

2023年6月1日
株式会社スカイコム

スカイコムと宮崎大学が共同で行った「ドキュメントの内容解析をAIを用いて行う研究」で特許取得

株式会社スカイコム（本社：東京都千代田区、代表取締役社長：川橋 郁夫、以下 スカイコム）は、宮崎大学と2021年度より共同研究してきた、AIを用いて文書内容や文章構成を解析し、章立てや段組み、画像の配置などの書式をブロック化してラベリングする処理技術手法が特許を取得したことをお知らせいたします。

この技術を利用することで、様々なフォーマットで作成された電子文書を人手に頼らず解析して分類し、効率的に文書管理・運用を行うことが可能となります。

【特許内容】

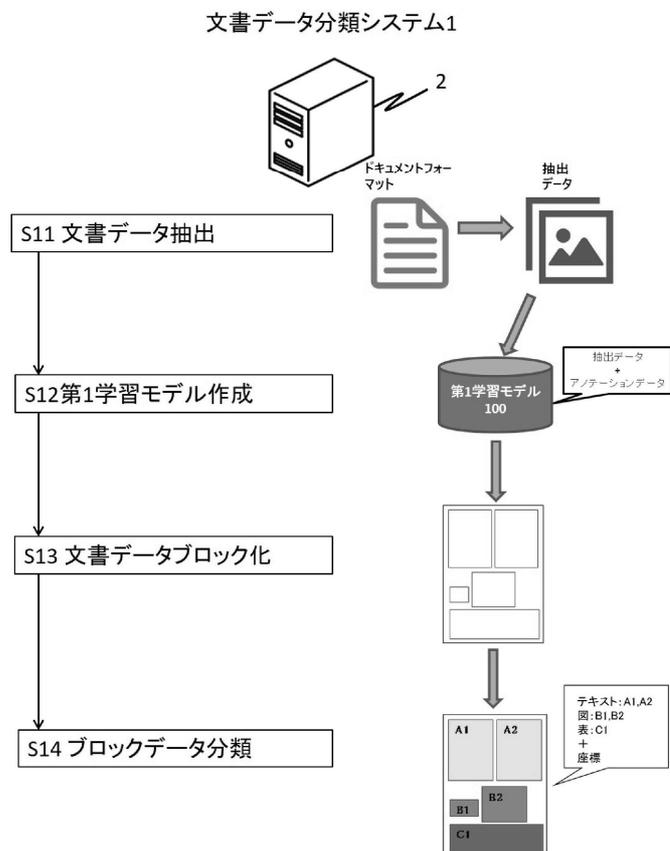
本発明の「文書データ分類システム」は、ドキュメントフォーマット*¹ からデータを抽出*² 後、分類する手段で、以下の3構成に分かれています。

■第1学習モデル作成手段

・ドキュメントフォーマットから抽出したデータにアノテーションデータ*³ を付与して、第1学習モデルを作成します。

■抽出したデータをブロックデータとしてブロック化手段

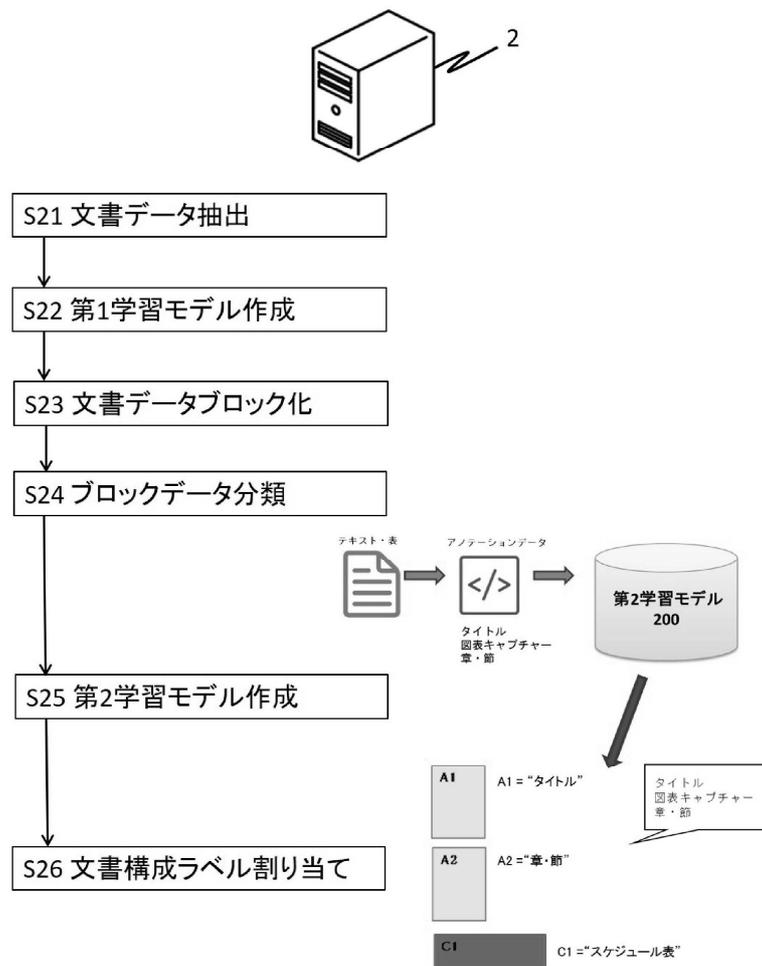
・第1学習モデルに基づいて抽出したデータをブロックデータとしてブロック化*⁴ します。さらにブロックデータをデータ種別毎に分類します。



■ブロックデータをデータ種別毎に分類する分類手段

- ・分類したブロックデータにアノテーションデータを付与して第2学習モデルを作成します。
- ・分類したブロックデータの文書要素がテキストの場合、内容を解析して章や節を検出して細分化します。表の場合も同様に解析し、表タイトルを検出します。
- ・第2学習モデルに基づいて、分類したブロックデータにラベルを割り当てます。

文書データ分類システム1



本発明によって、構成化されていない電子文書を解析結果に基づいて構成できるため、元の電子文書よりも内容の解析やラベルに基づいて、ユーザーが指定したキーワードに一致する文書を検索することができます。

- ※1 PDF、Word、PEG、PNG、GIF などのデータから構成されるファイル形式
- ※2 画像オブジェクト、表オブジェクト、テキストオブジェクトなどの要素
- ※3 画像、表、テキストなどの文書要素の種類や、サイズや座標などの配置情報
- ※4 電子文書を構成化に必要な文書要素に配置情報を関連付けてブロックデータを生成すること

【出願情報】

特許番号	特許第 7254314 号
発明名称	文書データ分類システム、文書データ分類方法及びプログラム
特許権者	株式会社スカイコム、国立大学法人 宮崎大学
出願番号	特願 2022-37668
出願日	令和 4 年 3 月 11 日
登録日	令和 5 年 3 月 31 日

・本文中に記載されている会社名、ブランド名および商品名は、すべて各所有者の商標または登録商標です。

株式会社スカイコムについて

- 社名 : 株式会社スカイコム
- 設立 : 1997 年 6 月
- 資本金 : 1 億円
- 代表者 : 代表取締役社長 川橋 郁夫
- 所在地 : 東京都千代田区神田松永町 19 秋葉原ビルディング 10 階
- ホームページ : <https://www.skycom.jp/>

選ばれる PDF メーカー、スカイコム。

『安全・安心な電子文書化社会の実現に貢献』をスローガンに掲げ、電子文書の世界標準フォーマットである「PDF」をコアに据えた PDF ソフトウェアプロダクト（ペーパーレス、業務改革推進のための製品・ソリューション）を提供しています。PDF 技術を自社内で開発する専門性と技術力で強みを発揮し、日本国内のみならず海外においても高い評価を得ています。

国際的な PDF 関連ベンダーの業界団体である PDF アソシエーションメンバーとして、これまで数多くの一般企業、金融機関、官公庁・自治体が利用するシステムやアプリケーションに当社の PDF テクノロジーを提供し、豊富な実績を築いています。

【本リリースに関するお問い合わせ】

株式会社スカイコム 営業本部 広報部 東、亀川

TEL: 03-5289-0780 FAX: 03-5289-0783

Email: pr@skycom.jp

©2023 SKYCOM Corporation., All Rights reserved.

本ニュースリリースに記載された内容は発表日現在のものです。その後予告なしに変更されることがあります。