論　説

イノベーション創発のための企業内情報システムについて

――ビッグデータ解析ツールの課題――

情報システム委員会*
第２小委員会

抄　録　現在，あらゆる業種，規模の企業において，イノベーションの創発が強く求められている。また以前より活用が望まれていたビッグデータも，活用事例が散見されるようになってきた。そこで，当委員会では，知財部門におけるイノベーションの創発を志向したビッグデータの活用をおこなうためには，どのような知財情報システムを構築する必要があるのか，更に，ビッグデータと知財情報とを組み合わせてどのようなイノベーションが創発できるかを調査・研究することにした。

2015年度の活動は，ビッグデータを活用するための知財情報システムの構築を検討するにあたり，ビッグデータ活用事例を整理することで，知財部門が取り扱うべきビッグデータを定義して，定義されたビッグデータを活用するためのツールの調査，分類を行うこととした。本論説では，得られた知見のうち，主にそれらツールの特徴，利用シーン，利用における課題について報告する。

目　次

1．はじめに

1. 1　研究目的

現在，自社の継続的な成長のために，あらゆる業種，規模の企業において，イノベーションの創発が強く求められており，知財部門も例外ではなくなっている。一方，情報通信技術（ICT）の進展により，以前では扱うことさえ不可能であった大量データや非構造化データ（２章参照），いわゆるビッグデータを活用した事例が散見されるようになってきた。そこで，当専門委員会では，企業内の情報システムにおいて，どのようなビッグデータが知財業務で扱えるか，ビッグデータを扱うことにより知財で何ができるか，ビッグデータを扱うときは何が必要

* 2015年度　The Second Subcommittee, Information System Committee

か，これらの現状を明らかにし，イノベーション創発へつなげるための情報システムの調査研究を行った。

1. 2　2015年度の活動概要

2015年度は一般的なビッグデータについて整理すると共に，知財分野において活用されるビッグデータの活用事例の収集と，既存の知財情報システムにおいてビッグデータを扱えるツールの調査と具体的な活用に向けての調査研究を行った。具体的には，次の手順で調査研究を進めた。

1）ビッグデータの定義を整理
2）ビッグデータ活用事例の探索
3）ビッグデータ活用目的の整理
4）ビッグデータ活用時の課題抽出
5）ビッグデータと知財情報を扱うツールの調査
6）これらの活動内容に基づく結論と今後の課題を抽出

2．ビッグデータについて

2. 1　ビッグデータが注目される背景

ビッグデータは次の３点の背景により，近年，急速に脚光を浴び，活用が進められている。

1）データの多様性

モバイルデバイスの普及や動画・ソーシャルメディアの普及，各種センサの普及などにより，データの種類が増えている。

2）データの蓄積量

蓄積され活用が可能なデータ量は，ストレージの大容量化などICTの進展に伴って，急激に増加している（図１）。

3）データの分析技術

非構造化データは，以前は取り扱いが困難であったが，テキストマイニング[2)]，人工知能（AI）[3)] およびセマンティック検索[4)] などの分析技術の発達により，分析できるようになってきている。

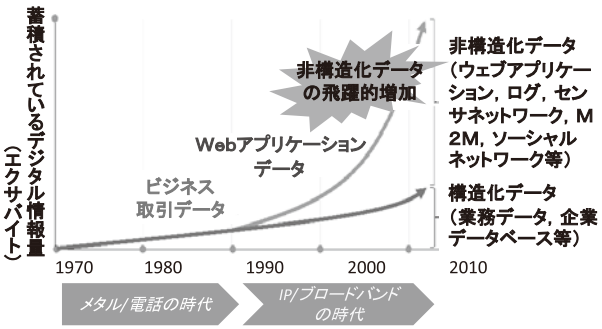

図１　ビッグデータ増加イメージ[1)]

2. 2　ビッグデータとは

ビッグデータとは，ICTの利用により情報の生成・収集・蓄積等が可能・容易になる多種・多量のデータを指す。ビッグデータは，通常，市販のデータ解析ツールやデータベースツールなどでの処理が困難な位，多量で複雑なデータであり，通常のデータ解析とは異なる処理を必要とする。

(1) ビッグデータの種類

ビッグデータとして取り扱われるデータは，構造を持つ構造化データと，構造を持たない非構造化データの２種類に大別される（図２）。

1）構造化データ

構造化データとは，汎用のデータベース等を利用することにより，データを整理することが可能なデータ（構造化して格納することが可能なデータ）の総称である。世間一般の例で具体的には，顧客情報や売上情報等の，表計算ソフトやデータベースで取り扱う表の構造を作っているデータなどを指す。

2）非構造化データ

非構造化データとは，構造定義を持たないデータの総称である。世間一般の例で具体的には，インターネット上でのブログ／SNS，新聞，電

子書籍など，表の構造を作っていないデータや，音声や画像などの，マルチメディアデータを指す。非構造化データは，IT分野における記録測定技術の急速な発展により，ビッグデータとして活用されつつある。

(2) 知財部門が対象とするビッグデータ

知財部門が取り扱うべきビッグデータは，営業，企画，生産管理部門などの一般的な部署が扱うビッグデータに加えて，各国の言語による特許公報のテキスト，図面・画像，特許の書誌的情報や審査経過情報，判例などの構造化データが対象となる。また，研究開発，品質管理，市場調査，顧客情報等の社内文書といった，通常のデータベースでは取扱いが困難な非構造化データについても重要な対象となる。

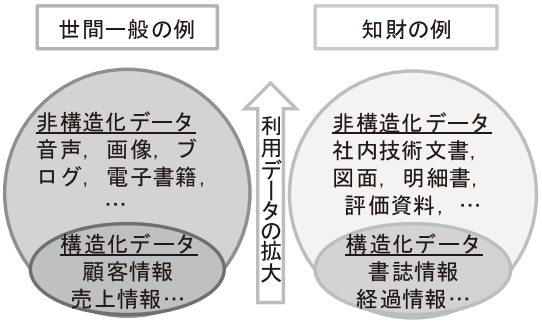

図２　ビッグデータのデータ種類イメージ

2. 3　本調査研究が着目するビッグデータ

本研究では，知財における企業内情報システムが取り扱うビッグデータを，知財情報，社外の情報，および社内の情報の３つに分類した（図３）。これらの各情報を既存のツールでどのように取り扱えるかを検証した。

(1) 知財情報

知財情報とは，管理システムで取り扱われる，出願情報，保有権利情報，費用の情報を指す。明細書，図面，社内技術資料（出願検討技術資料），自社分類，知財評価資料などもこれらに含まれる。また，調査・分析システムで取り扱われる，特許公報，経過情報，引用／被引用情報などもこの分類である。4. 3節で後述する特許データはこの分類に含まれる。

(2) 社外の情報

社外の情報とは，新聞，ニュース，雑誌，カタログ，学術文献，ソーシャルデータ（ブログ/SNS），行政情報（経済・気象・交通統計等）などである。4. 3節で後述する社外データはこの分類に含まれる。

(3) 社内の情報

社内の情報とは，事業情報，研究開発情報，製品情報，売上情報，顧客情報などである。

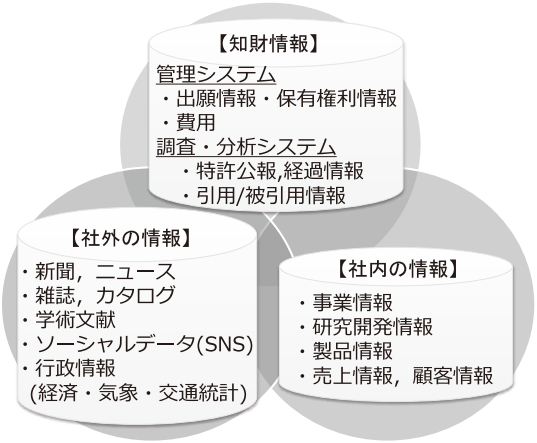

図３　知財におけるビッグデータの分類

3. ビッグデータの活用について

3. 1　活用事例

(1) 事例の調査方法

本研究では，ビッグデータを取り扱えるツールについて知財分野やそれ以外の分野においてシステムベンダ各社が公表している事例を調査した。探索方法については，インターネット上

の情報に対し，キーワード「ビッグデータ 活用 イノベーション 創発」等 を組み合わせにより抽出された，ビッグデータ活用事例，およびツール紹介事例を収集した。探索した業種，利用されているビッグデータ，および定義されているビッグデータの範囲は表１のとおりである。事例調査の結果，世間一般ではマーケティングを行っている営業部門や生産管理の現場など，事例は多数あることが分かった。また，知財部門や研究部門でのビッグデータの活用事例についても，既に存在することが分かった。一例を以下に示す。

(2) 世間一般の事例

事例１
対象範囲：生産現場
データ：生産ラインで発生するデータや製品の品質に関するデータ
成果：アナログデータを分析することにより，装置の異常を事前に発見する。

事例２
対象範囲：コンテンツビジネス
データ：視聴履歴
成果：ユーザの過去の視聴履歴を解析し，現在だけでなく将来の好みを予測し，その好みに合ったオリジナルコンテンツを作成する。

(3) 知財の事例

事例３
対象範囲：技術用途発見
データ：特許データおよびインターネット情報
成果：データを解析することで，技術の成熟度と顧客ニーズの変化を可視化し，両者の関係性が理解できるように表現している。

事例４
対象範囲：技術間の関係性
データ：自動運転に関する特許公報のテキスト
成果：技術をグループ化し，類似する技術どうしの距離を可視化して表現することにより，技術のホワイトスペース探索に取り組んでいる。

3. 2 活用目的

前項の事例調査を分析することにより，ビッグデータの活用目的と，課題解決に向けての期待について，世間一般の事例と知財の事例に整理し，各事例における課題について検討を行った（図４）。

表１ 探索した範囲

調査項目	事例
企業/ベンダ数	20社
業種	生産設備，通信，運輸，自動車，鉄道，コンテンツビジネス，ヘルスケア 等
利用されているビッグデータ	自社他社特許データ，自社が持つ大量の多種多様なデータ，性能診断結果，メンテナンス診断結果，交通系ICカード利用データ，センサ情報，検索システムの利用ログ，カーナビゲーション装着車の走行データ，CTスキャン装置の画像情報，視聴履歴 等
ビッグデータの定義	・ICTの進展により生成・収集・蓄積等が可能・容易になる多種多量のデータ ・高精度かつ連続した多数の位置データ ・実際の利用・移動に基づいた蓄積された膨大のデータ ・センサでとらえる人間の行動の情報 ・インターネットに繋がる様々な情報通信機器により作られたデータ集合体 ・複数のカテゴリそれぞれに属する大量のデータの組み合わせ

(1) 世間一般の事例

世間一般における事例においては，ビジネス上の課題として事務作業を効率化するといった身近な課題解決に期待がある一方で，より大きな課題として，業務精度の向上，無料データの利活用，正確な売上等の予想，新規のニーズ発掘などの解決においての期待がある。ビッグデータを分析活用することにより，新たな知見や価値を発見するという，大きな課題解決に対して貢献できることが期待されている。

(2) 知財の事例

世間一般の事例に加え，知財上の課題では，発明者への実績報償，拒絶対応などの庁手続きの効率化，特許予算の精度向上といった，知財分野に特化した課題から，技術動向や出願動向の正確な予測，ホワイトスペースの発見，イノベーションの創発や協創先を発見することなどといった，技術開発や知財戦略の課題解決に対して貢献できることが期待されている。

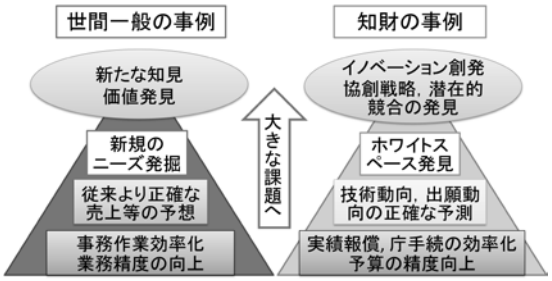

図４　ビッグデータの活用目的と課題

3. 3　活用時の課題

ビッグデータの活用において，データをツールで扱う際，ビッグデータが急速に普及した背景から生じる，以下の３つの課題に直面することが想定される。

１つ目は「データの多様性から生じる多言語対応」である。グローバル化の進展により，テキスト情報に加え，通貨や単位など，各国への対応が必須の課題となることが容易に推測される。

２つ目は，「膨大な蓄積量となるデータの取扱い」である。多言語にわたる特許公報や技術論文に加え，画像，音声といった膨大な容量を有するデータの取扱いをどのようにすべきか，という課題がある。

３つ目は，「データの分析技術から生じる品質・業務効率」である。多岐・多様なデータを分析，評価する際，統計処理やテキストマイニングといった様々なデータ処理や分析のスキルが必要となるため，得られる分析結果が作業担当者の経験や熟練度などに左右されてしまうことが課題である。

4．ビッグデータを取り扱うツール

4. 1　ツールの調査範囲

本調査研究では，ビッグデータと知財情報を利用でき，かつ3. 2（2）項に挙げている知財分野の課題を解決する助けになるツールを調査した。調査した範囲は以下のとおりである。

- 当委員会メンバーが知っているツール。
- 日本で宣伝，販売されているツール。
- 当委員会で想定した目的を達成できるツール。
- ビッグデータ活用事例をインターネット等で探すなかで，知財情報の利用が紹介されているツール。

4. 2　調査したツールの分類

(1) 分類の観点

調査したツールを整理するため，以下の４つの観点を設定した（図５）。

観点１）特許公報や整理標準化データなどの知財情報が既に取り込まれているかどうか（○：あり，×：なし）。

観点２）技術文献や企業データ，Web等で公開されている製品情報などの社外の公開情報が既に取り込まれているかどうか（○：あり，×：なし）。

観点３）社内外に係らず任意のデータをシステム上に取り込むことが可能である，または事業開発や製品，売上情報等を管理している社内外の任意のシステムと連携が可能であるかどうか（○：可能，△：一部可能，×：不可）。

観点４）操作する上での作業量の多さや作業に対しての熟練を要するなどといった，人への要求度（大，中，小）。

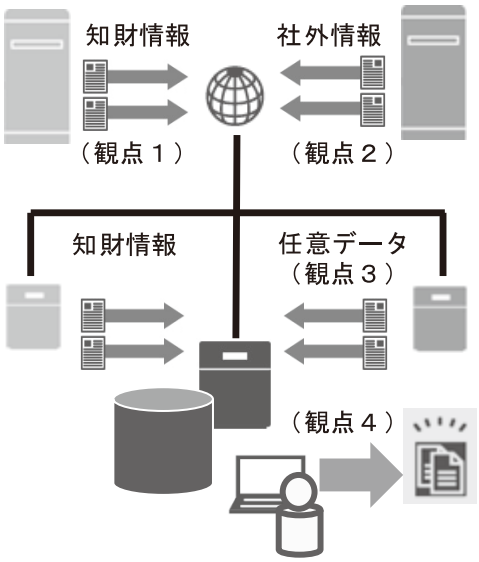

図５　調査したツールの観点

(2) 分類によるタイプ

上記４つの観点で対象としたツールを調査した結果，４つのタイプに分類されることが分かった。各タイプの利用イメージは以下のとおりである。

タイプ１）希望する任意のデータを取り込むことが可能である。

タイプ２）知財情報が既に取り込まれている。

タイプ３）知財情報と社外情報が既に取り込まれている。

タイプ４）人工知能技術などの利用により，従来手法よりも分析効率が向上している。

4. 3　分類したツールのタイプ別の特徴

以下に４つのタイプの特徴を示す（表２）。ここで，タイプ１とタイプ２は従来から広く利用されているツールである一方で，タイプ３もしくはタイプ４は，比較的新しいツールであり事例も少ないことから，本稿ではより具体的に紹介する。なお，特徴，利用イメージ，問題点・課題は各タイプを端的に表現したものであり，具体的なツール自体の評価を行ったものではないことを特記しておく。

表２　ツールのタイプ別特徴まとめ

タイプ	特許データ	社外データ	データ取り込み連携	人手介在量・人のスキル依存度
タイプ１	×	×	○	大
タイプ２	○	×	×	中
タイプ３	○	○	△	小
タイプ４	×	×	○	中

(1) タイプ１の特徴

【特徴】

・特許データの保有：なし

・社外データの保有：なし

・データ取り込み連携：可能

・人手介在量・人のスキル依存度：大

【利用イメージ】

特許公報データなどのテキストデータは別の検索ツールなどからあらかじめ出力しておき，そのデータをツールに取り込み，各種分析を行う。

【ツール例，活用事例】

・タイプ１Ａ（Ａ社ツール）：

テキストデータを数値化・分析し，意思決定に有益な形式にて表示。請求項のテキストマイニング等に活用される。

・タイプ１Ｂ（Ｂ社ツール）：

特許情報の自動集計と，任意の社内データを付与した特許情報マップの作成ができる。意思決定の判断材料として活用される。

【問題点・課題】

人的にデータを取り込む作業が必要となる，検索分析やマップ作成の際，作業者の熟練度により分析結果の完成度にバラつきが生じる可能性がある。

(2) タイプ２の特徴

【特徴】

・特許データの保有：あり
・社外データの保有：なし
・データ取り込み連携：不可
・人手介在量・人のスキル依存度：中

【利用イメージ】

特許公報がツールに備えられていて，Webサービスで分析が可能であり，マップ作成機能がツールに装備されていて可視化も容易である。

【ツール例，活用事例】

・タイプ２Ｃ（Ｃ社ツール）：

文書情報をクラスター化し，構造的な解析結果が提供される。文書情報を全体俯瞰として一枚の画に可視化できる。

・タイプ２Ｄ（Ｄ社ツール）：

特許文献に基づき，客観的な競争力分析，自社・他社開発動向の整理が可能である。テーマ探索や研究開発戦略策定等に活用される。

【問題点・課題】

特許情報などの予め用意されたデータ以外の分析はできない(データを任意に取り込めない)。

(3) タイプ３の特徴

【特徴】

・特許データの保有：あり
・社外データの保有：あり
・データ取り込み連携：一部可能（学術論文，ビジネス，業界，学会，米国裁判，企業情報，ニュース等）
・人手介在量・人のスキル依存度：小

【利用イメージ】

各国の特許公報や学術文献，企業情報や売上情報など，様々な社外データが予め用意されており，テキストデータに関してはセマンティック検索など，より高度な検索エンジンを利用することにより，効率的に情報を検索して可視化することができる。

取り扱うデータについては，グローバル対応として，予め人手や機械翻訳などにより統一された言語（多くは英語）に翻訳された上で分析に供されている。

【ツール例，活用事例】

・タイプ３Ｅ（Ｅ社ツール）：

競合分析全般に活用（技術動向，販売・特許戦略等）されている。各国言語は英語に機械翻訳の上収録されており，日本語への自動翻訳も可能である。公報等の誤り等は事前に修正の上収録している。データについては，ベンダ独自の分類を付与するなどの手当てもされており，提供するデータの信頼性，統一性が図られている。

・タイプ３Ｆ（Ｆ社ツール）：

ポートフォリオ管理や他社競合分析，ホワイトスペース探索などを目的に利用されている。セマンティック検索機能（キーワードではなく，文章の係受けを理解して検索）により精度の高い検索結果を得るエンジンを採用しており，加えて独自の特許評価（特許の強さ）の付加機能などにより分析結果の充実が図られている。またデータの正確性を高めるために，データ収録時に前処理として，子会社や関連会社のグループ化や企業名の表記の

揺れ修正，権利移転の追跡などのデータクレンジングやデータ間の関係性の修正を独自に行っている。

・タイプ３Ｇ（Ｇ社ツール）：

情報収集，調査～技術課題解決～アイデア創出までの一連の段階において，付加価値製品の導入，品質・信頼性改善などのイノベーションを実現する作業の支援に活用されている。セマンティック検索を使用することにより，効率的で正確性の高い検索を実現する。データの連係についてはODBC接続[5]により，社内システム・ローカルドライブ・任意のWebページ等を検索対象とできるなど，広範囲な連携を行うことが可能である。

【期待できること】

・品質，業務効率

従来ツールのように，分析データを収集する作業から始める必要が無く，あらかじめ用意されたグローバルの様々な特許や特許以外のデータを利用することが可能である。セマンティック検索などの高度な検索機能を有することや，修正されたデータを分析対象とすることによりデータ精度の向上が図られており，効率よく品質の高い分析ができる。

・多言語対応

人手や機械翻訳により予め英語に統一して変換されているデータを利用することにより，多言語への対応が可能となっている。出力結果の日本語への翻訳機能を搭載しているツールもあり，作業者の利便性も考慮されている。

・膨大なデータ量

特許公報だけではなく学術論文，ビジネス，業界，学会，米国裁判，企業情報，財務，標準技術，ニュースなどが予め格納されており，かつ高頻度で更新もされている。中には社内のデータベース等各種データソースと連携することができるものもあり，統合横断的な情報検索が可能なツールも存在する（図６）。

【問題点・課題】

セマンティック検索などの新しい検索エンジンそのものや，機械翻訳の信頼度について十分に検証していないため，ツールそのものの性能については未検証である。

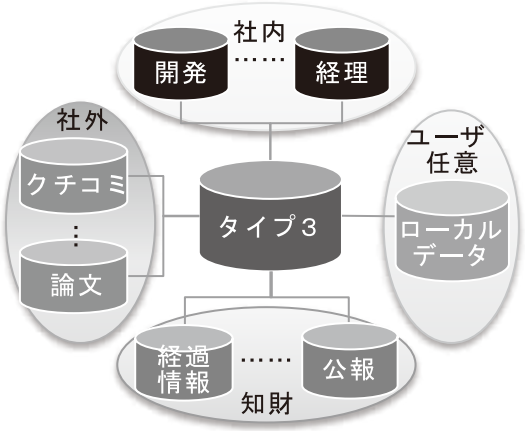

図６　横断的な情報検索が可能なタイプ３

(4) タイプ４の特徴

【特徴】

・特許データの保有：なし

・社外データの保有：なし

・データ取り込み連携：可能

・人手介在量・人のスキル依存度：中

【利用イメージ】

人工知能技術などの利用により，従来手法よりも分析効率が向上している事が期待される。

【ツール例，活用事例】

・タイプ４Ｈ（Ｈ社ツール）：

ランドスケイピング（少ない教師データでAIを学習させることを可能とした新しい機械学習の手法）により，テキストデータを精度と網羅性を高く解析することが可能である。公報以外にも発明提案書や無効化したい特許資料等についても対象として利用可能であり，対象テキストを教師データとしてAI

に学ばせ，解析・スコアリングして文書を仕分けた上に優先順位付けが可能である（図7）。

【期待できること】

人手によるスクリーニングに頼っていた文書の選別作業において AIが活用できることにより，大幅な省力化が見込まれる。またAIの判断により人の分析スキルに対する依存度が小さくなるため，結果的に品質の向上や均一化も期待できる。将来的にはAI自体の性能向上も相まってより多くのデータを組み合わせることが期待される。

【問題点・課題】

セマンティック検索と同様に，AIによる学習精度の信頼性や性能自体については十分な検証をおこなっていないため，ツールそのものの性能については未検証である。

5．おわりに

以上のように，本論説では企業内の情報システムにおいて，知財情報として扱えるビッグデータを特定し，知財分野においてビッグデータを扱うことにより何ができるかを現行のツールにおいて検証し，かつビッグデータを扱う際に必要な要素の抽出を行った。それぞれについて，以下に得られた結論を示す。

(1) どのようなビッグデータが知財で扱えるか

知財分野における情報は，文書として解析可能な非構造化データおよび文書から抽出され，紐付けられた構造化データが，ビッグデータとして扱われている。

(2) ビッグデータを扱うことにより知財で何ができるか

知財分野におけるビッグデータの活用は，特に業務効率化の場面で利用可能な環境が整備されつつある。たとえば，AI等を利用した業務自動化による精度向上と業務負担軽減，あるいは，ホワイトスペース探索における分析支援などがある。

(3) ビッグデータを扱うには何が必要か

膨大でかつ多言語のデータを扱うにあたり，データそのものの信頼性の担保とともに，「翻訳機能」，「セマンティック検索」や「AI」などを利用した効率化・分析の品質向上が必要で

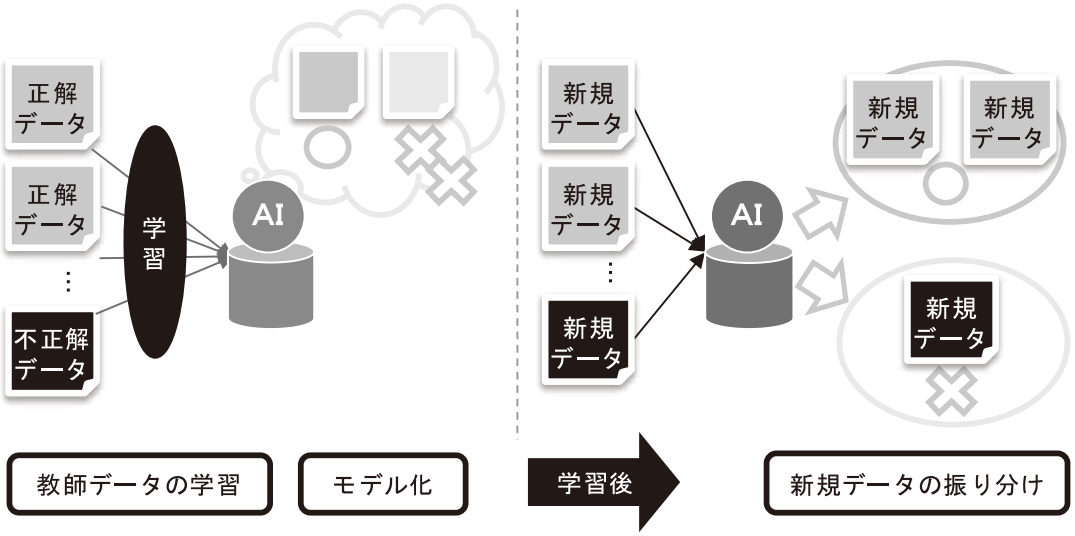

図7　AIによる学習と仕分け

あると考えられる。

知財分野におけるビッグデータを取り扱えるツールの飛躍的な性能向上が将来的に期待できる現状を踏まえ，2016年度以降では，未知の知見発掘をサポートするツールについて，具体的な使いこなし方やデータ判断方法の検討を重ね，システム化実現に際しての課題と解決策の具体化を検討する。

本報告は，2015年度情報システム委員会の第２小委員会メンバーである，若林宏明（エルゼビア・ジャパン），松本朋子（富士フイルム），落合昌孝（富士ゼロックス），椚原孝夫（村田製作所），小林幸信（サトーホールディングス），遠山正幸（三井造船），原口正義（バッファロー），古市将英（オムロン），和田智樹（東日本旅客鉄道）の執筆によるものである。

注　記

1） 総務省情報通信国際戦略局情報通信経済室，情報流通・蓄積量の計測手法に係る調査研究報告書，p.4（2013）
http://www.soumu.go.jp/johotsusintokei/linkdata/h25_03_houkoku.pdf（参照日：2016. 5. 12）

2） テキストマイニング
大量の文章データ（テキストデータ）から，有益な情報を取り出すことを総称してテキストマイニングと呼ぶ。自然言語解析の手法を使って，文章を単語（名詞，動詞，形容詞等）に分割し，それらの出現頻度や相関関係を分析することで有益な情報を抽出する。
http://www.trueteller.net/textmining/about.html（参照日：2016. 5. 12）

3） 人工知能
人の学習，推論，判断といった知能をIT技術により実現するもので，機械学習（人が自然と行うような学習をコンピュータが行うこと）から特徴量を抽出して，それをもとにデータに対して推論や判断を行う。従来より音声認識や画像認識，テキスト解析等に活用されており，人が扱うことが難しいビッグデータを有効に活用するために期待されている技術。

4） セマンティック検索
図８の検索事例のように，文章の係り受けや単語の意味を理解して検索する。

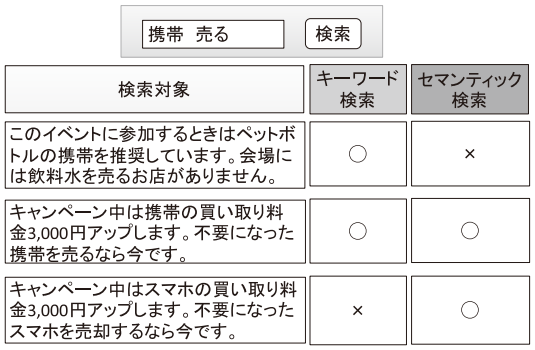

図８　セマンティック検索の例

5） ODBC（Open Database Connectivity）は，アプリケーションソフトウェアからMicrosoft SQL Serverなどの外部データソースに接続するために使用できるプロトコル。

（原稿受領日　2016年６月２日）