

日本語を処理する KIBIT の改良に成功

アジア言語処理の技術を強化

株式会社FRONTEO（本社：東京都港区、代表取締役社長：守本 正宏、以下 FRONTEO）は、独自開発のAIエンジン「KIBIT（読み：キビット）」において、自然言語処理における日本語処理の改良に成功したことをお知らせします。

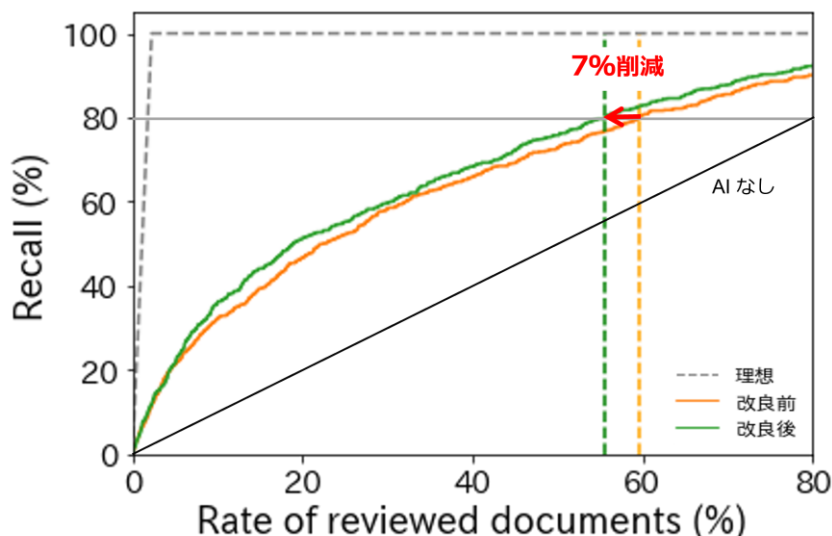
KIBITは、極めて少ない教師データで運用が可能な、弁護士や法務担当などの専門家の判断を支援するAIとして、2012年に開発されました。国際訴訟におけるeディスカバリ（電子証拠開示）では、証拠保持者（カストディアン）1人当たり数テラバイトに及ぶこともある膨大な文書データの中から、限られた期間内かつ調査に携わる弁護士の判断と同等の正確さで、証拠に関連する文書を発見することが求められます。KIBITを用いたAIレビューツール「KIBIT Automator（読み：キビットオートメーター）」は、この証拠発見の効率化に寄与するリーガルテックとして、国内外で活用されています。

日本語は、英語のように分かち書き（個々の単語間がスペースで区切られた表記）されず、また助詞などそれ単独では意味を持たない単語を伴うという特徴を持つ言語です。そのため、AIによる日本語の処理においては、文章を個々の単語などの形態素^{*1}に分解する技術（形態素解析）と、分解して得られた形態素を解析する技術の2つの技術を要します。さらに、後者の技術において、分かち書きをした際に抽出される1文字の形態素、例えば、「は」や「に」が何の品詞であるかを判断することは難しく、証拠への関連性の有無の判断にどの程度関係するのかを評価することが課題となっていました。

今回、FRONTEOの研究開発チームは、KIBITを構成するAIエンジン「Illumination Forest（読み：イルミネーションフォレスト）」において、1文字からなる単語を機械学習により自動的に取舍選択するようにアルゴリズムを改良しました。その結果、改良前と比べRecall^{*2}が向上し、証拠に関連する文書の80%を発見するのに必要な人がレビューする文書数の最大7%の削減を実現しました（図参照、FRONTEOのテストデータを使用）。

【改良前（日本語1文字の調整なし）と改良後（日本語1文字の調整あり）の比較】

日本語文書において証拠に関連のある文書の80%を得るのに必要なレビュー文書数を7%削減



本技術は、韓国語や中国語といった他の分かち書きされない言語にも適用が可能です。KIBITの強みの一つは処理が困難なアジア言語を得意とする点であり、今回の研究成果は、KIBITを搭載する幅広い製品群に一層の精度向上をもたらすものと見込まれます。

FRONTEOは今後も自然言語処理に強みを持つ独自のAIソリューションの高度化を推進し、専門家を支援するデジタルフォレンジック・ディスカバリベンダーとしてAIアルゴリズムの開発・改良に努めてまいります。

* 1 形態素：意味を有する最小の言語単位

* 2 Recall：再現率。証拠として関連性のあるすべてのデータの中で、関連性があることが正しく予測されたデータの割合。

■ **KIBITについて** URL: <https://www.fronteo.com/products/kibit/>

「KIBIT」は、専門家や業務熟練者が備える“暗黙知”を再現した独自の機械学習アルゴリズムを用い、キーワードに頼らずテキストを解析する人工知能です。高い自然言語処理技術を持ち、少量の教師データで短時間での高精度な解析が可能で、不正調査、知財戦略、技能伝承、危険予知等の分野で業務の効率化・高度化に貢献しています。

■ **KIBIT Automatorについて** URL: <https://legal.fronteo.com/products/kibit-automator/>

「KIBIT Automator」は、米国民事訴訟の公判手続きで必要となる証拠開示（ディスカバリ）の中でも特に、電子証拠開示（eディスカバリ）における文書レビュー作業の効率向上、作業担当者の負荷軽減、費用削減を目的として開発されたAIツールで、2019年3月にリリースされました。

ディスカバリで使われる調査手法を応用し、AIを活用して証拠資料である大容量の電子メールや電子ファイルの審査・分析を行います。近年、企業に求められている、短期での情報開示への対応も期待されます。

■FRONTEO について URL: <https://www.fronteo.com/>

FRONTEO は、自然言語処理に特化した自社開発 AI エンジン「KIBIT（読み：キビット）」と「Concept Encoder（商標：conceptencoder、読み：コンセプトエンコーダー）」、「Looca Cross（読み：ルーカクロス）」を用いて膨大な量のテキストデータの中から意味のある重要な情報を抽出し、企業のビジネスを支援する、データ解析企業です。2003年8月の創業以来、企業の国際訴訟を支援する「eディスカバリ（電子証拠開示）」や「デジタルフォレンジック調査」などのリーガルテック事業をメインに、日本、米国、韓国、台湾とグローバルに事業を展開してきました。同事業で培ったAI技術をもとに、2014年よりライフサイエンス分野、ビジネスインテリジェンス分野、経済安全保障へと事業のフィールドを拡大し、AIを用いて「テキストデータを知見に変える」ことで、創薬支援、認知症診断支援、金融・人事・営業支援など、様々な企業の課題解決に貢献しています。2007年6月26日東証マザーズ（現：東証グロース）上場。2021年1月第一種医療機器製造販売業許可を取得（許可番号：13B1X10350）、同9月管理医療機器販売業を届出（届出番号：3港み生機器第120号）。資本金3,034,846千円（2022年3月31日現在）。

※FRONTEO、KIBIT、KIBIT Automator、conceptencoder、Looca CrossはFRONTEOの日本における登録商標です。

<報道関係者のお問合せ先>

株式会社 FRONTEO 広報担当

Email: pr_contact@fronteo.com

<リーガルテック AI 事業に関するお問合せ先>

株式会社 FRONTEO リーガルテック AI 事業本部

Email: <https://legal.fronteo.com/contact/>