

多様で大規模な情報から『知識』を生産・活用するための基盤技術の創出 (平成20年度設定)

1. 戦略目標名

多様で大規模な情報から『知識』を生産・活用するための基盤技術の創出

2. 本戦略目標の具体的な内容

近年、センシング技術やインターネット等のインフラの高度化等により、大規模な情報へのアクセスが容易になってきている。このような大規模情報の取得・蓄積はインターネットによるサイバー世界だけでなく、実世界においても進展しており、これら大規模情報の中から、学術、医療、金融、防災、サービス等に有用な情報を発掘・獲得することは、今後益々重要な課題となってくる。

政府の長期戦略指針「イノベーション25」においても、「知識社会・ネットワーク社会及びグローバル化の爆発的進展」が予測されており、有用な情報を迅速かつ適切に得ることが我が国のあらゆる分野での国際競争力の強化に繋がる。

本戦略目標では、学術、医療、金融、防災、サービス分野等の多様なニーズに応じて、当該分野を高度化、効率化するための知的情報基盤の確立をめざし、様々な分野で生成・蓄積された多様で大規模な情報から『知識』を生産・活用するための基盤技術を創出する。なお、ここで言う『知識』とは、社会における人間の活動目的に応じて必要とされる有用な情報のことであり、計算機を使用した情報処理技術等により創出される。

多様な社会ニーズに対応した知識を生産・活用するためには、多様で大規模な情報を目的に合わせて柔軟に処理できる情報技術が求められる。これは、計算機の処理能力向上だけで得られる技術ではなく、知識を必要とし活用する個人や組織等のニーズや特性にも配慮しうる新たな技術を生み出すことによってはじめて得られるものである。

そこで、継続的にイノベーションを生み出すことを可能とするため、本戦略目標において、多様で大規模な情報から知識を生産・活用するための基盤技術の創出に取り組む。具体的には、整理・構造化した、多様で大規模な情報の分析・解析により知識を創出する技術について、応用分野における現実の課題を解決するための研究開発を、情報科学、統計数理科学、人文・社会科学等を融合して行う。これにより、知識の創出のための情報処理技術の研究開発とともに、個別の応用分野において知識の活用を可能とする技術の開発を行い、学術、医療、金融、防災、サービス分野等における実問題の解決や、競争力強化に繋がる新しい知見の発見を可能とする。

3. 政策上の位置付け

①第3期科学技術基本計画：分野別推進戦略 重要な研究課題「5. ヒューマンインターフェイス及びコンテンツ領域」【課題5】情報の巨大集積化とその活用

- ②戦略重点科学技術：世界と感動を共有するコンテンツ創造及び情報活用技術
- ③長期戦略指針「イノベーション25」：5章「イノベーション立国」に向けた政策ロードマップ 2) 次世代投資の充実と強化

4. 当該研究分野における研究振興方策の中での本研究事業の位置づけ、他の関連施策との切り分け、政策効果の違い

本戦略目標に関しては、第3期科学技術基本計画 分野別推進戦略を踏まえて文部科学省が平成18年7月にまとめた「情報科学技術に関する研究開発の推進方策について」の中で、「データベースと融合したスーパーコンピューティングの実現に必要な高度計算科学技術の開発」、「半構造的、連続性を有するマルチメディアコンテンツを組織化し、高度な機能を有するデータベースとして蓄積・管理する技術。生成されたコンテンツを研究や教育に生かすための利用技術」が取り上げられている。

関連する基礎研究として、文部科学省科学研究費補助金の特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」があるが、当該研究は、情報検索や自然言語処理等、情報科学における個々の要素技術の研究であり、本戦略目標のような、情報科学、統計数理科学、人文・社会科学等を融合した、新しい技術を創出するものとは異なる。

また「e-Society基盤ソフトウェアの総合開発」において、全世界の膨大なWeb情報の中から最新のものを自動的に収集・検索する技術を開発する「インターネット上の知識集約を可能にするプラットフォーム構築技術」と個々のWeb情報の関連性と時系列変化を解析する技術を開発する「先進的なストレージ技術およびWeb解析技術」が行われているが、これらはインターネット情報を対象とした基礎的な技術の研究開発であり、平成19年度で事業を終了する。

また、大規模な情報を検索・解析する技術開発として、経済産業省の「情報大航海プロジェクト」が関連するが、こちらは企業が顧客のニーズやサービスの品質に関係する様々な情報をビジネス目的に活用するための技術の研究開発と新たなモデルサービスの実証を行う事業である。

このように、本戦略目標の下で実施する研究開発と既存の研究開発施策には、内容に明確な違いがある。

5. この目標の下、将来実現しうる成果等のイメージ、他の戦略重点科学技術等に比して優先して実施しなければならない理由、緊急性、専門家や産業界のニーズ

(1) 必要性

近年、多様で大規模な情報が生成・蓄積されるようになっており、こうした情報は、社会の多様なニーズに対応した新しい経済・社会的価値を生み出す源泉となりうる資源である。世界最先端のIT国家を国是とする

我が国において、こうした情報資源の利活用技術はきわめて重要な技術である。一方で、

知識を創出するために現在使われているデータマイニング、Web検索などの情報技術では、扱える情報の属性（例：テキスト・画像・音声など）や情報量が限定されている。このため、一定以上の規模や多様性（属性）をもつデータを扱う場合、データフォーマットの違いから計算が困難であったり、計算量が指数関数的に増大して計算時間が膨大になり、処理が困難となっている。これは、計算機能力の向上のみで対応すべき問題ではなく、新しい情報技術（理論や方法論の体系的整備）が求められている。多様で複雑なために現在活用されていない大量の情報資源を、新しい情報技術によって知識創出に役立てることができれば、我々が獲得できる知識はより豊かなものになる。

また、現在多くの分野で理論、実験と並ぶ重要な方法として確固たる地位を築きつつあるスーパーコンピュータによるシミュレーションと融合することにより、実世界・サイバー世界情報の活用とシミュレーションによる予測情報が両輪となり、未だかつてない超高精度で広範な予測技術が創出され、様々な分野に波及する我が国発のイノベーションを誘発することが期待できる。このような基盤技術を構築するためには、計算機の処理能力の向上や情報技術の高度化のみでは困難である。知識を必要とし活用する個人や組織等のニーズや特性を考慮しつつ研究開発を行うことが効率的であり、さらには複数の実問題への取組事例の中から得られた知見を蓄積し汎用的な理論を導くことが望まれる。また、基盤技術の構築のような取組は、企業等での実施が困難な課題であり、国が主体となって推進することが重要である。

（2）緊急性

欧米諸国ではすでに関連プロジェクトが推進中である。大規模な情報を放置し、その活用技術を持たないことは、全ての分野における国際的な競争力を失いかねない。米国と欧州で進行中の主要な関連プロジェクトの例を以下に示す。米国では、ネットワーク・情報技術研究開発（NITRD）計画の2008年度予算要求のハイライトに、「Cyber-enabled Discovery and Innovation (CDI)」が挙げられている。これは、NSFが実施する"Computational Thinking"に関する研究プログラムとして、不均一なデータから知識を生産する技術、複数の相互作用する要素からなる複雑なシステムの理解、仮想組織の構築等に関する研究の推進が予定されており、既に公募が開始されている。

また、欧州委員会の「研究・技術フレームワークプログラム*1」において、ACGT（Advancing Clinico-Genomic Trials on Cancer）が2006年より5年間の計画で実施中である。これは、医療現場で日々生成される臨床情報と研究現場で生成されるゲノム情報の融合により、速やかな癌の診断と効果的な診療を実現することを目指し、データ処理方法、解析ツール、各種メタデータ群を開発・提供するものである。

*1 加盟国の研究者による共同研究を支援する5ヵ年プログラム。「産業の科学技術の基礎を強化し国際競争力を高め、欧州共同体政策を支援する研究活動を推進」を目的とする。

(3) 将来実現しうる成果等のイメージ

本戦略目標では、創薬、リスク管理、気象予測、サービス、ロボット制御等の応用分野における知識の生産・活用のための新しい基盤技術を提供し、以下のような成果を想定している。

- ①創薬における有機化合物の機能性の部分分子構造の解明、数百万件の診療記録から得られる診断及び最適な治療方法の予測、細胞中の遺伝子が関わる複数のプロセスの解明による難病の治療方法の確立、薬剤の副作用の迅速な把握による薬害の回避など。
- ②電力供給システムのような重要インフラの故障や異常の高精度な予測、工場の生産性向上のための効率的な指針の提示。
- ③センシングデータ、シミュレーション結果からの局所気象予報の精度向上。地震や竜巻の発生予測の精度向上。
- ④熟練技術者の経験、勘、直感を知識として体系化し、生産現場等に取り込む技術の確立など。

また、情報科学、統計数理科学、人文・社会科学などこれまでは別々の分野として研究が行われていたものを統合した新しい技術分野及び研究者コミュニティが形成され、実社会のデータを扱った様々なサービスの効率化に有用な人材が継続的に輩出される効果も期待できる。

6. 本研究事業実施期間中に達成を目指す研究対象の科学的裏付け

現在研究開発されている課題と今後発展が期待される課題を示す。

- ・大規模で多様なデータの情報処理を、現実的な時間内で達成するための超高速アルゴリズム
- ・Web情報、センサ情報、大規模シミュレーション結果等を融合した予測技術
- ・統計数理科学等を応用した相関分析技術とモデル化技術
- ・情報の構造・時系列分析による情報解析基盤技術
- ・人や組織が問題解決のシナリオを効果的に作り出す上で有用な情報を獲得・共有・統合するための技術
- ・複数のリソース（センサ情報、統計データ、Web情報、シミュレーション結果、組織内に蓄積されている情報等）から知識を創出する技術
- ・テキストデータや画像、音声などフォーマットの異なるデータを一括して管理し、取扱う技術等

7. この目標の下での研究実施にあたり、特に研究開発目標を達成するための留意点

研究の実施体制としては、応用分野における実問題に対応していく中で複数の事例を蓄積し、その中から汎用化できる方法論を探索するなど、様々な分野間の相互連携を進め、

研究開発を推進していく。

また、情報科学、統計数理科学、人文・社会科学などこれまでは別々の分野として研究が行われていたものを統合した、新しい技術分野及び研究者コミュニティーが形成され、実社会のデータを扱った様々なサービスの効率化に有用な人材が継続的に輩出される環境を構築するよう留意する。