

# Azure Cognitive Searchによる 企業内全文検索と 日本語強化パックのご紹介

アクロクエストテクノロジー株式会社

# 目次

1. 企業内全文検索の価値
2. 日本語文書検索の課題
3. Azure Cognitive Search
4. アクロクエストが提供するソリューション・サービス

# 1. 企業内全文検索の価値

# 企業活動で扱うデータの多くが自然言語によるもの

## データの種類

- 営業記録／顧客対応記録
- マニュアル／手順書
- ホワイトペーパー／論文
- 提案書／見積／契約書
- 報告書
- 商品説明／商品レビュー
- アンケート

## データの形式

- Office文書
- PDF
- HTML
- テキストファイル
- データベース
- 画像



**日々の業務やビジネスを  
圧倒的に変革するには**

**自然言語（テキスト）  
データの中にある価値を  
見つけ出し、活用する**

# 企業内全文検索とは

企業が扱うOffice文書、PDF、HTML、テキスト、システムのデータベース等のデータを統合検索可能にするシステム

『エンタープライズサーチ』  
と呼ばれる



# このようなことはありませんか？

あのファイルどこ  
にあったかなあ？

検索が遅い

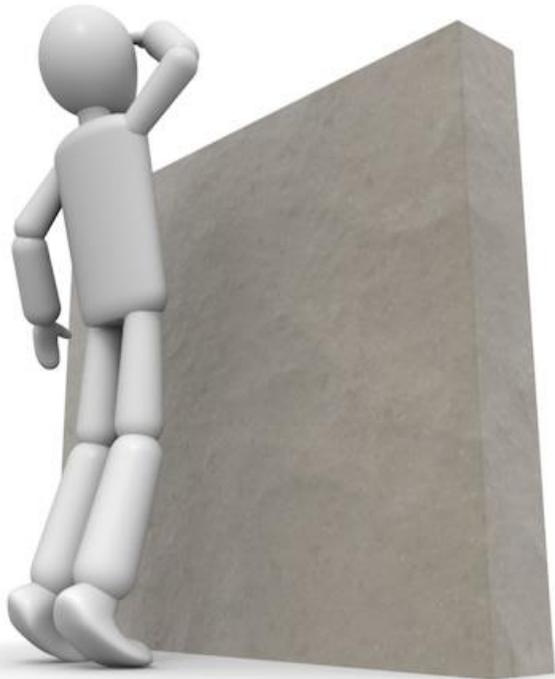
日本語の検索品質  
にイライラする



# このようなことはありませんか？

他部門、他システムの  
データにアクセスできない

デジタル化したのは良いが、  
活用できていない



# エンタープライズサーチ導入による効果

1日当たりの検索時間を以下と仮定する  
5分×6回=30分



1年間では一人当たり120時間=0.75人月を  
検索に費やしている



# エンタープライズサーチ導入による効果

エンタープライズサーチの導入により、  
このコストを3分の1に削減できるとしたら？



社員一人当たり、年間0.5人月分を削減できる

100人の企業なら、年間50人月  
1000人の企業なら、年間500人月



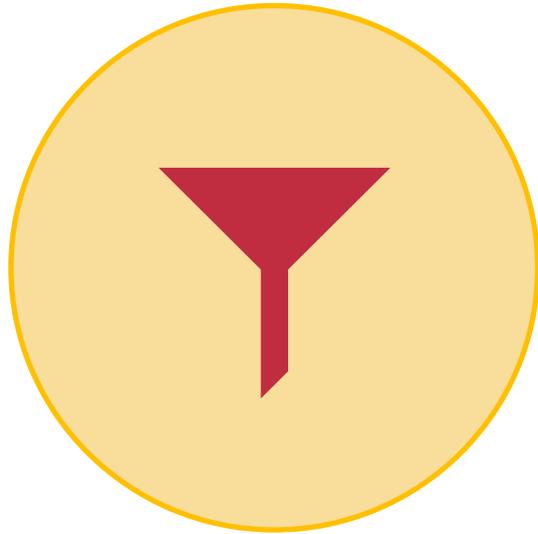
# エンタープライズサーチ導入による効果

エンタープライズサーチ導入がもたらすもの

- ① コストを削減できる
- ② 新たな価値を生み出すことに投資できる
- ③ 業務、部署、システムの壁を越えてデータを資産として活用できるため、ビジネスに圧倒的な効率化や変革をもたらすことが可能になる



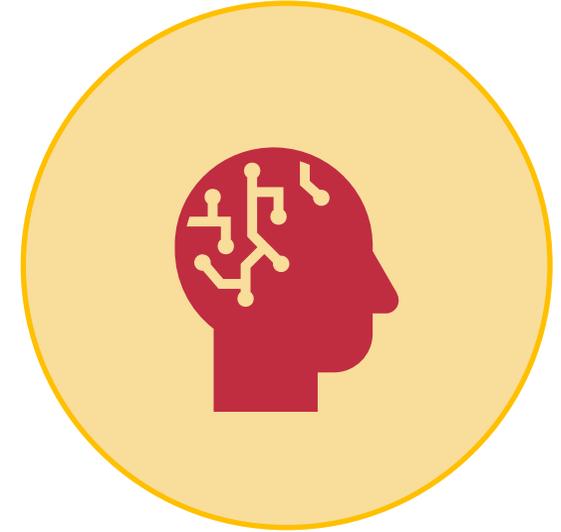
# 自然言語（テキスト）データを扱う時の課題



**テキスト抽出**



**検索**



**価値抽出**

# 自然言語（テキスト）データを扱う時の課題



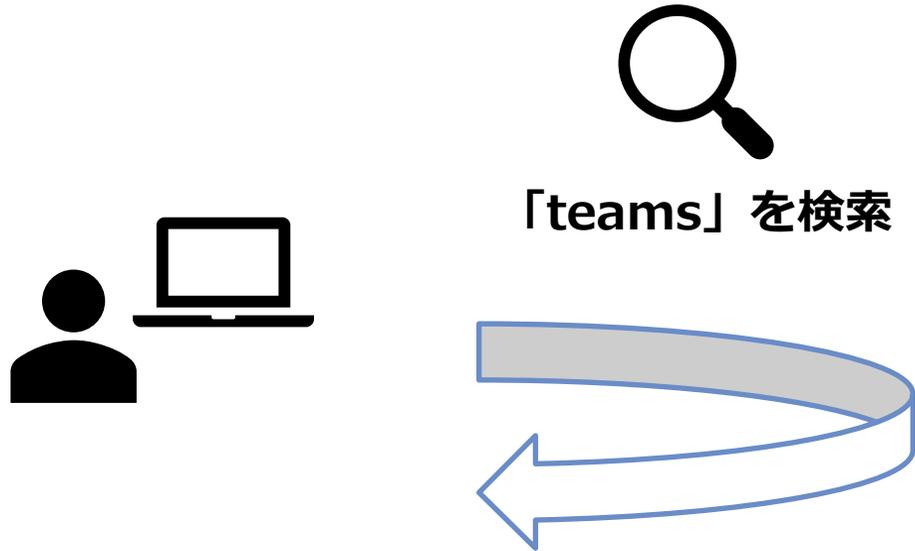
**検索**

**探せないものは  
存在しないものと同義**

**膨大なテキストから  
特定のワードを  
どのように探す？**

# 企業内全文検索がない場合

全てのドキュメントを  
先頭から最後まで探す



データ量が多くなるほど遅くなる



ドキュメント A

Azure はMicrosoftが提供するクラウドサービスです。



ドキュメント B

Microsoft365 ではTeamsを利用することができます。



ドキュメント C

Teamsを使ったオンライン会議を開催する手順は次の通りです。



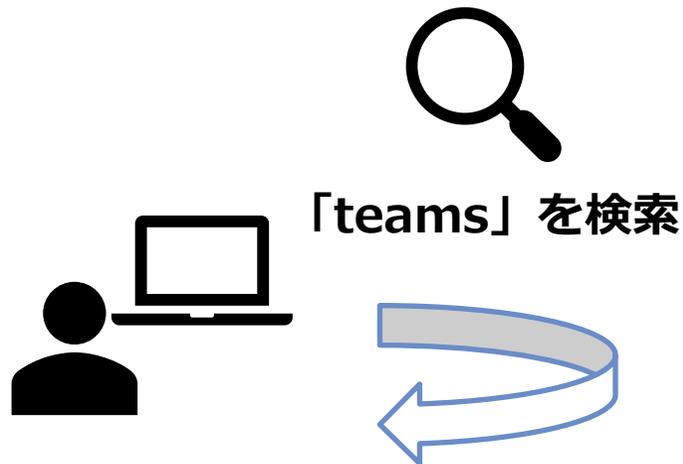
ドキュメント D

チャットが使える便利なソリューションを探しています。

⋮

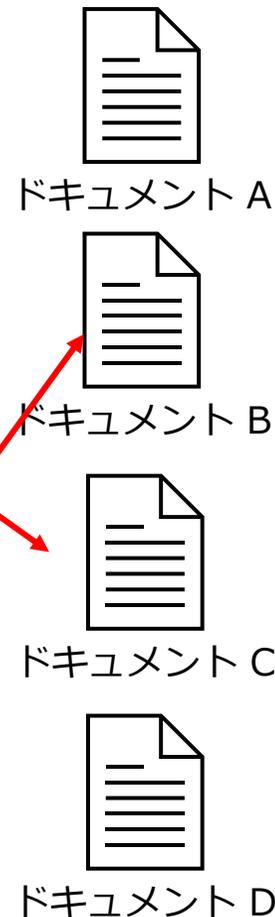


# 企業内全文検索を利用する場合



検索用の索引インデックス

単語	ドキュメント
azure	A
microsoft	A, B
クラウドサービス	A
365	B
teams	B, C
オンライン	C
会議	C
チャット	D
ソリューション	D
...	



Azure はMicrosoftが提供するクラウドサービスです。

Microsoft365 ではTeamsを利用することができます。

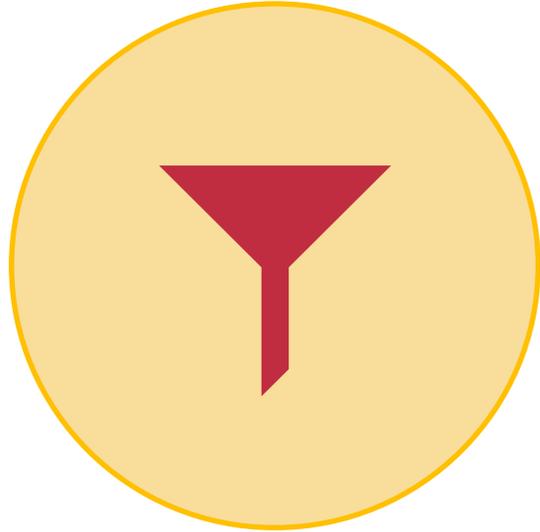
Teamsを使ったオンライン会議を開催する手順は次の通りです。

チャットが使える便利なソリューションを探しています。

膨大なデータでも高速に検索が可能

・  
・  
・

# 自然言語（テキスト）データを扱う時の課題



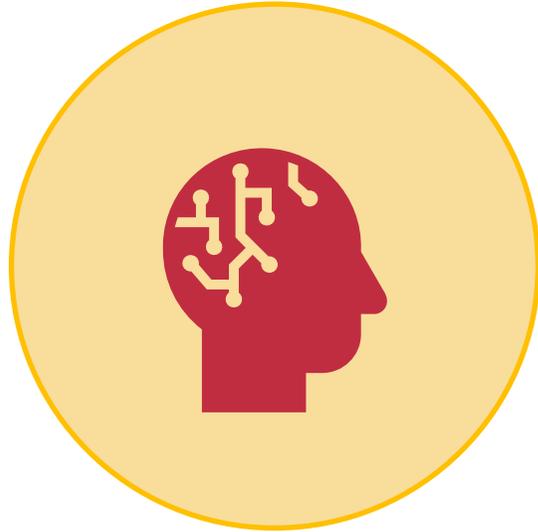
テキスト抽出

非構造化データから  
テキストデータを  
どのように抽出する？



# 自然言語（テキスト）データを扱う時の課題

検索だけではなく  
新たな価値をどう作る？



## 価値抽出



感情分析



場所抽出



人物抽出



組織抽出

Aあ

言語検出



キーワード抽出

#

タグ抽出

# 企業活動で扱う多くのデータが自然言語によるもの

## データの種類

- 営業記録／顧客対応記録
- マニュアル／手順書
- ホワイトペーパー／論文
- 提案書／見積／契約書
- 報告書
- 商品説明／商品レビュー
- アンケート

## データの形式

- Office文書
- PDF
- HTML
- テキストファイル
- システムのデータベース
- 画像

様々なデータ形式に対応し、膨大なテキストデータを  
高速に扱える企業内全文検索が必要

## 2. 日本語文書検索の課題

# 全文検索エンジンでの日本語検索でよくある話

昨日は東京タワーに行った。

**「今日」というキーワードで検索したら  
どうなるでしょうか？**

# 全文検索エンジンでの日本語検索でよくある話

昨日は東京タワーに行った。

今日

「昨」「日」「は」  
「東」「京」「タワー」  
「に」「行」「っ」「た」

「今」「日」

デフォルトの設定では . . .

ヒットして  
しまう!?

# 全文検索エンジンでの日本語検索でよくある話

全文検索エンジンのデフォルトの**Analyzer**は、基本的に空白で分かち書きされている英語用です。

日本語のように分かち書きされていない言語には向いていません。

日本語を対象にする場合には、日本語アナライザを利用し、さらに必要に応じたカスタマイズが必要となります。

# 日本語文書の全文検索における主な課題

#	項目	概要
1	大文字／小文字	Microsoft Office／microsoft office
2	全角／半角	A B C／ABC, 1 2 3／123, がギグゲゴ／ガギグゲゴ
3	記号の揺らぎ	マツコ・デラックス／マツコデラックス, きやり～ぱみゅぱみゅ／きやりーぱみゅぱみゅ
4	カタカナの揺らぎ	バイオリン／ヴァイオリン, インターフェース／インターフェイス／インタフェース
5	字種の違い	斉藤／斎藤／齊藤, 渡辺／渡邊／渡邊
6	同義語	ハンドル／ステアリング
7	意図しない分割	要件定義／要件 OR 定義
8	送り字の違い	見積／見積り／見積もり, 忌引／忌引き

# 日本語文書の全文検索における主な課題

#	項目	概要
9	内包	ライト/ヘッドライト/サテライト
10	キーワード自動補完	東京/とうきょう/tokyo
11	もしかして検索	サイゼリヤ/サイゼリア
12	不要な空白	OfficeファイルやPDFを読み込んだ場合やOCRで取り込んだテキストに空白が入ってしまう問題

- これらの日本語特有の課題に対応するには、  
専門知識を持った適切な設定とチューニングが必要
- 後述の『**Azure Cognitive Search日本語強化パック**』は  
これらの課題に対応可能

# 3. Azure Cognitive Search

# Azure Cognitive Search

モバイルおよび Web アプリ開発のための AI を活用したクラウド検索サービス

無料で始める >



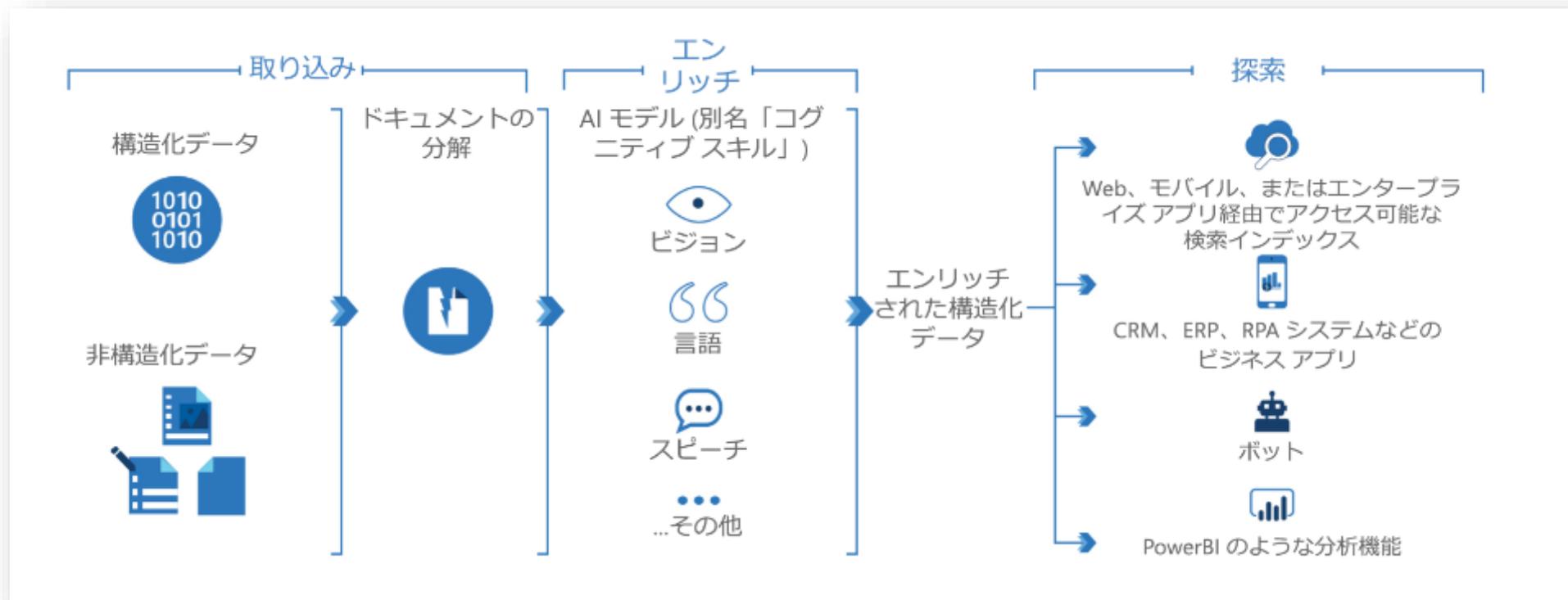
[製品の概要](#) [機能](#) [セキュリティ](#) [料金設定](#) [ドキュメント](#) [使用を開始するには](#) [お客様事例](#) [更新情報](#) [よく寄せられる質問](#)

## すべてのコンテンツから潜在的な洞察を引き出す

Azure Cognitive Search は、あらゆる種類の情報を拡充して関連コンテンツを簡単に特定し、大規模に検索する AI 機能が組み込まれた唯一のクラウド検索サービスです。以前は Azure Search として知られていたこのサービスでは、Bing と Office でも 10 年以上にわたって利用されている統合 Microsoft 自然言語スタックに加え、視覚、言語、音声の AI サービスが使用されます。イノベーションに費やす時間を増やし、複雑なクラウド検索ソリューションの保守に費やす時間を減らすことができます。

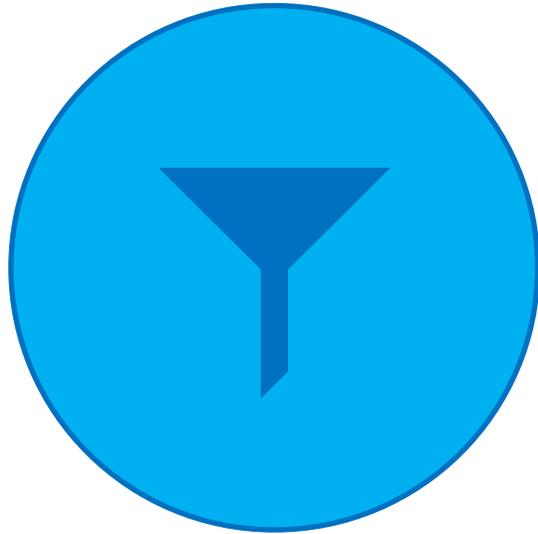
# Azure Cognitive Searchの紹介

1. Azureで提供されるマネージドな全文検索サービス
2. インデクサー、エンリッチメント、検索の仕組みを提供
3. 検索だけでなく、AIの活用による付加価値の追加が可能



マイクロソフト「AIを活用したナレッジマイニングソリューションを構築するための開発者ガイド」より引用

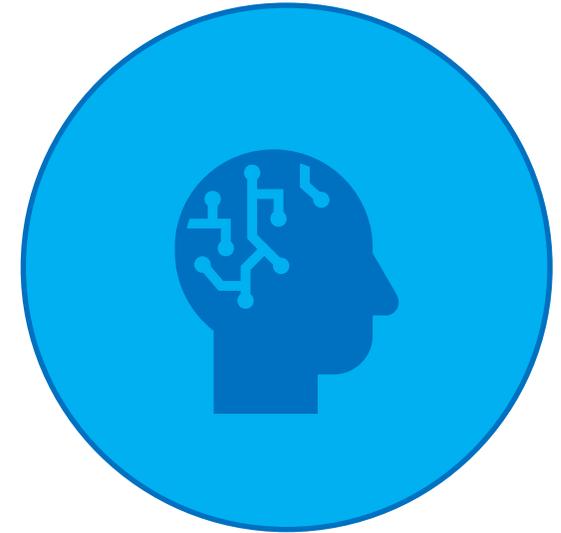
# Azure Cognitive Searchの3つの機能



インデクサー



検索



エンリッチメント

# Azure Cognitive Searchの検索機能

<https://azjobsdemo.azurewebsites.net>



# 検索

 Azure Cognitive Search  
JOB PORTAL DEMO

AVAILABLE JOBS (2558 jobs)

[Home](#)

[Jobs](#)

[About Azure Cognitive Search](#)

## SEARCH

Any distance from

## FILTER RESULTS

### BUSINESS TITLE

- > Auditor (17)
- > Project Manager (17)
- > College Aide (13)
- > Agency Attorney (12)
- > Assistant Corporation Counsel (12)
- > Construction Project Manager Intern (10)
- > Administrative Assistant (8)
- > Procurement Analyst (8)
- > City Research Scientist (7)
- > Watershed Maintainer (7)

### LOCATION

- > Internal (1341)
- > External (1217)



2558 AVAILABLE JOBS

Relevance

< 1 2 3 4 5 >



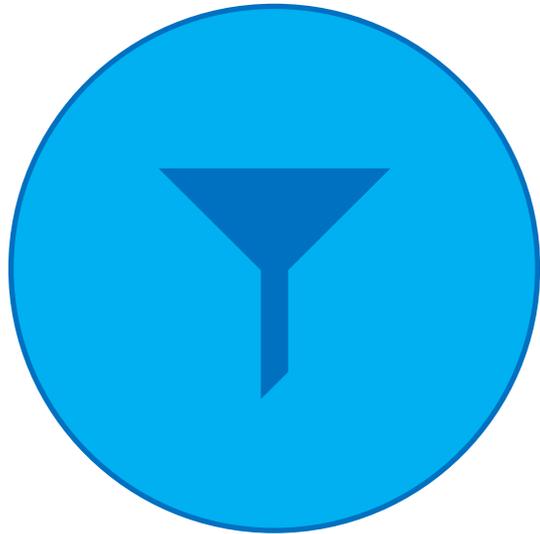
### 5 Business Analyst

Nov 150 William Street, New York N

Salary: \$56,937 to \$88,649 Annual

JOB DESCRIPTION: Under supervision of the Business Analyst Manager, with latitude for independent judgment and action, the candidate will serve as an ACS Business Analyst for the development, enhancement and maintenance of ACS systems and applications. The Business Analyst role is critical to the shared resources framework within MIS. Primary responsibilities include: conducting detailed business analysis, creating High-level, Functional and/or Business Requirements documents, producing forms ... [Read More](#)

# データの取り込み利用できるインデクサー



## インデクサー

### フォーマット

- PDF
- Microsoft Office
- HTML
- XML, JSON, CSV
- ZIP
- EML, RTF, TXT
- PNG, JPG

### データソース

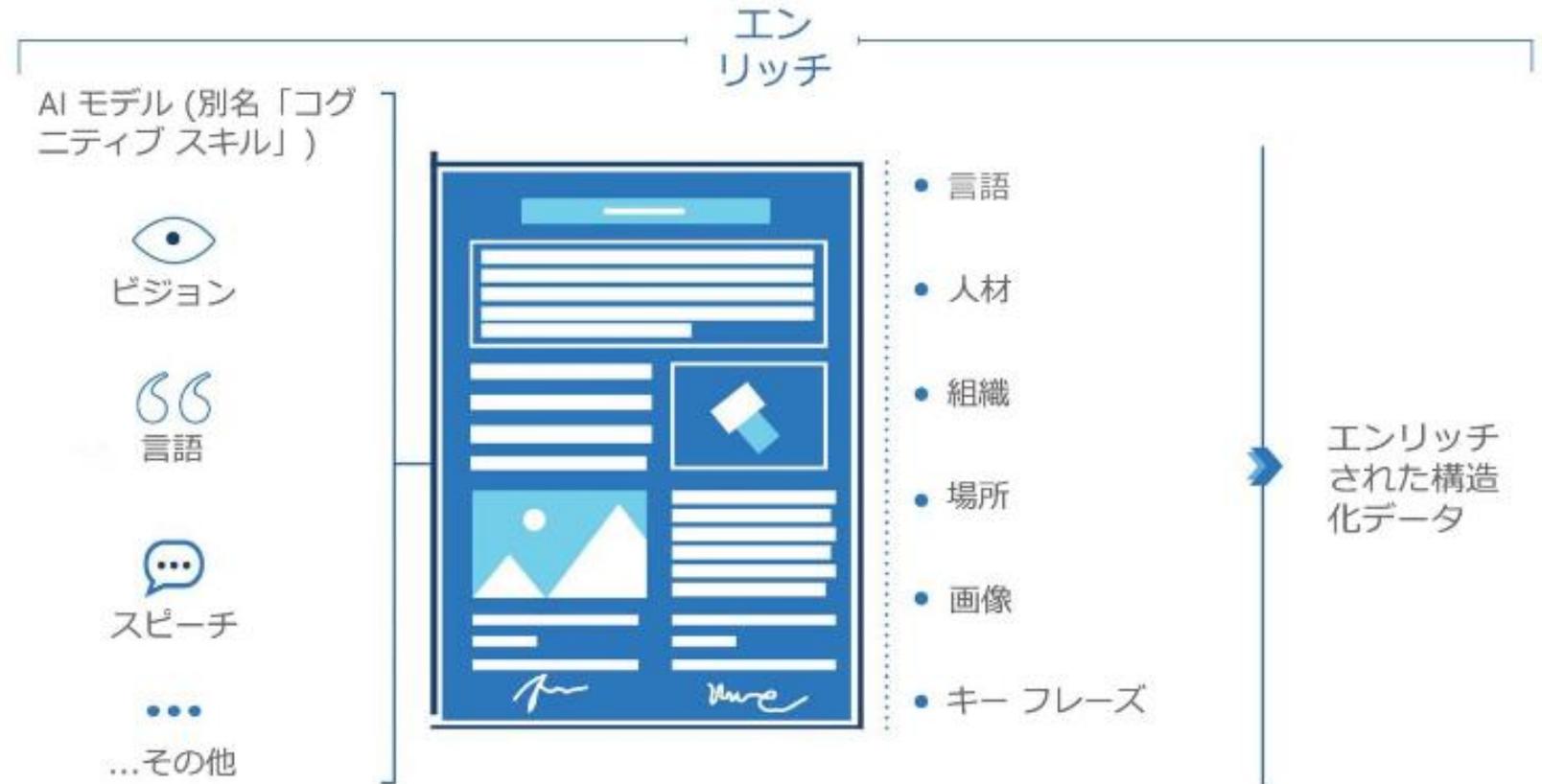
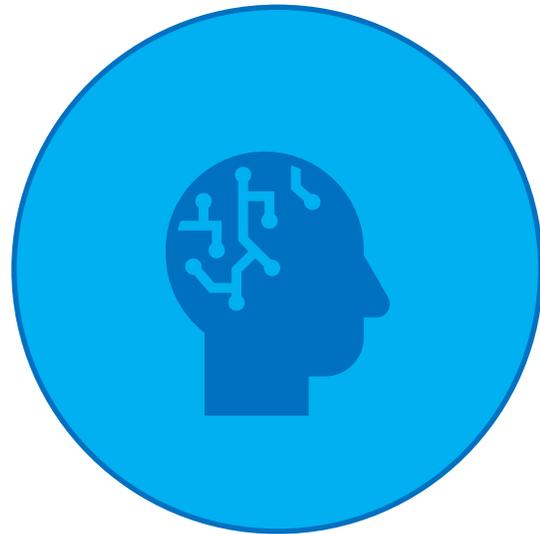
- Azure Blob Storage
- Azure Table Storage
- Azure Cosmos DB
- Azure SQL Database
- Azure VM上のSQL Server

インデクサーの同期間隔は最短で5分

REST APIによるデータ登録も可能

# データに付加価値を追加するエンリッチメント

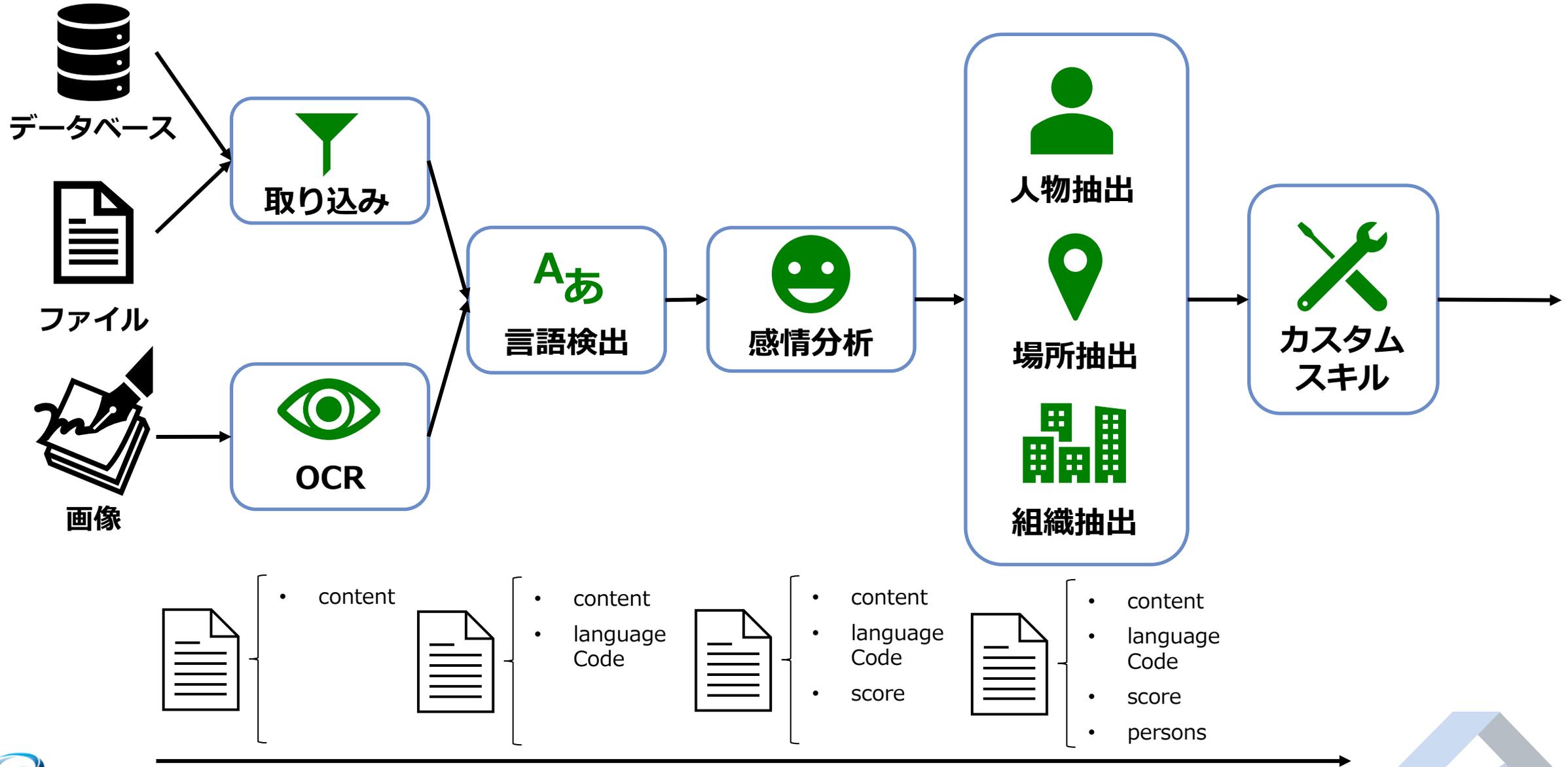
マイクロソフト「AIを活用したナレッジマイニングソリューションを構築するための開発者ガイド」より引用



## エンリッチメント

Azure Cognitive Serviceとして提供される機能を Cognitive Skillとして利用可能

# エンリッチメントパイプライン



# 4. アクロクエストが提供する ソリューション・サービス

# Azure Cognitive Search日本語強化パック

DX Search

Q インターフェイス

people

- 渡辺 一郎 (5)
- 齊藤 太郎 (3)
- 齋藤 太郎 (2)
- 齋藤太郎 (1)
- 渡邊 一郎 (1)
- 鈴木 太郎 (1)
- 齋藤太郎 (1)

locations

- 新宿区 (8)
- 東京都 (8)
- 新宿 (6)

organizations

- Office Software (2)

期間指定なし

関連度順

"インターフェイス"の検索結果 8件 1ページ目 1件目~8件目を表示

## 20200925注文書.pdf

様 No. 1 1 式 2 1 式 3 1 式 備考 ¥7,700,000合計 消費税 ¥7,000,000 ¥700,000 小計 担当: 合計金額 ¥7,700,000 (税込) 単価 ¥2,500,0002,500,000要件定義書 齋藤太郎 金額 インターフェイス仕様書 ソースファイル一式 ¥1,500,000 ¥3,000,000 1,500,000 3,000,000 摘要 数量 新宿第1ビル2階 TEL: 03-3333-3333 FAX: 03-4444-4444 E-Mail: testsamplemail@test.co.jp 納期: 別途ご相談 支払条件: 月末締め翌月末払い 見積No: EST-20200925-003 東京都新宿区新宿1-2-3 ORD-20200925-003 2020年9月25日ご担当: 渡辺 一郎 株式会社Office Software 〒222-0000 件名: 新業務システム効率化くん 開発 下記のとおり、発注致します。

## 20200925注文書02.pdf

様 No. 1 1 式 2 1 式 3 1 式 備考 ¥7,700,000合計 消費税 ¥7,000,000 ¥700,000 小計 担当: 合計金額 ¥7,700,000 (税込) 単価 ¥2,500,0002,500,000要件定義書 齋藤太郎 金額 インターフェイス仕様書 ソースファイル一式 ¥1,500,000 ¥3,000,000 1,500,000 3,000,000 摘要 数量 新宿第1ビル2階 TEL: 03-3333-3333 FAX: 03-4444-4444 E-Mail: testsamplemail@test.co.jp 納期: 別途ご相談 支払条件: 月末締め翌月末払い 見積No: EST-20200925-003 東京都新宿区新宿1-2-3 ORD-20200925-003 2020年9月25日ご担当: 渡辺 一郎 株式会社OFFICE SOFTWARE 〒222-0000 件名: 新・業務システム効率化くん 開発 下記のとおり、発注致します。

## 20200925注文書 – 作成中.pdf

摘要 数量 新宿第1ビル2階 TEL: 03-3333-3333 FAX: 03-4444-4444 E-Mail: testsamplemail@test.co.jp 納期: 別途ご相談 支払条件: 月末締め翌月末払い 見積No: EST-20200925-003 インターフェイス仕様書 プログラム一式 ¥1,500,000 ¥3,000,000 1,500,000 3,000,000 担当: 合計金額 ¥7,700,000 (税込) 単価 ¥2,500,0002,500,000要件定義書 鈴木 太郎 金額 消費税 ¥7,000,000 ¥700,000 小計 備考 ¥7,700,000合計 mailto:testsamplemail@test.co.jp mailto:testsamplemail@test.co.jp mailto:testsamplemail@test.co.jp mailto:testsamplemail@test.co.jp mailto:testsamplemail@test.co.jp

## 業務委託契約書.pdf

1. 甲の運用する「新・業務システム効率化くん」の要件定義、インターフェース設計、開発業務 2. 運用開始後のシステムメンテナンス 以上、本契約の成立を証するため、この契約書を2通作成して、双方署名捺印の上、各自1通を保有する。

## 業務委託契約書03.pdf

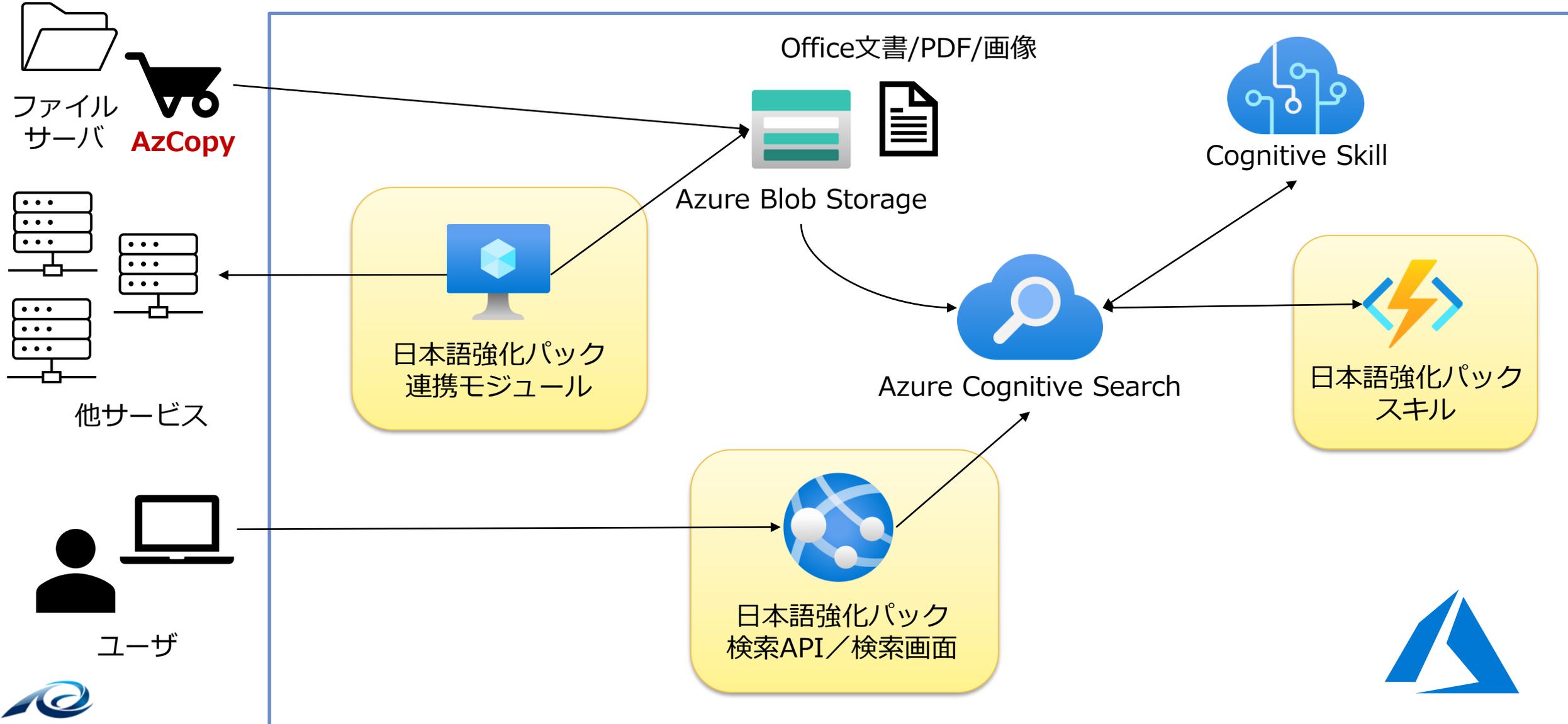
1. 甲の運用する「新業務システム効率化くん」の要件定義、インターフェース設計、開発業務 2. 運用開始後のシステムメンテナンス 以上、本契約の成立を証するため、この契約書を2通作成して、双方署名捺印の上、各自1通を保有する。

# Azure Cognitive Search日本語強化パックの特長

1. 標準で検索画面を提供するため、Azure Cognitive Searchを利用した全文検索システムの構築が簡単にできます
2. 全文検索システムが日本語文書を扱うときに発生する日本語固有の課題を強力にサポート。日本人にとって使い勝手のよい検索体験を提供します
3. Azure Cognitive Searchの便利な機能をフル活用できるため、Cognitive ServiceによるAI機能の活用やその他のAzureサービスとの連携が可能です

# Azure Cognitive Search 日本語強化パック

## アーキテクチャ



# アクロクエストの全文検索導入支援サービス

- Azure Cognitive Searchを利用した全文検索システムの導入を検討されている方は、ぜひお問い合わせください。

- ① PoC（実証検証）の実施
- ② 構築支援
- ③ 日本語の検索体験の改善支援



<https://www.endosnipe.com/azure-contact-form>

資料を最後までお読みいただき  
ありがとうございました



*Acroquest Technology*

*Evolve the Earth with Emotion of Technology*