生成AI活用のための非構造化データの構造化技術

Structuring Unstructured Data Utilizing Generative AI

田端　聡，西川高弘

要　約　大日本印刷株式会社（DNP）は，企業や個人が保有する非構造化データ（PDFや画像など）を構造化データに自動変換する技術の開発を進めている．本技術では，非構造化データから文書構造とレイアウトを自動解析し，タイトル，サブタイトル，本文，図版画像とそれに紐づくキャプションなどの利活用可能な構造化データ（XML形式のテキストと画像）に自動変換する．多種多様なレイアウトのドキュメントに対応し，未知のカテゴリのドキュメントに対しても数十枚の追加学習で対応することが可能である．本技術で得られた構造化データを検索拡張生成（RAG）に用いて生成AIに読み込ませることで，専門知識やノウハウを生成AIに効果的・効率的に追加できるようになり，生成AIの回答精度を大きく向上することが可能になる．

本稿では，この非構造化データの自動構造化技術と，構造化データを用いた検索拡張生成（RAG）に関する取り組みの背景，仕組み，現状の性能評価結果，今後の展望について紹介する．

Abstract　Dai Nippon Printing Co., Ltd. (hereafter, DNP) is developing technology that automatically converts unstructured data (PDF, images, etc.) held by companies and individuals into structured data. This technology automatically analyzes the document structure and layout of unstructured data and automatically converts it into more usable structured data (XML text and images), such as titles, subtitles, main text, and illustrations and their associated captions. It can handle documents with a wide variety of layouts, and it is also possible to handle documents in unknown categories with just a few dozen additional training examples. By using the structured data obtained from this technology for Retrieval-Augmented Generation (RAG) and loading it into the generative AI, it becomes possible to add specialized knowledge and expertise to the generative AI effectively and efficiently, and it also becomes possible to greatly improve the response accuracy of the generative AI.

In this paper, we will introduce the background of our efforts, the mechanisms, the current performance evaluation results, and future prospects about this technology for automatically structuring unstructured data and Retrieval-Augmented Generation (RAG) using structured data.

1. はじめに

大日本印刷株式会社（以下，DNP）は，企業や個人が保有する非構造化データ（PDFファイルなど）を構造化データに自動変換する技術（以下，DNPの構造化技術，あるいは，本技術）の開発（以下，本研究）を進めている．本技術で得られた構造化データを検索拡張生成（RAG）に用いて生成AIに読み込ませることで，専門知識やノウハウを生成AIに効果的・効率的に追加できるようになり，生成AIの回答精度を大きく向上することができる．

本稿では，まず，2章において，現在の生成AIの進展と社会実装の状況，関連する研究分

野を概観する．次に，3章でDNPの構造化技術の仕組みについて説明し，4章でその構造化データを用いた検索拡張生成（RAG）について述べる．5章において，構造化データを用いた検索拡張生成（RAG）の評価実験結果について述べる．最後に，6章で構造化技術と生成AIの今後の展望について述べる．

2. 生成AIの進展と現状，関連研究

2.1 生成AIの進展と現状

OpenAI社の対話型AI「ChatGPT」[1]の公開を契機に，生成AIブームが巻き起こっている．現在，世界中の企業や研究機関がさらに高度なAI技術の研究開発，社会実装のための応用研究開発に力を注いでおり，日々新たな技術や製品が生み出され，世の中を賑わせている．

生成AIの登場により，これからの人々の生活，および，労働市場に大きな影響が及ぶと予想されている．OpenAI社の分析によると，GPT3.5の段階でも80%の労働者が業務の10%で，19%の労働者が業務の50%で影響を受けるとされている[2]．また，現在のGPTの最新版であるo1は，法律，医療，プログラミングなどの専門分野を含め，様々なタスクで人間レベル以上の卓越した性能を示し，多くの知的労働を代替できる可能性が示されている．また，様々な分野にまたがる複雑な推論や感情認識などの能力も備えており，近い将来には汎用人工知能（AGI）の到来が期待されている[3]．

また，画像，音声，動画など複数のデータ形式を同時に処理できるようなマルチモーダル生成AI[4]や，生成AIに自律性を持たせて多様で複雑なタスクを自律的に処理できるエージェント技術[5]の研究開発も盛んに行われており，近く実用化が期待されている．

しかし現在，世の中の生成AIの盛り上がりとは対照的に，社会実装，ならびに，業務や生活への浸透は，特に国内では思うように進んでいないのが実情である．生成AIの社会実装の進みが遅い原因の一つとして，現行の生成AIは，インターネット上にある汎用的な知識やスキルしか保有していないという点が挙げられる．そのため，社会活動や企業活動に適用するには，インターネット上の汎用知識の他に，各企業や個人だけが持つ専門知識や能力を追加で与える必要がある．

生成AIに新たに知識を追加するためには，一般的に検索拡張生成（RAG）[6]と呼ばれる技術を使うことが多い．検索拡張生成（RAG）に関しては，2.2節で説明する．

また，生成AIに知識を追加するにあたり，現行の企業では，社内文書やマニュアル，手順書などの知的資産が非構造化データとして保存されており，デジタル化や構造化がされておらず簡単に活用することができないという問題がある．

文書画像から意味構造を認識する目的で，Document Layout Analysis（DLA）[7]という技術が長年にわたり研究されてきている．DNPの構造化技術もこの領域に属する技術である．2.3節で，Document Layout Analysis（DLA）の分野におけるこれまでの研究の流れを概観する．

2.2 検索拡張生成（RAG）

検索拡張生成（RAG）とは，生成AIが学習していない外部知識を，検索を用いて回答させる手法である．RAGの仕組みを図1に示す．

RAGでは，まず外部ドキュメントの情報を細かい単位，いわゆる「チャンク」に分割する．その後，分割された各チャンクにベクトルを付与し，それをベクトルデータベースに登録する．

ユーザーが質問をするときは，まずその質問内容をクエリに変換し，そのクエリに関連する情報をベクトルデータベースから検索する．そして，設定されたプロンプトと検索結果を大規模言語モデル（LLM）に入力し，それを基に回答を生成する．

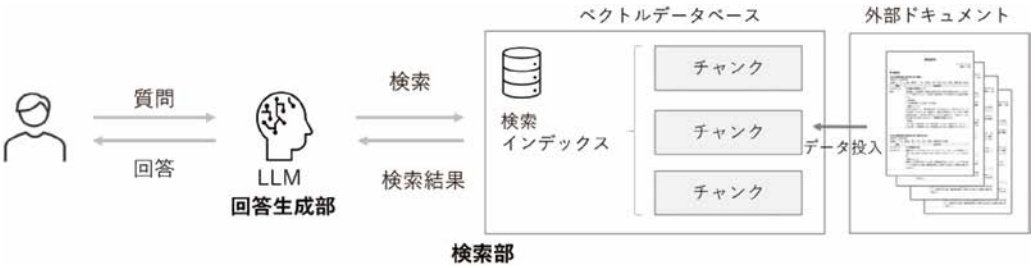

図 1　検索拡張生成（RAG）のしくみ

近年では，生成 AI の発展とともに RAG についても研究が進められている．Self-Reflective Retrieval-Augmented Generation（Self-RAG）[8]では質問文に対して，検索が必要かどうか判断させ回答を生成させる手法や，検索部分の結果が不十分であればインターネットから検索を行い検索結果を拡張させる Corrective Retrieval Augmented Generation（CRAG）[9]など様々な手法が検討されている．また，外部ドキュメントのデータの保持方法としてベクトルデータベースではなく，知識グラフを用いた Graph RAG[10]が提案されている．このような新しいアーキテクチャが提案される一方で，RAG においては質問に関連する情報の検索とその検索結果を基にして回答が生成されるため，適切なチャンクの分割が求められる．

従来のチャンク単位の分割手法として，特定の文字数やトークン数に基づき分割する手法があげられる．しかし，これらの手法では文章の意味の切れ目に基づかない分割が行われるため，質問に関連しない情報が含まれたり，質問に関連する内容が途中で分割されたりする可能性がある．その結果，余分な情報や必要なものが欠けた情報が回答生成部に入力され，質問に関連する情報の検索や回答の精度が低下する．

本研究では，3 章で述べる DNP の構造化技術を用いて文書の意味構造を反映させた構造化データを生成し，適切なチャンクの分割を行い，RAG で読み込ませることにより，回答精度の改善を図る．

2. 3　Document Layout Analysis（DLA）の研究の流れ

Document Layout Analysis（DLA）は，文書画像から意味構造を認識・理解する研究領域であり，長年にわたり研究開発が行われている．近年では CNN や Transformer などの Deep Learning 手法を用いた手法が数多く提案されている．

文書画像の意味構造認識を，CNN の物体検出タスクとして定義する手法[11]や，CNN の意味セグメンテーションタスクとして位置づけ文書構造の推論を行う手法[12]がある．さらに，Transformer を用いて，画像情報とテキスト情報をマルチモーダルで入力し，物体検出タスクとして推論を行う手法[13][14]もある．また，最近の事例として，Transformer モデルを用いて，テキスト領域検出，論理的役割分類，読み順予測などを関係予測問題として推論を行う手法[15]が，興味深い結果を示している．

しかし，これらの研究は，CNN や Transformer を用いて文書画像の意味構造を認識できる可能性を示しているものの，再利用可能な形式として内容を取得できるまでには至っていない．加えて，公開されているデータセットが学術論文や帳票などを中心とした偏ったものである影響もあり，雑誌，カタログ，社内文書やマニュアル，プレゼン資料など，多種多様かつ複雑なレイアウトに対応するのが難しいという問題がある．また，同様の理由で，日本語文書特有の縦書き，ルビ，割注といった特殊な形態を持つ文書に対応するのが難しい．

本研究では，既存の研究例のアプローチを参考にしつつ，小領域ドメインモデルによるアンサンブル戦略を用いて多種多様な文書に対応させるとともに，文書画像の意味構造をより深く認識・理解し，内容を再構成することで，再利用しやすい構造化データ（XML 形式のテキストと画像）を生成する．この構造化データを用いることで，検索拡張生成（RAG）をはじめとした人手作業が必要な非構造化データの再利用を大幅に効率化することが可能になる．

3. 非構造化データの構造化技術

本章では，DNP の構造化技術について説明する．構造化処理の概要を図 2 に示す．この処理では，非構造化データである印刷物のデータ（PDF）を入力すると，文書構造とレイアウトを自動解析し，タイトル，サブタイトル，本文，図版画像とそれに紐づくキャプションなど，利活用可能な構造化データ（XML 形式のテキストと画像）に自動変換する．文書内の要素となるオブジェクトおよび属性などの XML データ構造は，所定の形式[16]で定義する．

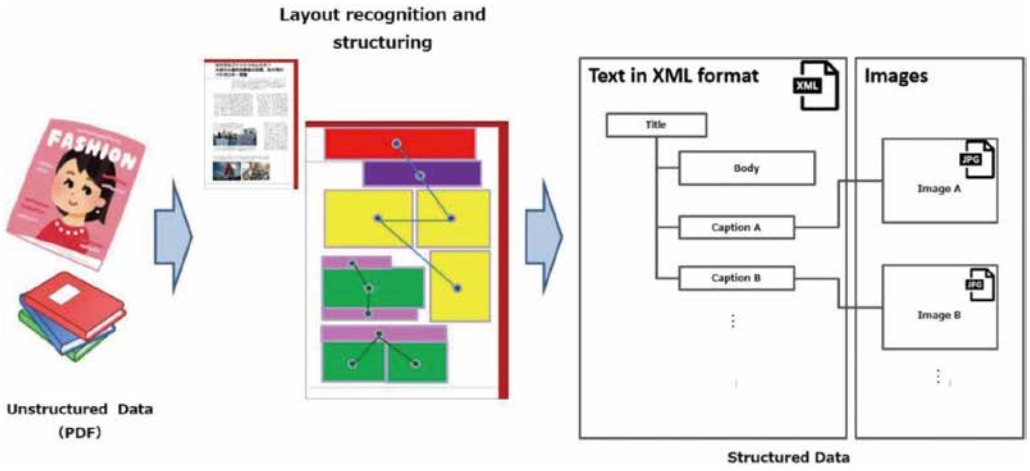

図 2　構造化処理の概要

3. 1　構造化の処理フロー

構造化の処理フローを図 3 に示す．図中の（3.2）～（3.7）は，本章の節番号に対応する．

対象となる非構造化データが入力されると，最初に，自動モデル選択（3.2）フェーズにて，入力データに最適なレイアウト認識モデルがモデルリストの中から選択される．

その後，選択したモデルを用いて，レイアウト抽出（3.3），テキストボックスの読み順推定（3.4），図表キャプション対応推定（3.5），テキスト再構築（3.6）の順番で処理を行う．

なお，入力は PDF データとし，テキストはあらかじめ抽出されているものとする．スキャ

ンされた文書画像等については，市販の OCR エンジンを用いてあらかじめテキスト抽出しておく必要がある．

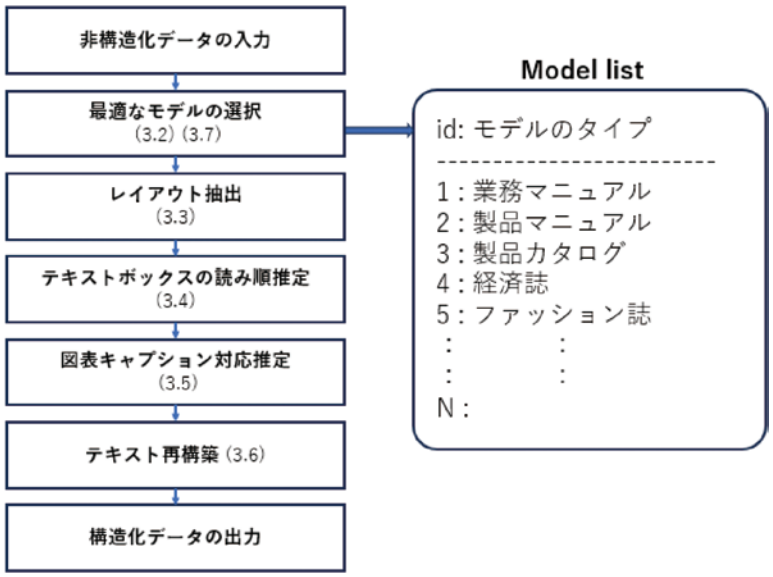

図 3　構造化の処理フロー

3. 2　最適なモデルの選択

このフェーズでは，入力データに最適なレイアウト認識モデルをモデルリストの中から選択する．モデル選択処理の概要を図 4 に示す．レイアウト認識モデルは，小領域ドメインのモデルによるアンサンブル戦略を用いて構築されており，業務マニュアル，製品マニュアル，製品カタログ，経済誌など，それぞれ数十枚からの教師データを用いて各ドメインに特化した形で学習されたモデルのリスト形式の集合体である．

入力データと各モデルの適合度を推定し，最も入力データに適合するモデルを選択して後段の処理で使用する．各レイアウト認識モデルの詳細は 3. 3 節で，モデル適合度推定モデルの詳

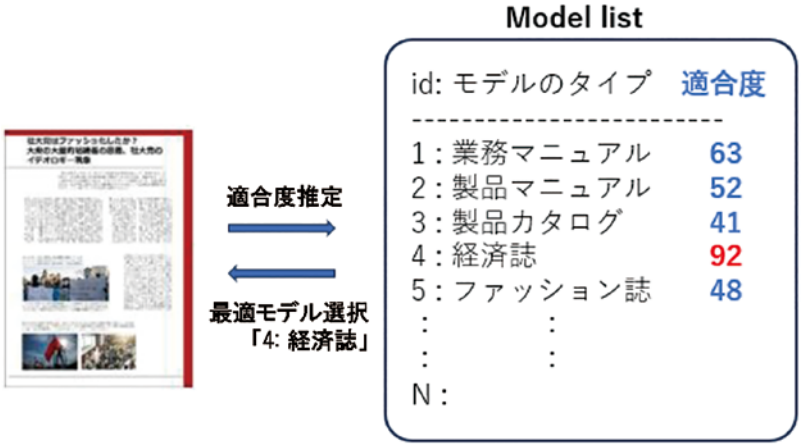

図 4　モデル選択処理の概要

細は 3.7 節で説明する．

3.3　レイアウト抽出

このフェーズでは，選択したレイアウト抽出モデルを用いて文書のページ画像を構造的および論理的な単位（タイトル，テキスト，図，キャプションなど）に分解し，各領域にタグ付けを行う．レイアウト抽出のモデルの構成を図 5 に示す．

レイアウト抽出モデルは，Transformer[17]ベースの Document Encoder と CornerNet[18]ベースの Structure Detector で構成されている．

我々は，文献[11]と同様に，物体検出タスクとして捉えることにした．理由としては，特に雑誌やカタログのような複雑なレイアウトの文書を取り扱う場合，セマンティックセグメンテーションのようにピクセル毎のカテゴリを推論するのではなく，物体検出タスクのようにテキスト領域，図領域などのインスタンスを同定することが重要だからである．また，物体検出ロジックに CornerNet ベースのアプローチを用いた理由は，アンカーフリーであるため，様々なアスペクト比やスケールを持つ文書オブジェクトを扱いやすいこと，CornerNet で使用されるオブジェクトのコーナー情報が文書オブジェクトにとって捉えやすく，抽出精度が高いことである．

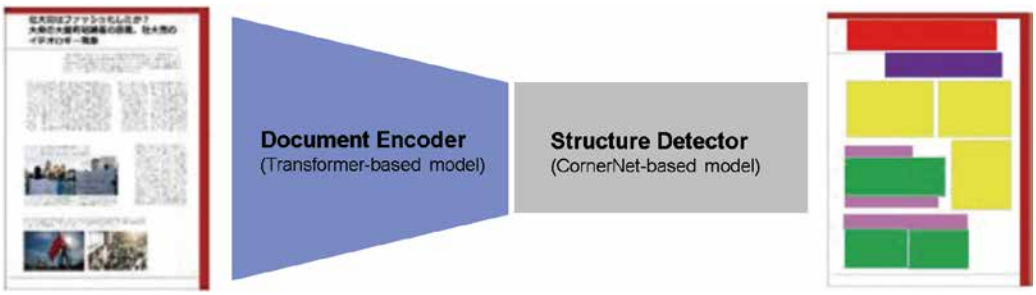

図 5　レイアウト抽出モデルの構成

3.4　テキストボックスの読み順推定

このフェーズでは，タグ付けされたテキストボックスの読み順を推定する．図 6 に処理の概要を示す．

一般的にテキストボックスの読み順には一定のルールがある．例えば，横書きの文書では左上から始まり右下で終わり，縦書きの文書では右上から始まり左下で終わる．しかし，複数段組レイアウトを考慮した場合，正しく読み順を推定するためには，段組の構造を推定する必要がある．

図 6 に示すように，本文内の文字の座標情報のセットを入力し，横書きの文書の場合は X 軸，縦書きの文書の場合は Y 軸に沿って 1 次元 Mean Shift Clustering 処理を行う．これにより，文字位置の密度を計算し，密度が山となる位置，数を算出することで，段組構造を推定することができる．段落構造を推定した後は，その同じ段組み内で，テキストボックスの順序をルールに従って決定する．

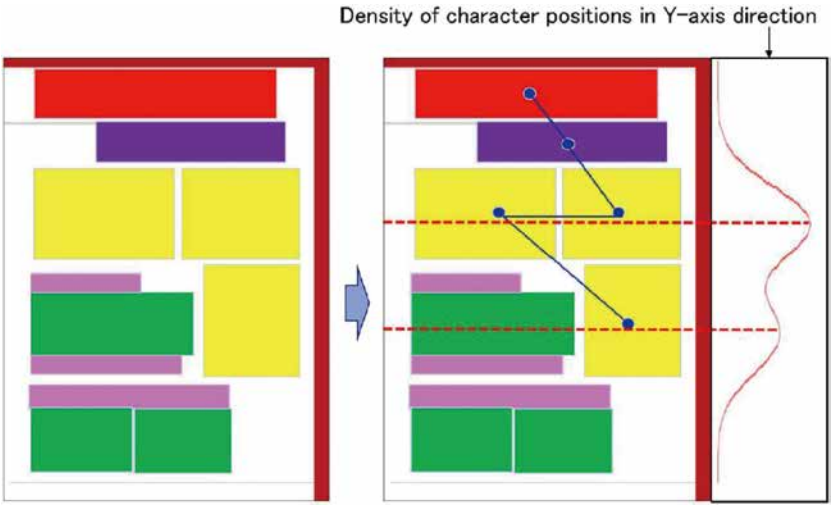

図 6　段組みの推定処理（縦書き 2 段組みの例）

3. 5　図表キャプション対応推定

このフェーズでは，図とキャプションの対応関係を推定する．処理の概要を図 7 に示す．

図とキャプションの対応関係は，基本的にはそれぞれ最も近いもの同士を結びつけるというルールが一般的である．しかし，ファッション雑誌やカタログなど，図を多く含む複雑なレイアウトの文書では，このルールを適用すると精度が大幅に低下することが検証の中でわかった．

図 7 に示すように，2 値画像「セマンティックマップ」を入力として，図がキャプションに該当するかどうかを 2 値で判別する CNN モデル（ResNet[19]）を学習して使用する．アブレーションの結果，CNN モデルを導入することにより，単純なルールを用いた場合と比較して対応関係の正答率が 23.1％向上した．

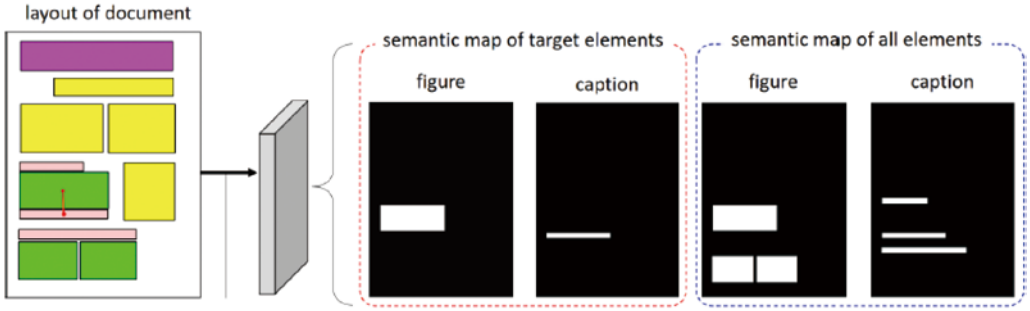

図 7　図表キャプション対応推定

3. 6　テキスト再構築

レイアウト抽出，テキストボックスの読み順推定，および，図表キャプション対応推定を行った後，テキストを連結・整形して XML フォーマットの出力データとして，再構築する．

タイトル，見出し，本文，リード文，などテキストのコンテンツについては，テキストボックスの順序通りに連結し，パラグラフごとにタグ付けを行い，一つの XML 要素として構築す

る．図や表については，それぞれがトリミングされ特定のフォルダに保存される．その後，関連するキャプションと画像パスを対応付け，XML 要素として構築する．

3. 7　モデル適合度推定モデルの詳細

本節では，3.2 節「最適なモデルの選択」で説明したモデル適合度推定モデルの詳細と学習方法について説明する．図 8 にモデル適合度推定モデルの構成を示す．

レイアウト抽出の AI モデルネットワークに対して，構造化モデルの適合度を予測するブランチを設け，MLP ベースのモデルを配置し，モデルの適合度を学習・予測する．

構造化モデルの適合度は，構造化結果の良し悪しを示す指標であるとともに，レイアウト抽出結果の mAP 値の回帰予測問題として定義する．適合度推論の回帰ネットワークの入力には，構造化 AI モデルネットワークのエンコーダの出力値と構造検出器の 1 層目の値を Global Average Pooling で正規化した中間ベクトル（256 次元）を用いる．この中間ベクトル（256 次元）は，構造化を実行する上でドキュメントの概念化・抽象化されたベクトルであるとみなし，モデル適合度の学習，推論に使用する．

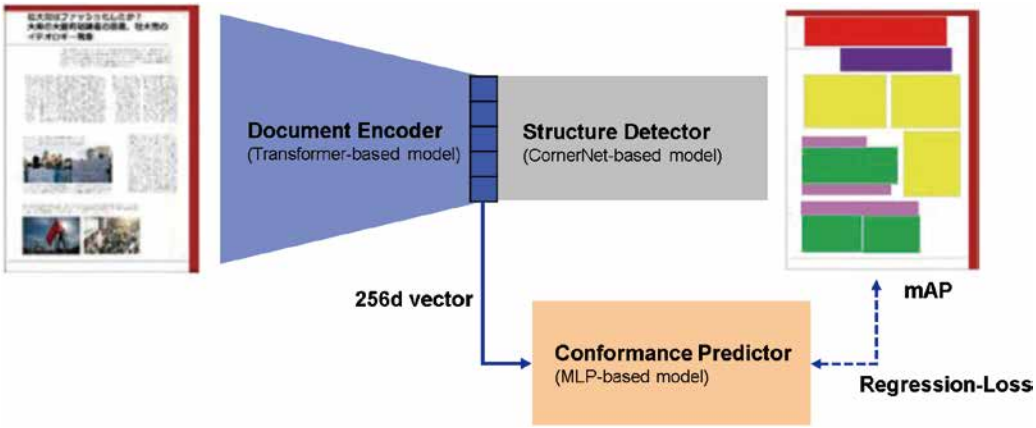

図 8　モデル適合度推定モデルの構成

3. 8　構造化処理の結果

図 9 に構造化処理の中でレイアウト抽出処理の結果を可視化した例を示す．サンプルとして，DNP の 2023 年 3 月期の決算短信[20]を使用する．レイアウト抽出処理を実行すると，文書の中で，見出し，本文，表などの各オブジェクトのカテゴリと領域が認識され，それぞれ異なる色で重畳表示される．

図 10 に構造化された XML ファイルの例を示す．タイトル，見出し，本文などテキストのコンテンツについては，パラグラフごとに一つの XML 要素として出力される．一方，図や表については，それぞれがトリミングされ特定フォルダに保存される．その後，関連するキャプションと画像パスを対応付け，XML 要素として出力される．

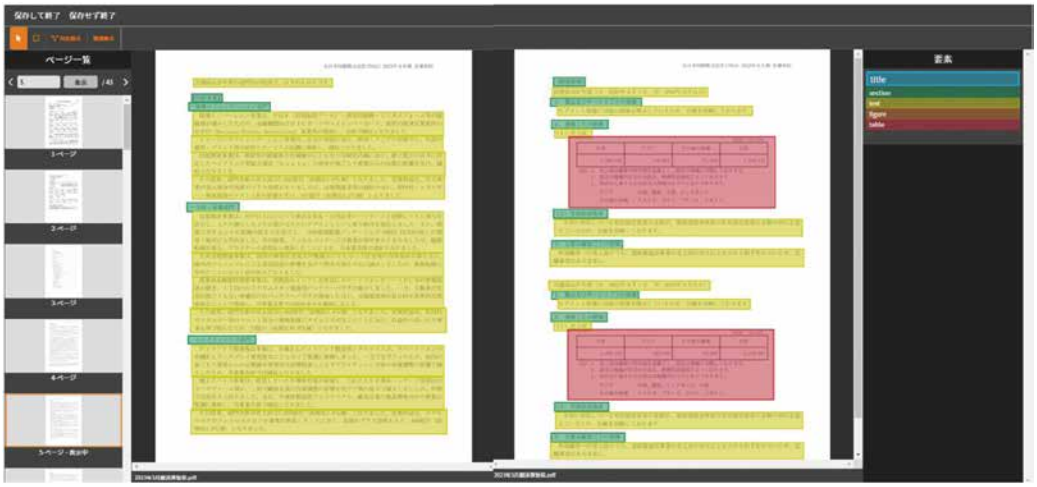

図 9　レイアウト抽出結果

```
<structural-data>
  <filename>00001.pdf</filename>
  <title>2023年3月期決算短信</title>
    :
  <section_1>【飲料事業】</section_1>
    <section_2>飲料部門 </section_2>
      <text>原材料価格や物流コストの上昇の影響にともない、大型PETボトル商品や小型
            パッケージ商品などの価格改定を実施しました。また、…</text>
      <text>その結果、部門全体の売上高は、コンビニエンスストアでの販売が回復したほか、飲
            料展やネット販売の伸長もあり、516億円（前期比3.8%増）となりました。…</text>
      :
  <section_1>【飲料事業】</section_1>
      <text>原材料価格や物流コストの上昇の影響にともない、大型PETボトル商品や小型
            パッケージ商品などの価格改定を実施しました。また、…</text>
      :
    <table>
      <image>table1.png</image>
      <caption>連結貸借対照表</caption>
    </table>
      :
    <figure>
      <image>figure1.png</image>
      <caption>事業系統図</caption>
    </figure>
      :
</structural-data>
```

図 10　構造化 XML ファイルの例

4.　構造化データを用いたチャンク分割手法

本章では，構造化データの活用例として，ドキュメントを構造化したデータを RAG 用のデータとして活用する方法について述べる．構造化データは，タイトル，見出しや本文などのタグ情報を含んでいる．それらの情報を参照し，チャンク単位に分割してベクトルデータベースに

登録する．構造化データを RAG 用データに変換する例を図 11 に示す．

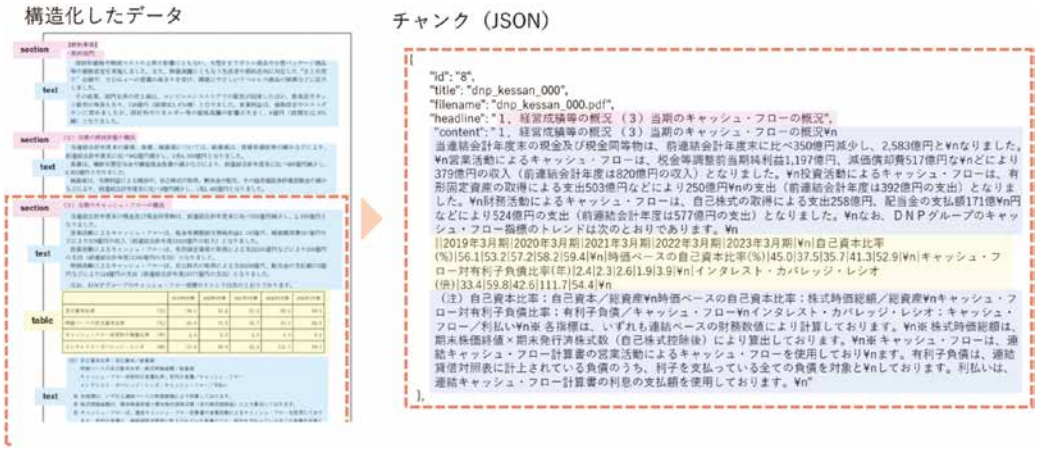

図 11　構造化データのチャンク分割の例

図 11 の例は，DNP の 2023 年 3 月期の決算短信[20]の一部である．チャンクの情報は JSON 形式で保持される．対象とするドキュメント全体に対して見出し単位でチャンクに分割し，作成されたスキーマの集合をベクトルデータベースに登録する．このように分割することで，文章の意味の切れ目を考慮しながら細分化された単位でデータベースに登録することができる．

5.　実験

本章では，構造化ありのデータと構造化なしのデータを用いた RAG の比較実験について述べる．

5. 1　比較実験の概要

構造化されたデータと文字数単位でチャンク分割された構造化されていないデータをそれぞれベクトルデータベースに登録し，それぞれの検索結果と RAG による回答結果を比較する．

検索の手法は，Azure AI Search のハイブリッド検索[21]を用いる．ハイブリッド検索は，BM25 によってランク付けされたキーワード検索の結果と，コサイン類似度によりランク付けされたベクトル検索の結果を，Reciprocal Rank Fusion（RRF）を用いて統合し，最終的な検索結果を求める手法である．キーワード検索を行う際には，質問文から GPT-3.5 を用いてキーワードの抽出を行う．また，ベクトル検索のためのエンベディングモデルは text-embeddings-ada-002[22]を用いる．

対象とするドキュメントとして，DNP の 2023 年 3 月期の決算短信[20]を用いる．実験に使用する質問は DNP の決算短信[20]を基に GPT-3.5 が生成したものを使用する．質問を生成するにあたり，GPT-3.5 に参照した箇所を明示させ，それを検索結果の正解と定義する．生成された質問文の中に誤りがあるものは事前に除外する．結果として，GPT-3.5 が生成した 500 件のデータから誤りを除外した後，残った 434 件のデータを検証に使用する．

生成された質問を用いて構造化ありのデータと構造化なしのデータの検索結果を定量評価し，RAG の結果として生成された回答の品質を定性評価して比較する．検索結果の評価指標

は，top-5 accuracy（チャンクの検索結果の上位 5 件の中に質問生成時に使用したドキュメントの部分が含まれていれば正解，含まれていなければ不正解）を用いる．

5. 2　比較実験の結果

構造化ありのデータと構造化なしのデータの検索結果の定量評価の比較を表 1 に示す．テキスト部に対する質問の検索結果は 11.14 ポイント，表に対する質問は 5.54 ポイント，全体の検索結果は 8.51 ポイント，それぞれ構造化なしの結果を上回った．また，RAG の結果の定性評価にて，回答で差異が見られた例を図 12 に示す．

表 1　構造化ありと構造化なしの検索結果の比較（top-5 accuracy）

	構造化あり	構造化なし
テキスト部に対する質問	95.89%	84.75%
表に対する質問	93.72%	88.18%
全体	**95.20%**	**86.69%**

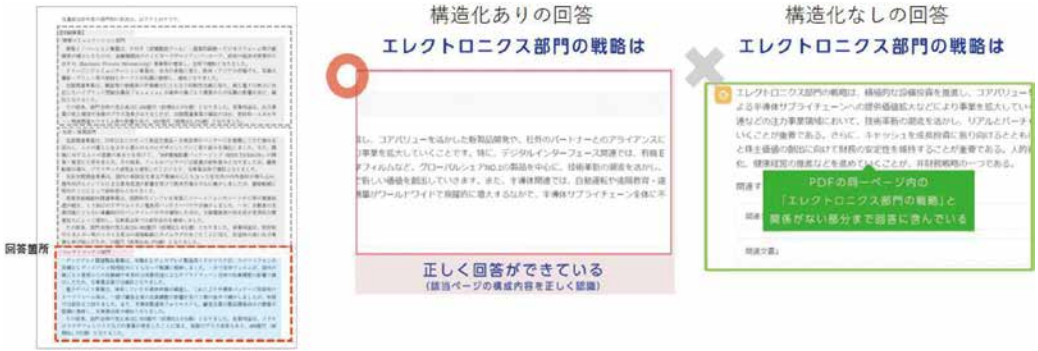

図 12　構造化ありと構造化なしの回答比較

図 12 に示す「エレクトロニクス部門の戦略は」という質問に対して，構造化ありのデータは見出しに基づいてチャンクの分割を行っているため，エレクトロニクス部門に関連する箇所のみを参照して正確に回答することができた．一方，構造化なしのデータは文章の意味の単位で区切られていないため，関連する内容以外の情報を含んだ回答が生成される結果となった．

比較実験の結果，構造化ありのデータが構造化なしのデータに対して優れた検索結果を示すことが確認できた．また，回答結果においても文章の意味の切れ目を考慮したチャンクの分割を行うことにより回答の精度が向上することが確認できた．

6. お　わ　り　に

本稿では，DNP で開発を進めている非構造化データの自動構造化技術と，構造化データを用いた検索拡張生成（RAG）について紹介した．本技術で得られた構造化データを検索拡張生成（RAG）に用いて生成 AI に読み込ませることで，専門知識やノウハウを生成 AI に効果的・効率的に追加して生成 AI の回答精度を大きく向上させ，人と AI のコミュニケーションを円滑にすることが可能になる．

今後の展望として，本技術をさまざまな業界や分野に適用することで，企業や個人に蓄積されている知識やノウハウの有効活用および具体的な問題解決に貢献していくとともに，より高度な構造化ができるように本技術を進化させようと考えている．具体的には，ドキュメント内の概念図やフローチャートの詳細な理解，動画や音声の構造化などに取り組んでいる．

現在，マルチモーダル生成 AI の進歩が著しく，それらのデータに関してもある程度の意味理解が可能になっている．しかし，現段階では性能・精度的に実用上十分ではない．それらのデータを AI に正しく理解させるためには，インターネット上の汎用知識だけでは不十分であり，企業や個人が保有しているドメイン知識や専門知識が必要で，それらを構造化して体系的に入力することが不可欠であると考えている．

今後も DNP は，データ構造化技術を核として AI の社会実装を推し進め，近い将来に訪れるであろう人と AI が真に共生する社会において，人と AI のコミュニケーションを円滑化し，より豊かで持続可能な未来を築いていくことを目指していきたい．

参考文献

［1］ ChatGPT, OpenAI, https://chat.openai.com/

［2］ Eloundou, Tyna, et al. “Gpts are gpts: An early look at the labor market impact potential of large language models.” arXiv preprint arXiv:2303.10130 (2023).

［3］ Zhong, Tianyang, et al. “Evaluation of openai o1: Opportunities and challenges of agi.” arXiv preprint arXiv:2409.18486 (2024).

［4］ Yin, Shukang, et al. “A survey on multimodal large language models.” arXiv preprint arXiv:2306.13549 (2023).

［5］ Wang, Lei, et al. “A survey on large language model based autonomous agents.”, Frontiers of Computer Science 18.6 (2024): 186345.

［6］ Lewis, Patrick, et al. “Retrieval-augmented generation for knowledge-intensive nlp tasks.” Advances in Neural Information Processing Systems 33 (2020): 9459-9474.

［7］ Binmakhashen, Galal M., and Sabri A. Mahmoud. “Document layout analysis: a comprehensive survey.” ACM Computing Surveys (CSUR) 52.6 (2019): 1-36.

［8］ Akari Asai, et al. “Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection” arXiv preprint arXiv:2310.11511 (2023).

［9］ Yan, Shi-Qi, et al. “Corrective retrieval augmented generation.” arXiv preprint arXiv:2401.15884 (2024).

［10］ Darren Edge, et al. “From Local to Global: A Graph RAG Approach to Query-Focused Summarization” arXiv preprint arXiv:2404.16130 (2024).

［11］ Li, Kai, et al. “Cross-domain document object detection: Benchmark suite and method.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

［12］ Yang, Xiao, et al. “Learning to extract semantic structure from documents using multimodal fully convolutional neural networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

［13］ Li, Junlong, et al. “Dit: Self-supervised pre-training for document image transformer.” Proceedings of the 30th ACM International Conference on Multimedia. 2022.

［14］ Huang, Yupan, et al. “Layoutlmv3: Pre-training for document ai with unified text and image masking.” Proceedings of the 30th ACM International Conference on Multimedia. 2022.

［15］ Wang, Jiawei, Kai Hu, and Qiang Huo. “DLAFormer: An End-to-End Transformer For Document Layout Analysis.” International Conference on Document Analysis and Recognition. Cham: Springer Nature Switzerland, 2024.

［16］ Sou Tabata, Haruka Maeda, Keigo Hirokawa, Kei Yokoyama. 2019. “Diverse Layout Generation for Graphical Design Magazines.” In ACM SIGGRAPH Asia 2019. 21:1-2

［17］ Dosovitskiy, Alexey. “An image is worth 16x16 words: Transformers for image recognition at scale.” arXiv preprint arXiv:2010.11929 (2020).
［18］ Law, Hei, and Jia Deng. “Cornernet: Detecting objects as paired keypoints.” Proceedings of the European conference on computer vision (ECCV). 2018.
［19］ He, Kaiming, et al. “Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
［20］ 2023 年 3 月期 決算短信〔日本基準〕(連結)，大日本印刷株式会社, https://www.dnp.co.jp/ir/ir-file/pdf/dnp_FY22Q4fin.pdf
［21］ Hybrid search using vectors and full text in Azure AI Search - Azure AI Search | Microsoft Learn, https://learn.microsoft.com/en-us/azure/search/hybrid-search-overview
［22］ New and improved embedding model - OpenAI, https://openai.com/ja-JP/index/new-and-improved-embedding-model/

※　上記参考文献に含まれる URL のリンク先は，2025 年 1 月 24 日時点での存在を確認.

執筆者紹介　田　端　　聡（Sou Tabata）
2005 年に大日本印刷(株)に入社．画像処理技術，AI，機械学習技術の研究開発とそれらを応用したアプリケーション，システムの研究開発に携わる．画像シミュレーションシステム，フォトプリント画像補正技術，デジタルサイネージ技術，デジタルアーカイブ技術，雑誌レイアウト自動生成 AI などの研究開発に従事．現在は，構造化技術の研究開発に従事．主幹研究員.

西　川　高　弘（Takahiro Nishikawa）
2023 年大日本印刷(株)に入社．AB センターにて自然言語処理．生成 AI を用いたサービスの研究開発に従事.