

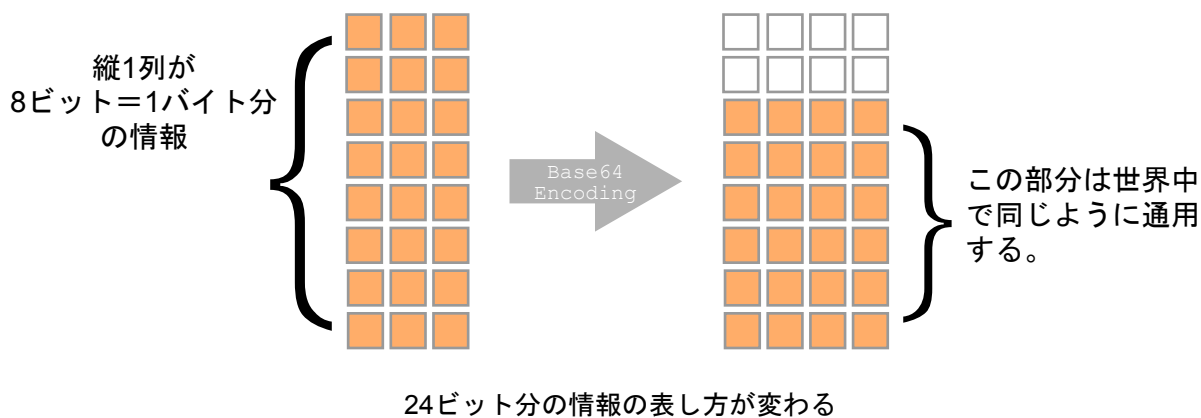
電子メールでは、基本的に文字情報しか送ることができません。だからみなさんが写真などを添付する場合は、このBase64という変換方法を使って文字情報に変換してからメールに貼り付けられます。この変換を行うことで、ファイルのサイズが増えてしまうのです。少し詳しく説明しましょう。

実はコンピュータ上の情報はすべて文字として表すことができます。だから写真などのデータをそのままメールで送ることも、不可能ではありません。ただ、そうした文字の中には、人間には意味不明なものや「特殊な機能を持った文字」も含まれているのです。こうした特殊な機能を持った文字などはそのシステムによって意味が異なる場合があり、世界中のどんなコンピュータにでも送ることのできるインターネットメールでは不具合の原因になったりします。例えばこちらが○という意味のつもりで送った文字が相手側では×という意味になったり、極端な場合は正しくメールが届かないこともあります。

そこで、どんなシステムでも扱うことのできる、「安全な」英数字や記号、合計64文字の組み合わせに情報の形を変換してしまう方法が考え出されました。これがBase64というものです。情報を64文字の組み合わせに変換すると聞くと難しそうですが、実はそれほど難しくはありません。その種はこうです。

コンピュータ上の情報の基本単位は1バイト=8ビットです。そして64文字の組み合わせというのは、実は6ビット分の情報に相当します。（6ビットというのは2の6乗のことです。2の6乗は64ですね。）だから、Base64というのは、8ビットを基本にしている情報の表し方を6ビットを基本とするものに直す作業になります。

しかし、8ビットの情報を単純に6ビットにしてしまうと、2ビット分の情報が欠けてしまいます。つまり、情報が壊れてしまいます。それでは困るので24ビットを一つの単位として変換を行います。24という数字は8と6の最小公倍数で、24ビットというのは8ビット単位では3個分の情報に、6ビット単位では4個分の情報になります。図を見てください。



色の塗られている■が24ビット分の情報です。8×3だったものが6×4になっています。これで、情報を壊すことなく、6ビット単位に、つまり64文字の組み合わせに情報を変換することができました。

ですが、このとき色の塗られていない□の分が無駄になります。コンピュータが文字を表現したりファイルに保存するときは、やっぱり8ビット単位を基本にしています。つまり、Base64という変換を行うと、今まで3列分で収まっていた情報が4列分が増えてしまうのです。このおかげで、実際の中身は変わりませんが、できあがるファイルのサイズが大きくなってしまいます。

どれくらい大きくなるのかというと、基本的には3列が4列になるので 1.333... 倍ですが、写真や音声にはない「改行」が数十文字おきに入りますから、結果として1.5倍くらいに大きくなってしまいます。

これが、「添付するとサイズが大きくなる」正体です。