

コンピュータ概論 A/B

-- -- 情報と検索エンジン -- --

数学科 栗野 俊一

(協力: 伊藤先生, TA: 鈴木大智 君 [院生 2 年])

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く
禁じます

2021/05/10 コンピュータ概

お知らせ

コンピュータ概論 A/B (2021/05/10)

お知らせ

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

お知らせ

□ 登校者へ

- 出席管理システムで出席を取りましょう

 - ▶ 出席管理システムに、学生証を提示します

- 座席は、QR コードのある所に座ります(座席指定はありません)

 - ▶ 教室の前にあるポスターに従い、携帯で QR コードを読み込み、座席の登録をしましょう

□ 講義が始まる前に

- PC の電源をいれ、WiFi に接続しましょう

- 本日の資料に、目を通しておきましょう

前回(2021/04/26)のまとめ

コンピュータ概論 A/B (2021/05/10)

前回(2021/04/26)のまとめ

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

前回(2021/04/26)のまとめ

□ 前回(2021/04/26)のまとめ

○ コンピュータ概論ガイダンス

▶ 兎に角、課題を出す

○ コンピュータの学び方

▶ 兎に角、沢山触る

○ 「学生」の心得

▶ 兎に角、自分で考える

○ CST Portal による課題提出方法 (ファイルの提出)

▶ 兎に角、操作をおぼえる

○ サクラエディタのインストール

本日(2021/05/10)の予定

- 本日(2021/05/10)の予定
 - 情報と検索エンジン
- 本日(2021/05/10)の目標
 - 「情報」という言葉の一次理解
 - 検索エンジンの利用
- 演習
 - CST Portal を利用した課題提出 (ファイルの提出)
 - ▶ 今回 (2021/05/10) の課題

本日(2021/05/10)の課題

コンピュータ概論 A/B (2021/05/10)

本日(2021/05/10)の課題

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

本日(2021/05/10)の課題

□ 出席パスワード : 20210510

□ 前回 (2021/04/26) の課題

○ CST Portal II での課題提出

▶ 提出形式 : ファイル (20210510-QQQQ.txt) とスクリーンショット (20210510-QQQ.png)

▶ 提出内容 : 前回の課題のメッセージを、サクラエディタとペイントで作成して、提出

□ 今回 (2021/05/10) の課題

○ CST Portal に以下の課題の結果を提出しなさい

▶ ファイル名 : 20210510-QQQQ.txt (QQQQ は学生番号)

▶ 表題 : URL の紹介

▶ 内容 : 最近、自分が気になっている単語 (3 個) に関する URL を三つずつ紹介する

▶ サンプル(20210510-9999.txt)を参照のこと

○ Wikipedia を紹介する場合は、一次情報も併記する事

情報とは

コンピュータ概論 A/B (2021/05/10)

情報とは

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

情報とは

□ 情報とは

○ 複数の可能性ある現象に対して、その内の一つあるいは一部を選んだもの(事象)

▷ (天気) 今日の午後は晴(雨の可能性があった)

▷ (数学) 二等辺三角形の内角和は180度だ (180度と異なる可能性があった)

○ 「正しい」情報と「正しくない」情報

▷ 正しい情報: 事実を記述したもの (じゃあ、「事実って何?」って話はしない)

▷ <<注意>> 「正しさ」は相対的: 「何と比較してか?」を考える必要がある

□ 情報の量(情報が多いか少いか)

○ 情報そのものから計る事ができる情報量(客観的/情報理論)

▷ その情報が表現する現象が「起きる可能性が低け」れば、それを得た時の「情報量が多い」

▷ (ギャンブル) 「穴馬が勝つ」と言う情報は、「本命が勝つ」と言う情報より情報量が多い

▷ (サイコロ) 「ゾロ目が出た」と言う情報は、「そうでない」と言う情報より情報量が多い

▷ 「学」としては、こちらを利用する

▷ bit : 共に 1/2 の確率の二つの事象 (Yes/No, 裏/表, 偶数/奇数) の内の一つが示された時の情報量(情報の最小単位)

○ 受け手の知識量から計られる情報「価値(主観的/一般的に利用される)」

▷ 受け手が知らない情報かどうか (報された時に知っていた情報は価値が低い)

▷ 受け手に価値のある情報かどうか (興味がない情報の価値は低い)

▷ 「現実の世界」ではこちらが利用される事が多い (主観が入る)

情報の量と質(メタ情報)

□ 情報「量」の定義

- 事象 E が確率 $P(E)$ で起きるとする
- 「事象 E が起きた」という「情報」の情報量($I(P)$)は次の式で表現される

$$I(E) = \log_2 \frac{1}{P(E)} = -\log_2 (P(E))$$

□ 情報の「質」

- 「情報」そのものに関する情報 (メタ情報)
 - ▶ その情報は「正しいかどうか (正しくなければ価値がない)」を判定する材料
- 例 1 : 誰から聞いた? (情報ソース)
 - ▶ 彼奴(あいつ)の情報じゃ、信頼できない
- 例 2 : 何語で書かれているか? (情報の形式)
 - ▶ 楔形文字で書かれているんじゃ、読み取れない
- 例 3 : どうやって入手した(情報の入手方法)
 - ▶ 不法な手段で入手した証拠は、裁判では取り上げられない
- 例 4 : それは何時の話? (情報の有効な範囲)
 - ▶ バーゲンは、昨日でお仕舞いだった..

□ 「情報(の内容)」だけでなく、その「量」と「質」にも気にする

「データ」と「(狭義な)情報」

□「データ」と「(狭義な)情報」:相対的な違いである事に注意

○データ:客観的な観測結果を記述した(形をしている)物

▶(科学) 新型コロナ感染者 全国で7000人超

▶(スポーツ) 昨夜のセ・リーグ、巨人7-5ヤクルト

▶(数学) 地面に三角形 ABC を描画、角度の測定結果は、約 30, 60, 90 度

○(狭義な)情報:一定の価値感(主観を含む)に基き、データを「解釈」した結果

▶(科学) 1月以来 各地で「過去最多」(感染は広がっている..)

▶(スポーツ) 巨人が勝った

▶(数学) 三つの内角 A, B, C の和は約 180 度

□一次情報/二次情報

○一次情報: そのデータ(情報)が最初に作られた形での情報

○二次情報: 一次情報や他の二次情報に基いて作られた情報

○「Wikipedia からの引用」が「ダメ」な理由

▶ Wikipedia は二次情報だから => 一次情報に当ろう (cf. 論文)

「データ」から「(狭義な)情報」へ

□「データ」から「(狭義な)情報」へ

○データを「処理(分析)する」事によって「(狭義な)情報」が得られる

▶ データは「取る」もので、情報は「作る」もの

○「処理する」には、「判断基準」が必要となる

▶ (科学) 感染が広がっているなので、より、自粛が必要

▶ (スポーツ) これぞ4番! 巨人・岡本和が2打席連発の逆転サヨナラ3ラン

▶ (数学) 平面幾何学の公理と論理を利用して証明できるので..

○「基準」が異れば、「同じデータから異なる情報」が得られる

▶ (科学) 未だ、最悪の事態にはなっていない、経済を重視すべき。

▶ (スポーツ) 残念、ヤクルト逆転負け、後少しだったのに..

▶ (数学) 球面幾何学では、180度より大きくなってしまう

以下、「データ」と「(狭義な)情報」をまとめて単に「(広義な)情報」と呼ぶ

情報の形

□ 情報の記述を構成する要素

○ 5W1H

- ▶ When : 何時
- ▶ Where : 何処で
- ▶ Who[主語] : 誰が
- ▶ What[述語] (Whom[目的語]) : (何に対して) 何を
- ▶ Why : なぜ (どんな目的で)
- ▶ How To : どうやって

○ 5W2H : 5W1H に 次の How を追加

- ▶ How Much (How Many) : いくらで (どのくらいで..)

□ 知識の形式

○ What : それは何であるか ?

- ▶ (数学) 定義/公理 (例 : 「連続」とは..)
- ▶ (情報) 知識 (例 : 「ファイル」とは..)

○ How to : それはどうやってやるのか ?

- ▶ (数学) 証明/計算 (例 : 関数 x^2 は $x=0$ の点で連続か?)
- ▶ (情報) 操作 (例 : ファイルをコピーするには?)

知識を身に付ける

□ 情報(問題の答え)の種類

○ 仮定/決め(定義)/事実：知らないと答えられない問題(一般と特殊がある)

▶ 「調べて」入手するしかない (検索する !!) / 考えてはならない(無駄)

○ 結論/導かれる物(定理)/予想：考えれば答えられる問題(複雑)

▶ 「考えて」入手する事が望ましい / 調べると時間がかかるかもしれない

▶ 調べて、入手できる事 *も* あるが、コストや確実性に問題がある

□ 知らないと答えられない情報の入手方法

○ 特定な対象と結びついており、その対象毎に異なる情報

▶ cf. 個人情報 (年齢など..)

▶ その情報の対象に確認する (典型的な質問の一つ)

○ 多くの人 (場合によっては全ての人) と関係があり、共通となる情報

▶ cf. 西暦 1998 年は、平成何年?

▶ Web を検索する

□ 考えれば答えられる情報の入手方法

○ 考えるのが一番 (「数学」そのもの)

○ 調べても良いが、身に付かない (cf. レポートの結果をコピーする)

▶ 「正しいかどうか？」の判定が必要(結局、考える事が必要に..)

検索エンジンの使い方

コンピュータ概論 A/B (2021/05/10)

検索エンジンの使い方

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

検索エンジンの使い方

□まずやること

- 「グーグルの上手な使い方」で検索する

- ▶「自分に合った」方法を身に付ける(まずは真似から..)

□「検索」とは

- 欲しい情報に関連するキーワードを入力し、ページを表示させる(だけでなく)

- ページの内容を「精査」して、「欲しい情報」が記載されたページを篩(ふる)う事

□良いキーワードとは

- 欲しい情報と関連が深い

- ▶欲しい情報が記載されているページには高い頻度で現れる

- ▶欲しい情報関係ないページに現れる可能性は低い

- 情報量が多い (cf. より危険なウイルスが広がっている)

- ▶専門用語は良いキーワードになる (cf. 変異株,インド株,L452R)

- ▶良く利用される言葉はキーワードとして不適切 (cf. 「新型」)

- まずは、良いキーワードを探せ (サンプルング)

- ▶思い付くキーワードで表示されるページから良いキーワードを探す

検索結果の利用方法

- 検索結果は「正しい」とは限らない
 - 内容が正しいかどうかを *確認* してから利用する
- 結果の「正しさ」のチェック
 - 複数の結果を比較しろ (より便利なページもある)
 - ▶ 検索リストのトップから幾つか(最低 3 できれば 10 位)を見て確認
 - 記載内容を確認し、「矛盾」が無いかを調べろ
 - ▶ ページ内のキーワードが適切に利用されているか
 - ▶ 「論理的」な説明になっているか
 - 検索の前に「答を予想」して、それと比較しろ
 - ▶ 予想とかけ離れているなら「変」かもしれない
 - ▶ 予想と逆の結果も検索してみる (それが間違っている事を確認する)
 - 一次情報を参照しているか
 - ▶ 根拠のない推論をしている可能性はないか

「コピー・ペースト」の良し悪し

□ 情報の特質

- 「情報」の最大の利点は「コピーできる」事
 - ▷ cf. 「活版印刷」は、「ルネッサンスの三大発明」の一つ
- 「情報」の効率的な利用法は「コピーする」事
 - ▷ cf. 「学習」とは、「過去の成果」を「頭にコピー」する事 / 「応用」とは、「頭の中の情報」を「現実にコピー」する事

□ 「コピペ」の問題点

- 「情報」と「情報の表現(コピペの対象)」は異なる
 - ▷ 「情報のコピー」と「情報の表現のコピー(コピペ)」は異なる
- 「情報の表現」は単独では意味を持たない(他の表現と関連して意味を持つ)
 - ▷ 文脈(他の情報)を無視した「表現のコピー」は、「元の情報」を表現しない
 - ▷ cf. 友達のやった課題を提出(ついでに、提出者の名前も友達の名前をコピペ) / 友達が結婚した、自分も結婚したいので友達の妻にプロポーズ
- 正しい「コピペ」をするには「文脈の理解」が不可欠
 - ▷ 正しく「情報をコピペ」するには、「表現の変更」が必要な場合がある(「表現のコピペ」でだめ)
 - ▷ cf. 「I like it」, 「Me too」は良いが、「I love you」, 「Me too」は駄目(「I love you too」が正しい)

□ 正しく内容が理解できているなら「コピペ」は強力な道具

- 「コピペ」するなら、「内容を理解して」から
 - ▷ 必要に応じて(コピペの後に..)「表現の変更」をする

演習

コンピュータ概論 A/B (2021/05/10)

演習

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

[演習 1] google を利用してみよう

□ google のページ

- url : <http://www.google.co.jp>

- ▶ お気に入り(ブックマーク)に入れておこう

□ グーグルの使い方

- 「グーグルの使い方」というキーワードで検索してみる

- 候補は複数に当たってみる : 「タブ」を有効利用しよう

□ Web Page の扱い方

- まずは、「ブラウジング(拾い読み)」をする

- ▶ いきなり冒頭から「読む」のは非効率 (当り外れがある [対] 教科書)

□ 本当に、そのページは、読む価値があるのか？

- ページの内容は、「求めている物事」に関係するか？

- 内容は適切？ (著作権に注意)

- ▶ 量(網羅性はあるか)、表現(理解り易いか)、便利性(転用可能か)

- 内容は *正しい* か？

- ▶ 論理的？ / 根拠はあるか？ / 他の情報と矛盾していないか？

- 最低、三つのページをブラウズし、比較する必要がある

- ▶ [悪例] 「I'm Feeling Lucky」を見て「ない」と言う...

[演習 2] 課題レポートを作成しよう

□ 今週 (2021/05/10) の課題

- 最近、自分が気になっている単語 (3 個) に関する URL を三つずつ紹介する

□ 解題を解く手順

- 検索キーワードを 3 個考える
- Google 等の検索エンジンを利用してキーワード検索する
 - ▶ 表示されたページを幾つかブラウズする
 - ▶ キーワードの説明として適切だと思われるページの URL をコピーする
- 「メモ帳」に、URL をコピーペーストする
- 形式は、サンプル(20210510-9999.txt)を真似して作成する
- ファイル名を「20210510-QQQQ.txt (QQQQ は学生番号)」にする
- CST Portal に提出する

CST Portal を利用したレポート提出

コンピュータ概論 A/B (2021/05/10)

CST Portal を利用したレポート提出

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます

CST Portal を利用したレポート提出

□ 課題の提出

- 主に、CST Portal を利用する
- 毎回、「講義連絡」で「レポート提出」を要求する
 - ▶ 課題を作成して CST Portal 経由で提出する事

□ URL

- <https://newportal.cst.nihon-u.ac.jp/cst/top.do>
 - ▶ 「コンピュータ概論」の Page のリンクからたどる
 - ▶ CST で検索して理工学部の Web Page からリンクをたどる

□ 出席

- 出席は成績に影響しないが、折角なので、やる
 - ▶ 「講義」を開くと、「出席ボタン」があるので押す
 - ▶ 「出席パスワード」を聞かれるので、「20210510」と入力する

おしまい

コンピュータ概論 A/B (2021/05/10)

おしまい

講義内容の静止画・動画での撮影、及び SNS 等への転載を固く禁じます