

PRESS RELEASE
報道関係各位

2025年7月23日
株式会社セゾンテクノロジー

セゾンテクノロジー、社内データを「AI Ready」化する スクリプトテンプレートを「HULFT Square」で提供開始

～生成 AI が理解しやすいデータへの事前加工を省力化、
回答精度も最大 90%まで向上～

株式会社セゾンテクノロジー（本社：東京都港区、代表取締役 社長執行役員：葉山 誠、以下セゾンテクノロジー）は、生成 AI の RAG（検索拡張生成）*1 として読み込む社内データを事前加工する「HULFT Square」向けデータ連携スクリプト、「AI 前処理テンプレートシリーズ」10 種類を、7 月 23 日より順次提供開始します。

本テンプレートを利用することにより、スプレッドシート、マニュアル、スキャン PDF、画像データ、音声データ、HTML など、社内存在するさまざまな業務データから、タグなどの不要なノイズを除去したテキストの抽出や、非構造化データの構造化データへの変換、およびデータに適切なりレーションなどを付与します。社内データを「HULFT Square」から生成 AI を経由して、生成 AI が参照しやすいデータに事前加工することで生成 AI の回答精度も最大 90%*2 まで向上し、データ加工にかかる作業工数も約 50～60%程度省力化*3 します。

背景

近年、企業における生成 AI の活用が進んでおり、大規模言語モデル（LLM）に新たに社内データの情報を組み合わせ、生成 AI の検索精度を向上させる RAG の手法が広がっています。一方で、社内データを生成 AI に参照させたものの、当初期待したよりも回答の精度が得られていないことが多いのも実情です。

社内データを生成 AI に最適な「AI Ready」なデータとするためには、生成 AI が理解しやすい構造に加工したり、データに適切な意味づけを付与したりするなど、エンジニアがデータの種類に応じてさまざまな手法で事前にデータを前処理する必要があります。このデータ加工処理に必要なノウハウや作業工数の確保が、社内データを迅速に生成 AI で活用し、高い回答精度も得るための課題となっています。

「AI 前処理テンプレートシリーズ」の概要

生成 AI に最適なデータに加工する「AI 前処理テンプレートシリーズ」は、データ連携プラットフォーム（iPaaS）「HULFT Square」上で利用できるデータ連携スクリプトである「HULFT Square アプリケーション」として提供されます。

社内データの種別に応じて 10 種類の「AI 前処理テンプレートシリーズ」を利用することで、「HULFT Square」経由で生成 AI とデータを連携させ、テキスト抽出やデータ変換、データの意味づけなどの事前加工の作業を生成 AI 自体に任せることで、RAG としてデータを読み込むために必要な生成 AI が理解しやすい状態に社内データを前処理することが可能です。

「AI 前処理テンプレートシリーズ」の特長と種別および処理内容の概要は次の通りです。

1. さまざまな社内データを生成 AI が理解しやすい状態に事前加工し回答精度が最大 90%^{*2}まで向上
2. ノウハウが必要なデータ加工の前処理を HULFT Square 経由で生成 AI 側が実行し加工作業を 50～60%程度^{*3}省力化

■回答精度の向上

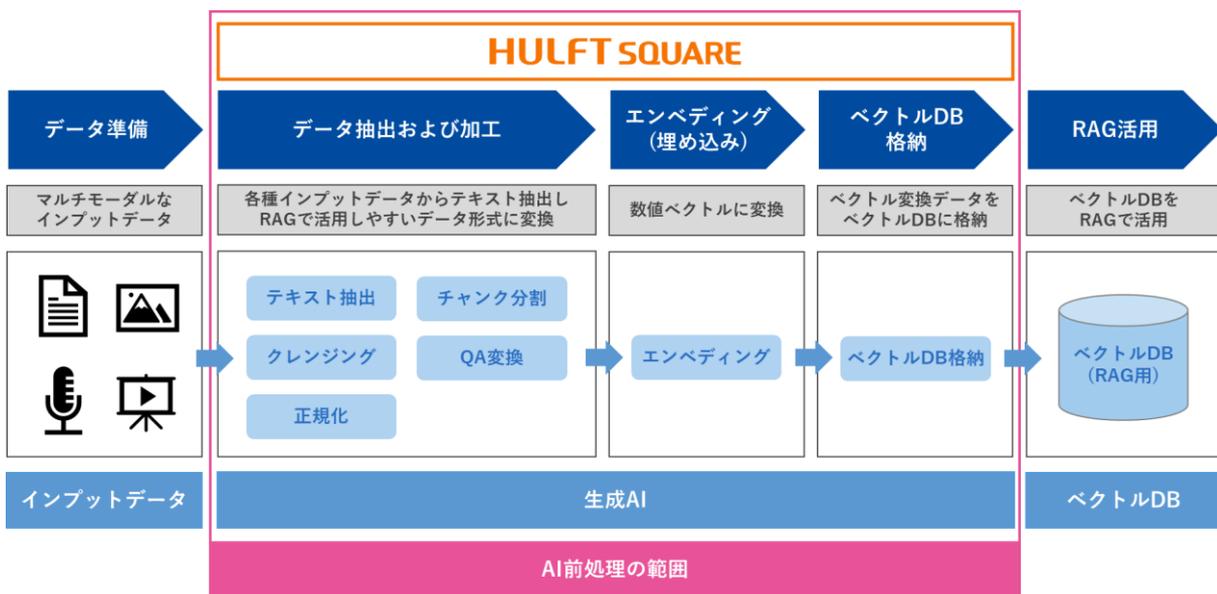
テンプレート種別	テンプレート名	効果とデータ処理の概要	提供開始
QA 形式への変換	AI 前処理 PDF から QA 表作成	マニュアル類の PDF を非構造化データから QA 形式に変換した構造化データとして CSV 形式で出力	2025 年 7 月 23 日
	AI 前処理 Excel から QA 表作成	Excel の表形式を QA 形式に変換した構造化データとして CSV 形式で出力	2025 年 9 月予定
	AI 前処理 JSON から QA 表作成	EC サイトなど外部システムの JSON 形式の複雑な階層構造のデータを QA 形式に変換した構造化データとして CSV 形式で出力	2025 年 9 月予定
	AI 前処理 XML から QA 表作成	階層構造やタグが複雑な XML を QA 形式に変換した構造化データとして CSV 形式で出力	2025 年 9 月予定
クレンジング ^{**4}	AI 前処理 HTML タグの削除	企業 Web サイトや社内ポータルサイトなどの HTML ドキュメントからタグを除去しマークダウン記法で出力	2025 年 8 月予定
	AI 前処理 特殊文字・記号の削除	HTML ドキュメントから特殊記号や記号を削除しテキスト形式で出力	2025 年 8 月予定

■データ抽出の省力化

テンプレート種別	テンプレート名	効果とデータ処理の概要	提供開始
ドキュメントからのテキスト抽出	AI 前処理 PDF からテキスト抽出	スキャンした PDF からテキストを抽出	2025 年 8 月予定
画像からのテキスト抽出	AI 前処理 画像からテキスト抽出	ホワイトボードを撮影した写真や手書きの議事録などの画像からテキストを抽出	2025 年 8 月予定
音声からのテキスト抽出	AI 前処理 音声からテキスト抽出	会議の録音などの音声データからテキストを抽出	2025 年 9 月予定

■データ格納の省力化

テンプレート種別	テンプレート名	効果とデータ処理の概要	提供開始
エンベディング	AI 前処理 エンベディング & ベクトル DB 格納	テキストや構造化データなどのインプットデータを数値ベクトルに変換	2025 年 9 月予定



「AI 前処理テンプレートシリーズ」のデータ加工処理イメージ

- *1 RAG (Retrieval-Augmented Generation) : 検索拡張生成とも呼ばれ、大規模言語モデル (LLM) に新たに外部の情報を組み合わせ、生成 AI の検索精度を向上させる技術
- *2 当社製品の価格表 PDF (HULFT10 : 表・テキスト混在の 12 ページのドキュメント) を変換対象として使用し、3 種類の生成 AI (Claude/Gemini/Qwen) で QA 形式に変換する前処理をした場合、無処理の場合は質問に対する正解率が 30~50%だったのに対し、前処理を行った場合は回答精度が 80~90%まで向上 (セゾンテクノロジー調べ)

- *3 PDF および画像からのテキスト抽出を行った結果、PDF からのテキスト抽出は平均約 60%、画像からのテキスト抽出は平均約 50%の作業時間省力化（セゾンテクノロジー調べ）
- *4 テンプレート種別「クレンジング」は生成 AI を経由せず「HULFT Square」のみでデータを加工処理

セゾンテクノロジーについて

「世界中のデータをつなぎ、誰もがデータを活用できる社会を作る」をミッションに、安全・安心の基盤となるデータ連携製品や IT サービス、金融や流通業をはじめとする多種多様な業種向けのシステム開発・運用をグローバルに展開しています。長年にわたり環境の変化に即応してきた強みを活かし、現在はクラウド型データ連携プラットフォーム（iPaaS）「HULFT Square」など先進事業の拡大に注力するほか、未来を切り拓くテクノロジーの実装に向けた取り組みを強化しています。

- セゾンテクノロジー：<https://www.saison-technology.com/>
- HULFT 製品サイト：<https://www.hulft.com/>

商標関連

- 「HULFT」は、セゾンテクノロジーの商標または登録商標です。
- その他の会社名、製品名、サービス名等は、各社の商標または登録商標です。

報道関係者様からのお問い合わせ先

株式会社セゾンテクノロジー

広報担当：山下・高橋

PR@saison-technology.com（広報代表）