

第3回目の情報爆発

関根 聡

ニューヨーク大学 准教授、楽天技術研究所 ニューヨーク所長

情報爆発やビッグデータという用語が世間をにぎわすようになって久しい。しかし、歴史をひもとくとこのような情報爆発は少なくとも過去に2回あり、今回の情報爆発は3回目ということになる。過去の2回の情報爆発は人類の歴史の潮流を大きく変えてきており、それらを顧みることは、過去2回の情報爆発に肩を並べると考えられる今回の情報爆発を考えるのに有益であろう。

歴史的な情報爆発の1回目は「文字の発明」である。文字の発明以前の人類は情報を主に人から人に文字以外の方法で伝えていた。そのため、人類の過去の経験や文化の伝搬・伝授は、一部の人間による口承などに頼っていた。そこには神話や伝説など宗教的・呪術的な要素までも含まれ、一部の知的な人々による独占的な文化伝達形態が存在していたと考えられる。文字の発見によって、人間が扱う情報の量は桁違いに拡大した。また、情報伝達の正確性が確保され、それまでは過去の経験は数世代しか遡れなかったのに対して、何世代も前まで遡ることができる「歴史」と、人類が効率的に生き延びていくのに有効な「知識」の蓄積が可能となった。

2回目はグーテンベルグの「活版印刷の発明」である。文字の発明によって人類の歴史と知識が蓄積できるようになったが、活版印刷以前の時代には、それらは限られた人間の間でしか流通せず、知識の独占が起こっていた。中世以前のヨーロッパの大学は閉じられた形の子弟制度で運営されていたり、知識を持つ者（封建者など）と持たない者（農民など）の階級制度を支えていたりした。それに対し、活版印刷は多くの人々が歴史や知識に触れる機会を実現し、知識の大衆化ともいうべき効果を生み出した。それは識字率を向上させ、中世から近代に向かう教育制度の変革を促した。また、大きな意味で、知識の大衆化は王制に対する革命や民主化にも貢献し、近代から現代に至る新しい政治制度を実現する基礎ともなった。

そして、我々が現在進行形で経験している3回目の情報爆発は「コンピューターとネットワークの発明」である。この発明には3つの大きな特徴があると考えられる。

(1) 膨大な情報を保存できること

コンピューターとネットワークの時代以前にも、情報は紙に書かれた文字として保存されてきた。しかし、紙の耐久性や文字を書くというスピードや困難さを考えると、紙は情報の入力や保存に最も適した形式とは言えない。コンピューターやネットワークが出現した後に保存された情報の量は桁違いに多い。

(2) 膨大な情報を人間以上のスピードで処理できること

膨大な量の情報が単に保存されているだけではその情報は人に役立つ知識とはならず、有効に活用できない。ここでは膨大な書籍を所蔵している図書館を想像すればいいだろう。図書館を利用して有効な知識にたどり着くには経験とノウハウが必要であり、その速度はコンピューターの検索とは比較にならないほど遅い。コンピューターでは、大規模な情報に対する高速な処理が可能であり、情報の活用効率（情報の知識化）が飛躍的に進んでいる。

1987年 東京工業大学応用物理学科卒業
1987年 松下電器東京研究所入社
1994年 ニューヨーク大学研究助手
1998年 同大学研究助教授
2007年 同大学准教授
2010年 楽天技術研究所ニューヨーク所長を兼任

自然言語処理、特に、情報抽出、知識獲得、言語解析の研究に従事。1992年に英国マンチェスター工科大学計算言語学科で理学修士（MSc）を取得。1998年にニューヨーク大学でPh.D.を取得。現在、言語処理学会理事、AFNLP（Asian Federation of Natural Language Processing）役員などを務める。



（3）膨大な情報を共有できること

膨大な情報の生産者と活用者の間に壁があっては情報の価値が最大化されない。もちろん、図書館は知識の共有を実現してはいるが、そのアクセシビリティの不便さは、情報や知識が膨大になればなるほど、せっかくの知識を埋もれさせる方向に向かわせていたと言わざるを得ない。それに対し、ネットワークの技術によって膨大な情報の共有が実現され、いつでもどこでも誰でも有益な知識を容易に活用できるようになった。また、情報の生産者と活用者の垣根がなくなり、全ての人々がどちらの立場にも立てるようになった。

このように、情報の蓄積、処理、アクセスの能力がコンピューターとネットワークの発見によって飛躍的に拡大され、それは明らかに人間の能力を超えるものになっている。もし、情報の知識化が高度に進み、膨大な情報の中から自分に有益な知識を獲得することが非常に容易になったら、将来、人間は知識を意識的に自分の脳の処理対象から分離し、知識に関する蓄積と処理をコンピューターとネットワークに任せるという方向になっていくのではないかと考えられる。既に、我々は難しい漢字やスペルの暗記は必要ではないと感じ始めているし、大抵の知識はウェブで獲得可能だと知っている。このように今後は知識そのものは徐々に軽視されていき、知識の活用能力や人とのコミュニケーション能力が重要視されていくのではないかと考えられる。つまり、3回目の情報爆発においては「人間と知識の分離」が実現されるのではないだろうか。

それぞれの情報爆発において、その情報爆発を実現するための重要な周辺技術の存在を忘れることはできない。1回目であれば、パピルスの作成技術であり文字形態のデザインである。2回目であれば、活字鑄造やインク作成の技術である。そして、3回目の情報爆発において重要な周辺技術とはどのようなものであろうか？もちろん、コンピューターとネットワークのハードウェア技術は重要ではあるが、3回目の情報爆発の焦点の1つが、膨大な情報を知識としていくための高速な処理ということであれば、自然言語処理技術も重要な技術の1つと言えるであろう。文字で書かれた情報を適切な人に適切な形式で提示するための技術である。情報抽出、検索、機械翻訳などの自然言語処理技術は全てこの視点から整理できる。

これらの技術がどのように世界を変えていくのかを予測するには、この3回目の情報爆発が「人間と知識の分離」を実現していくという流れに沿って考えていくのがいいのではないだろうか。過去2回の情報爆発では、それ以前の常識が全く通じない変革が起こった。例えば、2回目の情報爆発では中世の大学は全て崩壊し、知識の大衆化を目的とした大学に生まれ変わった。今回の情報爆発によって、現在ある大学の持つ役割は知識の獲得から、広い意味での知識の活用能力という観点に移行していくとも考えられる。我々は歴史の転換点に立っている。しかし、大きな変革は文字形態のデザインやインクの質の改善のような小さな技術革新の積み重ねによって実現されてきた。新しい時代に向けて、大きな流れを意識しながら、小さな進歩を実現していく日々の努力が切に求められている。