情報学専攻入学者選抜試験

行動情報学 問題冊子

注意事項

- 1. 試験開始の合図があるまで、この問題冊子の中を見てはいけません。
- 2. 問題冊子は、表紙を含めて全部で8ページあります。試験開始後に問題冊子のページ 数を確認して、落丁又は印刷が不鮮明な場合は直ちに申し出てください。
- 3. 解答は解答用紙に記入してください。問題冊子に記入しても採点されません。
- 4. 問題冊子、解答冊子は切り離さずにすべて提出してください。
- 5. 「情報処理」分野はすべての問題を、「情報システムとデータ分析」分野は大問3問から2問を選んで解答し、解答した問題に対応する解答欄の箇所の□内に印をつけてください。解答しない問題には印をつけてはいけません。

情報処理

※全問を回答すること。

(間1)

- (1) OS から見たときに 12GB の空き容量のある SD カードがある。画像 1 ピクセルを記憶するのに 4 バイト (B) 必要な場合、1 枚あたり 1024×768 ピクセルの画像を、この SD カードに何枚保存できるか。なお、1GB=1024MB、1MB=1024KB、1KB=1024B とする。
- (2) 以下 (i) \sim (iv) は通信に用いられる代表的なプロトコルの一部である。各プロトコルの説明として最も適切なものを選択肢 (a) \sim (e) の中から 1 つ選べ。
 - (i) HTTPS
 - (ii) FTP
 - (iii) SMTP
 - (iv) IMAP

選択肢

- (a) ハイパーテキストを安全に転送するためのプロトコル
- (b) ネットワーク上でファイル転送を行うためのプロトコル
- (c) メールサーバ上の電子メールにアクセスするためのプロトコル
- (d) インターネット上でメールを転送するためのプロトコル
- (e) 安全に遠隔のコンピュータと通信するためのプロトコル
- (3) 次の HTML 文書が標準的なブラウザ上でどのように表示されるか図示せよ。

- (4) インターネットとワールド・ワイド・ウェブ (WWW) の違いを説明せよ。
- (5) 人間にとって見やすく扱いやすいデータが計算機では非常に扱いにくいことがしばしばある。その種のデータは「バッドデータ」と呼ばれ、計算機を用いたデータ解析を行うためには前処理が必要となり、データ分析者を悩ませることになる。

以下は表計算ソフト Excel によって作成されたある科目の(架空の)成績データである。このデータを CSV ファイル形式(いくつかの項目を区切り文字であるカンマ「,」で区切ったテキストファイル)に書き出した後、計算機を用いて学生ごとに成績の平均点を計算するときに起こりえる問題点を3つ指摘せよ。

	А	В	C	D	E	F	G	Н	1
1	201X前期					201X後期			
2	グループID	履修学生ID	レポート提出日	評価		グループID	履修学生ID	レポート提出日	評価
3	1	201XA001	201X/7/10	۸		6	201XA001, 201XA002	201X/1/24	90
4	2	201XA002, 201XA003	201X/7/11]^		7	201XA003, 201XB001	201X/1/30	85
5	3	201XB001	201X/7/26	С		8	201XB002	201X/1/6	55
6	4	201XB002, 201XB011	201X/7/22	D		9	201XC333	201X年1月22日	70
7	5	201XC333	201X/7/16	В		10	201XB011	201X年1月16日	65
8									

(間2)

A さんは、Web サイトを巡回してデータを収集するプログラムを作っている。何度も同じ URL ヘデータ収集を行ってしまわないように、すでに巡回したことのある URL のデータを、連結リスト構造というデータ構造を使って保持することにした。

- (1) 初めて巡回した Web サイトの URL を1つこのデータに追加するとき、計算量を 考慮すると、連結リスト構造の先頭と末尾のどちらに追加したほうがよいか、その 理由もあわせて述べよ。ただし、この連結リスト構造は単方向連結リストであり、 先頭の要素への参照は保持されているが、リストの末尾の要素への参照は保持され ていないとする。
- (2) これから巡回しようとする Web サイトの URL が、すでに巡回済みの URL のデータに含まれないかどうかを調べたい。この連結リスト中に巡回済みの URL が n 個すでに格納されている時に、格納されている URL と一致するかどうかを調べるのにかかる計算量を n に対するオーダーO(..)の形で示せ。
- (3) 巡回済み URL データとこれから巡回しようとする URL データとの照合を高速に するにはどのような方法が考えられるか。 具体的な方法を擬似コードまたは Java コード等で示しながら述べよ。

情報システムとデータ分析

※問1~問3から2問を選択して解答すること。解答欄には選択した問題の \square に \square と記入すること。

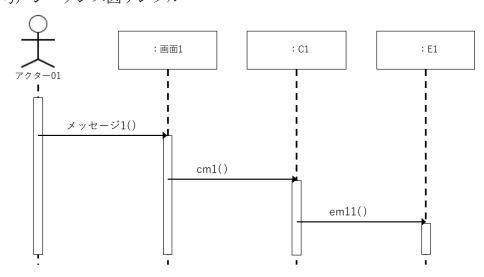
(間1)

ビジネスマネジメントと情報システムの関係について、以下の(1)から(2)の問いに答えよ。

- (1) NIST (米国国立標準技術研究所) は、ネットワーク経由で利用するクラウドサービスを、SaaS (Software as a Service)・PaaS (Platform as a Service)・IaaS (Infrastructure as a Service) という 3 形態のサービスモデルに分類している。 SaaS の具体的サービス例を挙げ、その内容を紹介すると共に、SaaS であると判断した根拠を説明しなさい。
- (2) ビジネスをモデル化する手法の一つとして Unified Modeling Language (以下 UML) が存在する。ビジネスプロセスを時系列で表現するシーケンス図は、情報システムの開発などで頻繁に利用されている。

オンラインショッピングサービスにユーザ登録を行うユースケースをシーケンス図で表しなさい。なお、ライフラインは、ユーザ・Web 画面・会員データベースを含む3つ以上は記載すること。

(参考) シーケンス図サンプル



(間2)

情報システム開発手法の代表的な開発モデルは、ウォータフォール開発モデルであるが、 最近は、仕様変更に柔軟に対応し、迅速に動くソフトウェアを開発することを目的とした アジャイル開発も、様々な企業で採り入れられるようになった。

ウォータフォール開発モデルは、上流工程から下流工程へ、システム企画、システム分析、システム設計、製作(プログラミング)、テスト、システム運用・保守という段階で順番に実行していくのが典型的である。

一方、アジャイル開発では、特に以下の4つの価値に重点を置き、短期間での反復開発 を行うものが典型的である。

- a. プロセスやツールよりも個人やメンバー間のコミュニケーションを優先する。
- b. 包括的なドキュメントより実際に動くソフトウェアを優先する。
- c. 契約の交渉よりも顧客との協調を優先する。
- d. 計画の遵守よりも変化への対応を優先する。

情報システム開発に当たっては、情報システムの適性に応じて開発モデルを適用することになる。

- (1) ウォータフォール開発モデルのそれぞれの工程で、システムの開発者が行うべき作業内容を簡単に説明しなさい。
- (2) ウォータフォール開発モデルで、システムの利用者が、システム分析及びテストで果たす役割について簡単に説明しなさい。
- (3) アジャイル開発モデルは、どのようなシステム開発に適しているのか、ウォータフォール開発モデルと異なる点を2点示して説明しなさい。

(問3)

表 1 は、24 年間以上の追跡調査が行われている Framingham Heart Study Data の一部で、特に喫煙と血圧の関係の調査に関して、バイオ統計学を学ぶ学生向けに一般公開されているデータの 1 つである[1][2]。表 2 はその変数表である。

- (1) 表1に基づき、変数「age」と「smoke」についてその分布を記述しなさい(質的 変数は各カテゴリの度数と割合を、量的変数は平均と中央値、不偏分散を求めなさい)。ただし、表1の「age」の和は509、偏差平方和は792であるとする。途中の計算経過も記述すること。
- (2) いま、喫煙の有無により収縮期血圧の分布が異なるのかどうかに興味がある。その 理解を助けるため、データを視覚化したい。このとき、どのようにデータを視覚化 すればよいであろうか。用いるグラフ名と変数名を含めて説明しなさい。

表 1. Framingham Heart Study Data (喫煙と血圧の部分)

id	sbp	bmi	age	tchol	female	smoke
1	106	27	39	195	1	0
2	107	23	38	223	1	1
3	121	28.7	46	250	0	0
4	105	29.4	52	260	0	0
5	108	28.5	58	237	0	0
6	128	25.3	48	245	1	1
7	141	25.3	54	283	1	1
8	150	28.6	61	225	0	1
9	183	30.2	67	232	0	1
10	130	23.1	46	285	0	1

表 2. 変数表

id	ID 番号				
sbp	収縮期血圧				
bmi	Body mass index				
age	調査開始時年齢				
tchol	調査開始時総コレステロール値				
female	性別(女=1, 男=0)				
smoke	調査開始時に喫煙者(=1), 非喫煙者(=0)				

出典

- [1] Dawber TR, Meadors GF, Moore FE Jr. Epidemiological Approaches to Heart Disease: the Framingham Study. Am J Public Health Nations Health. 1951;41(3):279-281.
- $[2] \quad \text{Teaching Datasets Public Use Datasets. https://biolincc.nhlbi.nih.gov/teaching/}.$