



過去50年分のドキュメントの検索性を高めるため ABBYY FineReader Serverを導入

企業紹介



企業名
千代田化工建設

所在地
神奈川県横浜市西区みなとみらい四丁目
6番2号みなとみらいグランドセントラルタワー

業界
建設業

企業URL
www.chiyodacorp.com

企業概要

千代田化工建設は1948年に創立した総合エンジニアリング企業。「エネルギーと環境の調和」を経営理念とし、エネルギー分野、化学・石油化学分野、医薬品分野、産業技術分野などのプラント建設を主軸に展開しています。

事例概要

同社では、保存するドキュメントを全文検索できるよう『ABBYY FineReader Server』を導入。『ABBYY FineReader Server』の構築や導入効果について、千代田化工建設の技術本部・エンジニアリングオペレーション業務部技術業務セクション2の川端真弘氏と、技術本部・エンジニアリングオペレーション業務部・技術業務セクション2技術情報サービスセンターの吉田純夫氏にお話を伺った。

課題

テキストデータ化されていない膨大なスキャンドキュメントの検索性の低さが課題

千代田化工建設は、豊富なプラント建設実績から得た知見と技術を活用し、プロセスプラントや社会インフラ整備における事業計画から設計・調達・建設・運転・保守までを担う「プロジェクト・ライフサイクル・エンジニアリング」を展開。「最先端AI技術と当社の知見を融合し、エンジニアリング技術のさらなる高度化も図っています」とエンジニアリングオペレーション業務部技術業務セクション2技術情報サービスセンターの吉田純夫氏は言う。

エンジニアリングオペレーション業務部は、同社の設計部門に属している部署だ。膨大な設計図や関連書類を作成・保存しており、それらの管理が大きな課題となっていた。特に、書類の検索性の低さには、長年頭を抱えていたという。

同社には事業プロジェクトごとに様々な資料が蓄積されている。プロジェクトの規模にもよるが、その数は数千から数万。書類の形式・含まれているページ数・用紙サイズもまちまちだ。CADや3Dモデル、オフィスファイルなどの電子データはもちろん、マイクロフィルムや紙の資料も少なくない。

これらの書類から必要な情報を検索するためには、紙に印刷されたリストから必要なマイクロフィルムを探し、目視で確認する、というようなアナログな手法をとるしかなく、その作業に膨大な手間と時間を費やしていた。

そこで、同社は書類を保存・管理するためにドキュメントアーカイブシステム「CDAS」(Chiyoda Document Archive System)を開発した。

CDASは、ストレージに保存されている膨大なドキュメントから必要な情報を瞬時に検索、管理できるように工夫されており、このシステムを使うことで、電子化した情報であれば簡単に検索でき、当初はこれで課題が解決できたと思われていた。しかし、その効果は同社内のユーザーが満足するものではなかった。

技術本部・エンジニアリングオペレーション業務部・技術業務セクション2の川端真弘氏は、「実際にシステムを稼働してみると、電子化されていても検索できないドキュメントが多いことに気がきました。これらは紙の書類をスキャンし電子化したもので、検索のためのテキスト情報を持っていません。システムを構築したものの、印刷されたリストからドキュメントを探す作業は従来の目視のまま、という状況になってしまったのです」と説明する。



技術本部
エンジニアリングオペレーション業務部
技術業務セクション2
川端 真弘 氏



技術本部
エンジニアリングオペレーション業務部
技術業務セクション2
技術情報サービスセンター
吉田 純夫 氏

ソリューション

大量のファイルを高速・高精度に処理するOCRの導入を検討

同社は、スキャンされたドキュメントをCDASで検索できるようにするため、OCR (Optical Character Recognition) の導入を検討した。OCRとは、画像から文字情報を読み取り、テキスト情報に変換するソフトウェアのこと。OCRを使えば、スキャンデータであっても、テキスト情報を抽出する事によってCDASから検索できるようになる。

そこで同社は、いくつかのOCRソリューションの検証を開始した。検証では、1ファイル・数千ページほどのPDFを処理するのに1時間30分から2時間程度の時間がかかった。「OCRごとに精度に大きく差がある上、1ファイルずつしか処理できないものもありました」(吉田氏)という。

「実際に検証することで、カタログスペックだけでは分からないさまざまな事が分かりました。当社の要件は、紙サイズにかかわらず、含まれているページ枚数の多い電子ファイルを高速で処理できてかつ精度が高いこと。この要件を満たした『ABBYY FineReader Server』(FineReader Server)を導入することにしました」(川端氏)

FineReader Serverを使って1年半で約10年分のドキュメントを全文検索可能に

- FineReader Server』とは、サーバーベースの大量文章変換用OCRソリューション。エンタープライズ向けとして提供されており、多くの企業で導入されている。
- FineReader Server』は、190を超える言語に対応しており、英語/日本語が混在するような文章であっても、正確なOCRを実現。また、PDFやJPEG、Word、TIFFなどさまざまな形式に対応するほか、ネットワークフォルダやMicrosoft SharePoint、エンタープライズコンテンツ管理 (ECM) システムなどに出力できる。ユーザー企業ごとの柔軟な運用を可能とするソリューションとなっているのだ。
- FineReader Serverを使い始めてから1年半が経ちますが、過去約10年分のドキュメントを全文検索できるようになりました。そのうちOCR化した枚数は、2千万枚はくだらないと思います。作成時期が古いドキュメントになればなるほど、OCRを必要とする資料が増えるので、非常に期待しています」

FineReader Serverを導入し、CDASから検索できるドキュメントが大幅に増えている。「全文検索できるようになり、ユーザーからも好評です」と吉田氏。ユーザーがドキュメントを検索するために要する時間も大幅に圧縮され、導入効果が現れている。

導入効果

チューニングを繰り返し、パフォーマンスの向上を実現

FineReader Serverを使って効率よくOCR処理ができるようになったが、導入当初から本格稼働まで、環境 (ソフトウェア・ハードウェア) のチューニングには多くの時間と工数を割いてきた。FineReader Serverはユーザーが設定できる項目が多いため、それらを調整することで変換速度や精度が大きく変わるのだ。

「作業ファイルが作成されるドライブを当社では通常の設定から変更しています。空き容量が少ないドライブでは、パフォーマンスが落ちてしまったからです。この変更には少々苦労しましたが、ABBYYさんのサポートを受けながら、ひとつひとつ工程を進めていきました」(川端氏)

また、FineReader Serverを構築していたサーバーについても変更した。当初は4コアのCPUと16GBのメモリーを搭載したサーバーを使用していたが、現在ではサポートとのやり取りで受領した資料を基に、12コアのCPUと32GBのメモリーを搭載したサーバーにしている。こうすることで、FineReader Serverの処理が大幅に向上し、同社が期待していた期間で大量のファイルを処理できるようになった。

CDASで検索できるドキュメントは、日増しに増えており、同社の生産性・作業効率の向上に寄与している。

FineReader Serverを使い、ドキュメントの検索効率が大幅に向上した千代田化工建設。同社の業務にFineReader Serverは欠かせないものとなった。ドキュメント管理に課題を持つ企業にとって、同社の事例は参考になる部分が多い。

ABBYY®

ABBYYジャパン株式会社
〒222-0033
神奈川県横浜市港北区新横浜2-3-12 新横浜スクエアビル14階 T 045.285.2410

WWW.ABBYY.COM/JA/