

特集

情報の価値化・ 知識化技術の実現 へ向けて



編集にあたって

山之内 徹

NEC サービスプラットフォーム研究所

情報化の浸透やネットワークの普及に伴い、ネットワークを介してアクセスできる情報の量や種類が爆発的に増大している。インターネット上では、マスコミによる日々のニュース情報、企業による製品やサービスの情報、大学による研究成果報告等だけでなく、Blog や掲示板などで消費者の情報も流通している。一方、企業や大学などの組織内部にも、顧客／学生から寄せられた問合せや要望、組織内で作成／流通されるドキュメント、個人の日報、週報などがある。

これらの情報は、検索エンジン等を用いて活用されているが、情報の質や規模と対比すると、まだその活用は限定的であり、より広範で高度な活用が期待される。爆発的に増大する情報を収集、分析し、人間によるアクションにつなげ、経済的価値、社会的価値、学問的価値などの価値を実現することである。ここで重要になるのが、情報を収集／抽出する技術、分析／統合する技術、そして、実社会においてこれを活用する技術である。

ネットワークを介してアクセスできる情報の形態は、多様である。データベース内のトランザクション情報、人手で作成されたテキスト情報、センサ情報、音声や画像などのマルチメディア情報などである。それぞれによって、意味のある情報を抽出するのに必要とされる技術

は異なる。トランザクション情報やシステムへのアクセスログの情報は構造化された情報であり、量の問題を解決する必要はあるが、比較的コンピュータで扱いやすい情報である。テキスト情報については、キーワードベースの全文検索が普及しているが、より高度な解析のために構文解析や意味解析、テキストマイニング等の技術が必要となる。センサ情報については、個々のデータを時間系列、空間系列で捉え、一段上のレイヤで意味のある情報に統合する技術が必要になる。一方、音声や画像などのマルチメディア情報については、全体情報の中から注目するオブジェクトに関する情報だけを切り出す技術が必要となる。

情報を分析／統合する技術も目的に応じて多様である。最もなじみの深いのが、必要とされる情報を探し出す検索技術である。ここでも、情報の中身に基づく検索技術、情報の生成／流通／利用状況に基づく検索技術、連想検索や検索ナビゲーションの技術などが必要となる。一方、ピンポイントの情報を探し出すのではなく、情報や知識の集合に対して、全体像を俯瞰するニーズも存在する。あらかじめ決められたグループに分類するクラシフィケーション技術、分類自身も同時に作り出していくクラスタリング技術、時空間に沿って情報の変化を解析す

る技術、複数の情報間の関係を抽出する技術などが必要となる。

実社会で情報を活用する技術では、情報分析の結果を可視化して提示し、隠れた情報の発見を促したり、システムやサービスの異常や改善点の発見を促したり、意思決定プロセスにおける仮説発見を促したりする方法が必要となる。ここで重要なことは、情報技術が人の意思決定の一部を置き換えると考えるのではなく、主役としての人間の創意工夫をいかに引き出すかを考えることである。人の意思決定プロセスにおいて、情報の見落としを補ったり、さまざまな視点での情報把握を可能にしたり、潜在化している情報間の関係性を仮説として提示したりして、人間に気付きを与え、価値創出プロセスをサポートすることである。

本特集では、情報の価値化・知識化技術の実現を目指したさまざまなアプローチをまとめて読者に伝えることを試みた。本領域は、学術研究、意思決定、問題解決などの人間の行動プロセスの中に情報処理技術を組み込むことによって、人間による価値創出を高度化することが重要になるため、適用領域に踏み込んだ記載をお願いした。特集の内容は以下の通りである。

松本洋一郎氏（東京大学）の「**学術創成としての知の構造化—東京大学工学系研究科における試み—**」では、学問領域知識の構造化について解説している。個々に細分化されて発展してきた学問領域の知識を融合する方法として、工学の複雑な事象を複数レイヤで解析した結果を融合する手法や、自然言語処理を用いて巨大な論文集合から有効な知識・情報を取り出す手法などを紹介している。これらは、東京大学が中心となって進めてきた、学術創成プロジェクトの成果である。今後は、知の構造化センター、情報の価値化・知識化技術協議会として継続されていく。

堀井秀之氏（東京大学）の「**社会問題解決のための知の構造化**」では、自然災害や大規模プラントの事故、大規模犯罪などの社会問題解決への知識構造化の適用について解説している。複雑で解決困難な社会問題の本質的な問題点を抽出するため情報や知識の全体像を構造化して可視化する手法と、問題の解決策の立案を支援するため問題の特性や対策の類似性に基づく構造化手法を紹介している。類似性を発見することにより、ある分野における問題解決策に関する知識を、異なる分野における問題の解決に活かすことができることを示している。

小池麻子氏（日立製作所）の「**テキストマイニングによる潜在的知識の発見支援**」では、医学生物学分野における、学術文献からの知識抽出について解説している。本分野における潜在的知識の発見／仮説生成に関する研究の歴史、課題を紹介した後、筆者らの潜在的知識発見支

援システムを紹介している。個々の文献に記載された断片的な知識を、構文解析や意味解析により、概念ネットワークの形で統合する。これを用いて、どの1つの文献にも陽には記載されていなかったが、それらを組み合わせることで発見できる潜在的知識を仮説として抽出可能なことを示している。

山西健司氏、森永聡氏、松村憲和氏（NEC）の「**CGMマイニングと知識化**」では、インターネット上に一般生活者が発信した情報（CGM: Consumer Generated Media）からの知識の抽出について解説している。CGM情報を時系列に分析する動的トピック分析手法、トピックごとに共通して語られている文脈を発見する文脈マイニング手法、異なる情報源からの情報を俯瞰し情報源ごとの特徴を知る分散協調トピック分析手法を紹介している。これらを用いることで、キャンペーンなどのマーケティング施策の効果を測定したり、異なる情報源ごとの特徴を抽出したりすることが可能なことを示している。

大澤幸生氏（東京大学）の「**チャンス発見からバリューセンシングへ**」では、知識や情報の構造化を人間の意思決定プロセスの中に活かすことについて解説している。意思決定のための環境データの収集・可視化と意思決定主体の思考シナリオの言語化・可視化を組み合わせることで、人が主体となって新しい事象の価値を見出す思考プロセスを支援する手法（チャンス発見）について紹介している。また、環境データの収集手段として、さまざまな物理環境を測定するセンサと、人の認知行動を測定するセンサを組み合わせ、バリューセンシングへと発展させることについて紹介している。

橋田浩一氏、和泉憲明氏（産業技術総合研究所）の「**オントロジーに基づく知識の構造化と活用**」では、人間とコンピュータが共有する意味に基づいてコンテンツやサービスを作成・流通・運用支援する技術について解説している。オントロジーを用いて意味構造を明示することでコンテンツの作成を支援する手法（セマンティックオーサリング）と、オントロジーに基づいて異種のコンテンツやサービスの連携を支援する手法などを紹介している。セマンティックオーサリングを用いることで、コンテンツの品質を高めることができることを示している。

情報の価値化・知識化技術については、その重要性、議論の方向性についてはほぼ合致してきているが、具体的なアプローチについてはまだ議論の途上にある。この特集をきっかけに、情報の価値化・知識化技術に関する議論をさらに深めることができれば、幸いである。

最後に、お忙しい中、快く執筆を引き受けてくださった著者の皆様にお礼申し上げます。

（平成19年7月16日）