

日税メールステーション 特別号

税理士が知っておくべき IT 知識～基本の"き"～

第9回「データと情報」

メールマガジンをお読みの皆様こんにちは、株式会社コアブリッジの柳です。

先号で「IoT が真価を發揮するのは、インターネットにつながったモノから集めた情報を効果的に使った時なのです」と書きました。

今回は、あらためて"情報"(information)や"データ"とは何であるか、その二つの違いと、コンピュータにおけるデータの表し方について記してみます。

■ データ→情報→知識→知恵

『DIKW モデル』あるいは『DIKW ピラミッド』と呼ばれる、データや情報に関する分類があります。Data - Information - Knowledge - Wisdom の頭文字をつなげたもので、それぞれおおまかに以下のような意味です。

- ・ Data(データ) : 単なる数字や記号の羅列
- ・ Information(情報) : データを分類や集計して意味を持つようにしたもの
- ・ Knowledge(知識) : 情報から規則性や見解などを導出したもの
- ・ Wisdom(知恵) : 知識を活用して判断したもの

データ → 情報 → 知識 → 知恵 の順に実際的な価値が昇華していくというものです。

たとえば、ある組織で未知の病気の蔓延が発覚した時のことを想像してみましょう(※あくまで説明のための例です)。

データ : 罹患者 20 人の体温は、37.6 度、38.2 度、37.7 度、37.9 度、・・・

情報 : 37 度未満はなし、37.0-37.4 度は 2 人、37.5-37.9 度は 8 人、38.0 度以上は 10 人

知識 : 発熱が 37.5 度以上の場合に感染の可能性が高い

知恵 : 毎朝検温し 37.5 度以上ある場合には傷病休暇を取るようになる

のように、意味(価値)が変化していきます。

コンピュータ(IT)は、データを収集・保存し、それらを集計する(情報処理)のは得意中の

得意ですね。

そこから知識を導き出すのは、以前ならば人間の頭で行っていましたが、AI(Artificial Intelligence)の進歩により、部分的にはコンピュータが代わりを行えるようになってきました。知恵の部分は、一部はAIに肩代わりしてもらえるものは出てきていますが、本来は人間の役目です(コンピュータに決められたくはないです、私は...)

■ コンピュータのデータ量の単位

PCやスマートフォンのカタログを見ると、メモリがxx GB(ギガバイト)、のような数字が必ず出てきます。

どのくらいの大きさのデータを記憶できるかを表しているものということをご承知の通りです。

「コンピュータは二進数(0と1で数値を表す)でできている」とよく言われます。これは、コンピュータを構成する要素(CPUとかメモリとかDVDなどの記憶媒体とかです)が、「電圧の高低」「電子が貯まっている/空になっている」「光を当てた時の反射の方向がまっすぐか曲がっているか」などによって2値を区別し、それによって処理を行うように作られているからです。

この、0か1かを保持するための最小の器のことをビット(bit)と呼びます。

計算等をするにはビットを束ねてある程度の大きさにしたほうが便利です。数ビットをまとめたものをバイト(byte)と呼びます。先程のギガバイトの"バイト"ですね。

蛇足ですが、iPhoneやMacを出しているApple社の林檎のロゴが"かじられた"形状になっているのは、byteとbite(かじる)を掛けたからだと言われていています。

現在は通常「1バイト=8ビット」ですが、コンピュータの黎明期は、1バイトが6ビットだったり9ビットだったりということもあり、「1バイト=8ビットと決めつけるのは知識不足」と言う人も以前はいました。このため8ビットのことをオクテット(octet)と呼ぶこともあります(octaはギリシャ語の8。オクトパス(タコ)、オクターブ(ドレミファソラシド)などでお馴染みですね)。

1ビットで0か1を表すことができ、2ビットでは00,01,10,11の4パターンを表すことができます。8ビット(1バイト)ならば2の8乗で256パターンを表すことが可能なため、1バイトのデータは0~255あるいは-127~127を保持することができます。2バイトならば2の16乗で65,536パターン、すなわち0~65,535あるいは-32,767~32,767を保持することができます。

2の10乗が1,024で、1,000に近い数値のため、2の10乗のことを1キロバイトとよび1KBと記します。

以下、

1MB (メガバイト) : 2 の 20 乗 : 百万に近い値 (厳密には 1,048,576)

1GB (ギガバイト) : 2 の 30 乗 : 十億に近い値 (厳密には 1,073,741,824)

1TB (テラバイト) : 2 の 40 乗 : 一兆に近い値 (厳密には 1,099,511,627,776)

1PB (ペタバイト) : 2 の 50 乗 : 千兆に近い値 (厳密には 1,125,899,906,842,624)

です。

ペタより後は、エクサ、ゼタ、ヨタと定義されています。

ハードディスク (Hard Disk Drive) や SSD (Solid State Drive) のような外部記憶装置の商品説明に「実際の記憶容量は表記の数値よりも少ない」旨が書かれているのは、例えば TB (テラバイト) 単位の場合には、上述のように一兆ぴったりではなく 10% くらいの誤差があるからです。使用可能な容量を確認した時に思ったよりも少なくてなんだか損した気分になりますね...

なお、外部記憶装置の未使用分も、裏ではコンピュータの動作に必要な色々な用途で使用されているので、少なくとも 10% 以上は空けておく必要があります。

■AI の発展においても

DIKW モデル(ピラミッド)を基に、生データに処理や判断を施すことによって最終的には知恵に昇華していくことをざっくりと説明しました。毎日のように取り上げられる"AI"も、現時点では知識への遷移まで行えるのも限定的です。この先さらに進歩し、さらに適用可能な範囲は広がってはいくでしょうが、データを情報化した後、知識ひいては知恵にまで導くためには、何らかの形であっても人の判断が必要不可欠です。

今号は以上です。

では、また次回お会いしましょう。

※本文中の情報、状況、数値等は執筆時点のものです

執筆者

柳 恵太 (やなぎ けいた)

株式会社コアブリッジ代表取締役。

ソフトウェア開発会社、メーカー、教育ベンダーを経て、2014年に株式会社コアブリッジを設立。これまでの、システム開発の上流から下流、受託側から発注側、エンジニアからプロジェクトマネージャー、ユーザーと開発者、企画・営業・開発・提供、日本と海外、社員から経営者といった、組織における幅広い役割を活かし、主にIT企業向けの人材育成やコンサルティング等のサービスを提供している。

情報提供元：



<https://www.corebridge.co.jp/>

※本コラムは、<https://www.corebridge.co.jp/column/> でもご覧になれます。