

【総括記事】



特定領域研究「情報爆発 (Info-plosion)」： 本格稼働から2年を経過して

喜連川 優^{*1}

*1 東京大学

ゼットバイトの時代へ

2008 年の CACM に Google 社の MapReduce 処理は 1 日に 20 ペタバイト (PB) と発表され¹⁾ その迫力に驚嘆した。AT&T のネットワークを 1 日に流れる情報量は近日 10 ペタバイトを超えるとされている。膨大な情報の流れと、膨大な情報の処理は、今、まさに情報爆発時代の真ただ中にあることを示していると言えよう。本特定領域「情報爆発 IT 基盤」の申請を行った 2004 年の秋、情報生成量はエクサバイトの単位で計量していた。2007 年および 2008 年 3 月に IDC より出されたレポート²⁾ によると 2010 年には人類の年間に創出する情報量は 1 ゼットバイトに達すると予測されている (図-1^{☆1})。情報爆発の勢いはますます加速している。

2008 年 2 月、米国 National Science Foundation (NSF) は Google, IBM と CluE (Cluster Exploratory Relationship) なるパートナーシップを結び、学だけでは構築・維持困難な大規模クラスター (1700 ノード) の利用を産学連携によって実現すると発表した。本プロジェクトにおいても、情報爆発時代においては爆発する情報量をさばくプラットフォーム技術が肝となることを認識し、2004 年、特定領域研究に新たに導入された支援班というスキームを利用し、研究者全員で共に構築し、運用する共創プラットフォームを提案した。2006 年より InTrigger と名付けた独自の分散計算環境を一步一步構築し、現在、約 1000CPU コアに達している。予算が限られており、規模は決して大きくできないことから、大学の有する大規模な計算資源、TSUBAME (東工大) や T2K (東大、筑波大、京大) とも連動可能とするように

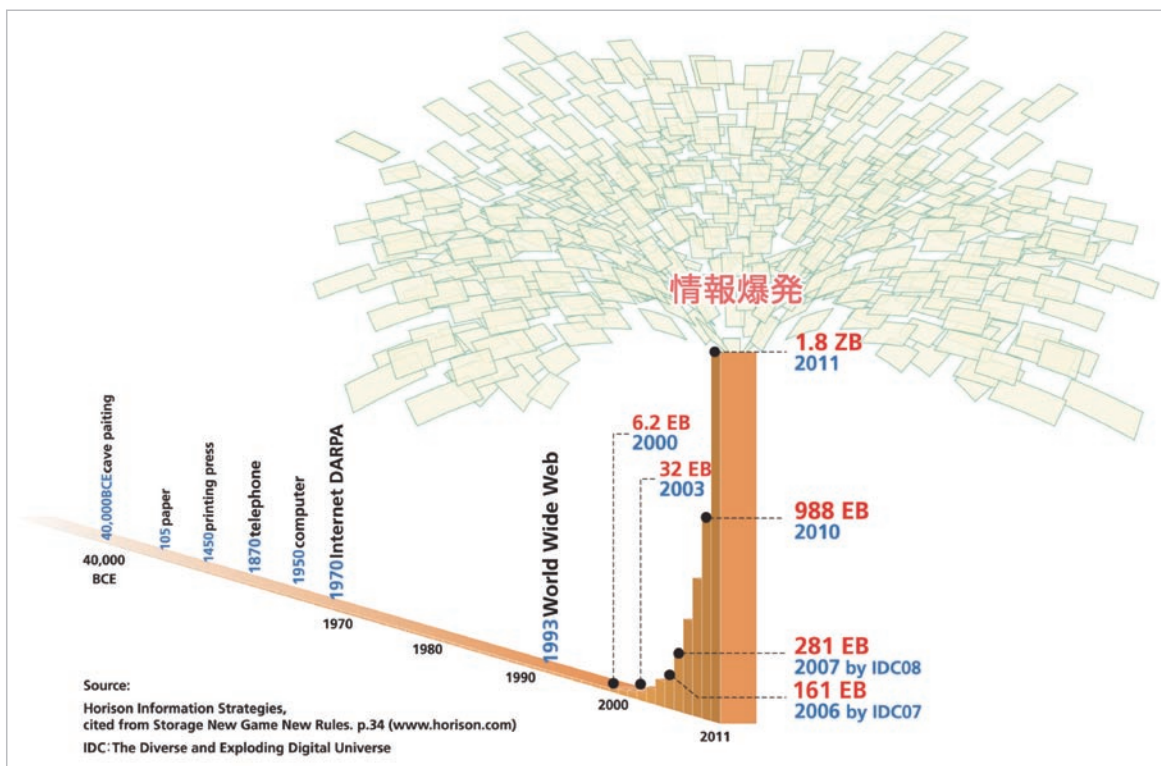


図-1 情報爆発

☆1 IDC の 2007 年予測値は 2008 年に修正されたが、本図は、両者とさらに 2004 年特定申請当時 Horison 社から公開された数字を合わせて描いた。

工夫し、いわゆる科学技術計算を目的とはしない、情報関連研究のための大規模システム研究環境が次第に整いつつある。民間企業の資源を借りるには至っていないものの、米国の後追いとならない形で、着々と基盤固めを進めてきた。また、情報爆発という現象とその問題意識は、2007年版現代用語の基礎知識の第1ページ目に取り上げられる等、非専門家にも広く着目されるに至っている。

2004年に情報爆発IT基盤なる特定領域研究を申請して以来、4年を経たこととなるが、もちろん至らぬ点、ご批判も多々あるかと感じつつも、研究の方向感は大きくは誤ってはいなかったのではないかと考えている。本稿では本情報爆発特定を目指すところ、現時点でのいくつかの成果、そして、今後について概観する。

情報爆発特定を目指すところとその現況

情報爆発特定は、安西慶應義塾長が率いられた情報学特定(01-05)の後継として、2005年に発足した。2005年は総括班のみの稼働となり、2006年より計画班、公募班からなるプロジェクト全体が始動した。その経緯や各柱の目指す方向に関する詳細については、文献3)に特集号としてまとめているので、ご参照いただければ幸いである。

本プロジェクトの目指すところを一言で表現するならば、情報爆発という『現象』に着目し、当該現象を情報分野の研究者の新たな課題と捉え、(1)情報爆発時代に生まれる種々の問題を同定しその解決を目指し先駆的に取り組むとともに、(2)情報爆発を問題とみなすのではなくむしろ積極的に見つめ、爆発する情報から新たな価値創出を目指す今までにない研究に挑戦しようというものである。一般に、情報分野におけるプロジェクトでは、対象の明確に定まったシステムやツールの開発を目指すことが多いが、本プロジェクトは現象に着目している点が特徴と言える。すなわち、そもそも、情報をこれだけ爆発させるに至ったのは情報関連研究者の寄与するところ大であり、我々こそが当該現象に立ち向かうべきと考えプロジェクトを推進してきた。

● 情報爆発に向けた新しい検索技術への挑戦とその研究開発の現状

上記(1)の情報爆発時代に生まれる最大の課題の1つは、情報のオーバーロディングである。まさに、日々膨大な情報に押し潰されんばかりの生活を送っておられる方々も少なくないであろう。個人の情報咀嚼能力の限界を超えた情報が氾濫することから、真に必要なとする情報からの距離が相対的に大きくなり、当該情報を的確に探

し出す手法が従前に比べてはるかに重要になっていると言える。情報の検索という研究課題はきわめて古典的であり、もはやそれ自体に高い新規性が感じられるものではないと見られがちであるが、情報爆発時代においては、状況が大きく変化しており、再度根本的に考え直す必要があると考えた。アナロジーとしてディスクを取り上げることができる。磁気ディスク技術は目覚ましい勢いで進展し、現在ではその容量は1ドライブ当たり1テラバイトに達する。しかしながら、読み出し速度はそれほど変化があるわけではない。すなわち、容量当たりの入出力性能は急激に低下しているとみなせる。ディスクを扱うデータベース技術の歴史は実は索引技術の歴史とも言え、当該問題を克服すべく多様な索引技術が綿々と研究されてきた。まさに状況は同じである。増大する情報に対し、人間の情報咀嚼速度は大きく変化はしていない。検索技術の一層の研究開発は不可欠と言えよう。

検索の技術開発に対し、容易に予想されるリアクションは、すでに便利な検索エンジンがあるではないかというものである。検索エンジンの誕生以来数多くの改良が重ねられ、いわゆる、ナビゲーションな問合せ(組織のWebサイトを検索するような、その存在をすでに知っている検索)はほぼ問題なく高い精度で検索可能となった。しかしながら、より複雑な、すなわち、検索語を3つ以上入れたくなるようなインフォメーションな問合せに対してはいまだに欲しい情報がなかなか出ないことを多く経験されていることであろう。Webを知の巨大図書館と見なす議論が多々展開されているが、Webを対象とした知の検索は単純な検索から次第に複雑な検索へと利用者の期待は当然のことながら膨らむであろう。知りたい情報を複数の検索語の集まりで表すことは必ずしも容易ではなく、情報爆発プロジェクトでは、困難な課題としてのインフォメーションな問合せに挑戦している。

ここで複雑な問合せとは『インドの経済発展を阻害する理由』『栄養のアンバランスが子供の心に与える影響』のようなものを指し、現行の検索エンジンではなかなか思ったものが探しにくいのに対し、本プロジェクトでは当該問題に対して、従来に比べてはるかに深い自然言語処理を行うことにより抜本的な質的向上を狙っている。黒橋・新里(京大)が開発したエンジンは、たとえば上述の問合せに関して、商用検索エンジンに比較してすでに飛躍的に高い精度を実現しつつある。なお、申請時点では耳にしなかったものの、米国においてもMSに買収されたPowerSetなど多様な企業が水面下で同様のシステム開発を進めているとされており、当然のことながら熾烈な戦いの場でもある。ただ、ローカリゼーションは重要な要素であり、とりわけ、中国も、韓国もアジア



は米国の大手検索エンジン会社が苦手であることから分かるように、言語固有の難しさがあがり、勝算を見込んでいる。

複雑な検索とはやや異なる問題として、意見分布の捕捉を挙げることができる。一般に次々と新しく生まれる社会の問題に対しては多様な意見が存在するのが常であり、それらを整理して閲覧したいというニーズはきわめて高い。『赤ちゃんポスト』『後期高齢者医療』『株式会社による病院の経営』など、正しい考えを提示するというより、いろいろな視点からの意見分布の提示自体にきわめて高い価値があると言える。現在のランキング手法ではマジョリティの意見が上位にランクされる。一方で、マイノリティではあるものの新鮮な視座を与える意見も少なくないが、現在のサーチエンジンでは少数意見は存在すら把握することが容易でない。対立する意見を、ロングテイルの部分も含めて分かりやすく整理するエンジンは見当たらない。情報爆発時代において、正に必須なツールになるはずであるとの判断から、藤井（筑波大）が取り組み、OpinionReaderなるシステムを開発した。本特集別稿に赤ちゃんポストの例を示すが、現状のサーチエンジンに比べ意見分布の把握を大幅に容易にすることができている。

さて、先に示した複雑な問合せに耐えるだけの深い文法処理をしようとする、現行の商用検索エンジンを利用してその上に独自に開発した処理系を載せるというアプローチはとれない。ロングテイルの情報まで見せるには、そもそも先頭から1000ページしか提供されない、すなわち、テイル部分を見ることのできない現在の商用エンジンは利用できない。我々の目指す根幹的研究をするには、自ら膨大なページのクロールをし、大量のパーズをしなくてはならないわけであるが、これには膨大な労力が必要となり、1人の研究者ではそれだけで疲弊してしまう。骨太の基礎研究を行うには、骨太の研究基盤が必須であるとの判断から、後に述べるように、本プロジェクトでは支援班の枠組みを利用し、検索エンジン共創プラットフォーム『TSUBAKI』を構築している。ページの収集や大量ページに対する基本的な文法処理など低位の作業を一元化することにより研究者の負担を大幅に軽減し、存分に上位の多様な検索技術の開発に集中することを可能とする環境が、2年を経てようやくできつつある。当該プラットフォームの構築に対して、個々の検索技術に関する研究成果と同等あるいはそれ以上に、力を投入してきており、本プロジェクトの大きな出力である。特定研究参加者以外にも緩やかにオープンにしておき、新しい検索を目指す多様な研究の場としてぜひ、ご活用いただきたい。

情報技術から距離感のあるユーザが情報爆発に圧倒さ

れることなく快適な生活を送れる技術開発にも注力している。河原、松山（京大）らは、コミュニケーションこそが、情報を「知る」という段階から「分かる」という段階へ、飛躍的な理解増進を可能とするキー技術となると考え、現在、「情報コンシェルジュ」なる対話システムを開発している。詳細は別稿に譲るが、コンテキストを正確に同定し、システム側がプロアクティブに情報提供をする新しい対話方式が、自然な対話促進にきわめて有効であることを定量的に確認することができた。今後、当該対話エンジンを検索エンジンのフロントとすべく精力的に研究を推進してゆく予定である。

● 情報爆発からの価値創出への挑戦とその研究開発の現状

筆者の学生時代である25年前を振り返ってみると、情報がなかなか手に入らず、情報に飢えていた時代と感じる。爆発するほど大量の情報が利用可能になった今日、発想を転換し、むしろそれを積極的に活用する多様な研究が展開すべきと考えた。

とりわけ情報量の増大はセンサ技術の進展の寄与するところ大と言える。土壌センサ技術の多大な進歩はDigital soil mappingなる手法を生み出し、農地の土壌成分を精細、かつ容易に定量化し、適地適作を可能としつつある。食糧問題が大きく取り上げられる昨今、膨大なセンサ情報の戦略的活用はきわめて大きな価値を生む。グリーン by IT、トレースによる物流効率化、製品安全のためのPLMなど、現時点においてもきわめて多くの情報爆発からの価値創出をまさに戦略的にビジネス化するシナリオが数多く生まれてきている。また、将来を見据えるとさらなる計り知れない可能性を秘めている。

本プロジェクトでは、運動量センサにフォーカスし、爆発するセンサ情報を活用した新たな生活習慣病の予防・治療ソリューションについて研究を進めている。須藤（東大）、中島、井上（九大）は、九州大学医学部との共同により、予備軍も入れると1700万人にも上るとされている糖尿病患者に対し、センサ技術を活用した、リアルタイムでコンテキストアウェアな生活指導システムの実証実験を試みつつある。働き盛りの壮年層こそが国を支える活力源であるにもかかわらず、超多忙な生活の中で待ち時間の長い病院に向かうことはきわめて困難であり、仮に、診断を受けたとしても、生活習慣病では重篤な事態が急に発生することが少ないことから、5分程度のわずかな問診を月に1度する程度ではそもそも生活改善のインセンティブの維持は容易ではなく脱落してしまうことがほとんどである。本年4月より、保険制度が改正され、保険事業者が健康指導をすることが義務付けられることとなったのもこれらが背景にある。本情報爆発プロジェクトでは、最新のセンサネットワーク技術

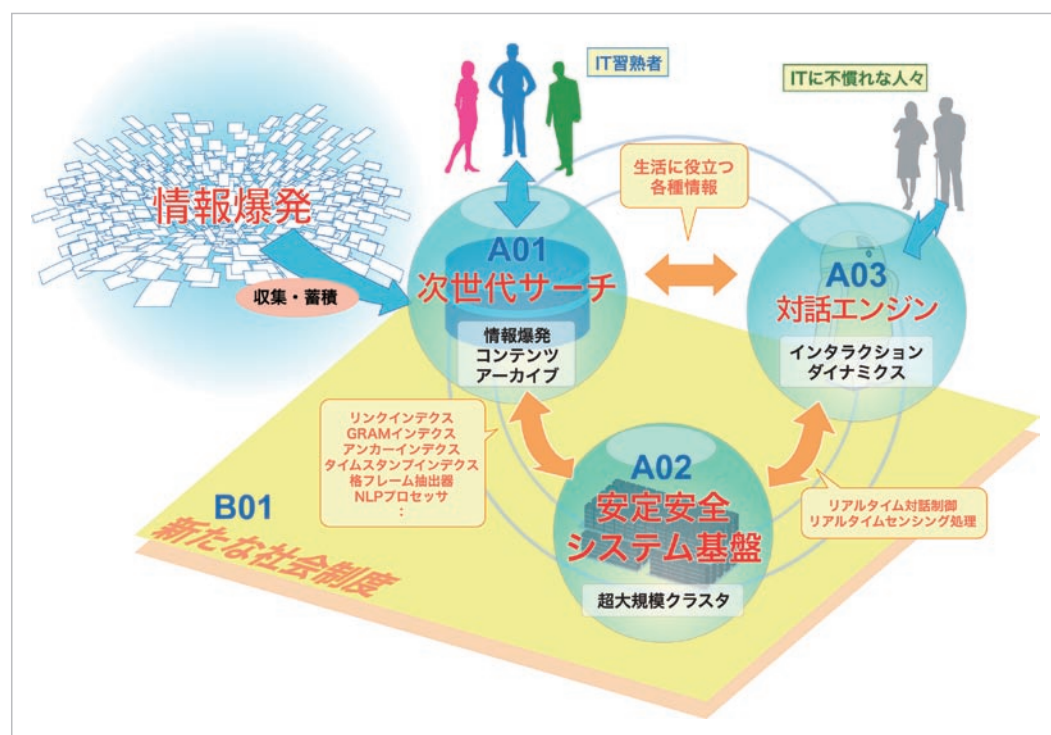


図-2 情報爆発プロジェクトの構成

導入するとともに、爆発するセンサ情報を積極的に利用し、1日の運動量を定量化すると同時に「見える化」する。これにより、医師は一月単位ではなくきわめて高頻度にセンサ情報をチェックすることが可能となり、従来に比べて飛躍的に精緻な生活指導を行うことができるようになる。同時に、患者は自ら自身の運動量を直接把握することが可能となり、運動継続のインセンティブが大きく改善される。生活指導は紙に書いて一月に一度手渡しても、忙しい生活の中ではうっかり忘れてしまい、なかなか実際に実行することは容易ではない。センサ情報からの高度なコンテキスト推定技術により今の状況を捉え、その時点時点で、今、これをやってくださいという指示をタイムリに提示することは、患者にとってきわめて有難く、大きな生活改善が期待される。多くのセンサから生み出される爆発する情報は、実時間コンテキスト同定解析に利用される以外にも、センサに蓄積され、患者指導手順の継続的な改善のために大規模マイニングの対象となる。最先端のセンサ技術を利用したセンサ情報爆発からの価値創出を目指し、本プロジェクトでは、生活習慣病抑止支援システムの実証を着々と推進している。eヘルスなる名称で種々のプロジェクトが国際的にも立ち上がりつつあるが、医師と広範な情報研究者の密な連携は必ずしも多くなく、先駆的な成果の手ごたえを得つつある。また、糖尿病はアジア圏固有の課題であり、我が国において実施する意義はきわめて高い。

さて、このような真に体感可能な実験を行うためには、医師と情報爆発研究者の意志疎通、医療従事者による新

たな生活指導の精細なルール化、医師会との連携、センサ技術の実世界への適用時の問題把握、膨大なセンサデータの蓄積とその解析、被験者のインセンティブ設計など多様な技術の融合、チームの連携、膨大な実地作業が必須となり、少数の研究者では到底手に負えないと躊躇することが通例であろう。総勢250名を超す多様なメンバを擁する情報爆発プロジェクトでこそ可能な規模感のある実証実験と言えよう。最新のIT技術を実際に社会に適用し、そこで、新たな問題を自ら肌で感じることは、きわめて貴重な体験であり、そのような体験から次への骨太の研究指針が得られると信じている。

情報爆発プロジェクトの組織

本プロジェクトは4つの柱、A01「情報爆発時代における情報管理・融合・活用基盤」(計画班、喜連川優、田中克己、下條真司、安達淳、黒橋禎夫)、A02「情報爆発時代における安全・安心ITシステム基盤」(計画班、近山隆、柴山悦哉、松岡聡、中島達夫)、A03「情報爆発時代におけるヒューマンコミュニケーション基盤」(計画班、國吉康夫、西田豊明、松山隆司)、B01「情報爆発時代における知識社会形成ガバナンス」(計画班、須藤修)ならびに支援班から構成される(図-2)。A01は、爆発する情報に対する検索ならびに情報爆発からの価値創出を、A03は、情報の専門家でないユーザに対しての情報提示に関して、人間の対話の深層解明に向けて研究を進めるとともに、人に優しい対話エンジンの実現を

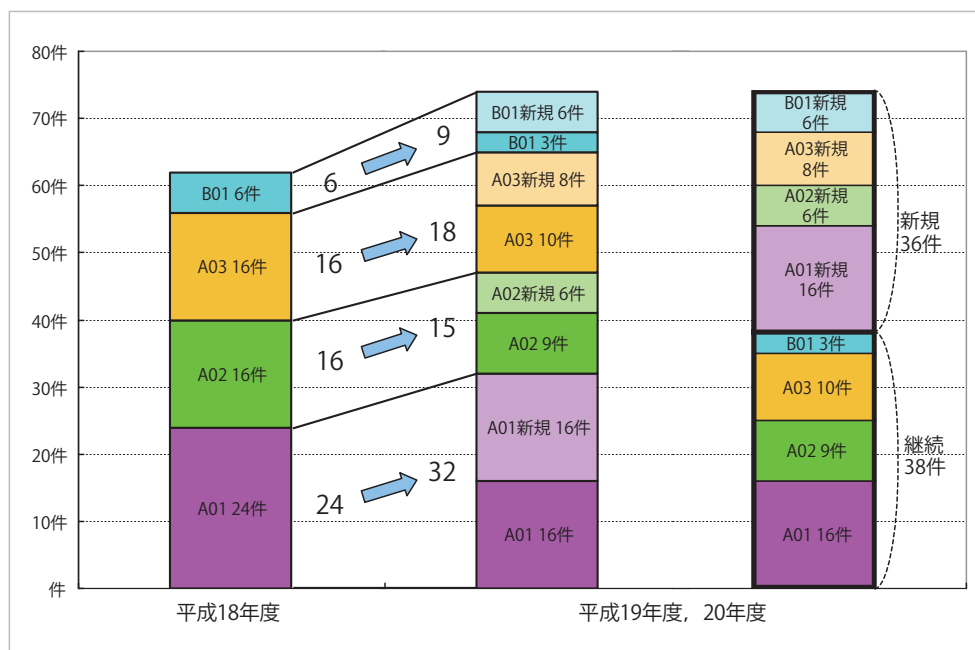


図-3 H18 から H19, 20 への公募研究の分布

試みる。A02 は A01 ならびに A03 のソリューションを実現するためのシステム技術の研究を担う。近年情報システムの脆弱性が広く指摘されているところであるが、システム自体の挙動モニタリングを行うことにより、すなわちソフトウェアセンサからの情報爆発を利用することによりシステムのレジリエンスの向上を目指す。B01 は学問的な成果だけを目指すのではなく、社会還元を実現すべく、社会において実際に深刻化している問題に対して、本プロジェクトがいかに貢献できるかを真摯に取り組む場である。A01 から A03 までの技術を結集し、制度設計の検討も含め、有効性の実証を行う。

計画班に加え、予算的にはそれを凌駕する公募班が組織化されている。図-3 に平成 18 年度ならびに平成 19 ～ 20 年度の採択された公募研究の柱ごとの数を示す(平成 18 年度に 1 年間の研究を公募し、平成 19 年度に 2 年間の研究の公募を行った。本年秋に、21, 22 年度の 2 年間の公募を予定している^{☆2})。図が示すように、プロジェクトの規模は拡大している。また、約半数の入れ替わりが発生していることから分かるように、流動性の高い柔軟な運営が実現されている。

戦略的大規模実験を速やかに可能とする 情報爆発共創プラットフォームの構築

骨太の成果を出すためには、やや遠回りであったとし

てもしっかりと研究基盤を構築することが不可欠であるとの考えから、先に紹介したように、本特定では、支援班の枠組みを活用し、「情報爆発共創プラットフォーム」の構築を進めてきた。情報系のすべての研究が該当するわけではないが、多くの分野において、実験の規模が大きくなり、1 人の研究者では研究環境の構築・維持が困難になりつつある。物理・生物などにおけるビッグサイエンスと同様の戦略が情報系においても重要であると感じる次第である。昨今、クラウドとも呼ばれる超巨大クラスタは、システム研究者にとっての新しい研究ターゲットであり、また、巨大データを対象とする多様な情報研究者にとってのエキサイティングなプラットフォームでもある。MapReduce, Hadoop, PigLatin など大規模クラスタを想定したミドルウェアの開発が海外において活発になされているが、これらは、きわめて大容量のデータ操作を特徴としている。データ当たりの計算量はわずかな場合が多く、いわゆるスパコンが対象とする科学技術計算とは負荷特性が大きく異なる。新しいワークロードに対応した研究の機動的な立ち上げは不可欠であるものの、個々の研究者では研究環境の構築はままならないのが現状である。情報爆発時代においては、まさにこのような処理が今後主流になると想定し、システム研究者と大規模データの利用研究者が連携して研究が進められるべく工夫をしている。すなわち、田浦(東大)を中心として、InTrigger なる実利用に耐える分散計算プラットフォームを構築してきた。システムソフトウェアを研究するための我が国では希少なプラットフォームと言え、本システムに関しても、ぜひ多くの関連研究者の利用を期待したい。

☆2 なお、この 1 年、2 年、2 年の区切りは特定研究の制度において規定されている。公募研究の選定にあたっては、領域代表が委員長を務めない審査委員会が設けられ厳正に審査される。

TSUBAKI サーチプラットフォーム研究者は InTrigger を利用し大規模な言語解析を実施している。もちろんこのような研究スタイルの現実は容易ではない。システム開発者にとって、論文を書くためのソフト開発と実運用に耐えるソフト開発とは大きなギャップがあることは当然であり、多くの細かい苦情が寄せられることにもなるが、トップレベルの成果を出すパートナーとの緊張のある連携は大きな励みでもあり、得難い刺激でもある。プラットフォームは重層構造を成している。TSUBAKI はさらに、多くの自然言語・情報検索研究者によって利用されている。今後もますます多様な好循環が生まれることを期待している。さらに重要なことは、30 代から 40 代前半の若手研究者がその構築に多大な情熱を傾けているということである。多様な分野の若手研究者の交流が共創プラットフォームを介して実現されつつある。

A03 柱の角・坊農・西田（京大）らによる IMADE センサルームは複数の人間のインタラクションの精細な補足を可能とする最新鋭の共通プラットフォームであり、研究者が固有の機器を持ち込み実験することも可能としている。もちろん、InTrigger とも接続されている。

こんなサーチがあると便利だなと感じることは少なくない。しかし、それを一から始めるのでは息が切れてしまう。大島・中村・田中（京大）らによる SlothLib は、自分のサーチアプリケーションを数時間から 1 日程度で作りに上げるためのツールキットである。プログラミングは不要であり、チュートリアルを受けるだけですぐに利用可能である。現行の商用サーチエンジンの結果を利用し、手際よく、自分の思いを実現し、その有効性を簡単にチェックすることができる。サーチに関する競争は熾烈をきわめており、自然言語処理をベースとする基礎研究（TSUBAKI）は重要である一方、キュートな機能の実現によるネット利用の大きな流動も日常茶飯事である今日、使い心地の良い新鮮なニーズの吸収とわずかな時間での実現による次世代のサービスを探るツール構築プラットフォームも本プロジェクトの大きな特徴である。紙面の都合上、他の活動は、本特集の他の記事に委ねることとしたい。

他プロジェクトとの連携：世界トップレベルの学の成果を産へ

●情報大航海プロジェクト、情報信憑性プロジェクト、NICT との連携

情報爆発プロジェクトは 2005 年に発足し、2006 年より実質的に始動した。経済産業省による情報大航海プロジェクト、また、総務省による情報信憑性プロジェクトが 2007 年 4 月に発足した。筆者が「情報大航海時代

における経済・社会・文化のあり方に関する研究会」座長を仰せつかっていることもあり、情報大航海プロジェクトとの連携は積極的に行ってきた。たとえば、本年 3 月 17 日に開催された情報大航海シンポジウムでは、情報爆発プロジェクトの検索技術に関する成果を黒橋、藤井から紹介した⁵⁾。また、2007 年度は 10 の実証実験が実施されたが、A01 柱の研究者である中川（東大）、武田（NII）、中島（奈良先端大）など多くのメンバが情報大航海プロジェクトにも参画し、最新の学の基礎研究成果を産に展開することが有機的に実現されている。また、NICT からの研究者鳥澤、灘本、木俣などが情報爆発に公募班研究代表者として参画しており、NICT との密な連携がなされている。情報信憑性プロジェクトに関しても、田中（克）（京大）がその公募プロジェクトに採択され、検索の信憑性に関する研究を推進している。また、黒橋（京大）は NICT が開発する信頼度付きサーチエンジン wisdom の開発にも携わっている。内閣府によって作成されたイノベーションパイプライン網⁴⁾（コンテンツ領域）では、情報大航海プロジェクトや情報信憑性プロジェクトがそれぞれ実用化研究、応用研究と分類されているのに対して情報爆発プロジェクトはこれらの基礎を支えるプロジェクトと位置付けられている。我が国の情報産業は国際的見地からは必ずしも優位とは言えず、省を越えた機動的連携が必須であり、まさに、それを実現すべく努力している次第である。

●学からの世界レベルの先進的基盤共通技術の提供

近年「頻出パターンマイニング」なる技術の研究が急速に進展し、非常に多くの局面で利用する手法となった。性能コンテストも実施され、宇野（NII）の LCM は世界的に突出した性能を発揮しているが、さらに湊（北大）が BDD 手法を適用することにより、出力コストを低減し一層の大幅な性能を達成した。まさに本プロジェクトにおける機動的連携の典型と言える。平成 19 年度における情報大航海プロジェクトの 1 つとして、ドコモは「マイライフアシスト」と名付けたサービスの創出を企画している。通常、ブラウザのクッキーを利用することにより、ユーザを同定し、取得した Web 空間のログを解析し、提示するコンテンツをパーソナライズすることはこれまでも数多くなされてきた。Web の利用が著しく進みサイバー空間に滞留する時間が急激に増えてきているのは事実ではあるものの、我々の「実世界」での生活時間がなくなることは有り得ないことも事実である。ブラウザクッキーに相当するような、実世界において我々の動きを補足するのに最も効率の良い手段とは何か？ ドコモは生活の中で最も肌身離さず持っている携帯で攻める戦略を大航海プロジェクトに提案した。携帯は GPS

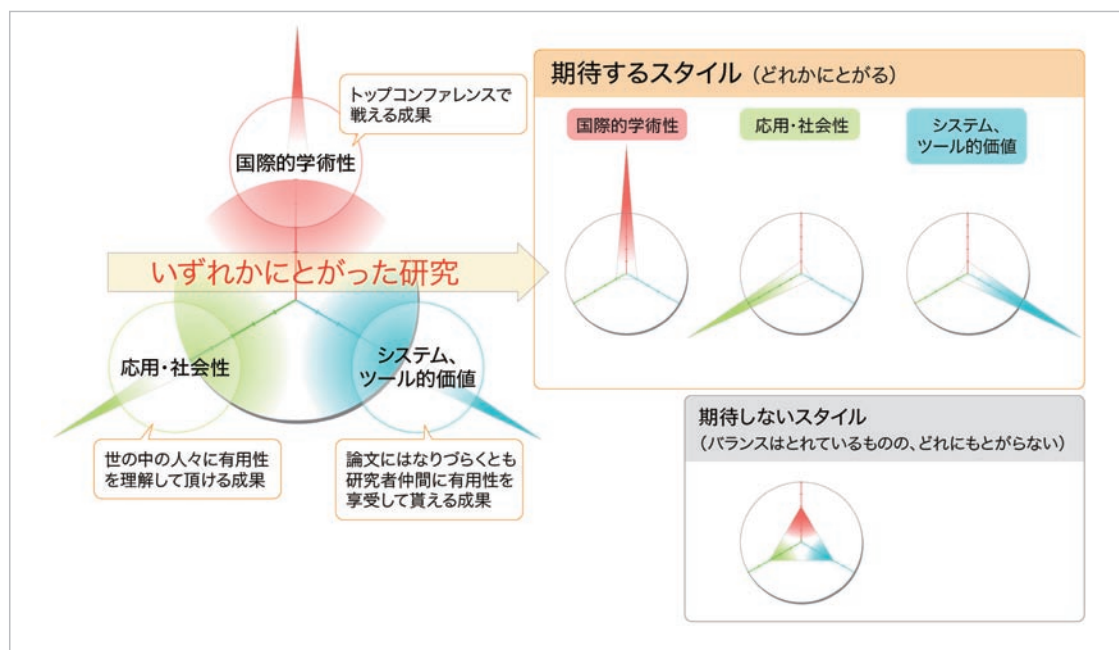


図-4 自らの評価軸

に加え、モバイル Suica 等による移動、購買、写真、メールなど今や非常に多様な役割を担っており、個人情報情報を匿名化するなどは当然必要であるが、当該情報の利用により、生活に密着した情報提示が可能となる大きな可能性を秘めていると言える。一方、個人の活動情報は膨大なものとなることは明白であり、その高速マイニングは種々のサービスの基盤となることは間違いない。基礎となるアルゴリズムの開発を企業に期待することは現時点では厳しいのが実情であり、情報爆発プロジェクトにおける宇野・湊のアルゴリズムは情報大航海プロジェクトにおける基盤共通技術として、利用しやすくツール化し提供されている。

竹田（九大）らは世界トップレベルの超高効率圧縮文字列検索アルゴリズムを開発している。圧縮率の向上と圧縮した状態での検索効率の向上の両立は容易でない。従来は圧縮率を高くできなかったのに対し、通常の gzip レベルの圧縮率で動作する基盤アルゴリズムを開発した。このような学のトップレベルの手法については、今後とも、順次、使いやすい形で、産へ提供してゆきたいと考えている。

自らの評価軸

本プロジェクトは、研究代表者、研究分担者、連携研究者を含めると総勢 200 名を超える大所帯となり、プロジェクト全体としての明確な方向性の共有が必須と考えた。情報分野の職種に魅力が薄れ、大学における進学においても情報離れが加速する中で「研究者自らがワクワクできること」を大前提とした。自分の研究がいかに

ワクワクするものかを情熱を持って伝えることができるこそが要と感じている。その際には、際立ったエビデンスこそが必要との判断から、プロジェクトとして、研究成果に対して以下の3つの評価軸を設けることとした。

- 1) 国際的レベルの学術成果
- 2) 社会に貢献できる成果
- 3) 情報分野の研究者にとって価値の高い成果

1) は当然であり議論するまでもないが、国際的トップカンファレンスでプレゼンスを示すことを指す。国際会議は現在供給過多であり論文数の意味はもはや大きくないと考える。2) については、技術を社会に貢献できる形にするには、通常大きな努力が必要となるものの、その行為自体は必ずしも学術性の高いものではないことが多い。しかしながら、情報技術が広く社会に浸透しつつある今日、逆に、学に閉じた成果のビジビリティは必ずしも高くならない傾向にあり、積極的に社会への展開を進めることも重要であるとの考えからこの軸を設けた。オープンソフトウェアツールなど、情報系研究者にとっての価値を3番目の軸とした。一般にツール開発はその使いやすさ、完成度が問われることが多く、学術的な先進性とは独立である場合が多い。また、一般社会に理解が得られるものでもない。しかしながら、情報関連分野の研究者にとってその研究を大きく加速することができるようなツール/システムの開発はきわめて重要であり、高く評価されるべきものと考えた。ここで、重要な点はこれら3つの軸すべてに対して中庸な達成度を目指すのではなく、1つの軸だけを選択し、ずば抜けた成果を目指すことを目的としている点である。残りの2つの軸については一切問わない(図-4)。個々の研究は、

限られた予算しか配分されておらず、選択と集中を心がけている。

おわりに

Web の誕生後ページ数の増大とともに、サーチエンジンが先端技術の象徴として開発が進められてきたが、現在、インターネットビジネス全体から見ると、ソーシャルネットワークや動画提供を始めとする多様なサービスが展開され、いわゆる Web サーチはもはやその一部にしか位置付けられない。今や、検索エンジン会社が保有するデータの大半はログであり、インデクスや Web キャッシュの量はわずかでしかない。システムの核は人の動きを捕捉する膨大なログ解析部にあると言える。

クリックスルー率を指標とするマネタイゼーションという明快なゴールと、効果の即座的な可観測環境は情報技術者にきわめてエキサイティングな研究開発のターゲットを与えた。検索エンジン会社への学からの多くの人材流動のゆえんであろう。将来にわたり情報技術の重要なマイルストーンと位置付けられよう。

しかし、情報爆発を高精細オブザベーションのイネイブラと捉えるならば、上記ほど明快ではないにしても、広大な潜在的「ワクワク領域」が感じられる。情報爆発は非常に多くの人間社会・文化・経済活動に影響を与えており、広い視野と高度な戦略に基づき研究を進めたいと考えている。

本稿では、情報爆発から生じる問題への挑戦として次世代サーチへの取組みについて、また、情報爆発を積極

的に活用し価値創出に結びつける研究として大量のセンサ情報を活用したヘルスケア実証実験について、また、これらの研究を支える情報爆発共創プラットフォームについて、現在の活動を述べた。紙面の都合から本稿で紹介した成果は一部にすぎず、種々の機会で情報発信を進めたい。

謝辞 本プロジェクトの推進にあたり多くの方々より多大なご支援を頂戴している。Yahoo! JAPAN には頻出 1 万個の検索語リストをご提供いただいている。情報爆発組織運営に関しては国立情報学研究所に多大なご力添えを頂戴している。

参考文献

- 1) Dean, J. and Ghemawat, S. : MapReduce : Simplified Data Processing on Large Clusters, CACM, Vol.51, Issue 1 (Jan. 2008).
- 2) IDC white paper : The Diverse and Exploding Digital Universe (2008). <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>
- 3) 特集：情報爆発時代に向けた新しい IT 基盤の研究、人工知能学会誌、Vol.22, No.2, pp.208-240 (Mar. 2007).
- 4) イノベーションパイプライン網：<http://www8.cao.go.jp/cstp/project/bunyabetu2006/jyoho/1kai/haihu1.html>
- 5) あふれる情報、賢く利用、新たな価値を創出、日本経済新聞 朝刊 26・27 面(2008 年 4 月 20 日)。

(平成 20 年 7 月 9 日受付)

喜連川 優 (正会員) : kitsure@tkl.iis.u-tokyo.ac.jp

東京大学生産技術研究所教授、戦略情報融合国際研究センター長、本会副会長、本会フェロー、電子情報通信学会フェロー、ACM SIGMOD Advisory Board Member、IEEE TCDE Asian Coordinator、10 年間にわたる大規模 Web アーカイブの構築、次世代超高速データベースエンジンの研究開発に従事。