KNOWLEDGE MINING

KNOWLEDGE MINING

- Search and learn from vast amounts of information.

- Searching for information online has never been easier
- It is difficult to find information from documents that are not in a search index.

- Knowledge mining is the term used to describe solutions that involve extracting information from large volumes of often unstructured data.

AZURE COGNITIVE SEARCH

- Private, enterprise, search solution that has tools for building indexes.

- The indexes can then be used for internal only use, or to enable searchable content on public facing internet assets.

- Azure Cognitive Search can utilize the built-in AI capabilities of Azure Cognitive Services such as
  - image processing
  - content extraction
  - natural language processing to perform knowledge mining of documents.

- The product's AI capabilities makes it possible to index previously unsearchable documents and to extract and surface insights from large amounts of data quickly.

WHAT IS AZURE COGNITIVE SEARCH?

- It provides the infrastructure and tools to create search solutions that **extract data from** a variety of structured, semi-structured, and non-structured documents.

- Azure Cognitive Search **results contain** only your data, which can include text inferred or extracted from images, or new entities and key phrases detection through text analytics.

- It's a Platform as a Service (PaaS) solution so Microsoft manages the infrastructure and availability, allowing your organization to benefit without the need to purchase or manage additional hardware resources.

AZURE COGNITIVE SEARCH FEATURES

- Data from any source
- Full text search and analysis
- AI powered search
- Multi-lingual
- Geo-enabled
- Configurable user experience

IDENTIFY ELEMENTS OF A SEARCH SOLUTION

- Data source that contains the data artifacts you want to search

- After extracting the basic data and metadata fields from the source, you can define a skillset that applies a sequence of AI skills to enrich the data.

- An indexer is used to run the process of extracting the source data and applying the AI skills to iteratively build a hierarchical JSON structure of fields for each document.

- The fields extracted by the indexer are persisted in an index, which can be searched by client applications.

- The fields are used for searching, filtering, and sorting to generate a set of results that can be displayed or otherwise used by the client application.

- Optionally, you can also create a knowledge store, which stores output from an AI enrichment pipeline in tables and blobs in Azure Storage for independent analysis or downstream processing.

CREATE AN INDEX IN THE AZURE PORTAL

- Before using the Azure Search service, you'll first need to extract your data to an Azure data source.

- Supported data storage sources include:
  - Azure SQL Database
  - SQL Server on an Azure VM
  - Cosmos DB
  - Azure Blob storage
  - Azure Table storage

USING THE AZURE PORTAL'S IMPORT DATA WIZARD

- Data Source
- Index
- Indexer
- Skillset
- Knowledge store

- To use Azure Cognitive Search, you'll need an Azure Cognitive Search resource

QUERY DATA IN AN AZURE COGNITIVE SEARCH INDEX

- After we build the index, we can perform queries.

- The schema of the index determines what queries can be answered.

- **Queries** can be submitted as an **HTTP or REST API request**

- **Response** coming back as **JSON**.

- Queries can specify
  - what fields are searched and returned
  - how search results are shaped
  - how the results should be filtered or sorted.

- A query that doesn't specify the field to search will execute against all the searchable fields within the index.

- Azure Cognitive Search supports two types of syntax:
  - **Simple syntax** covers all of the common query scenarios
  - **Full Lucene** is useful for advanced scenarios.

SIMPLE QUERY REQUESTS

- coffee (-"busy" + "wifi")