

データを人知れず短くしているのは置き換えパズル! (実験)

所属 一般科目
担当者 中村 博文

この出展は資料や実験を皆で見たり聞いたりしながら、データを短くしているしくみのいくつかを確かめていきます。参加者は手ではなくて頭を働かせます。

①通信や保存でデータが機器内などでどうなっているかおさらいします。(→0と1)

②画像を例に、0と1でどう表しているかや、短く表す工夫を複数確かめます。

基本のひとつは、よく現れる内容を短かい内容に、あまり現れない内容を長い内容に表し直すことです。使うときに元に戻します。

  ⇒ ⇒ 10110101...

③音声についても確かめます。携帯電話などで使われている、近いものが得られる内容に表し直す方法も確かめます。

)))  ⇒ ⇒ 01101011...

対象 小学4年生くらい ~ 大人まで。

所要時間 25分くらい。

実施人数 20人くらい。同伴者を含む。各回とも先着の着席者。自由席。

その他 現代社会を人知れず支えている「計算を使った役立つしかけ」の一部について、今日、自分の頭で、そのひらめきを追体験してみませんか。

性能を上げるための工夫によっては、パズルがより複雑や面倒になります。

発展 (もしもしかったら、いろいろ調べて、もう少しきわめてみてください。自由研究になるかは提出先に確認を。)

- データは、いろいろの機器で共通する表し方として多くの場合に0と1を使って表します。0と1の並びを短く表し直すこと(一種の符号化)をデータ圧縮や圧縮や情報源符号化と言います。あちこちでデータを短く表し直していて伝送時間や記憶場所を節約しています。言い換えると、時間や記憶場所を何割増しや何倍にも有効活用しています。使うときに元に戻します。それを復号や伸長や解凍などと言います。
- 0~9を使って通し番号を付ける十進数の様に、情報機器では0と1で通し番号を付ける2進数で考えることが多いです。値だけでなく文字や音や画像もそうです。0や1のことをビットと言います。2進数のけた数はビット数とも言います。確率(現れる度合い)が分かるときの表し直す最適な長さが知られています。例えば、確率が半分なら1ビット、1/4なら2ビット、...、1/256なら8ビットです(早いと高校で習うことのある対数で表すと、確率がpなら $-\log_2 p$ ビット)。ぴったりにできなくても圧縮はある程度できます。
- 従来の電話の音質は音の強さを256段階に区別(8ビットでせる)して1秒あたり8千回データを取った程度です。携帯電話などでは、音声のデータそのものの圧縮ではなくて、音声の特徴データまたは過去の内容との差の特徴データを送ります(音声圧縮よりは音声符号化と言います)。受信側はそれをもとに計算で作った数値を音にして聞きます。特徴データから作った内容ですので話者の声ではありません。携帯電話では段階数もデータを取る頻度も増やしたのを、例えば0.02秒毎に百数十ビットにして送ります。今日の実験では単純化して複数の音の高さ(周波数)と強さ(振幅)のデータだけで作った音を聞いてみます。
- 今日は触れられませんが調べてみては：データ圧縮、文字やファイルや動画の圧縮法、平均情報量や情報エントロピー、データ量と情報量の違い、可逆圧縮と非可逆圧縮、ハフマン符号、算術符号、LZ77、情報理論。
- 圧縮は、現れ方に偏りがあって偏りの予測がある程度当てられれば全体を短く表し直せますが、もし予測が大きく外れた際の膨張する可能性と引き換えです。データの種類ごとに方法は沢山ありますが、もし皆さんが考えた工夫に利点があれば広く使われていく可能性があります。備考:aha! はマーチン・ガードナーの著書より。