

「調べれば分かること」と 「調べなければ分からないこと」

青山学院大学経営学部

助教授 荒木 万寿夫

私は、googleなどに代表される検索エンジンに大変お世話になっています。大学の初等情報教育でも検索エンジンの効率的な利用は、WordやExcelの使い方を覚えること以上に重要な位置づけとなっています。検索エンジンサイトは、ロボット（クローラ）と呼ばれるWeb巡回プログラムによって数多あるホームページの内容を収集し、検索用インデックスを生成していることも、今ではよく知られるところです。ロボットは無数のホームページ間にクモの巣のように張り巡らされたリンクをたどり、訪れた先々のページを片端から収集します。

ロボットとは、文字通り機械的にホームページを収集してくるプログラムであり、事前に「こういう情報を見つけてこい」とか「これについて調べてこい」等の目的を与え、その任務を遂行してくれるものではありません。しかし、蓄積された膨大な情報塊に対して簡単なキーワード検索だけで「調べれば分かること」も、時に私たちの関心や疑問に対する答えそのものとなり、あるいは少なくともそのヒントとなって大いに役立ちます。

私達もその気になれば、関心のある企業等のホームページをすべて見て回って、それらファイルをパソコンに保存することができます。手作業に依らず、指定したサイトをまるごと自動的に保存してくれるロボットのような無償プログラムも、ネット上では容易に見つかります。つまり、当該サイトはどれくらいファイルを持っているのか、多少の手間はかかるにせよ、それは技術的には思い立って「調べれば分かること」の例と言えます。

私はある研究プロジェクトに関連して、東証一部上場企業の中から180社ほどを選んでコンテンツ量(KB)を実際に調べた実験結果をみたことがあります。サイトを構成するファイル総量に占める画像ファイルの割合に着目すると、画像を多用して視覚訴求力を志向するか等について、業種によってかなりその傾向に違いがあることが分かりました。例えば、壮麗な建造物を実績としてアピールしたい建築業と、文字情報ベースでも兎に角正確な情報を伝えることを旨とする金融保険業といった性格の違いをみることができました。

実は上の試験調査では、企業規模とサイト充実度との関係を調べたいという意図が実施前に先ずありました。上場企業には有価証券報告書の提出義務がありますから、企業名で照合してコンテンツ量と売上高の関係をみるすることができます。横軸に売上高を、縦軸にコンテンツ量（いずれも常用対数値）をそれぞれにとって散布図を眺めてみたところ、確かに正の相関が確認できました。このような分析結果は、



初めから目的意識をもって「調べなければ分からなかったこと」の例と言えます。

ところで、平成17年実施の国勢調査に係わる調査票の未回収問題は、マスコミも大きく取り上げ社会的関心をよびました。居留守を含む調査拒否等、調査環境悪化の背景として、プライバシーや個人情報に関する国民の意識の高まりが指摘されました。こうした状況を受けて、本年1月から竹内啓東京大学名誉教授を座長とする「国勢調査の実施に関する有識者懇談会」が計7回開かれ、先ごろ、調査票の原則郵送回収やインターネット申告を認める等の内容を盛り込んだ報告がまとめられました。この報告の中では、「業務の効率化や精度向上のために行政情報を積極的に活用していく方策は十分に検討すべきである。」としながらも、同時に実地調査を行うことの重要性が強調されています。

従来から住民基本台帳などの行政記録を利用すれば、国勢調査は必要ないのではないかという議論がありました。住民基本台帳からは住所、氏名、性別、年齢、世帯構成等の「調べれば分かる」情報は確かに得られますが、施策立案の目的に添った「調べなければ分からない」社会的経済的属性に係わる情報はありません。それでは個人に関する他の行政記録も併用できれば良さそうですが、(国民総背番号制の導入がない現状では)個人コードに依らず氏名情報等で各種の行政記録をリンクすることは技術的に困難でしょう。上で例示した企業情報と有証データの照合のように、上手くいきません。

分析目的にかなう世帯や個人に係わる統計データを、行政記録などの既存情報だけでまかなっていくことは、企業情報に比べてもさらに困難だと思います。当該目的に沿った調査を設計し、実査を経て初めて得られる「調べなければ分からなかったこと」には、それだけに大きな価値があります。価値ある統計は国民が共有する資産なのだという認識を、私達はいま明確に確認する必要があるのではないのでしょうか。その資産を形成し活かし、そして将来の世代に継承していくことが、統計に携わる者の仕事なのだと思います。