

「ビッグデータ」に思う

群馬大学 小林春夫

30年以上も前になるが、コンピュータサイエンスの専門家から次のように教わる。
「プログラムはデータとアルゴリズム（ファンクション）から成る。
これまではファンクションを中心に考えてきたが、データを中心に考えるとより良いプログラム（変更容易、バグが少なくなる）が作成できる。
この考え方がオブジェクト指向プログラミングである」

現在、ビッグデータ(big data)がIT分野のバズワード(buzzword)として世の中に跋扈している。次のような表現も最近知る。

AI fuels data. (AI はデータを燃料とする)

しかしながらビッグデータの話は、その名の通りデータの「量」は意識しているが「質」の議論はほとんど聞いたことがない。良質の正しいデータであることを暗黙の前提にし、それで何ができるか、どのように役に立つかを議論していることがほとんどである。

データの「質」が今後重要になってくると思う。

インターネットに書かれていることには明らかに間違っていることも散見する。
「ジャーナリズム」「現場主義」などと難しいことは言わなくても、自分が実際に現場を見る、当事者に話を聞いて書けば間違いは少なくなる。明らかに現場を見ていない、想像で書いているので間違っているというものが目につく。

2者の紛争の場合は 両方から話を聞くことでより正確な情報になる。
個人の現在の立場は直接本人に確認することが必要である。

間違った情報(False Information)がある程度あっても正しい解析結果を出してくるデータ解析技術(False Information Tolerant)が必要であろう。データは時間とともに情報の価値を失う、誤りになることもある。意図的に（悪意をもって）間違ったデータを与えて間違った結果を出すような攻撃、データ改竄に対するセキュリティ（Security）も近い将来問題になってくるのではないか。

「三人市に虎を成す」（戦国策）