**データの伝達とデジタル化，圧縮**

**－デジタル化された情報の伝達方法と圧縮技術－**

年 組 番 氏名

目標：（この課題でできるようにしたい目標を書きなさい）

**はじめに**

|  |
| --- |
| 文字情報にはアルファベット・数字・ひらがな・カタカナ・漢字・その他の記号など，様々なものがあるが，コンピュータではすべてのデータを０と１という２つの信号で表している。  コンピュータで文字情報を表現することを考え，文字データを０と１に置き換えて伝達する方法と，効率よく伝達するための圧縮技術を考えてみよう。 |

１　コンピュータ上の文字情報を遠く離れた場所に伝達する方法を考える。

次の文

|  |
| --- |
| Ｉ＿ＡＭ＿Ａ＿ＳＴＵＤＥＮＴ |

といったアルファベットを送る場合を考えてみる。

なお，０～９の数字だけが伝達することができるとし，情報を受け渡しする双方が同じ変換表を持っているとする。

アルファベットの大文字26文字と空白の合計27文字だから，次のような変換表を作る。

表１：０～９の数字に置き換える変換表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ａ | 01 |  | Ｈ | 08 |  | Ｏ | 15 |  | Ｖ | 22 |
| Ｂ | 02 | Ｉ | 09 | Ｐ | 16 | Ｗ | 23 |
| Ｃ | 03 | Ｊ | 10 | Ｑ | 17 | Ｘ | 24 |
| Ｄ | 04 | Ｋ | 11 | Ｒ | 18 | Ｙ | 25 |
| Ｅ | 05 | Ｌ | 12 | Ｓ | 19 | Ｚ | 26 |
| Ｆ | 06 | Ｍ | 13 | Ｔ | 20 | 空白 | 00 |
| Ｇ | 07 | Ｎ | 14 | Ｕ | 21 |  |  |

この変換表に基づいて，問題の文を数字に置き換えると次のようになる。

0900011300010019202104051420（全部で28文字）

「Ｉ＿ＡＭ＿Ａ＿ＳＴＵＤＥＮＴ」を下記のコード表（２進数）に従い置き換え，データ量としては何桁になるか答えよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ａ | 00001 | Ｂ | 00010 | Ｃ | 00011 | Ｄ | 00100 | Ｅ | 00101 | Ｆ | 00110 | Ｇ | 00111 | Ｈ | 00111 |
| Ｉ | 01001 | Ｊ | 01010 | Ｋ | 01011 | Ｌ | 01100 | Ｍ | 01101 | Ｎ | 01110 | Ｏ | 01111 | Ｐ | 10000 |
| Ｑ | 10001 | Ｒ | 10010 | Ｓ | 10011 | Ｔ | 10100 | Ｕ | 10101 | Ｖ | 10110 | Ｗ | 10111 | Ｘ | 11000 |
| Ｙ | 11001 | Ｚ | 11010 | 空白 | | 00000 | |  | | | | | | | |

〔01001 00000 00001 01101 00000 00001 00000 10011 10100 10101 00100 00101 01110 10100〕

全部で何文字〔70〕文字

こうした，変換作業に対して，自分の考えを述べよ。

〔　桁数が多い，処理が大変，表以外の文字はどう表現すればいいか…など　〕

**実習課題**

以下の課題について答えよ。

１　はじめにの１の解答例において，Ａ～Ｉを１～９としないで01～09と考えた理由を答えよ。

|  |
| --- |
| 文字の区切りがわからなくなるから。 |

２　実際の情報機器の伝達では２進法が使われている。それでは，「はじめに」の１の

00～26までは，０と１を用いて何桁の２進数で表したらよいか考えよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 情報機器では，＋と－，オンとオフ，電圧の高低で表すため，２進表現を使う。   |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 10進法 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | … | | ２進法 | 0 | 1 | 10 | 11 | 100 | 101 | 110 | 111 | 1000 | 1001 | 1010 | … |   いま考えているのはアルファベットの大文字と空白の合計27文字だから，  「27」は２進法で「11011」と表すことができる。  ５桁の２進数（２進法で表された数）で考えればよいことになる。  A.［　５　］桁 |

３　実習課題１，２のような例を基にするとアルファベットはどのように表現されるか考えよ。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ５桁の２進数（２進法で表された数）でアルファベットを考える場合，２進数で表された「1111」は５桁で考えることにすれば「01111」と表記する。  　下記の変換表が作成できる。この変換表で表された２進数を**「文字コード」**と呼ぶ。   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | Ａ | 00001 | Ｈ | 01000 | Ｏ | 01111 | Ｖ | 10110 | | Ｂ | 00010 | Ｉ | 01001 | Ｐ | 10000 | Ｗ | 10111 | | Ｃ | 00011 | Ｊ | 01010 | Ｑ | 10001 | Ｘ | 11000 | | Ｄ | 00100 | Ｋ | 01011 | Ｒ | 10010 | Ｙ | 11001 | | Ｅ | 00101 | Ｌ | 01100 | Ｓ | 10011 | Ｚ | 11010 | | Ｆ | 00110 | Ｍ | 01101 | Ｔ | 10100 | 空白 | 11011 | | Ｇ | 00111 | Ｎ | 01110 | Ｕ | 10101 |  |  | |

４　文字コードを”01001”のように表すと，区切りが分かりづらい。人間が分かりやすいように文字コードはどのように表記を工夫しているか調べよ。

|  |
| --- |
| ２進法の４桁で表現される数は１０進法で「０～15」となる。このとき，変換された「10～15」は２桁の数であって紛らわしいので，これを「Ａ～Ｆ」と表すことにすると，「０～９，Ａ，Ｂ，Ｃ，Ｄ，Ｅ，Ｆ」と表し直すことができる。  　この方法にしたがうと，先ほどの「２進法の４桁」を「16進数の１桁」で表現できる。  　この変換を16進法といい，変換された「１～Ｆ」を16進数という。  注　文字の「Ａ」と16進数の「Ａ」は違うものなので注意！！ |

５　例えば「ＵＵＵＶＶＶＷＷＷＸＸＹＹＺＺＺＺ」という文字列を短くして伝達する工夫に「圧縮」という考え方がある。どのように圧縮すればよいか考えよ。

|  |
| --- |
| 「Ｕ３Ｖ３Ｗ３Ｘ２Ｙ２Ｚ４」というように圧縮してから２進数のデータに変換することが考えられる。これをランレングス圧縮という。  　ただし，問題１などの文字の重なりが少ないものは効果がない場合や，逆に情報量が増えてしまうこともある。  　これに対応した，「PickBits」といった技術もある。 |

６　日本語の文字コードにはどのようなものがあるか調べよ。

|  |
| --- |
| JISコード，シフトJISコード，EUC，UNICODE |

**確認テスト**

次の各問いに答えよ。

(1) アルファベットを２進数で表すコード表を作成したい。桁数は何桁以上にする必要があるか。

〔５桁〕

(2) ひらがな（濁点・半濁点，小文字なし）で書かれた文を２進数で表すために作成するコード表の桁数は，何桁以上にする必要があるか答えよ。

〔６桁〕

(3) 文字コードを０と１を使って表す理由を答えよ。

〔コンピュータ上ではデータを＋と－（オンとオフ）で表すため，２進表現を使う〕

(4) 実際のアルファベットの文字コードは８桁の２進数で表される。しかし，(3)のように０と１で表されると人にとっては非常に見づらい。このため実際のコード表にはどのような工夫がなされているか。

〔16進数で表す〕

(5) (4)の方法の利点を簡単に答えよ。

〔16進数を用いると，２進数の４桁ずつと変換を対応させることができる。〕

(6) データを圧縮する考え方として，ランレングス法がある。この方法について簡単に説明せよ。

ランレングス法：〔”ＡＡＡ”と同じ文字が続く場合，何文字続いているかを計算し”Ａ３”というように圧縮してから変換を行う方法〕

(7) ひらがなと漢字（常用漢字2136字）で書かれた文を０と１の数字に置き換えるとき，変換表の数字は理論上何桁必要になるか理由をつけて答えよ。

〔２１２＝4096となるので，12桁あれば十分表現できる〕

(8) また実際にはどのような変換表が用いられているか述べよ。

〔実際には，２バイト（１バイト＝８ビットなので16桁）で表されており，２16＝65536種の文字が表現できる〕

**自己評価シート**

あてはまるものにチェックを入れなさい。（自分が何について理解し，できるようになったかを振り返るものです。正直に答えること）

|  |  |
| --- | --- |
| なぜ文字データを０と１に置き換えなければならないか説明できた。 |  |
| アルファベット大文字＋空白の27文字分のコード表を作成できた。 |  |
| 解答例でＡのコードを１としないで01とした理由を説明できた。 |  |
| なぜ16進数表記を使うか説明できた。 |  |
| ランレングス圧縮の方法を理解できた。 |  |
| ランレングス圧縮の改良方法を調べた。 |  |
| 日本語の変換コードについて調べた。 |  |
| 確認テストの半数以上の内容について理解できた。 |  |
| 確認テストの大半の内容について理解できた。 |  |