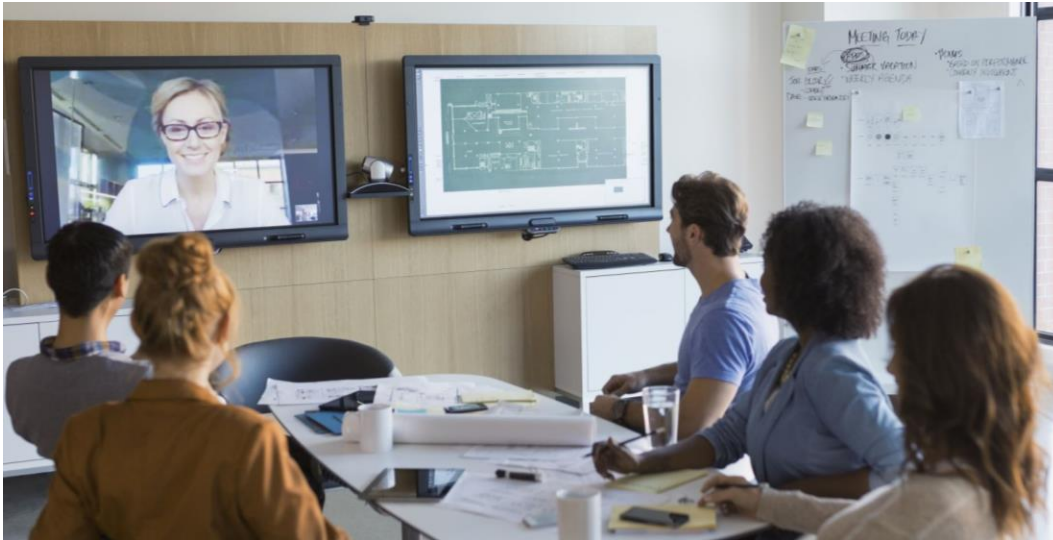




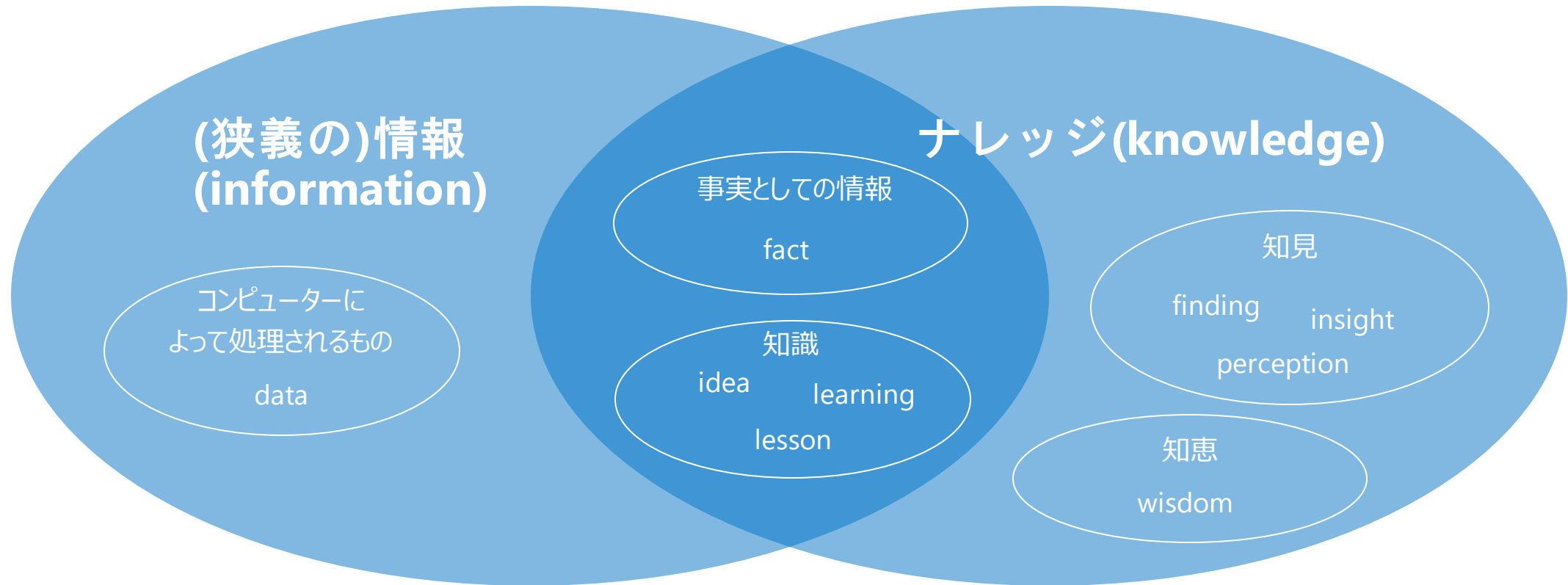
Azure Cognitive Search が実現する 情報のナレッジ化と全文検索システム

日本マイクロソフト株式会社
クラウドソリューションアーキテクト
高橋 敬祐

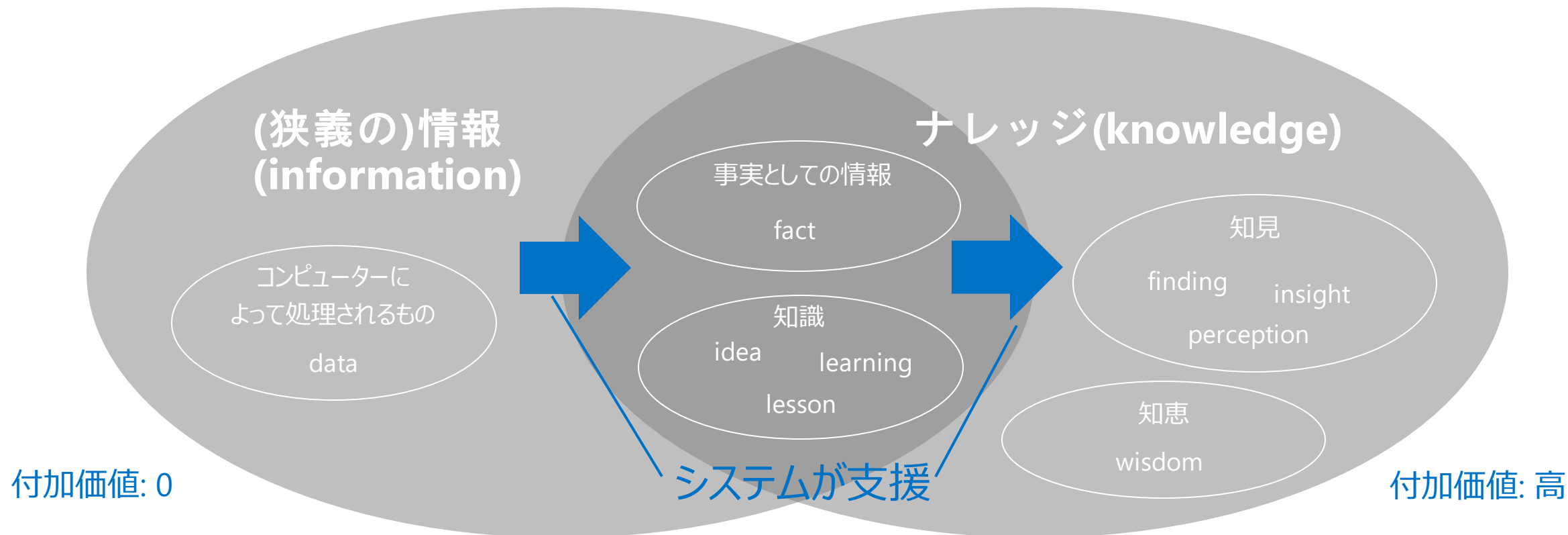
情報？ナレッジ？



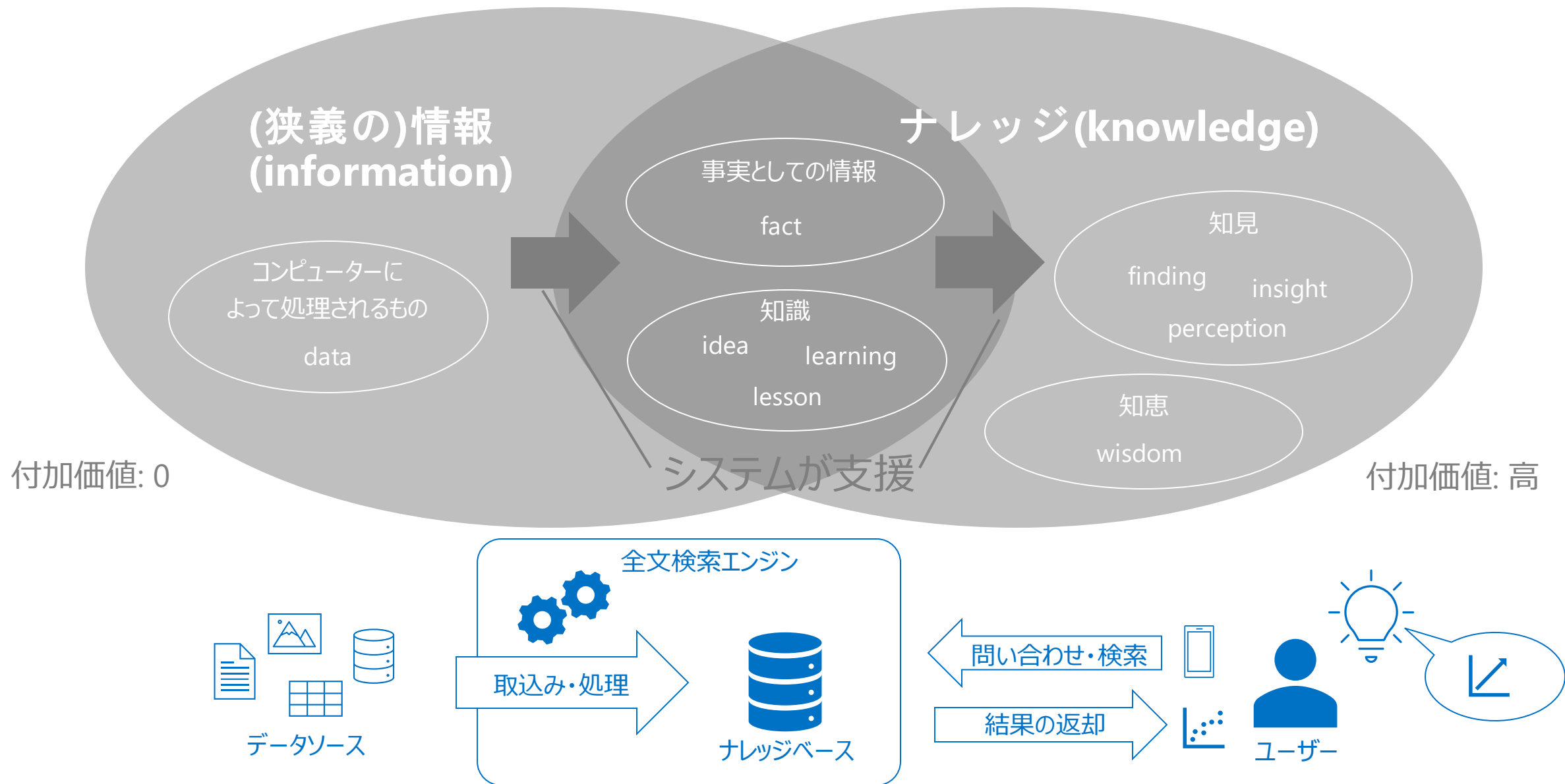
情報とナレッジ



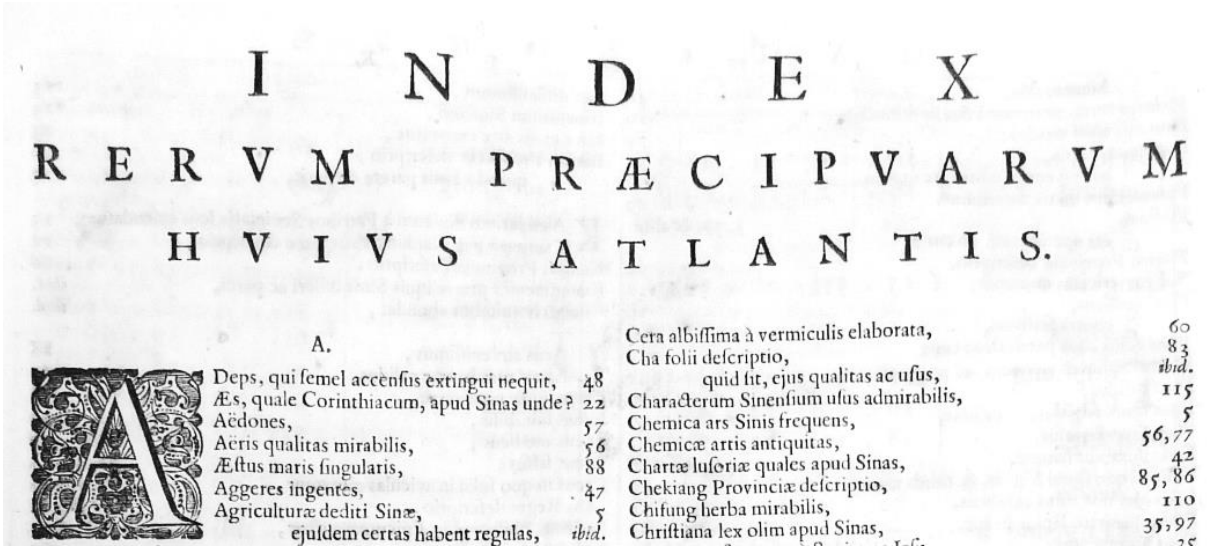
情報からナレッジを引き出す



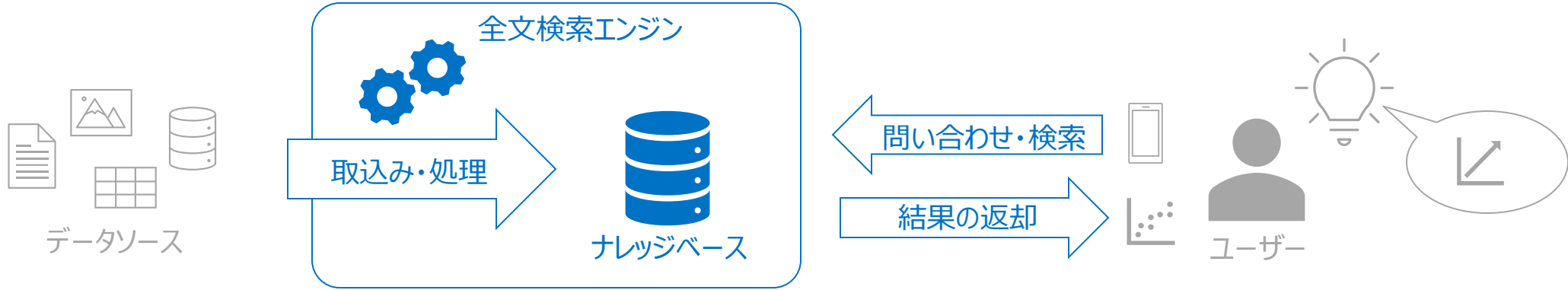
情報からナレッジを引き出すためのシステム



全文検索エンジンのしくみ(典型例)



引用元: <https://ja.wikipedia.org/wiki/%E7%B4%A2%E5%BC%95>



コラム: プリミティブな全文検索エンジンの作り方

データ収集処理

インデックス生成処理

転置インデックス生成

検索処理



取込み・加工処理



```
SELECT c FROM t  
LIKE '%foo%';
```

問い合わせ・検索

結果の返却



コラム: プリミティブな全文検索エンジンのペインポイント

データ収集処理

インデックス生成処理

転置インデックス生成

検索処理



取込み・加工処理



データベース

```
SELECT c FROM t  
LIKE '%foo%';
```

問い合わせ・検索

結果の返却



ユーザー

プレーンテキスト化
できるもの
しか取り込めない

データ量が増加すると
検索時間が長くなる

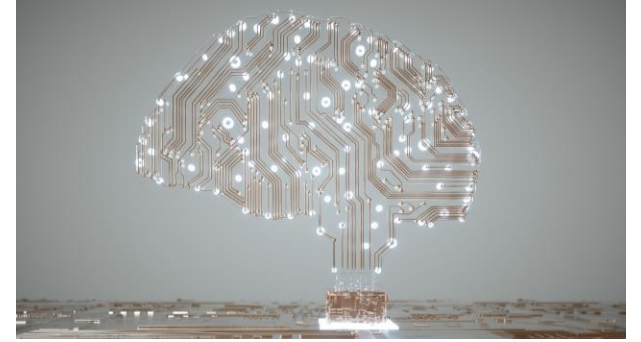
インデックスを作ると
取込み時間が長くなる

プロビジョニングした
容量しか使えない

DB分割を始めると
複雑化・ガラパゴス化

言葉の多様性や
柔軟性への追従
・対応が困難

クラウド・ビッグデータ・AI時代の全文検索エンジン



- ✓ フルマネージド
- ✓ PaaS / SaaS
- ✓ CapEx から OpEx へ
- ✓ 伸縮性
- ✓ 可用性
- ✓ 冪等性
- ✓ 他サービスとのAPI連携

- ✓ スケールアウト
- ✓ データ量に依らない性能
- ✓ Apache ファミリーとの互換性

- ✓ 言語の特定
- ✓ キーフレーズの抽出
- ✓ 画像からのOCR
- ✓ 画像からの推定
- ✓ 他のAIとのAPI連携

Azure Cognitive Search 紹介

Azure Cognitive Search

全文検索機能を提供する PaaS 型の検索サービス

検索の専門的な知識なしで、検索システムの構築・利用が可能

- 豊富な検索機能をシンプルなプロトコルと API で利用可能
- フルマネージドサービス & スケーラブル
- 充実した自然言語解析、ランキング調整機能
- 事前構成/カスタム構成の AI 連携



Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



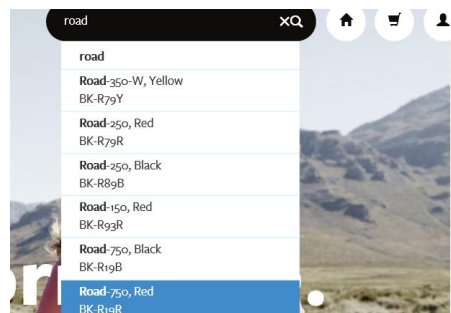
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

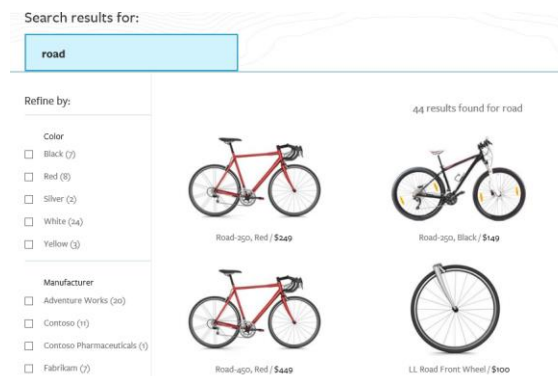
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



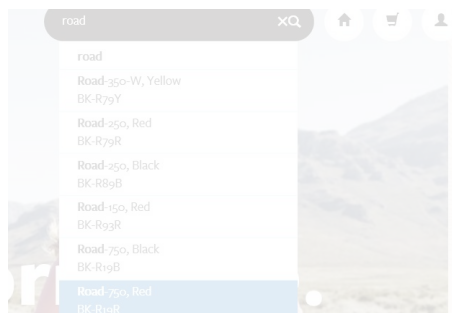
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

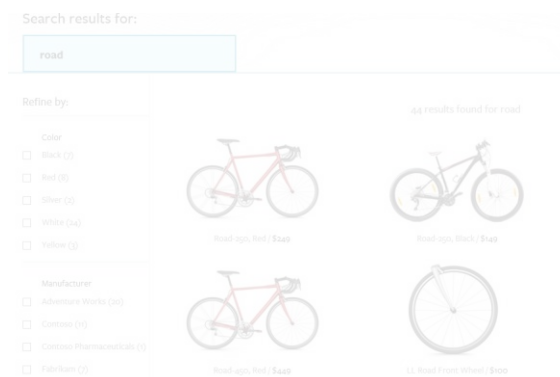
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

検索スコア

検索語句との一致 or
検索語句の
インスタンス
の含有量で
大小が決定

```
"value": [  
  {  
    "@search.score": 1.3097656,  
    "question": "Azure では IPv6 がサポートさ  
  },  
  {  
    "@search.score": 1.3056309,  
    "question": "Azure サポートの購入方法は?"  
  },  
  {  
    "@search.score": 1.2120581,  
    "question": "Azure サポート特典を使用して
```


Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

検索スコア

検索語句との一致 or
検索語句の
インスタンス
の含有量で
大小が決定

```
"value": [  
  {  
    "@search.score": 1.3097656,  
    "question": "Azure では IPv6 がサポートさ  
  },  
  {  
    "@search.score": 1.3056309,  
    "question": "Azure サポートの購入方法は?"  
  },  
  {  
    "@search.score": 1.2120581,  
    "question": "Azure サポート特典を使用して
```

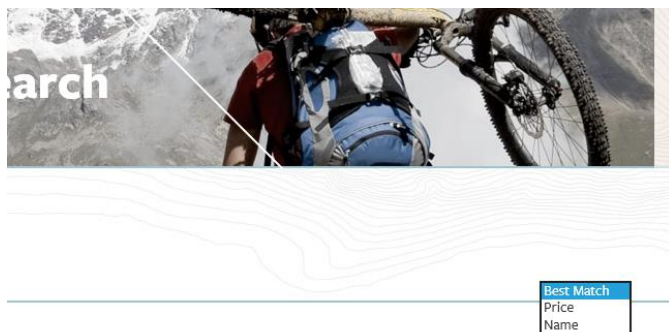
- TF-IDF (Term Frequency-Inverse Document Frequency)

Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



検索スコア

検索語句との一致 or 検索語句のインスタンスの含有量で大小が決定

```
"value": [
  {
    "@search.score": 1.3097656,
    "question": "Azure では IPv6 がサポートさ"
  },
  {
    "@search.score": 1.3056309,
    "question": "Azure サポートの購入方法は?"
  },
  {
    "@search.score": 1.2120581,
    "question": "Azure サポート特典を使用して"
```

BM25Similarity (BM25)

- TF-IDF (Term Frequency-Inverse Document Frequency)タイプの関数
- より直感的な結果が得られる

Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



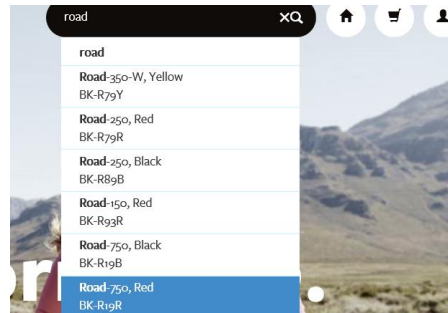
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

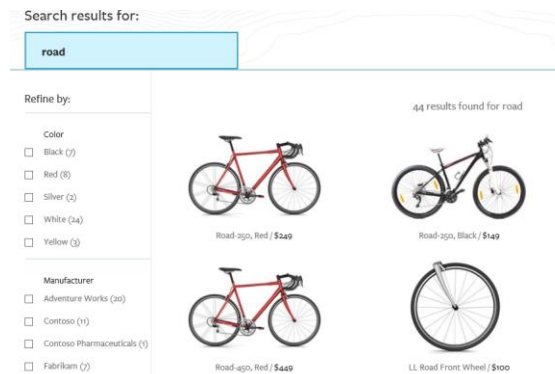
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示

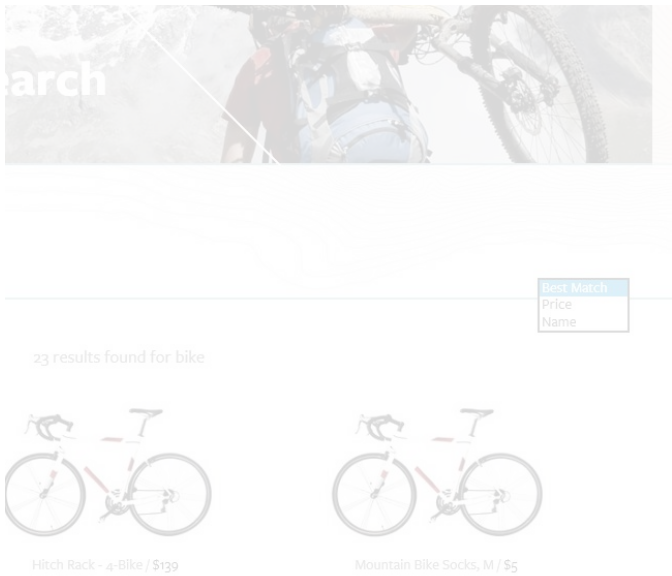


Azure Cognitive Search の検索エクスペリエンス

スコアリング

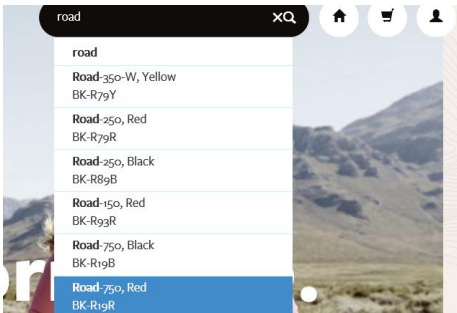
検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



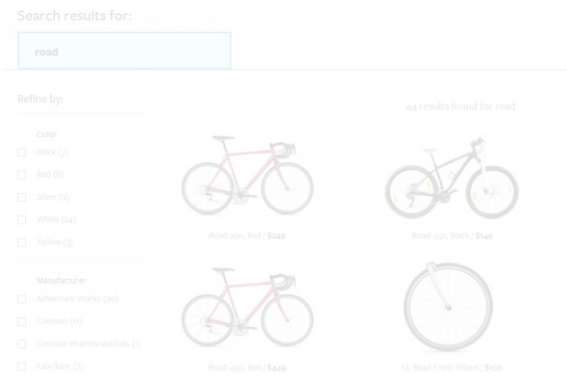
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



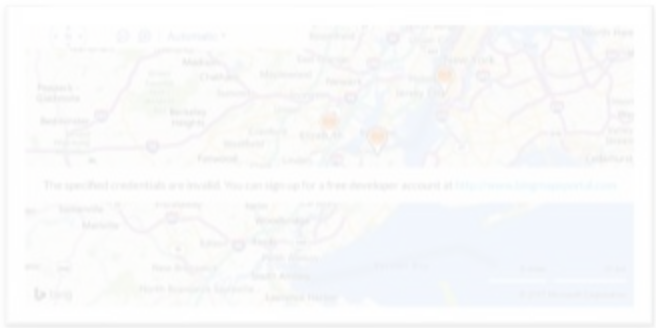
ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

検索スコア

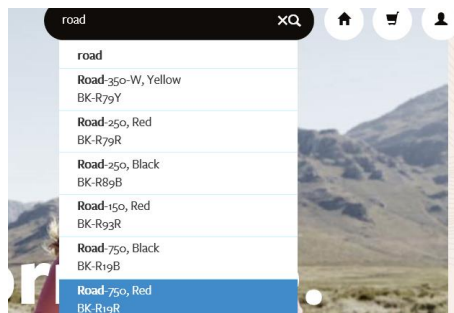
+

セマンティック検索スコア

※ 2022/11 時点でパブリックプレビュー

サジェスト

入力した検索文字列に対して、関連の
高い検索候補の表示



- ①セマンティックの再ランク付け
- ②セマンティックキャプションのハイライト
- ③セマンティック回答

Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



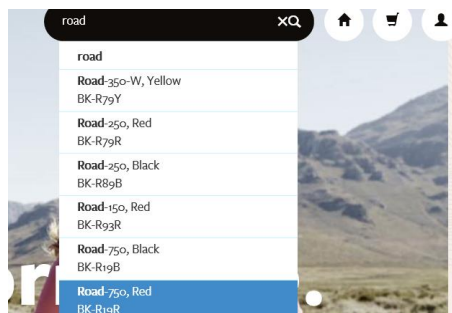
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

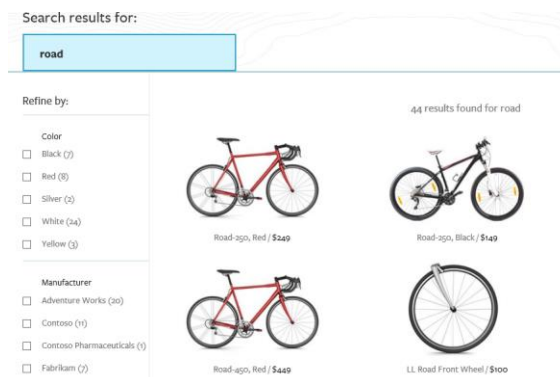
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示

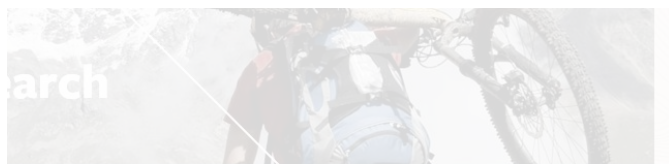


Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



23 results found for bike



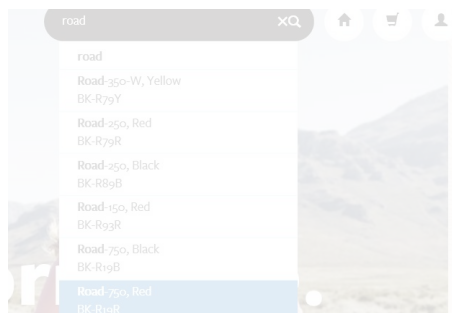
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

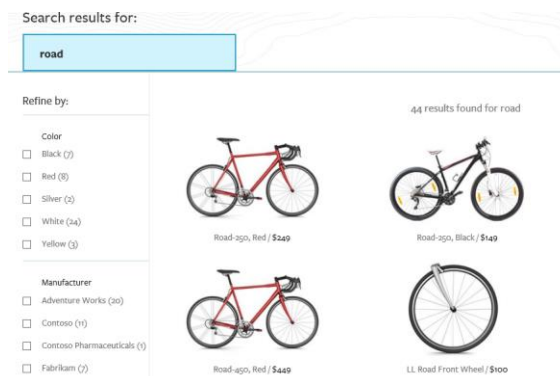
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

ファセット = 検索結果に対するドリルダウンフィルター処理を行うためのナビゲーション機能

Hotels Search

wifi

19 Results

Category:

- Budget (5)
- Luxury (5)
- Resort and Spa (5)
- Boutique (3)
- Extended-Stay (1)

Amenities:

- free wifi (11)
- laundry service (6)
- 24-hour front desk service (5)
- pool (5)
- restaurant (5)
- concierge (4)
- continental breakfast (4)
- free parking (4)
- view (4)
- coffee in lobby (3)
- air conditioning (2)
- bar (1)

Super Deluxe Inn & Suites

Complimentary Airport Shuttle & WiFi. Indoor/Outdoor Pool, Fitness Center, Fl...

Category: Boutique

Amenities: bar, free wifi, free wifi

Double Sanctuary Resort

5* Luxury Hotel - Biggest Rooms in the Traveler. Free WiFi, Flexible check in/o...

Category: Resort and Spa

Amenities: view, laundry service, free w...

Pull'r Inn Motel

The hotel rooms and suites offer the pe... will elevate your stay, whether you're tr... or just looking for a remarkable getaway right from your suite, your stay will be n...

ファセット

ファセットパラメータを使用して、複数の検索結果を表示

Search results for: road

44 results found for road

Refine by:

Color

- ☐ Black (7)
- ☐ Red (5)
- ☐ Silver (2)
- ☐ White (24)
- ☐ Yellow (3)

Manufacturer

- ☐ Adventure Works (20)
- ☐ Contoso (11)
- ☐ Contoso Pharmaceuticals (1)
- ☐ Fabrikam (7)

Road-100, Red / \$249

Road-100, Black / \$149

Road-100, Red / \$449

LL Road Front Wheel / \$100

データ構造

```
{
  "@odata.context": "https://demo-search-...",
  "@odata.count": 50,
  "@search.facets": {
    "Category": [
      {
        "count": 13,
        "value": "Budget"
      },
      {
        "count": 12,
        "value": "Resort and Spa"
      },
      {
        "count": 9,
        "value": "Luxury"
      },
      {
        "count": 7,
        "value": "Boutique"
      },
      {
        "count": 5,
        "value": "Suite"
      },
      {
        "count": 4,
        "value": "Extended-Stay"
      }
    ]
  }
}
```

Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



Best Match
Price
Name

23 results found for bike



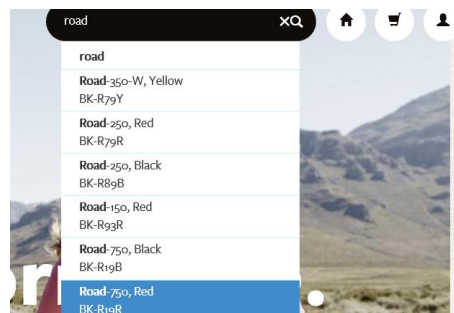
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

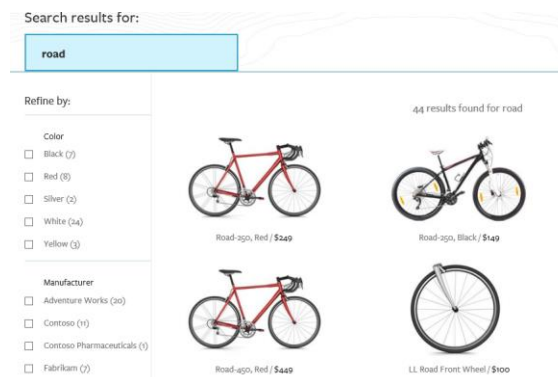
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

スコアリング

検索語に対して関連性の高さを数値化したスコア順にドキュメントを表示

スコアリングプロファイルで、スコアのチューニングが可能



23 results found for bike



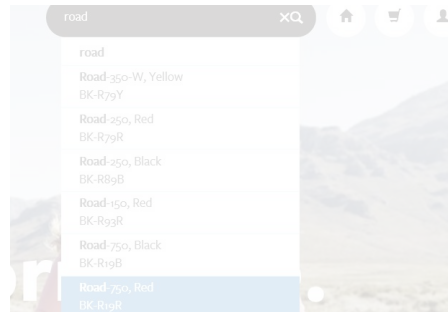
Hitch Rack - 4-Bike / \$139



Mountain Bike Socks, M / \$5

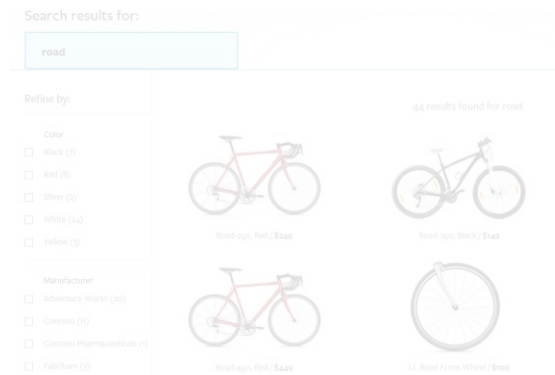
サジェスト

入力した検索文字列に対して、関連の高い検索候補の表示



ファセット

ファセットパラメータを使用して、複数の検索結果を表示



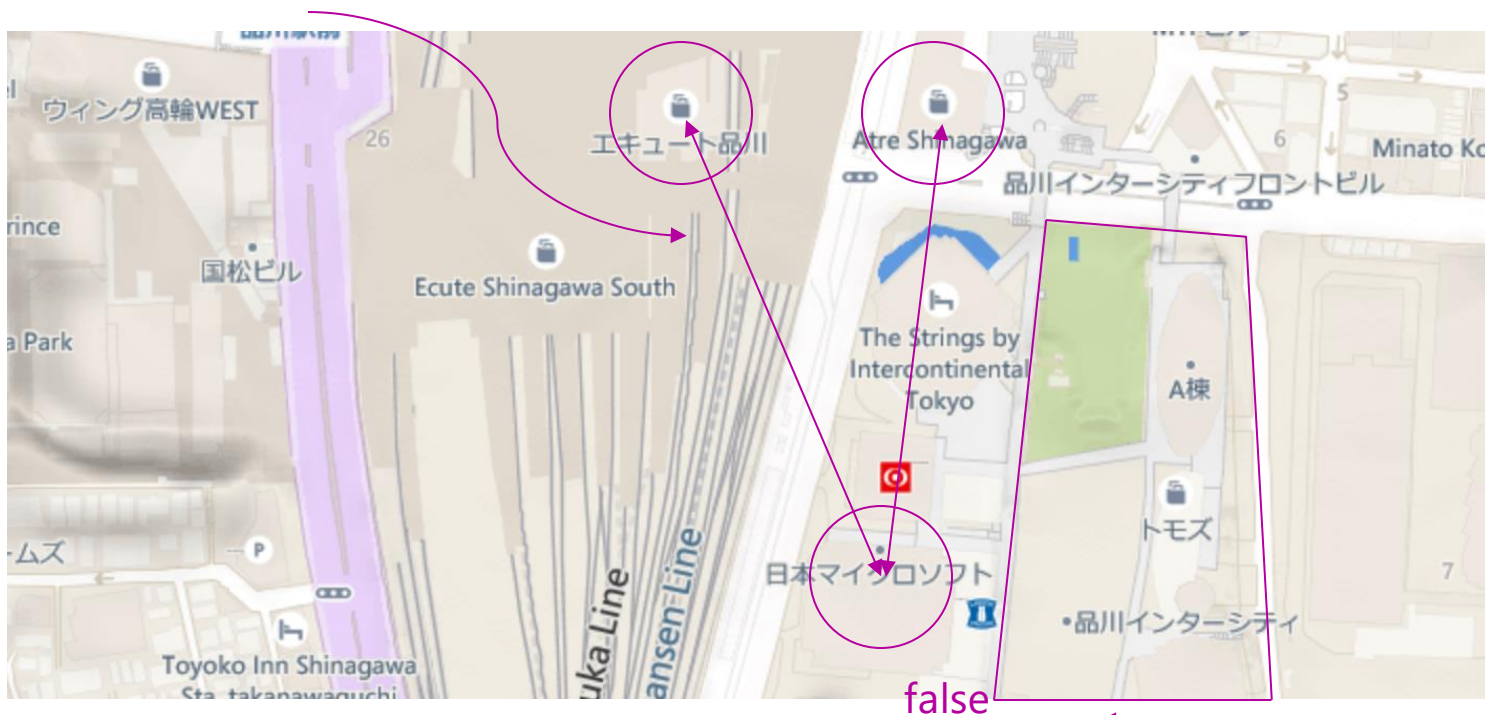
位置情報検索

位置情報に基づいた検索結果を表示



Azure Cognitive Search の検索エクスペリエンス

ユースケース①: 2点間の距離(km)



位置情報検索

位置情報に基づいた検索結果を表示



ユースケース②: ある地点と多角形との内外判定

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いものの順に
取得することができる

特徴1：検索速度

データ量が多くなっても高速かつ安定

検索エンジンの特徴

- 1. 高速かつ安定した**速度**で
- 2. **高い精度**の検索結果を
- 3. **関連性**の高いもの順に
取得することができる

特徴1：検索速度

データ量が多くなっても高速かつ安定

検索エンジンの特徴

- 1. 高速かつ安定した**速度**で
- 2. **高い精度**の検索結果を
- 3. **関連性**の高いもの順に
取得することができる

転置インデックス：トークンからドキュメントを引き当てるデータ構造

特徴1：検索速度

データ量が多くなっても高速かつ安定

検索エンジンの特徴

- 1. 高速かつ安定した**速度**で
- 2. **高い精度**の検索結果を
- 3. **関連性**の高いもの順に
取得することができる

転置インデックス：トークンからドキュメントを引き当てるデータ構造

Doc#	ドキュメント内容
1	Microsoft is introducing SQL Server
2	Windows Server on Azure
3	Microsoft is introducing Azure
4	Application programming on Microsoft Azure

特徴1：検索速度

データ量が多くなっても高速かつ安定

検索エンジンの特徴

- 1. 高速かつ安定した**速度**で
- 2. **高い精度**の検索結果を
- 3. **関連性**の高いもの順に
取得することができる

転置インデックス：トークンからドキュメントを引き当てるデータ構造

Doc#	ドキュメント内容
1	Microsoft is introducing SQL Server
2	Windows Server on Azure
3	Microsoft is introducing Azure
4	Application programming on Microsoft Azure

テキスト解析
インデクシング

特徴1：検索速度

データ量が多くなっても高速かつ安定

検索エンジンの特徴

- 1. 高速かつ安定した**速度**で
- 2. **高い精度**の検索結果を
- 3. **関連性**の高いもの順に
取得することができる

転置インデックス：トークンからドキュメントを引き当てるデータ構造

Doc#	ドキュメント内容
1	Microsoft is introducing SQL Server
2	Windows Server on Azure
3	Microsoft is introducing Azure
4	Application programming on Microsoft Azure




単語（トークン）	含まれるドキュメント
microsoft	1, 3, 4
introducing	1, 3
sql	1
server	1, 2
Windows	2
azure	2, 3, 4
application	4
programming	4

特徴2：検索結果の精度

精度が高い =>
文字のマッチング

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2.  **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる


特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2.  **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

「京都」⇒

✖ 「東**京都**庁」

「**京都**観光」

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

「京都」⇒

✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

「京都」⇒


✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

「サイゼリア」「キャノン」「ビッグカメラ」
⇒「サイゼリヤ」「キャノン」「ビックカメラ」もヒットさせたい

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2.  **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

「京都」⇒

✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

「サイゼリア」「キャノン」「ビッグカメラ」

⇒「サイゼリヤ」「キャノン」「ビックカメラ」もヒットさせたい

検索精度向上のための主要技術・ソリューション

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

「京都」⇒

✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

「サイゼリア」「キャノン」「ビッグカメラ」

⇒「サイゼリヤ」「キャノン」「ビックカメラ」もヒットさせたい

検索精度向上のための主要技術・ソリューション

- N-gram
- 形態素解析

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

「京都」⇒

✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

「サイゼリア」「キャノン」「ビッグカメラ」

⇒「サイゼリヤ」「キャノン」「ビックカメラ」もヒットさせたい

検索精度向上のための主要技術・ソリューション

- N-gram
- 形態素解析

- 語幹変化
- 見出し語変化
- 同義語展開
- 正規化
- ストップワード除去
- アンチフレージング

特徴2：検索結果の精度

精度が高い =>

× 文字のマッチング

○ 正しい言葉を理解してヒットさせる

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

「京都」⇒

✖ 「東**京都**庁」
「**京都**観光」

「キング」⇒

「バーガー**キング**」「**キング**牧師」「**キング**ダム」
✖ 「コインパー**キング**」「クック**キング**」「ステー**キング**」

「サイゼリア」「キャノン」「ビッグカメラ」

⇒「サイゼリヤ」「キャノン」「ビックカメラ」もヒットさせたい

検索精度向上のための主要技術・ソリューション

- N-gram
- 形態素解析

- 語幹変化
- 見出し語変化
- 同義語展開
- 正規化
- ストップワード除去
- アンチフレージング

- スpellチェック
- クエリサジェスト
- ファセット

特徴3：検索結果の関連性

関連性 (ランキング) によるソート

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

特徴3：検索結果の関連性

検索エンジンの特徴

1. 高速かつ安定した**速度**で
2. **高い精度**の検索結果を
3. **関連性**の高いもの順に
取得することができる

関連性 (ランキング) によるソート

- **クエリとドキュメントの関連性**を評価してソート
- データベースのORDER BY句による結果ソートは異なる評価手法

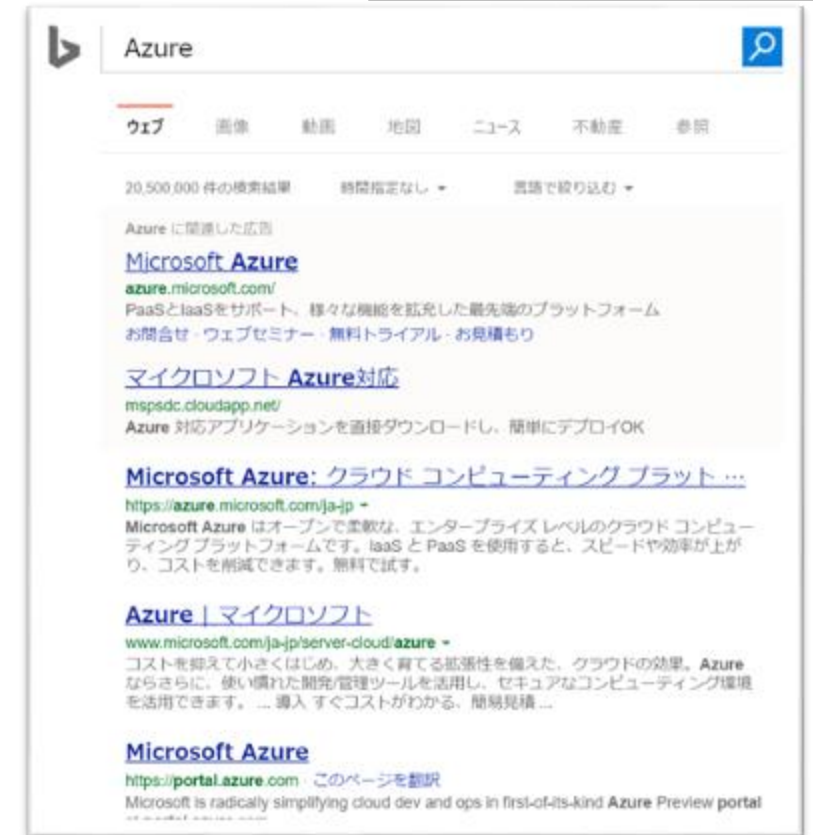
特徴3：検索結果の関連性

検索エンジンの特徴

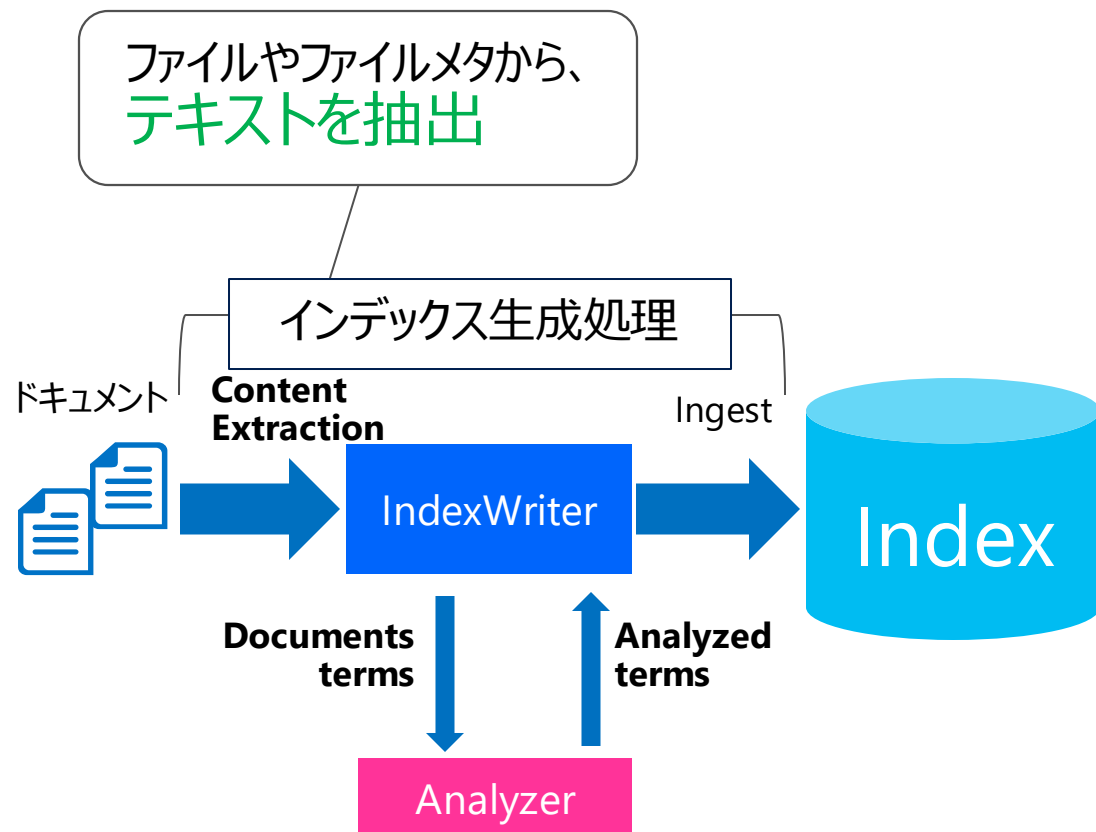
1. 高速かつ安定した速度で
2. 高い精度の検索結果を
3. 関連性の高いもの順に取得することができる

関連性 (ランキング) によるソート

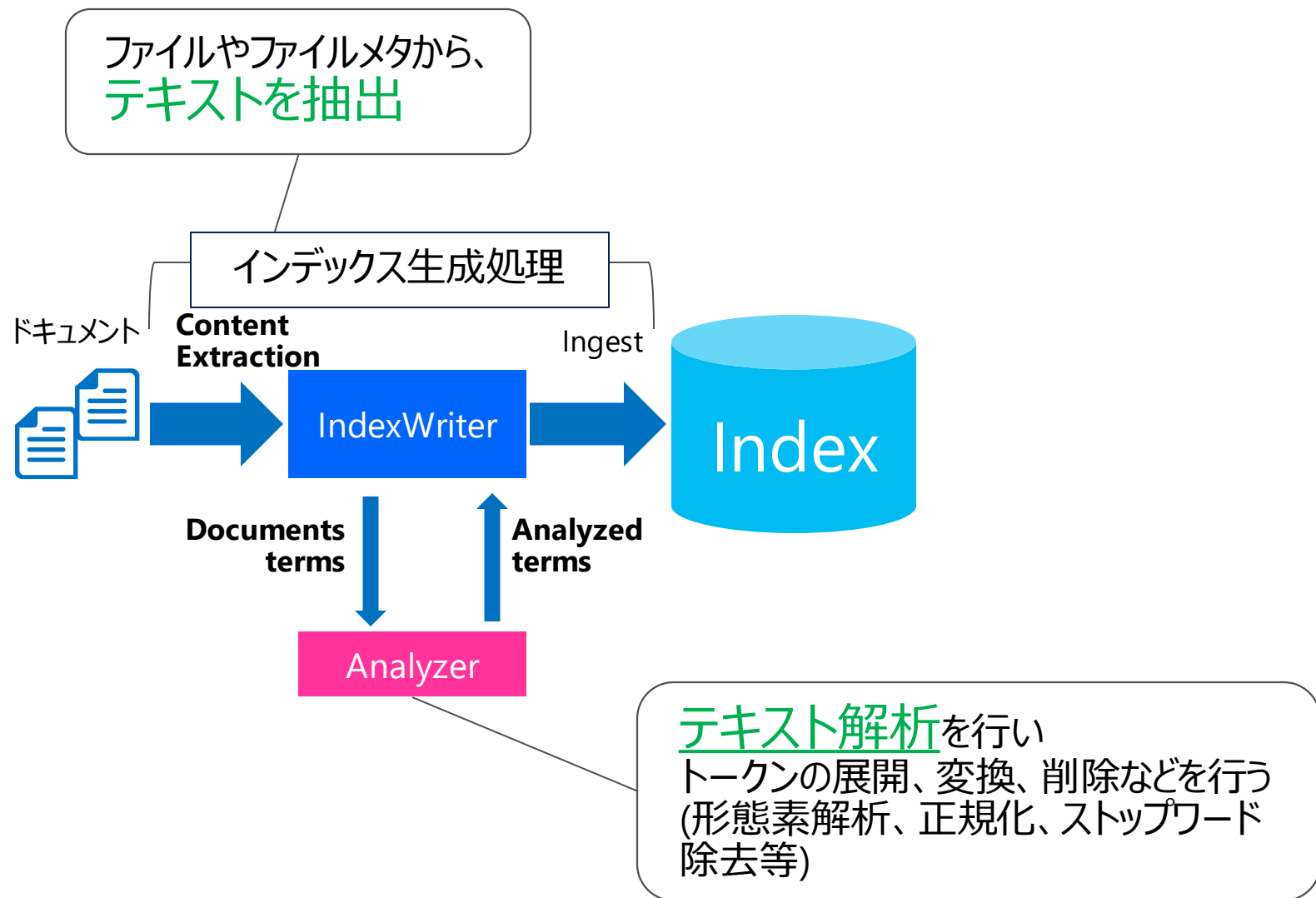
- クエリとドキュメントの関連性を評価してソート
- データベースのORDER BY句による結果ソートは異なる評価手法



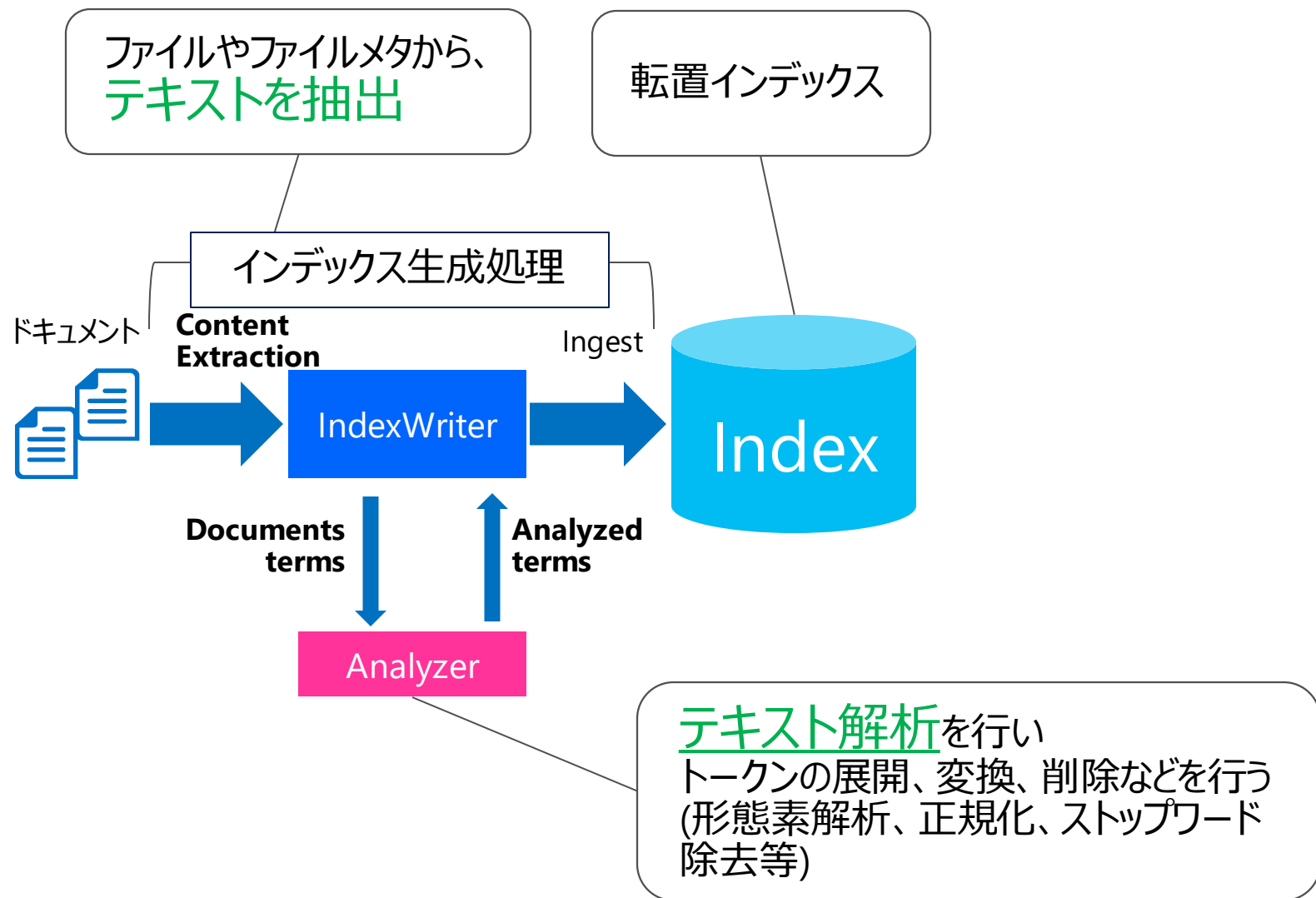
検索の全体処理フロー



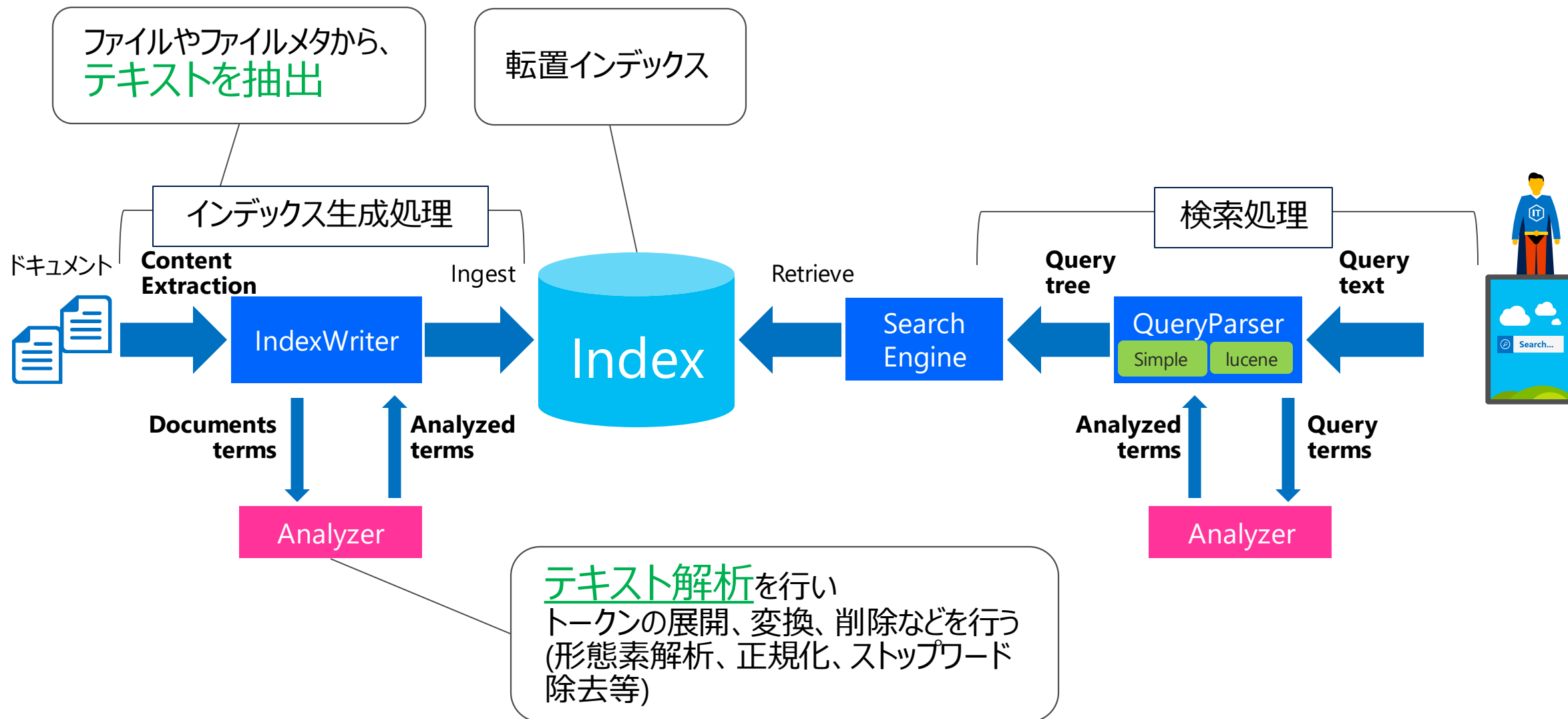
検索の全体処理フロー



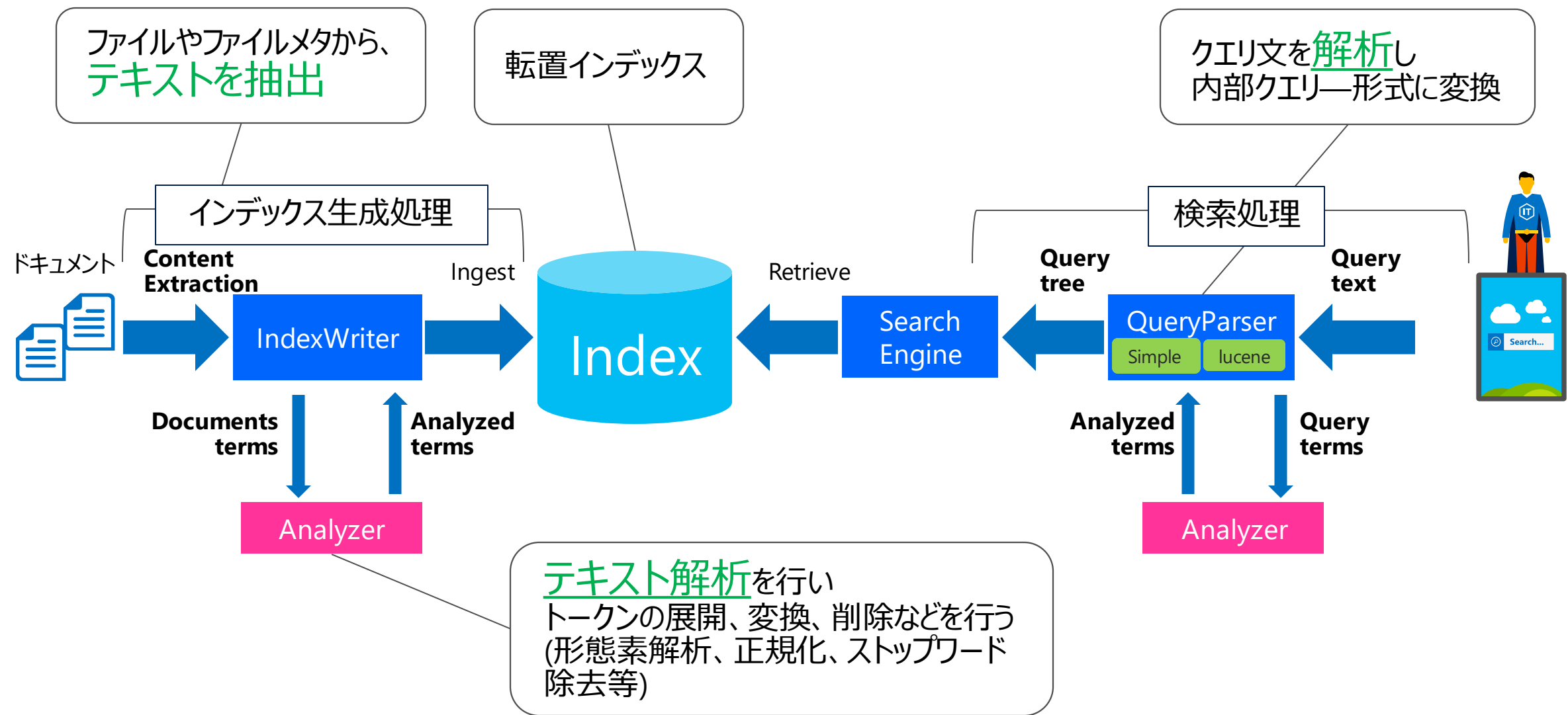
検索の全体処理フロー



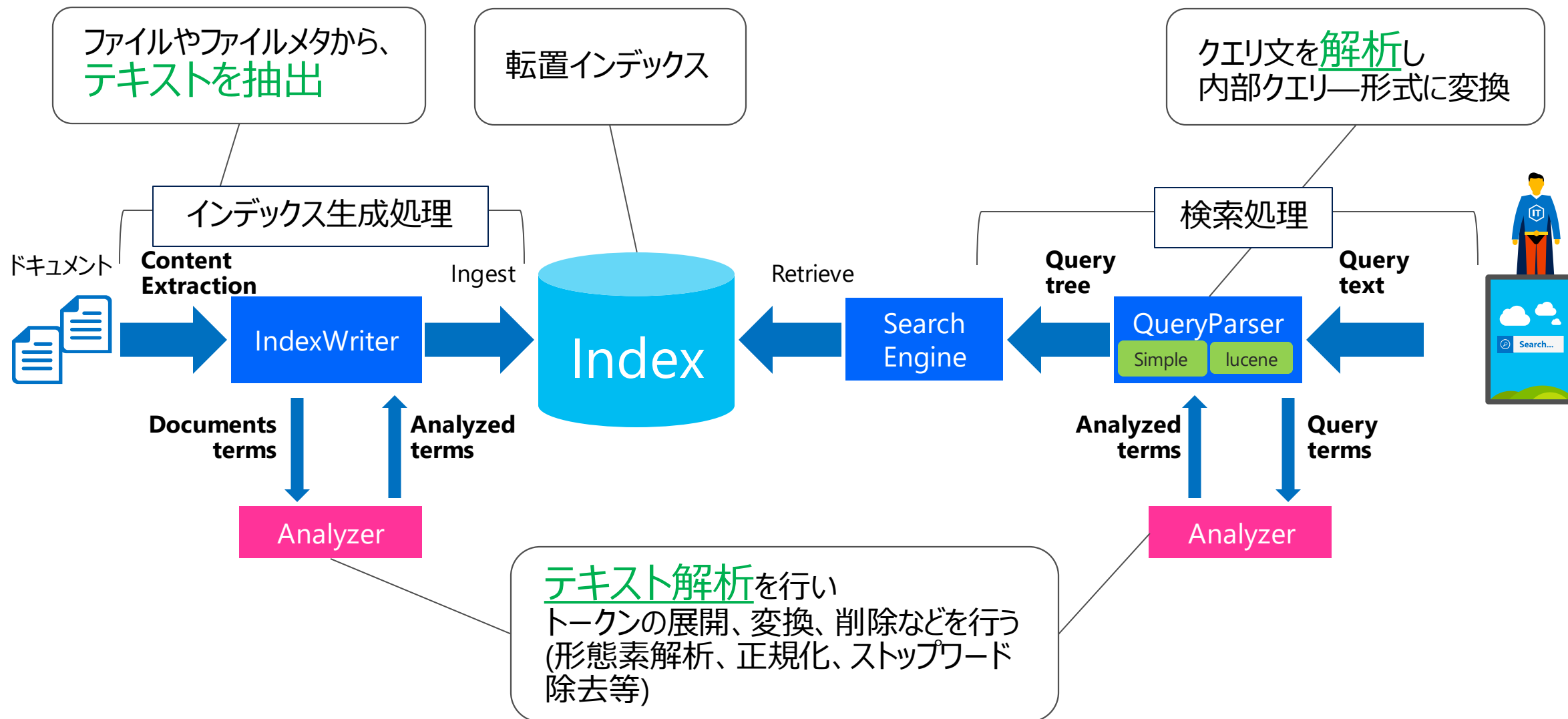
検索の全体処理フロー



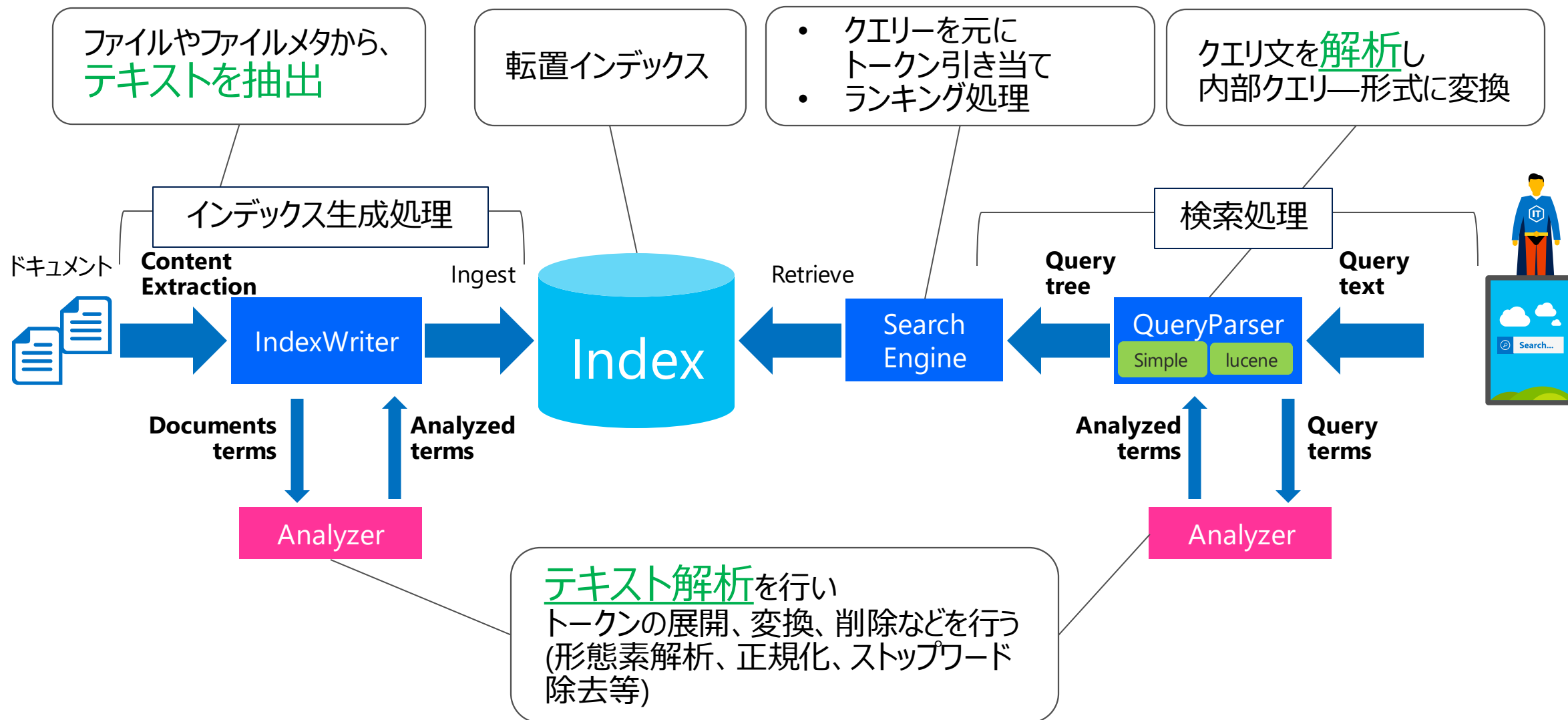
検索の全体処理フロー



検索の全体処理フロー



検索の全体処理フロー



インデックス作成：従来の検索の場合

ファイルやファイルメタから、**テキストを抽出**

転置インデックス

- クエリを元にトークン引き当て
- ランキング処理

クエリ文を**解析**し
内部クエリ形式に変換

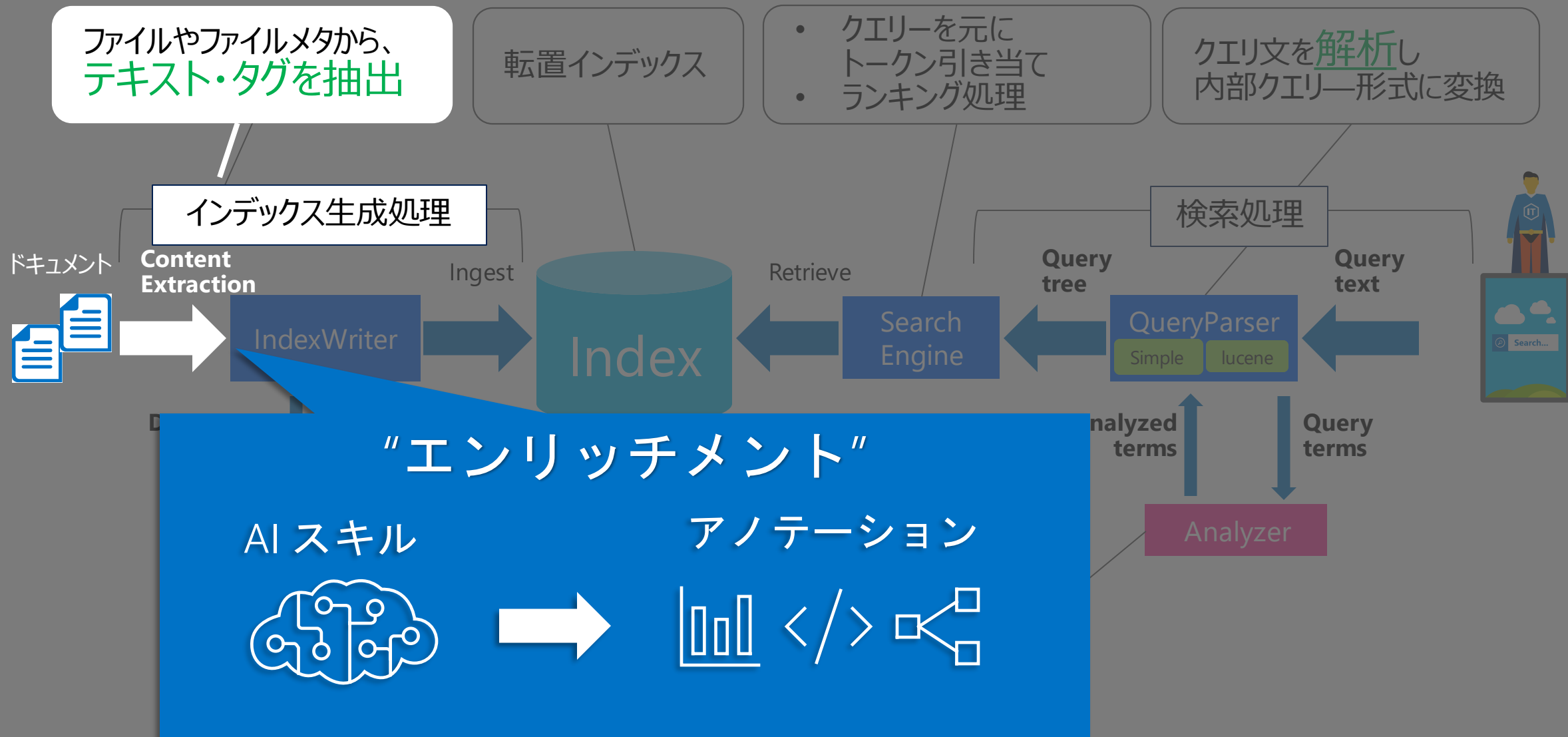
Doc#	ドキュメント内容
1	Microsoft is introducing SQL Server
2	Windows Server on Azure
3	Microsoft is introducing Azure
4	Application programming on Microsoft Azure

テキスト解析
インデクシング

単語（トークン）	含まれるドキュメント
microsoft	1, 3, 4
introducing	1, 3
sql	1
server	1, 2
windows	2
azure	2, 3, 4
application	4
programming	4

トークンの展開、変換、削除などを行う

インデックス作成：AI 連携の場合



Azure Cognitive Search の AI 連携

Cognitive Services と連携し、文書や画像からテキストやタグ情報を自動抽出
=> 検索時には、**自動抽出された情報を検索対象に**

文書・画像

Cognitive skills

一括テキスト検索と
絞り込み




キーフレーズ抽出


ロケーション情報抽出


感情分析


組織情報抽出


人情報抽出

Aあ
言語特定


顔特定

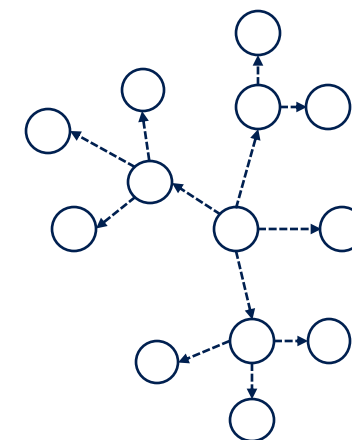

有名人認識

タグ抽出


カスタム スキルの実装


ランドマーク特定


印刷文字認識



REST API による Cognitive Search の操作

REST API インターフェース

<https://<アカウント名>.search.windows.net>

インデックス作成・更新	/indexes/<indexname>	PUT
インデックス取得	/indexes/<indexname>	GET
インデックス統計情報取得	/indexes/<indexname>/stats	GET
インデックスの削除	/indexes/<indexname>	DELETE
ドキュメント追加・更新・削除	/indexes/<indexname>/docs/index	POST
ドキュメントの検索	/indexes/<indexname>/docs	GET
ドキュメントの参照	/indexes/<indexname>/docs/<key>	GET
ドキュメント数取得	/indexes/<indexname>/docs/\$count	GET
テキストの分析	/indexes/<indexname>/analyze	POST
シノニムマップの作成	/synonymmaps	POST
シノニムマップの更新	/synonymmaps/<synonymmapname>	PUT
シノニムマップの削除	/synonymmaps/<synonymmapname>	DELETE
シノニムマップの取得	/synonymmaps/<synonymmapname>	GET
....		

レスポンス(JSON)

```
{
  "@odata.context":
    "https://yoichikademo.search.windows.net/indexes('messages')/$metadata#Collection(Microsoft.Azure.Search.V2016_09_01.IndexResult)",
  "value": [
    { "errorMessage": null, "key": "1", "status": true,
      "statusCode": 201 },
    { "errorMessage": null, "key": "2", "status": true,
      "statusCode": 201 },
    { "errorMessage": null, "key": "3", "status": true,
      "statusCode": 201 }
  ]
}
```

API による検索イメージ

エンドポイント

<https://<アカウント名>.search.windows.net/indexes/<インデックス名>/docs>

\$filterパラメータ

- 絞込み用
- アナライザとランキング処理共に無し

?api-version
=2020-06-30

共通

&search="xxx"
&searchMode=all
&queryType=full

&\$filter
= xxx

&\$count=true
&\$top=5
&\$skip=10
&\$select=title,speaker
&\$orderby=level desc
&facet=tag
&highlight=title

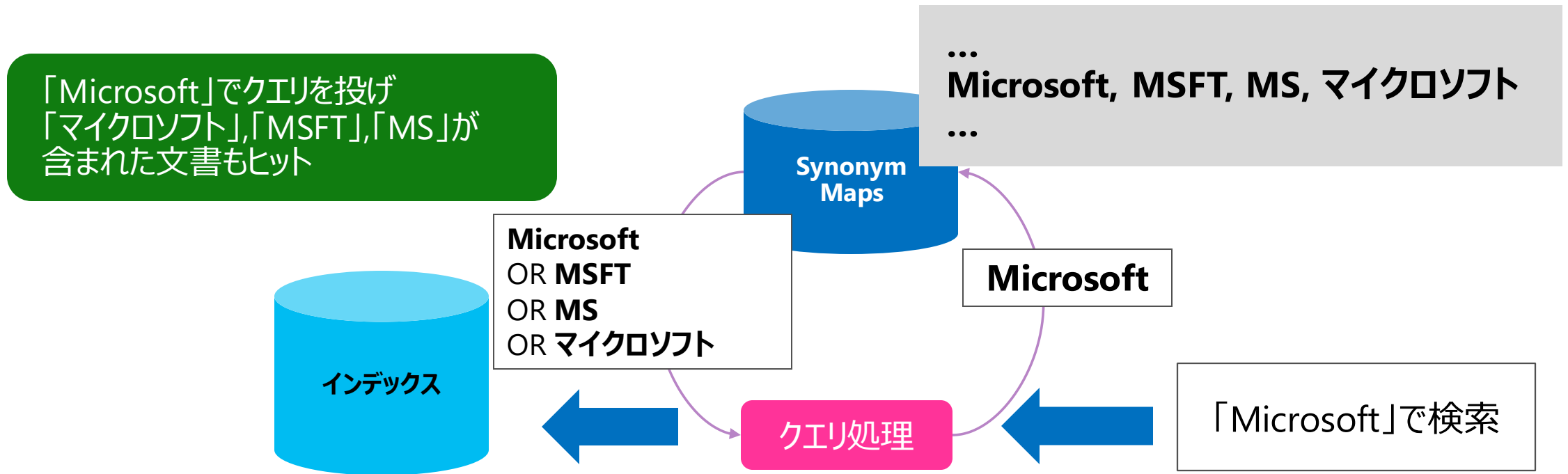
結果表示用

search関連パラメータ	
search	<ul style="list-style-type: none">クエリ文字列クエリ文字列にアナライザー処理ランキング処理
searchMode	<ul style="list-style-type: none">クエリ評価方法を決定(all any)
queryType	<ul style="list-style-type: none">クエリパーサーを選択 (simple full)

検索精度向上の施策：シノニムマップ

同義語、類義語展開などが実現可能

- インデックスに対して設定することで、**クエリ実行時**の同義語展開が行われる



精度向上の施策：ランキングとスコアリングプロファイル

- クエリに対する関連性の数値化
= スコアリング
- ランクスコアの構成

TF-IDF
ベース
のスコア



スコアリング
プロファイルによる
調整



ランクスコア

```
"value": [  
  {  
    "@search.score": 1.3097656,  
    "question": "Azure では IPv6 がサポートさ  
  },  
  {  
    "@search.score": 1.3056309,  
    "question": "Azure サポートの購入方法は?"  
  },  
  {  
    "@search.score": 1.2120581,  
    "question": "Azure サポート特典を使用して
```

ランキングのチューニングは
スコアリングプロファイルで行う

スコアリングプロファイルの設定

特定の基準に基づいてブーストすることで、
ランクスコアをチューニング

Magnitude

- 数値の大小に基づいたブースト

Freshness

- 日付の新しさに基づいたブースト

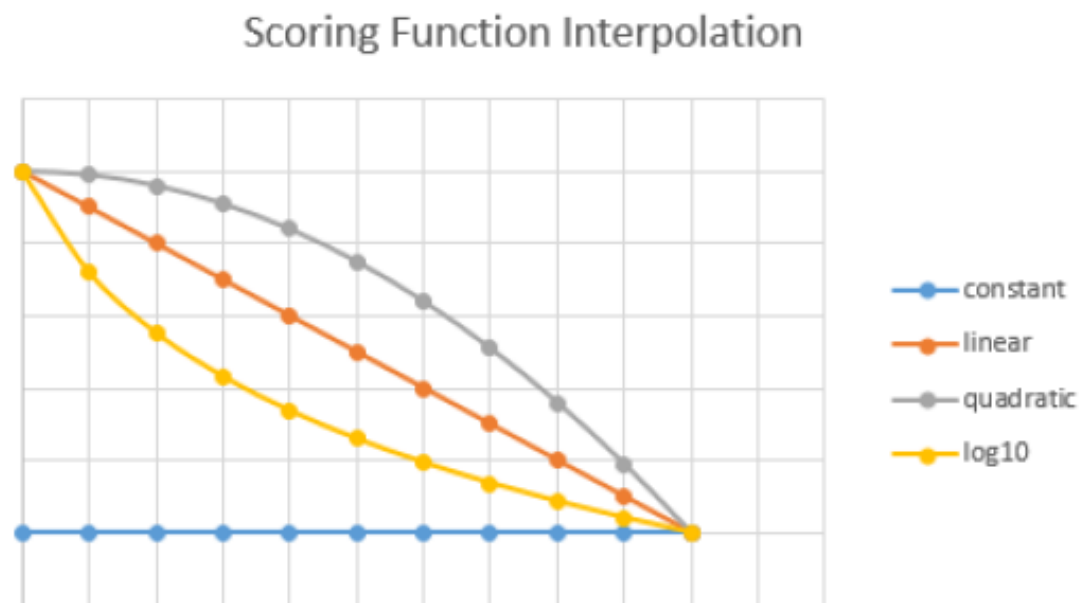
Distance

- 地理的な距離に基づいたブースト

Tags

- 付与したタグに基づいたブースト

4種類の補完曲線により、ブーストの傾きを調整



JFK Files Demo



John F. Kennedy (JFK)

November 22, 1963

<http://aka.ms/jfkfiles>

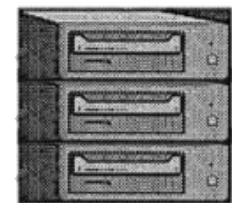
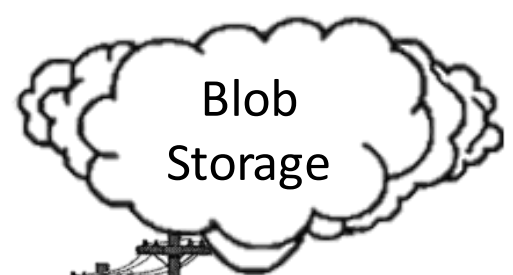
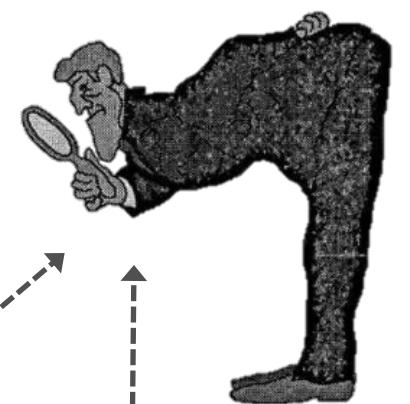
JFK FILES
COGNITIVE SEARCH
ARCHITECTURE

DECLASSIFIED



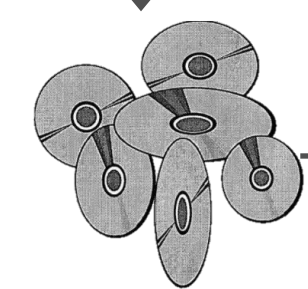
Custom skill: CIA Cryptonyms
Cognitive Services:
Computer Vision
OCR + Handwriting
Entity Linking

Azure Cognitive Search



Azure Function

Cognitive
skill set



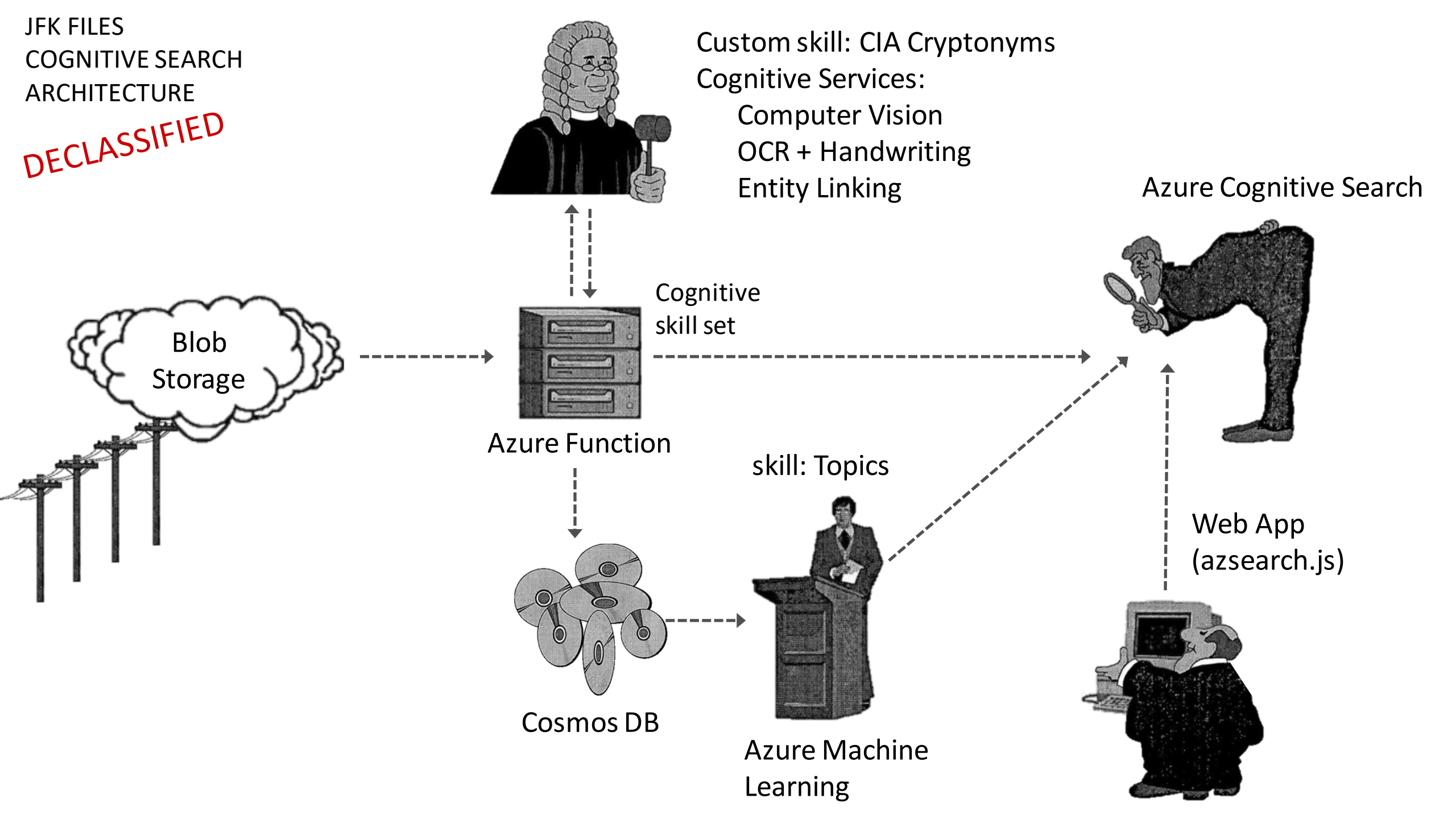
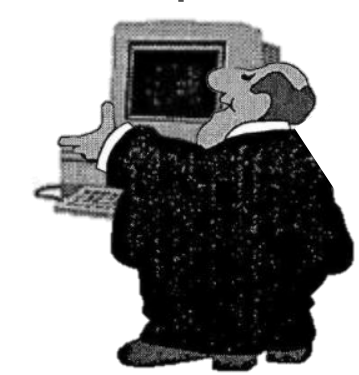
Cosmos DB

skill: Topics



Azure Machine
Learning

Web App
(azsearch.js)



Job Portal Demo

SEARCH



Any distance from ▾

10001

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)



2558 AVAILABLE JOBS

Relevance ▾

< 1 2 3 4 5 >



5
Nov

Business Analyst

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 resources framework within MIS. Primary responsibilities include: conducting detail

<https://azjobsdemo.azurewebsites.net/>

Private

Another Knowledge Mining Demo

Knowledge Mining Demo

Microsoft

エンティティマップを見る

結果件数：2 件

現在適用している絞り込み

タグ

Microsoft

Microsoft Corporation

結果に含まれるタグ

企業・組織

車種

キーフレーズ

ファイルの種類

☐ PDF

☐ JPEG

☐ PNG

manual 1

タグ：TE FXCSV Adobe NAA IBM Nissan Microsoft Oracle NISSAN PROFIT その他 セダン バス image 検査入力起動 入力方法 手入力 タッチ入力 キー入力 文字入力 手書き入力 検査情報 車両情報 車両検査結果 日産車両状態証明書発行システム検査入力編 目次 数車両 車両諸元情報CSV

[image: image169.jpg] DURL Client VIP Access HUND/18 EPSON IBM DB2 HISSAH Nissan Menu a KeySQL for ORACLE KeySQL @ Live Picture Office 10#/!F5% Microsoft Office Office (*1BR< Microsoft Virtual Server Oracle - OraHome92 Windows Update Oracle Installation Products PROFIT Windows 150 Q

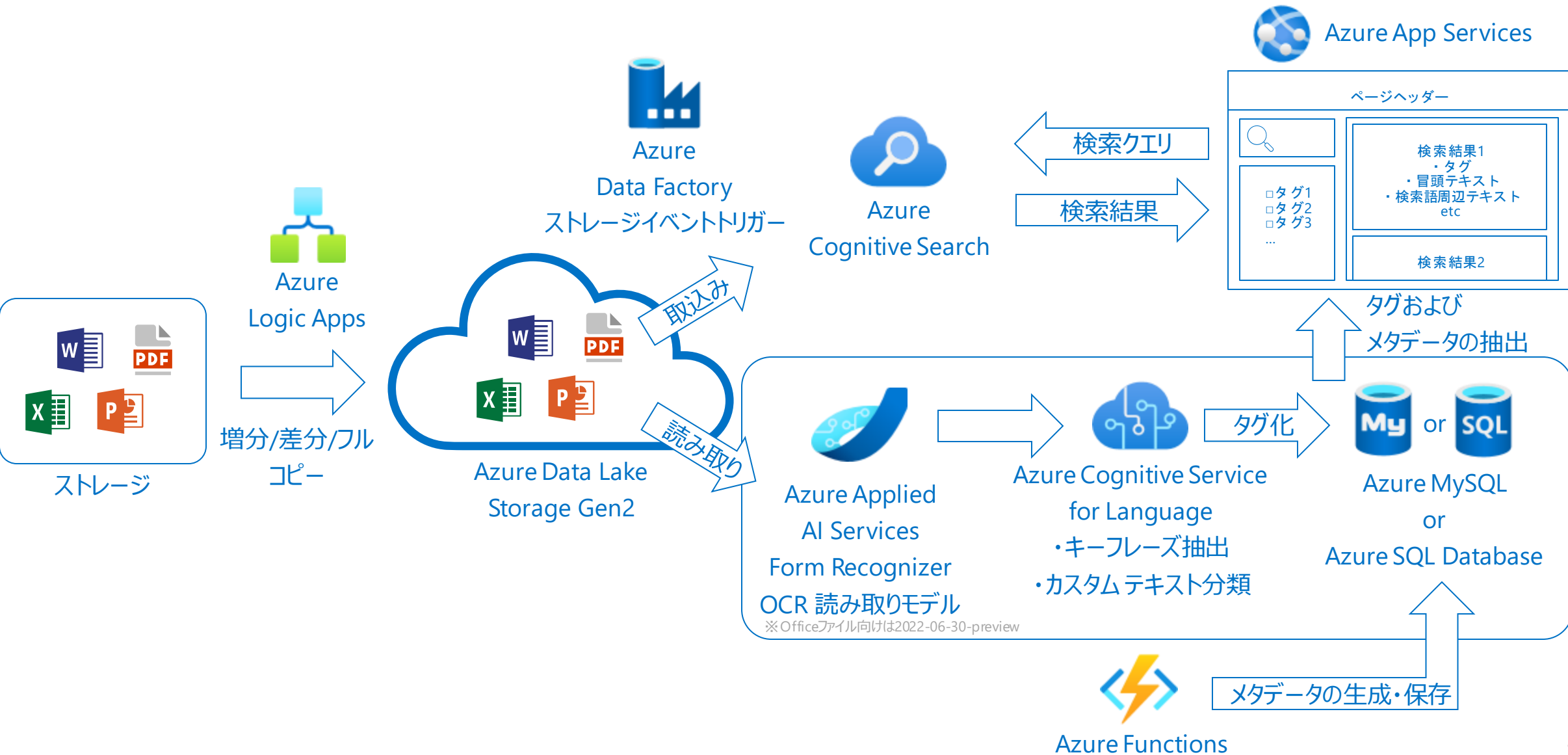
operation com

タグ：Microsoft Adobe Microsoft Corporation アドビシステムズ その他 パスワード変更画面 キーワード変更画面 ログインパスワード 画面説明 ログイン画面 キーワード延長確認画面 キーワード延長結果画面 キーワード延長申込画面 管理ユーザー 編集画面 請求先情報変更画面 メニュー画面

・ Windows 8.1 Update：Internet Explorer 11 ・ Windows 10：Internet Explorer 11, Microsoft Edge ・ Adobe Acrobat Reader DC （注）Windows、Internet Explorer および Microsoft Edge は、米国 Microsoft Corporation の、米国およびその他の

ドキュメント検索システムへの 適用イメージ

テキストとタグからドキュメントを抽出するMVP方式案





本書に記載した情報は、本書各項目に関する発行日現在の Microsoft の見解を表明するものです。Microsoft は絶えず変化する市場に対応しなければならないため、ここに記載した情報に対していかなる責務を負うものではなく、提示された情報の信憑性については保証できません。

本書は情報提供のみを目的としています。Microsoft は、明示的または暗示的を問わず、本書にいかなる保証も与えるものではありません。

すべての当該著作権法を遵守することはお客様の責務です。Microsoft の書面による明確な許可なく、本書の如何なる部分についても、転載や検索システムへの格納または挿入を行うことは、どのような形式または手段（電子的、機械的、複写、レコーディング、その他）、および目的であっても禁じられています。これらは著作権保護された権利を制限するものではありません。

Microsoft は、本書の内容を保護する特許、特許出願書、商標、著作権、またはその他の知的財産権を保有する場合があります。Microsoft から書面によるライセンス契約が明確に供給される場合を除いて、本書の提供はこれらの特許、商標、著作権、またはその他の知的財産へのライセンスを与えるものではありません。

© 2022 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.