

情報過去問解答解説

2008 年度版

共通問題 1

問題解説

データの扱いの一例として「階層モデル」というものがあります (p.84 「4.3.3 階層モデル」)。このモデルの特徴としては

- 枝分かれ構造を表現することができる。
- 枝分かれの幹の方を高位のものとして、データの順位を付けられる。
- 低位のデータを指定すると、その所属する(ないしはそれが派生した元の)データを遡ることができる。
- 高位のデータを指定すると、以下に含まれるデータの集合を把握できる。
- 全てのデータを、もれなく、重複無く配置することができる。

などがあります。その例としては、

➤ 住所

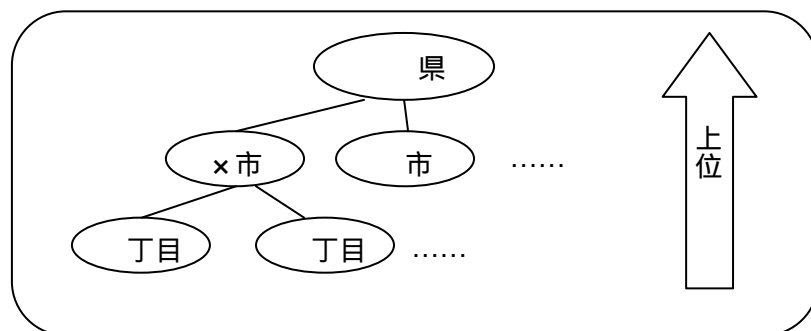
「県」は「市」よりも上位である。例として、一つの県を指定すれば以下に含まれる市区町村を全て把握できる。また、Aさんの現在位置が「 番地」であるとする、そこから「 丁目」「××市」「 県」とユニークに上位のデータを指定できる。そして結局 Aさんがどこの県に所属していたのかが把握できる。

➤ パソコンにおけるフォルダ分け

上位のフォルダを指定すれば、それ以下のフォルダを指定できる。以下似たようなもの

➤ 推移図

上から下へと状態が推移していくものとします。ある操作を指定するとその操作に従って下位の状態へと推移していきます。ある状態に対して複数の操作が指定できる場合がほとんどで、枝分かれ構造を形成します。オートマトンはこの例とも言えますが、上位に戻ることもできる点で少し異なります。



階層モデルのイメージ

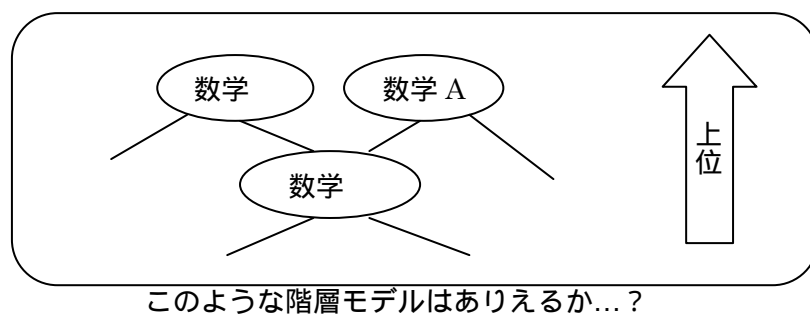
などがあります。他にも例はいくらでも思いつくはずですが、この問題は、階層モデルの特徴を一通り把握しておけば解ける問題です。(b)や(e)は非常に典型的な階層モデルの例と言えるでしょう。なのでこの解説は特にしません。

では一つずつ検証していくことにします。まず(d)をよく読むと、「ビデオ予約録画」という大まかなくくりから、「開始時刻の設定」という下位のくりに階層分けする、とあります。このような文章が与えられたときに大切なのは、**具体例を幾つか考える**ことです。ぱっと見ただけでは「ビデオ予約録画」と「開始時刻の設定」があるだけで、それらがどのように関係し合っているのかがよくわかりません。しかし、「ビデオ予約録画」の下位に位置する「開始時刻の設定」と同等な作業として例えば「録画チャンネルの指定」や「画質（圧縮率）の設定」を挙げることができることに気づけば、これが階層モデルを用いることのできる例として認識できるでしょう。それ以下には例えば「 のボタンを押す」とか「ディスクを取り出す」とか、さらに具体的な作業が続きます。なので、上位の作業をするためにはその下位に続く行為をしていけば良いということになります。さらにその具体的な操作が「 のボタンを押す」という最も原始的な行為にまで達すると枝は終点(最下位)となり、そこまでに通った経路の本数を数えることで、手順の回数を知ることも可能です。つまり(d)は**階層モデルを用いるのが有効な例**ですね。

ではもう一つ、少しわかりづらい(a)を考えてみることにしましょう。こういうときは**実際に図に描いてみることも必要**です。ここでは「階層モデルとしてあるべき条件を満たすように図を描いていき、その条件の一つに反するような描き込みをしなければならなくなったときに、この例が階層モデルに適しないという結論を出す」という手法を使います。

ひとまず、大学の科目一つ一つを描いていくわけですが、その際に**全ての科目は一つずつしか存在しない**(の五つ目)ことに注意しましょう。そして、相対的に上位に位置するのは、「先に履修しなければならない科目」です。例えば『「情報科学」を履修するためには「情報」が既履修』とか『「数学」を履修するためには「数学」と「数学 A」が既履修』とか、『「電磁気学」や「物性科学」を履修するためには「力学」が既履修』などの条件です。これらは全て階層モデルの特性を満たすでしょうか。

一つ目は「情報」を下位、「情報科学」を上位にすればいい話です。二つ目は「数学」を下位、その上位に「数学」と「数学 A」を置けばいい話に見えますが、よく考えるとこのような例はあり得るでしょうか。

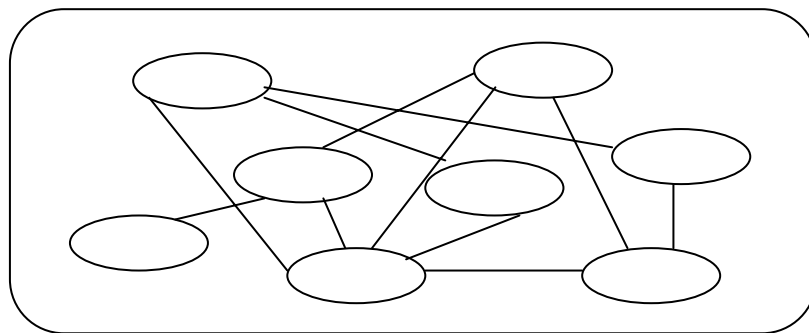


住所の例で言えば、「山口市」が「志村県」と「堀内県」に含まれているということになります。なんだかまずいような気がしませんか？

とまあ、ここまで誘導しておいてあれなのですが、結論としては問題ありません。ごめんなさい。住所の場合はそれぞれが属する上位集団は一つに限られる必要があります(なぜなら住所は「上位をたどって**唯一**の集合にたどり着く」ことが必要な仕組みだからです)。この例では、ある科目を選択する際に予め履修しておかなければならない科目を把握できれば良いので、一つの要素が二つ以上の集合に属していても問題ありません。「数学」に焦点を当てたときに、予め履修しておくべき「数

学」と「数学 A」が(さらに、それぞれを履修するために必要な科目まで)知れるのですから。なので、この例を階層モデルを用いて表すのは有効であると言えます。

最後に残った(c)を考えます。ぱっと見たところで違うように思えば、データモデルがそれなりに理解できているということです。この問題を解くには階層モデルの知識だけでは足りません。というのも、インターネットがどのように形成されているかを知っている必要があるからです。インターネットは「WWW(ワールドワイドウェブ)」と呼ばれるように、コンピュータとコンピュータ(正確には少し違いますが)が網目状につながっています。これは元々、ある一つの経路が遮断された時でも別の経路を通ることで目的地にたどり着くことを可能にするために組まれた構造なのですが、この説明はインターネットのところでやります。つまりインターネットの構造は、上位と下位のように順位づけられた階層ではなく、**全てのデータが網目状に組み合わさったものである**ということです。下の図を見るとよりわかりやすいと思います。これより、(c)は階層モデルを用いるのが適切でない例であり、その理由としては、一つのルータやコンピュータから双方向の通信をしているためそれぞれに順位が存在しないことや、もっと言ってしまえば幹から派生していくような木構造ではなく、等価なデータが網目状に組み合わさった構造をしているから、という理由が挙げられるでしょう。



インターネットの構造例

以上長くなりましたが、これで共通問題1の解説は終わりです。データモデルについては詳しく説明したつもりですが、階層モデルに偏っているので、その他のモデルも把握しておく必要があります。実際にこのような「使用が適切かどうか」の問題が出たときに、説明も含めて解答できるようにしておきましょう。また、「特徴を述べよ」と聞かれたときも決定的な特徴を二つくらい挙げられるようにしておくことが大切です。

共通問題 2

問題解説

これは単なる計算問題です。情報量の計算は非常に典型ですので、得点源にしましょう。

情報量、というと漠然としていて、さらにそれを数値にするとますます実感のわかないものになってしまいますが、何かの情報を知ったときそれがどの程度価値の有るものを表す指標と思ってもらえればほぼ間違いないと思います。曖昧さがどれだけ減ったか、とも考えられます。そしてその基準は「二択の(同様に確からしい)選択肢がどちらであるかを知ったとき」の情報量を1とします。この感覚は非常に重要なので体にたたき込みましょう。これがわかれば「四択の選択肢が一つに絞れたとき」の情報量は2であることは容易に予想できるでしょう。

次にその定義式です。その前にちょっと説明を。上で「容易に予想」と書きましたが、その計算の過程は「まず四つから二つにするのに情報量1、そしてその二つを一つにするのに情報量1。合わせて2」と考えます。つまり、情報量は加法性を満たす必要があるということです。具体的に言えば、情報量 x を持っているときに、情報量 y を得たとすると、持っている情報量が $(x+y)$ になって欲しいという要求です。これを満たした情報量の定義として、以下のようになります。

$$\text{情報量} = -\log_2(\text{知った情報が実現する確率})$$

これによって、確率という「重ね合わせは乗法で行われる値」が加法性を満たすように修正したのです。右辺にマイナスが付いていることと、初めに述べた基準が成り立つように底が2になっていることにも注目しましょう。これで、情報量を求めるという作業は結局事象が起こる確率を求める作業だということがわかるはずですが、問題文の「情報量」を全て「確率」に置き換えれば、センターレベルの問題になります。あとは日本一優秀な東大生の頭で何とかしてください。答えは簡単に載せておきます。情報量は I で表します。

(1) $p(\text{日}) = 1/16$

(2) $\bar{p}(\text{東}) = 1 - p(\text{東}) = 1 - 5/16 = 9/16$

(3) $p(\text{日 or 東}) = p(\text{日}) + p(\text{東}) = 1/16 + 5/16 = 3/8$

(4) $p < \text{日 or 東} | \text{日} > = (p(\text{日})) / (p(\text{日 or 東})) = (1/16) / (3/8) = 1/6$

(5) $I_1 = I_3 + I_4$ (つまり、(1)だけで知ったときと(3)を知ってから(4)を知るのは情報として等価)

(6) 1問目と2問目の選び方は独立なので、日本史が2問などとなる場合もあることに注意しましょう。
 $p(\text{どちらか日}) = p(\text{日}) * p(\text{世界史}) = (1/16)(5/16) * 3 = 15/256$

(7) $p < \bar{\text{米}} | \text{日 and 東} > = \frac{p(\text{日}) * p(\text{東})}{1 - p(\text{米})} = \frac{(1/16) * (5/16)}{11/16} = 5/176$ 。これもやはり、情報量の加法性が確認で

きます。

共通問題 3

問題解説

問題 A

1.

GUI (Graphical User Interface) と CUI (Character User Interface) についての問題。GUI とは名前通り「グラフィックを用いてパソコンを操作するしくみ」のことで、パソコンの画面でマウスを動かし、アイコンをクリックし...、といったより直感に基づいた操作を可能にします。これのおかげでパソコンが一般人にも使いやすくなりました。一方 CUI はキーボードから文字を打ち込んで情報を処理する仕組みで、演習で苦戦した Mac のターミナルや Windows ではコマンドプロンプトの操作がこれに当たります。「こんなことをして何が得なんだよ...」と思った人が多数だと思いますが、CUIの方がGUIより遙かに優れている点があります。何よりそのシンプルさです。GUIではマウスを動かし、アイコンを選択して、それが新しいウィンドウとして開くまでに膨大なプログラムが動いています。それに対して CUI では、一つのファイルを開くには、ファイルを開くための数文字のコマンドを入力し、ファイル名を指定すればそれで終わりです。一つのファイルでは感動が少ないかもしれませんが、大量のファイルを操作したりある種の作業を繰り返したりする際には CUI は GUI に比べて圧倒的に有利になるのです。まあ、一般的に使う分にはその直感性から GUI の方が便利なのは確かです。

では問題に入ります。前半は「線分を描く際の操作方法」で、後半は「手間の違いを説明」です。A はキーボードのみしか使えないので、線分を描く際には「始点の座標」と「終点の座標」などを指定していくことになります。B ではマウスしか使えないので「始点をポインタでクリックしてから、終点でもういちどクリックする」とか「始点でマウスのボタンを押し、終点までドラッグしてボタンを離す」などの方法があります。これは特に問題ないように思います。

手間については、これは二種類に分かれます。まず「とにかく線を引きたい」場合には A では手を動かすことが B よりも多くなってしまい、B の方が手間が掛かりません。しかし「正確な長さの線分を正確な場所に入力したい」ときは A の方が手間が掛からないと思います。GUI は必然的に体で操作をするため、微妙な感覚のずれが線分の始点、終点、長さなどのずれに繋がり、さらにある決められた座標から座標をつなぐ線分を描く際には、必死にディスプレイを凝視し、手を震わさないように神経を研ぎ澄ませながらクリックとドラッグを行わなければなりません。

9 章の試験範囲はこの GUI と CUI のところだけです。それぞれの特徴と、有利な点、不利な点をいくつか挙げられるようにしておけば十分だと思います。

2.

これは「不正アクセス防止法」を知っているかどうかの問題です。専門的知識もあった方が良いでしょうと思いましたが、教科書にもあまり詳しく書かれていないし、ぐっふのページからは条文そのものに飛んでしまうのでいまいちです。なので、一般常識的な範囲で答えることにしましょう。教科書には p.244 あたりで解説されています。

この法律で禁止されているのは、ざっくり言えば「不正にアクセスすること」です。もうちょっと言えば、「禁止されていそうな行為をして、サイトやコンピュータに接続(侵入)すること」です。これ以上細かく言うと定義文になってしまいそうなのでこれくらいで。なので、(a)のように「不正なサイ

トにアクセスすること」でもなく、(b)のように「**他人のアクセスを可能にすること**」でもありません。(d)は最近の「ダウンロード違法化」によって違法行為になりつつありますが(これまではダウンロードは違法でなかった)ここでの不正アクセスには当たりません。サーバには認証機構(パスワードなどを入れないとログインできない仕組み)が入っていないのでアクセスすることは問題ありません。

不正アクセス行為と条文に定義されている内容を書き出してみます。全部で三種類あります。ちなみに全てインターネット上の話です。

他人のアカウントやパスワードを用いて、制限された内容を利用すること

上記以外の方法で、制限を突破すること(セキュリティホールやウイルスなどを用いる)

他のサーバやパソコン側に侵入し、制限を解除してしまうこと。(これもセキュリティホールなどを用いる)

抜け穴をなくすためにえらくややこしく書いてありますが、結局は制限された内容を勝手に見てはだめよ、という内容です。それ以上のことはありません。

10.3で触れられている内容は、「著作権」と「プライバシーとセキュリティ」があります。

著作権について。有形の著作物(文学、音楽、写真など)は他人のものをパクったかどうかで決まるので扱いが比較的楽です。一方デジタルデータになると扱いが微妙になってきます。明らかに著作権を侵害しているように「思える」行為でもそれを取り締まる法律がなかったり、大して問題なさそうな行為でも解釈によっては違法になったりします。その複雑さが顕わになったのが Winny に関する訴訟で、教科書とぐっふのページにも結構詳しく書いてあります。この辺りの問題は専門者でも解釈に困るところがあり、正誤を二分するような問題は作りにくく、出しても解釈を述べるような問題につきるように思います。しかしそれでは採点が難しいので結局はこの分野は出題されないような気がします(ここ三年の問題で著作権を扱った問題はありません)。まあこれは予想でしかないので外れてもおとがめ無しで。

プライバシーについても出題されるかどうかは微妙なところで、最低限「個人情報保護法」についてそれで規定されているルール(p.244~245 参照)を把握しておけば問題ないはずです。ただ条文を一字一句覚える必要はなく、概要を読んでどのような行為が禁止されているかを把握し、事例に合わせて判断できるようにしておけば大丈夫でしょう。

セキュリティは不正アクセス禁止法に尽きます。p.246 はさらっと読んでおくだけで十分です。

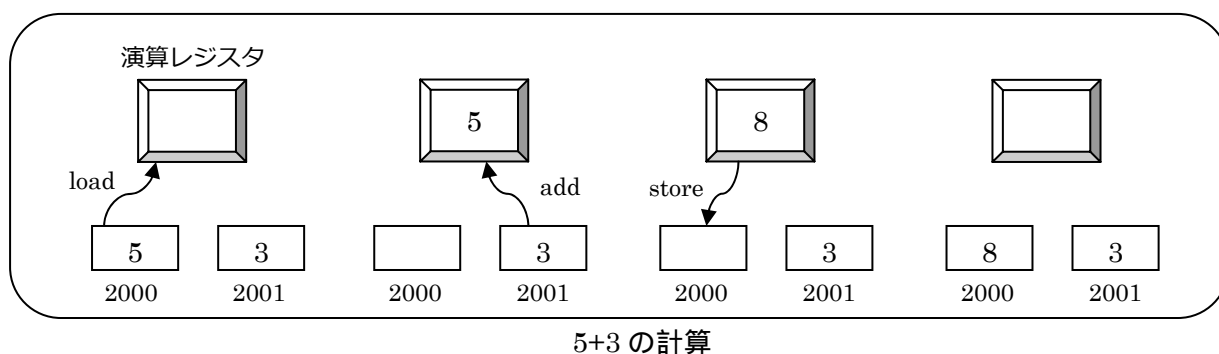
問題 B

計算プログラムの問題です。教科書の p.154~に載っているプログラム例は、足し算と引き算しかできない(逆に言えば、足し算と引き算はできる)機械を仮定して、それにどのように命令を与え計算をさせるかを考えてみたものです。ここで大事なのは、**命令を与えられている機械になりきる**ということです。余計な考えや混乱を起こさずに愚直に計算していけば解けます。なので A の問題より解きやすいのです。

ここで必要なのは「演算レジスタ」と「アドレス」という概念です。演算レジスタとは言ってしまうと作業台のようなもので、命令によってある数字をその上に持ってきたり、作業台の上にある数字にある数字を加えたりすることができます。一方アドレスには基本的に二種類あり、命令が書かれているところ(問題で言えば 1001~1011)と数字をストックしておくところ(2001~2004)がありま

す。命令が書かれている部分は基本的に改変を受けることはありませんが、数字をストックしておくところは計算が進むに従って変わっていきます。

注意して欲しいのは演算レジスタとアドレスの違いです。**アドレスにある数字を直接演算することはできません**。例えばアドレス 2000 に 5 という数が入っているとします。これに 3 を加えたいとすると、まず空いているアドレス（ここでは 2001 とします）に 3 を記録しておいて、一旦「2000 のデータを演算レジスタにぶち込む」（記号で書けば『load 2000』となります）という命令を与えてから、「演算レジスタにある数に、アドレス 2001 にある数を加える」（『add 2001』）という命令を与え、最後にもう一度、「演算レジスタにある数をアドレス 2000 に戻す」（『store 2000』）という命令をすることになります。以下の図を見ればもう少しわかりやすいかも知れません。



この図がイメージできれば、あとは計算を追っていただくだけです。プログラムが始まると、機械は番号の若い順に命令を実行していきます。（機械にとって）複雑な計算をする際には、途中で『jump』が出てきて番号の若いアドレスに戻り、同じ命令を何度か繰り返すことをします。その際に使われるのが「回数カウンタ」のように使われるアドレスで、繰り返す回数を予め保存しておいて、計算サイクル一回ごとに 1 ずつ減らしていくのです。問題でいう 2003 がこれに当たります。初め 8 だったデータが、1004、1005、1006 の命令によって、1 ずつ減っていくのがわかります。最後に『jumpzero』の命令を使うことで、初めに決めた回数だけ計算をしたら結果を表示し（write）計算を終了（halt）します。一通り手を動かして「こんなものか」と思ってしまうと、問題なく解けるはずですよ。

ちなみに問題で、アドレス 2001~2004 のデータ（2001,2002,2003,2004）は

(0,7,8,1) (7,7,7,1) (14,7,6,1) (21,7,5,1) ... (49,7,1,1) (56,7,0,1)

と推移していき、最終的には 8×7 の答えである 56 を出力し、停止します。