

日本の展望

—学術からの提言2010

情報革命 —情報爆発と情報融合炉—

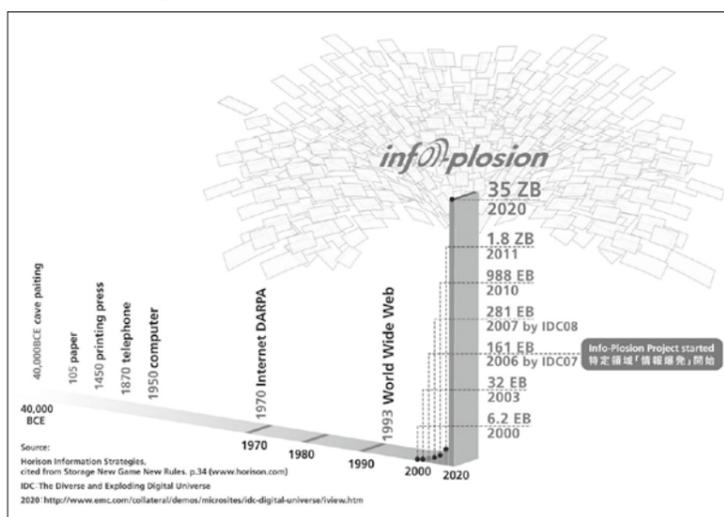
喜連川 優



■情報爆発

時代とともに人類の創出する情報量は増大し続けており（図1、縦軸：情報量、横軸：年代）、取分け、21世紀以降、爆発的な増大が見て取れる。筆者らは当該現象、即ち「情報爆発」を今世紀における情報分野における最も重要な研究課題の一つと捉え、そこから派生する新たな課題の解決に取り組むと共に、情報爆発を克服する新しい技術の創出を目指し文部科学省科学研究費特定領域研究「情報爆発時代に向けた新しいIT基盤に関する研究」を2005年より推進している。

図1 情報爆発



■ウェブの登場と検索エンジン

情報が爆発的に増大していることは多くの人々が日頃感じていることであろう。取分けウェブの登場によって、国際会議の発表論文を始め昔はなかなか手に入りにくかった資料も最新の情報が容易に入手可能となる等恩恵もあるものの、その量は咀嚼可能な限界を遥かに超える状況に到り、情報過多な時代においては、真に必要な情報を効率良く見つけ出すことを可能とする検索手段の重要性が従前に比べて大きく増したと言える。商用検索エンジンの利用度合いは、米国、ヨーロッパ等では圧倒的にGoogleが首位に立っているものの、ア

ジア圏では必ずしもそうではない。中国では百度、韓国ではNaver、ロシアではYandexが極めて強い。これは所謂ローカリゼーションが重要であることを意味しており、まだまだ日本語に強い検索エンジンの開発意識はあると考えられる。現在の検索エンジンは、検索語をシステムに投入することにより、当該語を含むページを返す仕組みであるが、そもそも、我々

が欲するものは、自然な質問に対してのより直接的な回答である。現行の検索エンジンは、簡単な問合せ、即ち、ブックマークの替わりに利用するような場合には機能するものの、「高齢化に伴い成長が期待されるビジネス」のような複雑な調べ物をした場合には、未だ非力である。単語に分けて羅列したとしてもそれらの語が含まれていることしか保証されない。しかも、与えた語に強く反応し過ぎる傾向があり、同義語、類語は考慮されない。これは質問者にとって的確な検索語を考え出さねばならないことを意味し、悩ましい限りである。検索エンジンは現時点ではまだまだ未成熟なレベルにあり、長期的な研究開発が不可欠な領域と言える。複雑な問いに耐えるべく情報爆発特定領域ではTSUBAKIと名付けた日本語の為の次世代オープンソース検索エンジンを開発している。また、「子供手当」「赤ちゃんポスト」等、定義よりも、その施策に対しての意見分布を見たい場合も多い。現在の商用検索エンジンでは多数派の意見が検索結果の上位を占めることとなり、少数派の意見はそもそも見ることが容易ではない。何が正しいと一概には言えないトピックに関しての意見分布を抽出する新しいシステムの開発も進めている。これらの研究を実施するには、既存の商用検索エンジンを用いその検索結果に手を加えるだけでは限界があり、自らウェブページを収集することが必須となる。膨大なページを収集・クレンジング・索引付けするには多大なエネルギーを必要とし、前処理だけで

疲弊してしまうことから、多くの研究者が協創可能な共用研究開発基盤が不可欠と考え、一步一步その構築を継続的に推進し、ようやく分散コンピューティング基盤上に検索エンジン開発プラットフォームが出来つつある。

■非ウェブ情報を利用した新しいITサービス

ウェブが重要な情報源であることは論を待たないものの、ウェブ上の情報が全てではないことも事実である。2008年7月末にグーグル社は1兆URLの存在を発表した。平均ページサイズ等の統計を利用してウェブ情報の量を推計してみると、所謂検索エンジンの対象となるサーフィス（表層）ウェブの量は情報空間全体からするとまだまだ少ない。即ち、爆発する非ウェブ情報の活用に関する研究も極めて重要と言える。人が一日に書くことの出来る文章は有限であり、センサー情報等多種多様な非ウェブ情報が今後重要となってくることは間違いない。情報爆発には二つの側面が考えられる。一つは、情報過多な中で情報に溺れ、必要な情報を探すのに益々時間を費やすようになるというネガティブな捉え方である。前述の検索エンジンはこの為の武器である。一方で、従来、情報が少なかった時代に比べると、この溢れんばかりの膨大な情報を多様に活用することにより、従来にはない新しいサービスを生み出すことが可能となるチャンス到来の時代というポジティブな捉え方もできる。例えば、加速度センサー技術の進歩により日々の運動量を

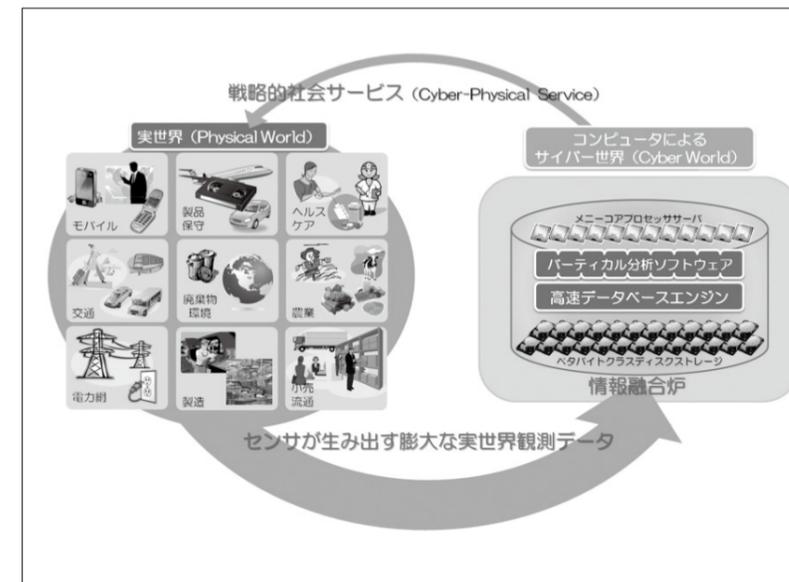
計測することが可能となりつつある。平成20年4月より特定健康審査・特定保健指導が開始された。メタボ症候群の対象者数は膨大であり、例えば、適切なタイミングで、「今週は運動量が少ないので、ひと駅前で降りて自宅まで歩いてはどうでしょう」というような示唆を個人に適した形で提供するモバイルサービスが考えられよう。経済産業省情報大航海プロジェクトにおいて、「情報薬」（適切なタイミングで与えられる情報は薬と同様の効果が得られる）の有効性を明らかにすべく実証実験を行ったところ大変有望な結果が得られた。又、携帯のGPS機能により取得可能な人間の空間的行動情報等も非ウェブ情報の代表例と言えるが、該情報を利用した個に寄り添う新しい情報提供サービスの実証実験を2,000人規模で行ったところ、大変高いコンバージョン率が得られることも分かった。現状の検索エンジンが言葉を売買する広告にその収益を依存しているのに対し、該サービスでは行動パターンがマネタイズの原因となり新しいビジネスモデルが生まれる。この様なセンサー情報を利活用したサービスは我々が調査を行った限りまだ他国では実施されていない。尚行動に関する個人情報は慎重かつ丁寧に扱い、利用者による提供許諾範囲の編集機能も提供した。当初、自分の位置情報の提供を粗く設定する利用者が多かったが、実証実験を進めるに従い、より詳細な情報を提供する傾向があるという極めて興味深い実験結果も得られた。より適切な情報が提供されること

がより詳細な自己情報の提供動機になったと想定される。非ウェブ情報の活用は今後大きな広がりが見込まれる。

■ 爆発する情報の分析による価値創出

2008年Googleは1日に20ペタバイトのデータを処理していると学術論文に発表した。2010年e-Bayは1日に50ペタバイトを処理すると報告した。不況下においてもe-Bay、Amazon等サイバー産業は良好な成長を遂げている。巨大ITインフラを用い、従来では考えられなかった膨大な量の情報を解析することにより機動的な企業戦略を練る時代となってきた。又、精緻な大規模シミュレーションと観測データの融合によって、より正確な天気の前予測を行い、その結果に基づきダムの事前放流を制御することで豪雨に伴う洪水被害の軽減を図ろうとする試みも始まっている。所謂スマートグリッドでは、スマートメータから収集される膨大なデータを解析することによりピークカットを実現する電力料金プランが設計可能となる。農業では、土壌水分をセンサーでモニタし大量情報を解析することにより散水量を大幅に節約する計画が進んでいる。機器の状況をセンサーにより詳細にモニタ・解析することで、保守コストの低減を図る等、その応用対象はエネルギー、水、道路、ビル、街、配送、廃棄物管理等多岐にわたる。これらは要するに大規模な系を可観測とし、先進的センサー観測から得られる膨大なデジタル情報を徹底的に解析す

図2 情報融合炉



ることで系の抜本的な効率化と新しいサービスの創出を狙おうとしており、共通する所が多い。筆者は膨大な観測データを格納する高性能データベースエンジンと多様なパーティカル分析ソフトウェアから構成されるシステムが、次世代の社会プラットフォームとして中核的な役割を果たすと考えており、情報融合炉と名付けた(図2)。米国においてもサイバーフィジカルシステムなる同様の研究が精力的に推進されつつある。

■ おわりに

最近「by IT」という言葉が良く使われる。情報技術そのものの更なる探求も重要である一方、最新のITを駆使することにより、必ずしもITの活用が十分に進んでいない社会シス

テム等を飛躍的に高度化する研究の重要性が謳われている。多くの場合爆発する大量情報の解析が重要な役割を果たす。紙面の都合上、解析基盤を形成するクラウド等の技術動向について述べる事が出来なかったが、高性能コンピュータシステムが従来にないスピード感のある新しいメディアやソーシャルネットを生み出す核技術となる時代

にもなっている。例えば、ツイッターなるリアルタイムブログが我国においても人気を集めている。ハドソン川の旅客機不時着もtwitpicなるシステムに投稿された画像が瞬時に広まり、その後TVのニュース報道で利用された。1日に6,000万件投入されるつぶやきは、一秒当たり120万件の速度で配信される。その実現には、膨大な情報を超高速にさばく最先端のITが肝となっている。情報爆発を中心として多様な先進的技術が今後益々精力的に生み出されてゆくと考えられる。

Profile

喜連川 優 (きつれがわ まさる)
日本学術会議第三部連携会員、東京大学地球観測データ統融合連携研究機構長、生産技術研究所教授、戦略情報融合国際研究センター長
専門：データベース工学