

# 複雑で大量なデータをリアルタイムで分析し、迅速で効果的な意思決定へ

## 「米国人気クイズ番組に挑戦したIBMの質問応答システム、Watson」

2011年2月14日から16日にかけて、情報科学の歴史に新たな1ページが刻まれました。IBMリサーチが開発した質問応答システムWatson（ワトソン）が、米国の人気クイズ番組「Jeopardy!（ジョパディ!）」で過去最強のチャンピオン2名に挑戦し、3日間（2ゲーム）の獲得賞金総額で首位となる快挙を成し遂げたのです。ここで「質問応答」という技術が人との対戦の鍵となりました。クイズ番組では質問に対して正しく解答することを競いますが、これに相当する技術は情報科学の分野では質問応答と呼ばれ、先進的な情報検索の手法として過去10年間に盛んに研究されてきました。

### 質問応答システム Watsonとは？

Watsonとは、質問文の内容を分析して、事前に収集された大量のテキスト情報から質問の解答候補とその根拠・確信度を計算し、高い確信度の候補が得られた場合に解答する、という一連の知的処理を高速に実行するコンピューター・システムです（写真）。人間が質問に解答する場合には、長年にわたり蓄積された知識を体系化して記憶しており、質問を正しく理解することができ、解答を思いつくかどうかは直感的

にわかります。人間と同じ仕組みをコンピューター上でそのまま再現することは、知識の整理と体系化に人手による膨大な作業を要する点から現実的ではありません。また質問を人間のように正確に理解することも技術的に不可能です。

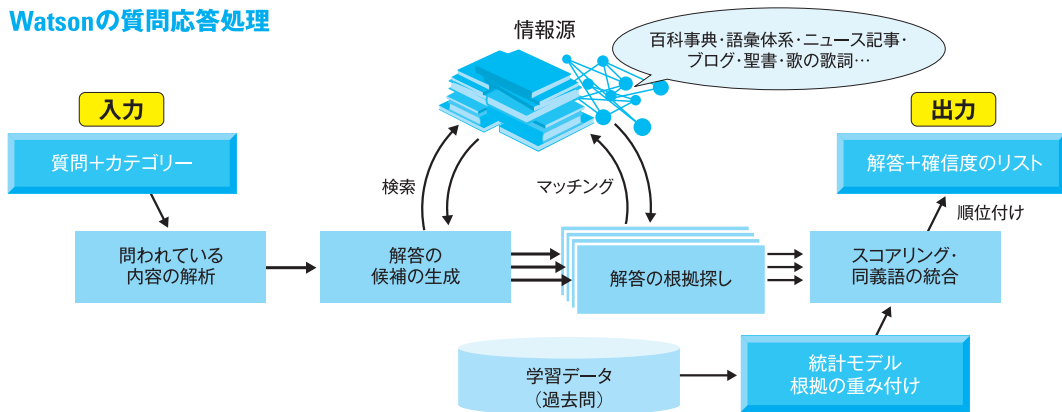
このため、Watsonでは、人間とは違う方法で質問に対する解答を求める、図（P.19）のようなアーキテクチャー（設計方式）が採用されました。質問文から何が解答として問われていそうか、また解答の手がかりとなる情報が何かを推定するために言語解析の手法が駆使されています。次に解答の候補となりそうな語句を、事前に収集された大量の文書やデータベースなどの情報源（ニュース記事、百科事典、ブログ記事、語彙の体系など）と質問文から推定し、100-1,000個程度の候補を列挙します。ここで集められた解答候補をそれぞれ元の質問文の手がかりと組み合わせて仮説を作成し、再度情報源にこの仮説を満足するような「根拠」があるかどうかを調べます。根拠の確からしさは、仮説と根拠との合致度に応じて数値化されます。有望な解答候補や仮説は情報源の多数の根拠で支持されやすく、その総和が確信度として計算されます。最も高い確信度の解答候補を選択し、その確信度が十分な値以上であるときに、Watsonはその質問に解



上がWatsonの実機。右がクイズ番組での対戦の様様。写真中央がWatsonの回答席。



## Watsonの質問応答処理



複雑な質問を解析、解答候補を推定し、根拠に基づき正しい解を導き出す、質問応答処理のプロセス。

答します。

Jeopardy!は早押しクイズ番組であるため、質問が表示されてからボタンを押す判断をするまで、数秒程度の時間で処理を終えなければなりません。これを実現したのがP.18左下の写真にあるような約3,000コアのIBM Power 750で構成された大規模な並列処理プラットフォームです。

## 質問応答技術のさまざまな分野への応用の可能性

Watsonで実現された質問応答技術は、人が知りたい情報を質問として表現したときに、膨大なテキスト情報に基づいてその解答を瞬時に獲得できる能力を提供します。Watsonでは、搭載されているデータやアルゴリズムを入れ替えるだけでさまざまな分野に適用可能な一般的なアーキテクチャーを採用しており、ニュース記事や個人をつぶやきなどのテキスト情報だけでなく、センサーから得られるデータ、健康情報など個人のデータ、そのほかさまざまな情報を大量に処理して多様な情報要求に対する解答をその確信度とともに提示することができます。現在ではまだ英語で記述された質問や情報源のみを対象としています。Watsonの研究開発にはIBM東京基礎研究所の2名の研究者も参画しており、日本語を対象とした質問応答技術に取り組むとともに、大学や研究機関との交流を推進しています。

例えばこのような技術は、製造業で部品の改良に

必要な手段を探したり、医療の現場で患者の潜在的な病気を発見したり、金融の世界で金利上昇の影響を確信度付きで予測したり、といった極めて広い範囲の分野に応用できる可能性があります。医療分野への応用については、既にニュアンス・コミュニケーションズ・インク社と共同で開発を進めています。このプロジェクトには、コロムビア大学メディカル・センターとメリーランド大学医学部が協力しており、膨大な量の情報の中に埋もれている重要な知識や事実を明らかにし、思いつかなかったかもしれない解答を提示することにより、医師や看護師など意思決定者のアイデアや仮説の立証を支援することが期待されています。他の分野でも同様に、今まで有効活用できていなかった膨大な自然言語コンテンツやデジタル・データの中に埋もれている重要な知識を発見することを可能にし、また情報を効率的に、人間の利用用途に合わせて加工することができるようになります。問題の解決にどのようなデータが利用可能であるか、どのような質問に解答すれば業務上の課題を克服できるか、といった観点で新しいシステムを設計していくことが今後の研究の一つの柱となっていくことでしょう。

なおIBMは、今回Watsonが獲得した賞金100万ドルの全額を慈善事業に寄付しています。そのうち半分の50万ドルが、IBMが仮想スーパーコンピュータ・システム「ワールド・コミュニティ・グリッド(WCG)」を用いて支援を行っている世界の人道的研究プロジェクトに寄付され、その中には千葉県がんセンターと千葉大学による神経芽腫(小児がんの一種)の新しい治療薬開発を目的とした「ファイト! 小児がんプロジェクト」が含まれています。