

企業内情報システムについて

～ビッグデータ解析を踏まえて～

2016年4月度 関東・関西部会
情報システム委員会
第2小委員会

2015年度 第2小委員会メンバー

| 役職 | 氏名 | JIPA会員名 |
|--------------|-------|-----------------|
| 小委員長 | 若林 宏明 | エルゼビア・ジャパン株式会社 |
| 小委員長補佐 | 松本 朋子 | 富士フイルム株式会社 |
| 委員 | 落合 昌孝 | 富士ゼロックス株式会社 |
| 委員 | 梶原 孝夫 | 株式会社村田製作所 |
| 委員 | 小林 幸信 | サトーホールディングス株式会社 |
| 委員 | 遠山 正幸 | 三井造船株式会社 |
| 委員 | 原口 正義 | 株式会社バッファロー |
| 委員 | 古市 将英 | オムロン株式会社 |
| 委員 (期中退任) | 和田 智樹 | 東日本旅客鉄道株式会社 |

(敬称略、委員五十音順)

発表内容目次

1. 本研究の狙い
2. ビッグデータの定義
3. ビッグデータとその活用事例の調査
4. ビッグデータを扱うツールの調査
5. 結論

発表内容目次

1. 本研究の狙い
2. ビッグデータの定義
3. ビッグデータとその活用事例の調査
4. ビッグデータを扱うツールの調査
5. 結論

1. 本研究の狙い

1) 2015年度の研究目的

2) 研究活動サマリー

1) 2015年度の研究目的

企業内情報システムにおいて

- ・どのようなビッグデータが知財で扱えるか (情報)
- ・ビッグデータを扱うことにより知財で何ができるか (目的)
- ・ビッグデータを扱うには何が必要か (手段)

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を
調査・研究した。

2) 研究活動サマリー

2015年度は以下の内容で活動を行った。



- ビッグデータの定義の整理



- ビッグデータ活用事例の収集



- ビッグデータ活用目的の整理



- ビッグデータ活用時の課題抽出



- ビッグデータと知財情報を扱うツールの調査



- 上記活動内容に基づく結論と今後の課題を抽出

発表内容目次

1. 本研究の狙い
- 2. ビッグデータの定義**
3. ビッグデータとその活用事例の調査
4. ビッグデータを扱うツールの調査
5. 結論

2.ビッグデータの定義について

- 1)ビッグデータが注目される背景
- 2)ビッグデータとは何か
- 3)本研究におけるビッグデータの種類

1) ビッグデータが注目される背景

ビッグデータは以下の背景により、近年急速に脚光を浴びており、その活用が進められている。

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、各種センサの普及等により、データの種類が増加している。

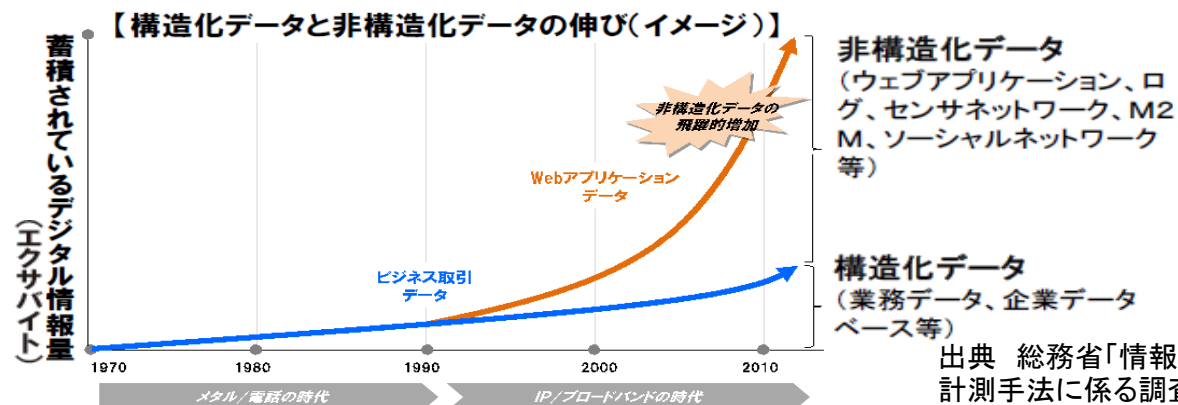
データの蓄積量

- ・ ICT*の進展に伴い、蓄積・活用可能なデータ量は急激に増加している。

*ICT: Information and Communication Technology

データの分析技術

- ・ 従来取扱いが困難であった非構造化データは、テキストマイニング、人工知能及びセマンティック検索等の分析技術の発達により分析可能になってきている。

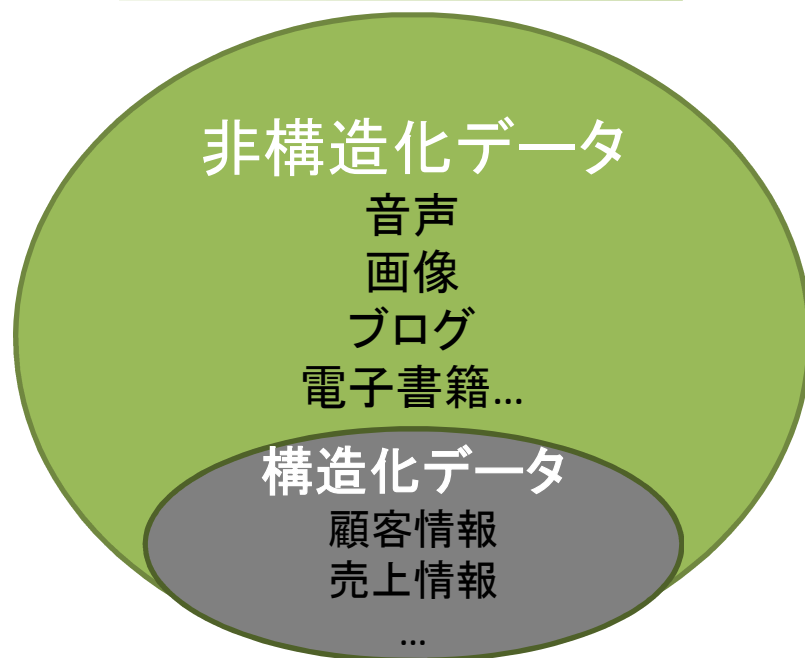


出典 総務省「情報流通・蓄積量の計測手法に係る調査研究 報告書」

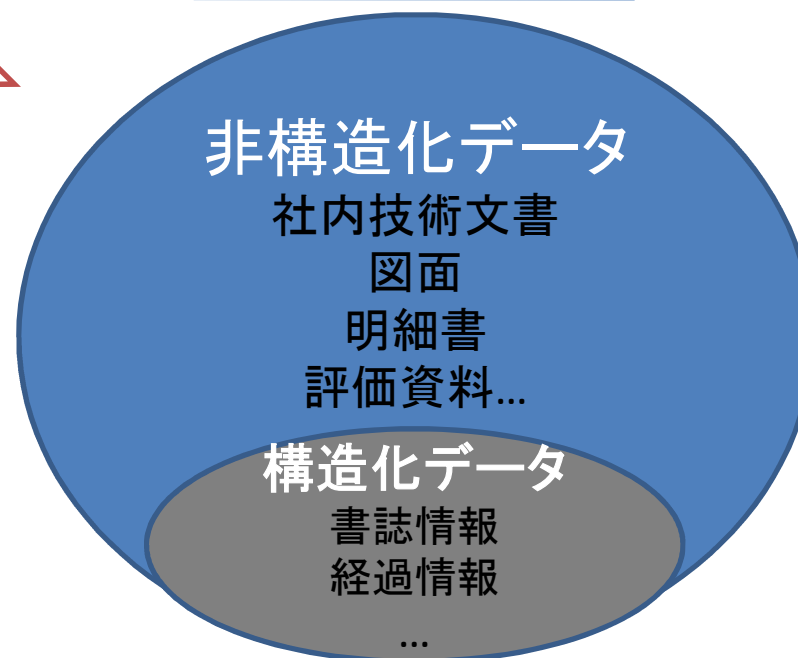
2) ビッグデータとは何か

ビッグデータは、ICT技術の利用により情報の生成・収集・蓄積が可能・容易になる多種多様のデータを指す。データの種類で大別した場合、非構造化データと構造化データに分かれる。※1

世間一般の例※2



知財の例

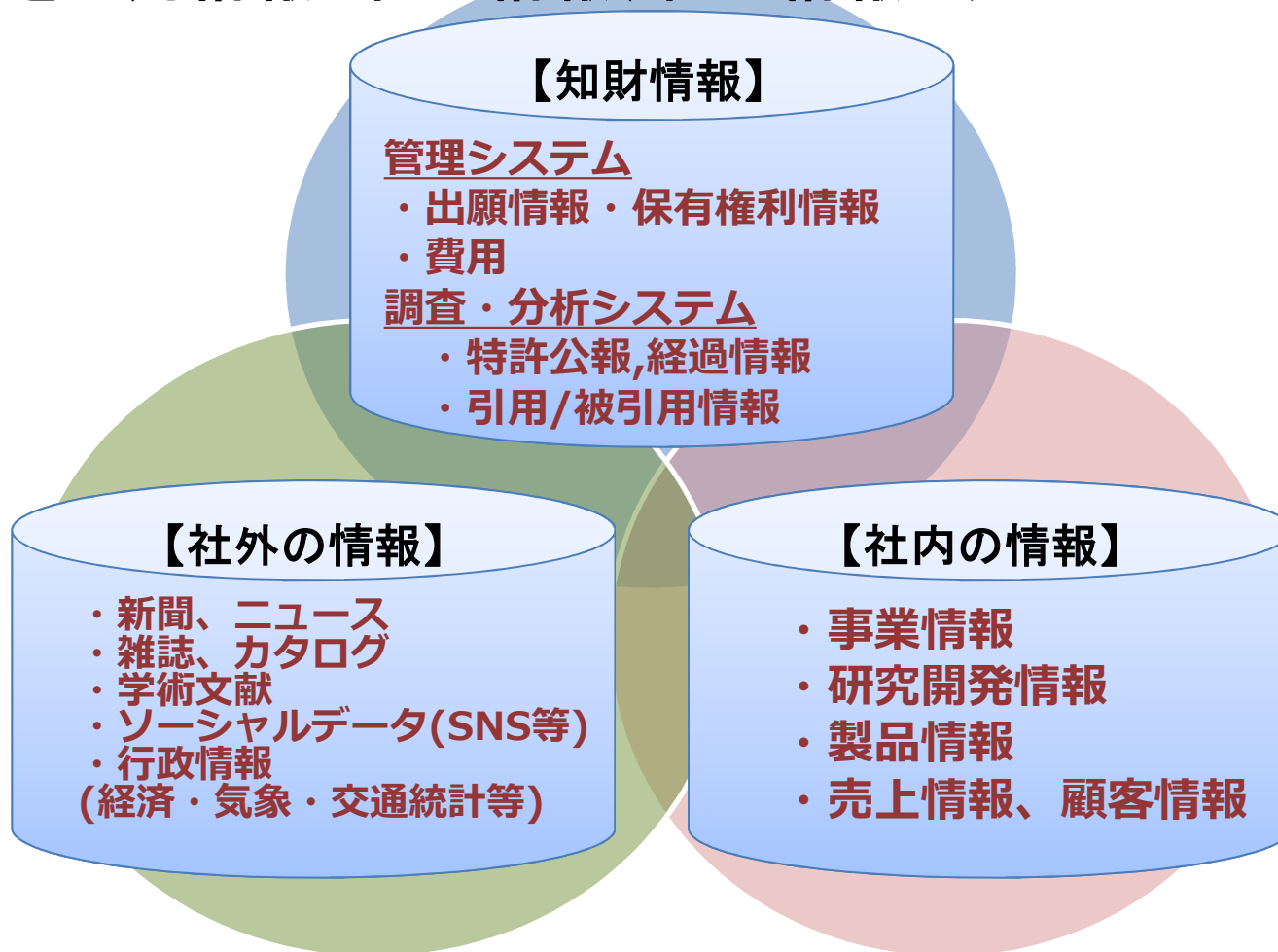


↑
利用データの拡大

※1 出典 総務省「情報流通・蓄積量の計測手法の検討に係る調査研究」(平成25年)による
※2 マーケティングが必要な部門(営業)や生産管理の現場等の部門を指す

3) 本研究におけるビッグデータの種類

知財における企業内情報システムに関する本研究では、ビッグデータを知財情報と社内情報、社外情報に分けることができる。



発表内容目次

1. 本研究の狙い
2. ビッグデータの定義
- 3. ビッグデータとその活用事例の調査**
4. ビッグデータを扱うツールの調査
5. 結論

3.ビッグデータとその活用事例の調査

- 1)ビッグデータ活用の事例
- 2)ビッグデータ活用の目的
- 3)ビッグデータ活用時の課題

1) ビッグデータ活用の事例

ビッグデータ活用に関し、世間一般および知財の事例を調査した結果、既に多数の例が存在することが分かった。次頁で抜粋例を紹介する。

| 調査結果 | 事例数 |
|---------------|---|
| 企業/ベンダー数 | 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、センサー情報、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 | <ul style="list-style-type: none">・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・センサー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蓄積の組み合わせ |

調査方法: 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」等を組合せて検索を行い、ビッグデータを活用している世間一般または知財の事例、もしくはツールの紹介に用いられている事例を収集した

ビッグデータ活用の事例 — 調査の抜粋例 —

世間一般と同様に、知財においても様々な場面でビッグデータの利用・検討が進んでいる。

(※インターネット調査に基づく)

世間一般の例

電機メーカー

対象範囲 : 生産現場

利用データ : 生産ライン、品質データ等
(年間 ○ Exa Byte)

成果 : アナログデータの分析により
装置の異常を事前に発見する

コンテンツ提供サービス

対象範囲 : コンテンツビジネス

利用データ : 過去の視聴履歴

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、
オリジナルコンテンツを作成する

知財の例

特許調査分析サービス

対象範囲 : 技術用途発見

利用データ : 特許データ、市場情報、
インターネット情報

成果 : 技術の成熟度 及び顧客ニーズ
変化の関係を可視化する

自動車業界に特化した調査サービス

対象範囲 : 技術間の関係性

利用データ : 自動運転に関する特許公報
(数万件)の特許全文

成果 : 技術をグループ化、関係性を
距離で表現して可視化する

2) ビッグデータ活用の目的

ビッグデータ活用には、今までのデータ分析では解決できなかった大小様々な課題解決への貢献が期待されている。

世間一般の例

新たな知見
価値発見

新規のニーズ
発掘

従来より正確な売上
等の予想

事務作業効率化
業務精度の向上

知財の例

イノベーション創発
協創戦略、潜在的競合
の発見

ホワイトス
ペース発見

技術動向、出願動向の
正確な予測

実績報償、庁手続等の効率化
予算の精度向上

大きな課題へ

3) ビッグデータ活用時の課題

ビッグデータを活用する場合、データを扱うツールはビッグデータが急速に普及した背景から生じる課題に直面すると想定される。

＜普及の背景 → 想定される課題＞

データの多様性 → 多言語対応

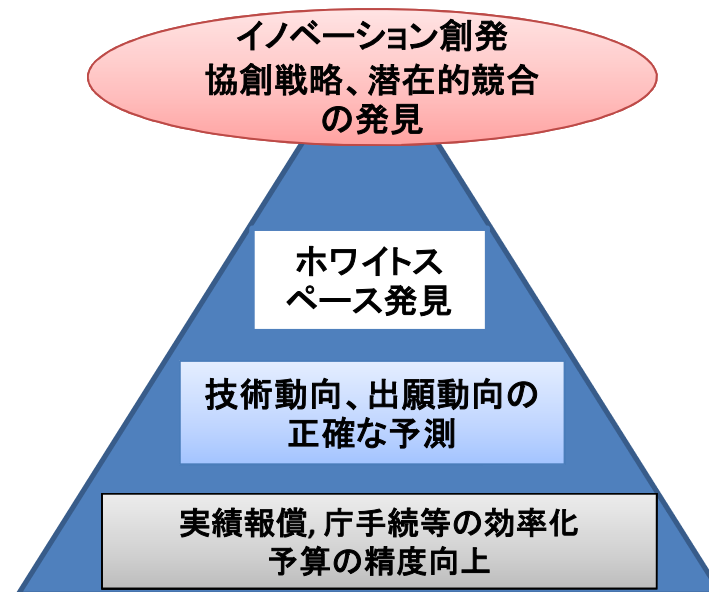
- グローバル化の進展により多くの言語への対応必要性

データの蓄積量 → 膨大なデータ量

- 特許公報だけではなく技術論文や画像音声の取扱い

データの分析技術 → 品質・業務効率

- 利用者の経験や熟練度によって分析や評価の精度にバラつき



発表内容目次

1. 本研究の狙い
2. ビッグデータの定義
3. ビッグデータとその活用事例の調査
- 4. ビッグデータを扱うツールの調査**
5. 結論

4. ビッグデータを扱うツールの調査

- 1) ビッグデータを扱えるツールの調査範囲
- 2) 調査したツールの分類
- 3) 分類したツールのタイプ別の特徴

1)ビッグデータを扱えるツールの調査範囲

ビッグデータと知財情報を利用でき、知財部門の目的を達成する助けになりうるツールを調査した。

調査範囲は以下3点とし、2015年度はベンダー7社にヒアリング、インターネットやカタログ、特許情報フェア等で5社調査を実施した。

- 1.当委員会メンバーが既知であるもの
- 2.日本で宣伝、販売されているもの
- 3.ツール活用事例で知財情報の利用が紹介されているもの

2) 調査したツールの分類 —分類の観点—

調査したツールを以下の項目で整理し分類した。

項目1. 知財情報が既に取り込まれている

(公報データ、整理標準化データ等)

項目2. 社外情報が既に取り込まれている

(一般の技術文献、製品情報、Web情報等)

項目3. 社内の任意のデータを取り込むか

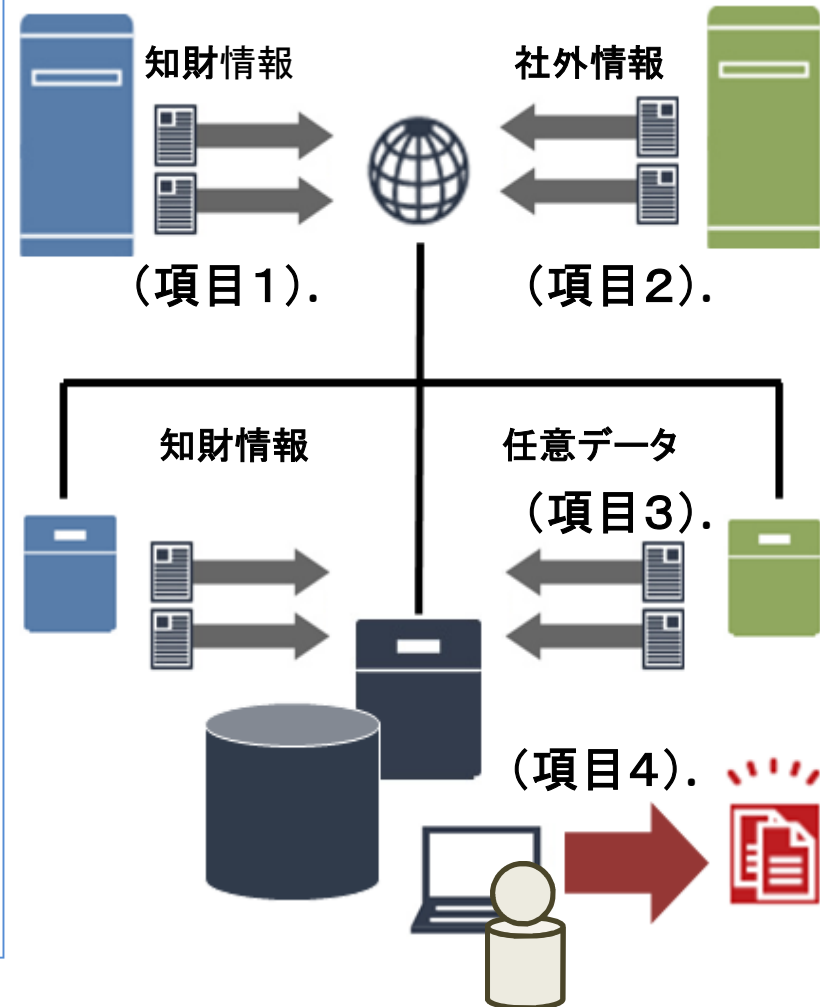
又は任意のシステムと連携可能である

(事業/開発/製品/売上情報やその管理システム等)

項目4. 人への要求度(介在の度合い)

人の介在の量(作業の多さ)

人の介在の質(分析・解析スキル)



2) 調査したツールの分類-4タイプ別

| タイプ | 特許データ | 社外データ | データ取込連携 | 人への要求度 (依存度) | 利用イメージ |
|------|-------|-------|---------|-----------------|---|
| タイプ1 | × | × | ○ | 大 | ・出力データをツールに取込み各種分析を行う |
| タイプ2 | ○ | × | × | 中 | ・特許公報などプリセットデータをWebシステム経由で利用し、容易に可視化ができる |
| タイプ3 | ○ | ○ | △ | 小 | ・予め各国の特許公報を始め様々なデータが用意されており、セマンティック検索などを利用し効率的に情報を検索して可視化ができる |
| タイプ4 | ○ | × | ○ | 小 | ・分類と数値データを組み合わせがある程度自動化されており、人工知能技術の利用により分析効率の向上が期待されている |

2) 調査したツールの分類-4タイプ別

| タイプ | 特許データ | 社外データ | データ取込連携 | 人への要求度 (依存度) | 利用イメージ |
|------|-------|-------|---------|-----------------|---|
| タイプ1 | × | × | ○ | 大 | ・出力データをツールに取込み各種分析を行う |
| タイプ2 | ○ | × | × | 中 | ・特許公報などプリセットデータをWebシステム経由で利用し、容易に可視化ができる |
| タイプ3 | ○ | ○ | △ | 小 | ・予め各国の特許公報を始め様々なデータが用意されており、セマンティック検索などを利用し効率的に情報を検索して可視化ができる |
| タイプ4 | ○ | × | ○ | 小 | ・分類と数値データを組み合わせがある程度自動化されており、人工知能技術の利用により分析効率の向上が期待されている |

当委員会ではここに注目

3) 分類したツールのタイプ別の特徴

【タイプ3】

| タイプ | 特許データ | 社外データ | データ取込連携 | 人への要求度(依存度) | 利用イメージ |
|------|-------|-------|---------|-------------|--|
| タイプ3 | ○ | ○ | △ | 小 | ・予め各国の特許公報をはじめ様々なデータが用意されており、セマンティック検索などを利用し効率的に情報を検索して可視化ができる |



3) 分類したツールのタイプ別の特徴

タイプ3の特徴①

【特徴】

| 特許データ | 社外データ | データ 取込連携 | 人への要求度 (依存度) |
|--|--|---------------|-----------------|
| Type3A(A社ツール) : 90カ国以上、うち6カ国全文。 Type3C(C社ツール) : 7カ国 | 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、news等 ※ツールにより用意されて いるデータソースは異なる。 | 一部のツールは 可能 | 小 |

【利用イメージ】

ベンダにて各国の特許公報をはじめ様々な社外データが予め用意されており、セマンティック検索などを利用し効率的に情報を検索して可視化。

ホワイトスペース探索や競合他社分析、新規プレーヤの発見、協業相手の探索等を目的として利用される。

データに関しても、予め人手や機械翻訳などにより英語に翻訳されている。

3) 分類したツールのタイプ別の特徴

タイプ3の特徴②

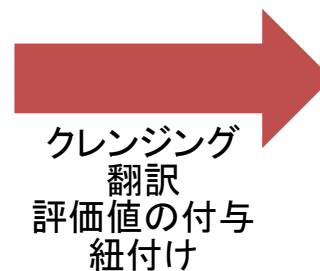
【ツール例、活用事例】

① Type3A (A社ツール) :

競合分析全般(技術動向、販売・特許戦略等)。各国言語を英語翻訳し収録。日本語への自動翻訳も可能。公報の誤りを修正し収録、独自分類を付与するなど、データの信頼性、統一性が図られている。

② Type3B (B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や、独自の特許評価(特許の強さ)の付加機能などにより分析を支援する。またデータ格納の前処理として、データの正確性を高めるために、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データクレンジングを行ったり、ルールに従い異なるデータ間の紐付けを行っている。



信頼性・利便性向上

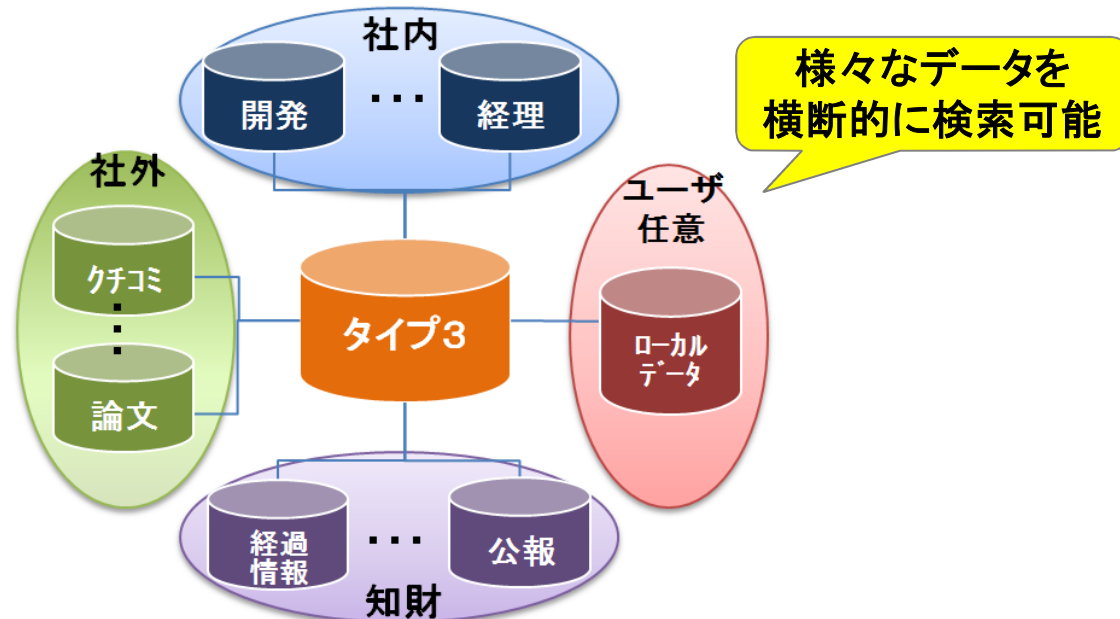
3) 分類したツールのタイプ別の特徴

タイプ3の特徴③

【ツール例、活用事例】

③Type3C(C社ツール) :

持続的イノベーション実現を目的に、情報収集、調査～技術課題解決～アイデア創出までを支援。セマンティック検索(キーワードではなく、文章の係受けを理解して検索)をすることで、効率的で正確性の高い検索を実現。ODBC接続により、社内システム・ローカルドライブ・Notes・任意のWEBページ等を検索対象・関係が可能。付加価値製品導入、品質・信頼性改善に対するの発想を支援。



3) 分類したツールのタイプ別の特徴

タイプ3の特徴④

【期待できること】

品質・業務効率

- 予め用意されたグローバルの様々な知財や知財以外のデータを利用可能
- セマンティック検索機能による検索精度向上
- 予め修正されたデータ格納によりデータ精度向上

多言語対応

- 人手や機械翻訳により予め英語に変換
- 日本語への翻訳機能(※ツールによる)

膨大なデータ量

- 特許公報だけではなく学术论文、ビジネス、業界、学会、米国裁判、企業情報、財務、標準技術、newsなどを予め格納
- 高頻度でのアップデート
- 社内のデータベース等各種データソースと連携

【問題点・課題】

セマンティックや翻訳の信頼度については未検証

3) 分類したツールのタイプ別の特徴

【タイプ4】

| タイプ | 特許データ | 社外データ | データ取込連携 | 人への要求度 (依存度) | 利用イメージ |
|------|-------|-------|---------|-----------------|--|
| タイプ4 | ○ | × | ○ | 小 | ・分類と数値データを組み合わせがある程度自動化されており、人工知能技術の利用により分析効率の向上が期待されている |

3) 分類したツールのタイプ別の特徴

タイプ4の特徴①

【特徴】

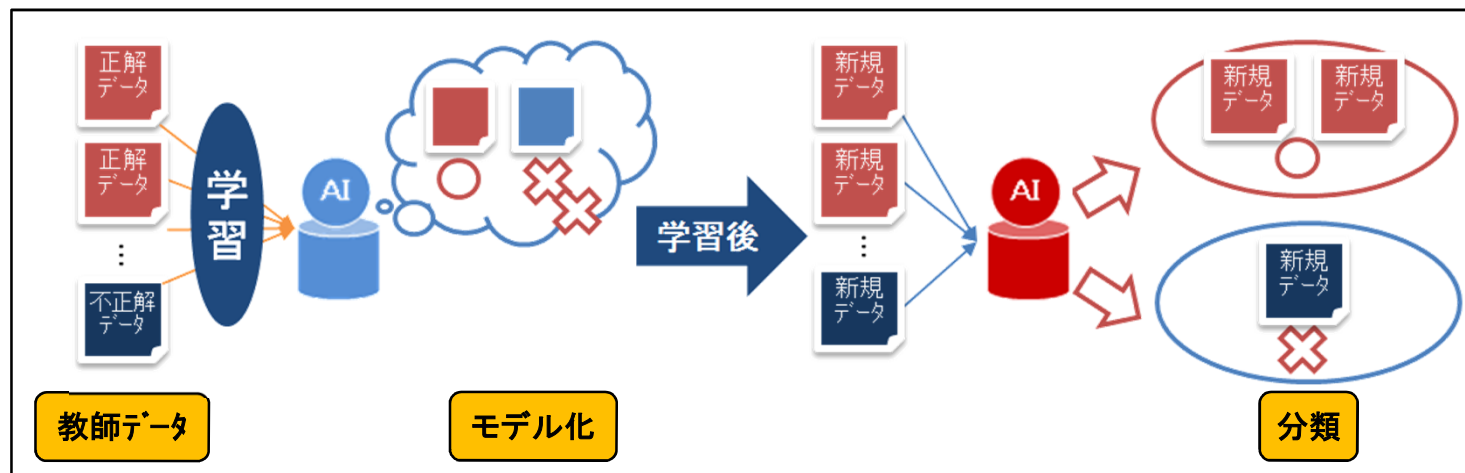
| 特許データ | 社外データ | データ 取込連携 | 人への要求度 (依存度) |
|-------|-------|-------------|-----------------|
| ○ | × | ○ | 小 |

【ツール例、活用事例】

①Type4D(D社ツール) :

ランドスケイピング(機械学習の手法)により、精度と網羅性を高く文書を解析。

公報以外にも発明提案書や無効化したい特許資料等のテキストを人工知能(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。



3) 分類したツールのタイプ別の特徴

タイプ4の特徴②

【期待できること】

品質・業務効率

- 文書解析・スコアリングに基づく文書の優先順位付けにより効率的に必要な文書を探すことが可能となり、特許関連文書のレビュー効率の向上が期待される

【問題点・課題】

AIによる学習精度の信頼性は未検証

5. 結論

- ・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及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。
- ・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。
例: 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
ホワイトスペース探索における分析支援
- ・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機能」などを利用した
効率化・分析の品質向上が必要とされると考える。

以上

2016年度
情報システム第2小委員会

ご清聴有難うございました

～世界から期待され、世界をリードするJIPA～