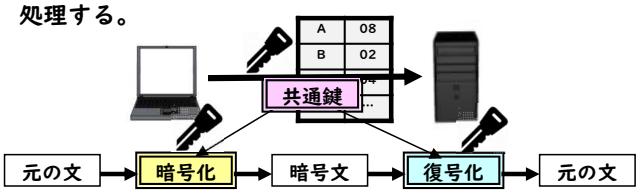
暗号資産などの利用にあたり、不正に情報が改ざん、利用されていないか防ぐ技術
・情報をホストコンピュータなどでかなりするのではなく、個々の利用者の端末で同じ情報を相互に監視しながら保持する。
・情報は単一のものではなく、過去の情報とのリンクがあり、長い期間のデータの処理の記録が管理できる。

暗号化とその応用

暗号化の二つの方式

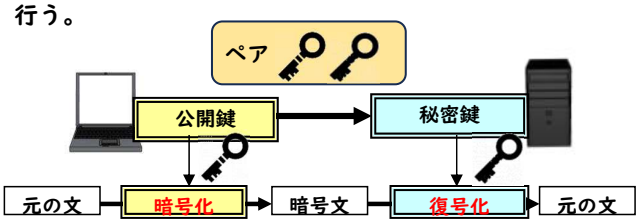
○ 共通鍵方式

暗号化側と復号化側で、暗号表のような同じ共通鍵を持ち処理する。



○ 公開鍵方式

ペアになった公開鍵と秘密鍵を使って、暗号化と復号化を行う。



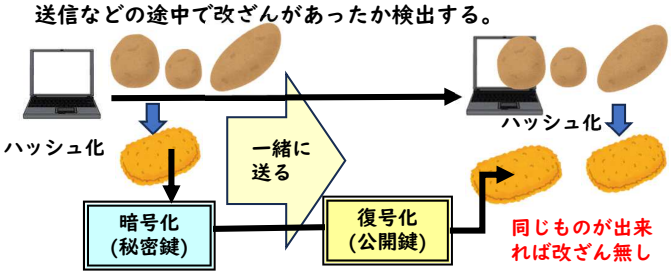
公開鍵: 多くの人に配布する。
秘密鍵: 特定の人、企業のみ保持する。
暗号化と復号化については、相互で行える(対照的)。
・ある公開鍵で暗号化したものは、ペアの秘密鍵でのみ復号化可
・ある秘密鍵で復号化したものは、ペアの公開鍵でのみ復号化可

	共通鍵方式	公開鍵方式
長所	・処理が高速	・一つの公開鍵で多くのユーザが利用できる
短所	・複数のユーザがいる場合、複数の共通鍵が必要 ・鍵を他者に知られないよう渡す必要がある	・処理に時間がかかる ・公開鍵の受け渡しが楽

暗号化技術の応用

○デジタル署名

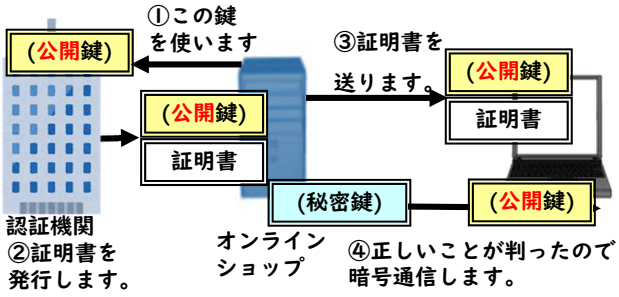
送信などの途中で改ざんがあったか検出する。



○デジタル証明書

公開鍵自体が正しいものであるか証明する仕組み(デジタル署名でもともと悪人が送っているのは気がつかない)

事前に信頼のおける機関に公開鍵の登録



OSSL/TLS(HTTPSで使用)

公開鍵方式と共通鍵方式を長所の組み合わせ
① ユーザ側で共通鍵の生成
② ユーザ側で公開鍵を使って共通鍵を暗号化してサーバ側に送る。
③サーバ側では秘密鍵で復号化して公開鍵を取り出す。
④公開鍵を使ってユーザとサーバーで高速に通信

データと基本統計量

データの尺度

	尺度水準	
質的	名義尺度	データの区別だけある:血液型、クラス名
	順序尺度	順番だけに意味がある:テストの成績順、満足度
量的	間隔尺度	数値の間隔が等しい:西暦、摂氏音頭
	比例尺度	数値で原点の0があり比例関係が成り立つ: 身長、速度、時間

データの整理(クリーニング)

調査後のデータには不備により不適切なものを含まれているので、それらを処理する必要がある。

欠損値: 何らかの原因でデータの取得ができなかった。
外れ値: 想定している範囲から著しく大きい/小さい
異常値: 外れ値の中で測定ミス、記入ミニなど原因が明確

基本統計量:代表値

データの集まり全体を表すもの

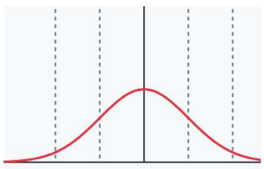
	概要
平均	全ての値を足して要素の数で割ったもの
最大値	最も大きい値
最小値	最も小さい値
中央値	データを小さい順に並べた時に、丁度真ん中の順番にある値
最頻値	同じ値のデータが最も多い値
第1四分位値	データを小さい順に並べた時、それぞれ順番が1/4, 2/4, 3/4にある値(第2四分位値 = 中央値)
第2四分位値	
第3四分位値	

基本統計量:散布度

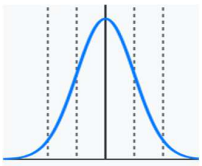
データの散らばりを表すもの

	概要
範囲	最大値-最小値の値
分散	平均値と個々のデータの差(偏差)を求め、偏差を2乗した値の平均値
標準偏差	分散の平方根

標準偏差の大きさとデータの散らばり



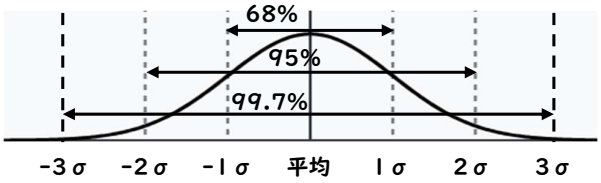
標準偏差が大きい
(散らばりが大きい)



標準偏差が小さい
(散らばりが小さい)

正規分布と標準偏差(σ)の関係

元のデータが正規分布に従っていると仮定した場合、例えば、平均+標準偏差x2と平均-標準偏差x2の値の範囲で全体のデータの95%が入っている。



偏差値

平均が異なるデータを比較するため、全体の平均値を50として、全体の中でどれくらいの位置にいるのかを表した値。

偏差値 =(データ－平均点)/標準偏差×10+50

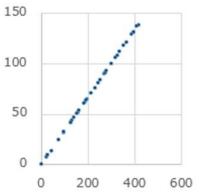
上記で説明したように、偏差値70の場合、2σの順位にいるため、50+95/2=97.5%のとろにいることになる。

相関と回帰分析

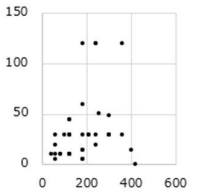
図解ノート

相関

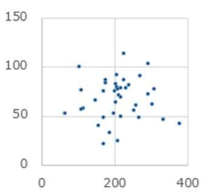
2つの要素が密接に関係し合い、一方が変化すると他方も変化するかの関係



完全に相関がある: y = ax



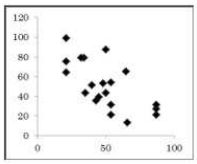
相関がある



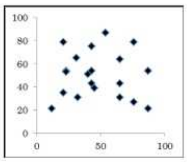
相関が全くない: バラバラ

相関係数 r

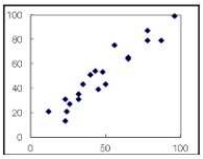
相関の強さを表す数値。 -1 ≦ r ≦ 1の値をとる。≦



負の相関がある
r ≦ 0



相関がある

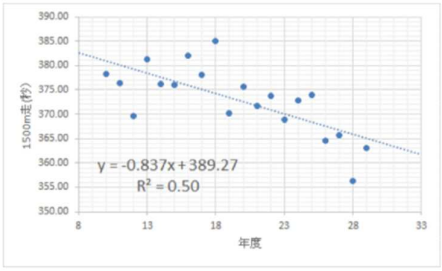


正の相関がある
r ≧ 0

絶対値	相関の強さ
0.0～0.2	ほとんど相関関係がない
0.2～0.4	やや相関関係がある
0.4～0.7	かなり相関関係がある
0.7～1.0	強い相関関係

回帰分析

関係性を数式にして、現状の傾向の把握や予測を行う方法

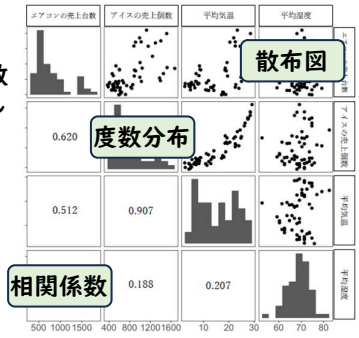


直線回帰: 2変数のデータを一次関数としして表現できると近似すること。相関係数が大きい時に有効

回帰直線: 直線回帰において近似された直線

散布図・相関行列

複数の変数間の、散布図、相関係数(、度数分布)を一つの図で表したもの。



相関と因果関係

因果関係: ある変数が原因となり、他の変数がその結果となる関係。

疑似相関: 相関が大きい場合でも因果関係があるとは限内。例えばアイスの売り上げとビールの売り上げの相関は大きいが、本来の原因は気温が上がったことであり。このような因果関係がないが、相関が大きい時は疑似相関である。

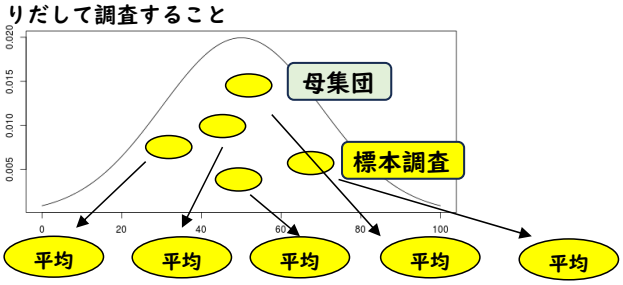
標本調査と仮説検定

図解ノート

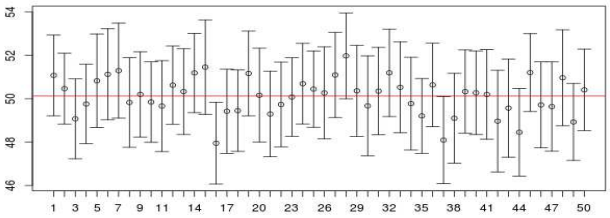
全数調査と標本調査

母集団と全数調査: 調査の対象となるすべての対象が母集団であり、対象すべてを調査することが全数調査

標本調査: 全数調査は手間がかかるため、一部の対象をとりだして調査すること



標本調査の代表値



標本調査の場合、母集団のどの対象を対象としたかで、その平均や標準偏差は、本来の母集団の代表値とは異なる。実際は複数回、標本調査を行うとこれらのバラバラになる。

t分布: 複数の標本調査の分布

上図のように、複数の標本調査を行った場合は、各調査の平均は正規分布に類似したt分布になる。この場合、標本調査の対象数が増えるほど正規分布に近づく。

95%信頼区間

標本調査の結果は真の母集団と平均が異なるため、ある、表法調査の結果からある程度程度の幅の信頼区間を設定する。多くの場合95%の幅を設定する。

この場合、母集団から標本を取ってきて、その平均から95%信頼区間を求める、という作業を100回やったときに、95回はその区間の中に母平均が含まれる。

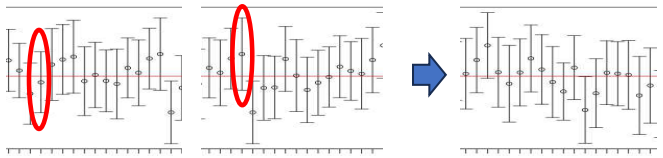
t検定:異なる標本調査結果の仮説検定

仮説検定: 標本調査から母集団に対する仮説の真偽を判断する

t検定: 2つの標本調査結果がある場合に、それぞれの元になった母集団が異なる集団であるか検討する。(母集団の分散がわからない場合に仕様)

	概要	t検定の場合
帰無仮説	最初に立てる仮説 検定の対象	標本調査は、同じ母集団からとったものである。
対立仮説	帰無仮説に対立するもので、本来証明したいこと	標本調査は異なる母集団からとったものである。

例: t検定の考え方



A方法の成績の標本調査　B方法の成績の標本調査　A方法とB方法の任意の二つの標本調査の平均の差

帰無仮説: A方法のB方法の勉強で成績に違いは無い

検定の考えた: A方法とB方法で標本調査を複数回行って、その平均の差を考える。この時複数回行えば、その差もバラバラになる(この差もt分布に従う)

帰無仮説の棄却の判断: もしA方法とB方法の差が非常に大きければ、それはめったに起こらないことなので帰無仮説が成りたたないと判断。A方法とB方法で違いがある。

有意水準5%: 通常の仮説検定で良く利用される基準。標本調査の組み合わせを100回やっても、5回しかないぐらいに差が大きいこと。