CS410 Technology Review

Azure Search – Explore And Analyze Azure Cognitive Search That Is With Built-in AI Capabilities

Guoyong Yang (guoyong2)

Introduction

Azure Cognitive Search is a cloud search service provided by Microsoft to help users especially developers to build rich search experience over structured and unstructured data from web, mobile and enterprise applications.

Microsoft Azure uses Natural Language Processing (NLP), AI, Apache Lucene, and Language Understanding (LUIS) to implement search abilities on text retrieval and text analysis.

The knowledge mining includes:

- Create Azure resources
- Extract data from a data source
- Enrich data with AI skills
- Use Azure's indexer in the Azure portal
- Query your search index
- Review results saved to a Knowledge Store

Body

Azure cognitive search is a cloud search service on knowledge mining that has tools for building user-managed indexes upon large volumes of unstructured data. It's a Platform as a Service (PaaS) solution. It can quickly identify the main points in unstructured text to get a list of relevant phrases that best describe a passage using key phrase extraction. And it can identify sentences that best convey the main idea of a document with extractive summarization.

The data can be text inferred or extracted from images, or new entities and key phrases detection through text analytics.

Key Phrase Extraction is to use a pre-trained model to detect important phrases based on term placement, linguistic rules, proximity to other terms, and how unusual the term is within the source data.

Text Translation Skill is to use a pre-trained model to translate the input text into various languages for normalization or localization use cases.

Image processing skills is to create text representations of image content, making it searchable using the query capabilities of Azure Cognitive Search.Optical Character Recognition Skill

Azure Cognitive Search comes with the following features:

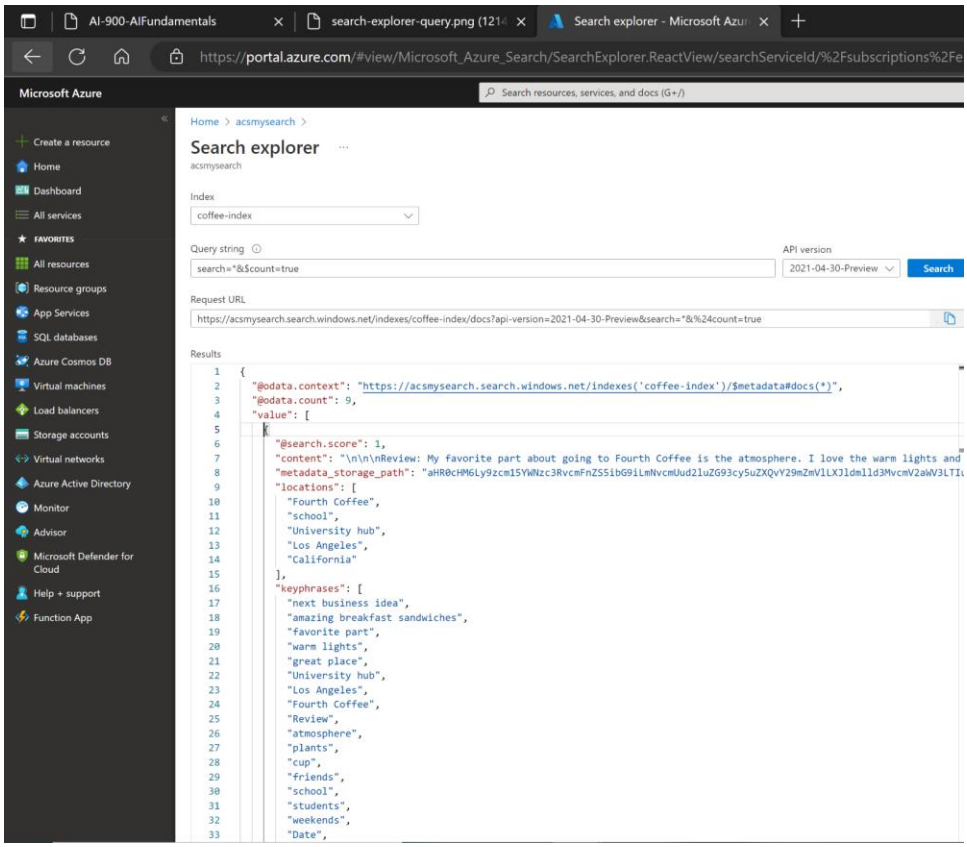
- Data from any source: Azure Cognitive Search accepts data from any source provided in JSON format, with auto crawling support for selected data sources in Azure.
- Full text search and analysis: Azure Cognitive Search offers full text search capabilities supporting both simple query and full Lucene query syntax.
- AI powered search: Azure Cognitive Search has Cognitive AI capabilities built in for image and text analysis from raw content.
- Multi-lingual: Azure Cognitive Search offers linguistic analysis for 56 languages to intelligently handle phonetic matching or language-specific linguistics. Natural language processors available in Azure Cognitive Search are also used by Bing and Office.
- Geo-enabled: Azure Cognitive Search supports geo-search filtering based on proximity to a physical location.
- Configurable user experience: Azure Cognitive Search has several features to improve the user experience including autocomplete, autosuggest, pagination, and hit highlighting.
- The data format that Cognitive Search supports is JSON.
- indexers also support AI enrichment. You can attach a skillset that applies a sequence of AI skills to enrich the data, making it more searchable.

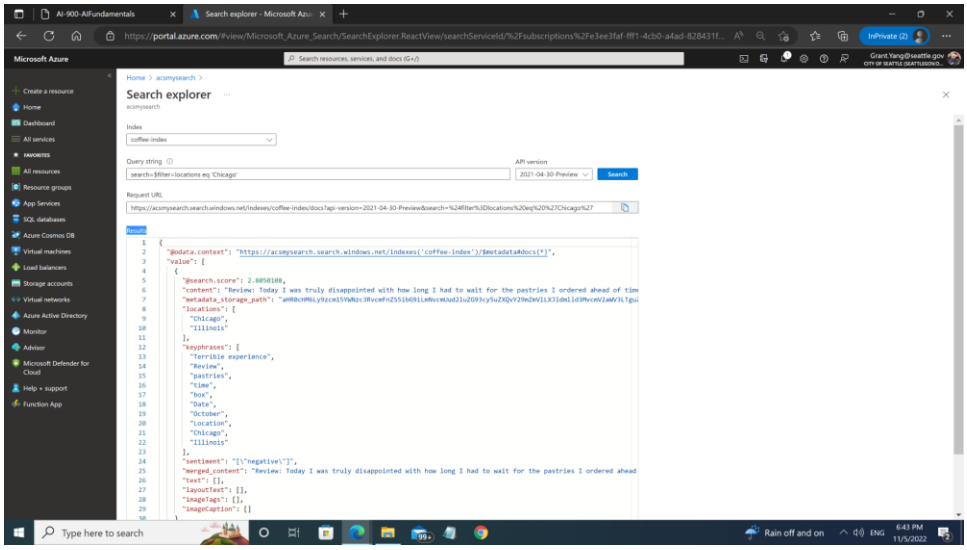
To practice the azure cognitive search, I follow the url below to practice it:

Introduction to Azure Cognitive Search - Training | Microsoft Learn

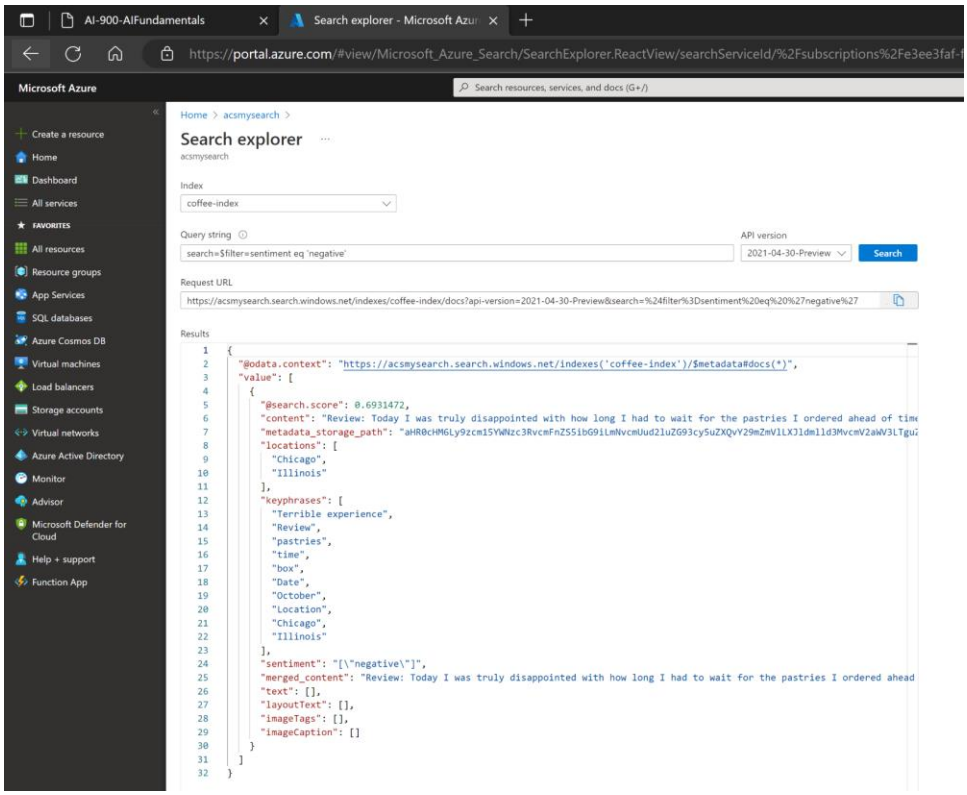
The data includes unstructured, typed, image-based, and hand-written documents. It used indexer to build an index in portal. Using pull method, I loaded data with an indexer. With NLP feature the unstructured data is mapped as searchable and filterable fields in an index. Then, I can run some search queries to get the results as snapshot below:

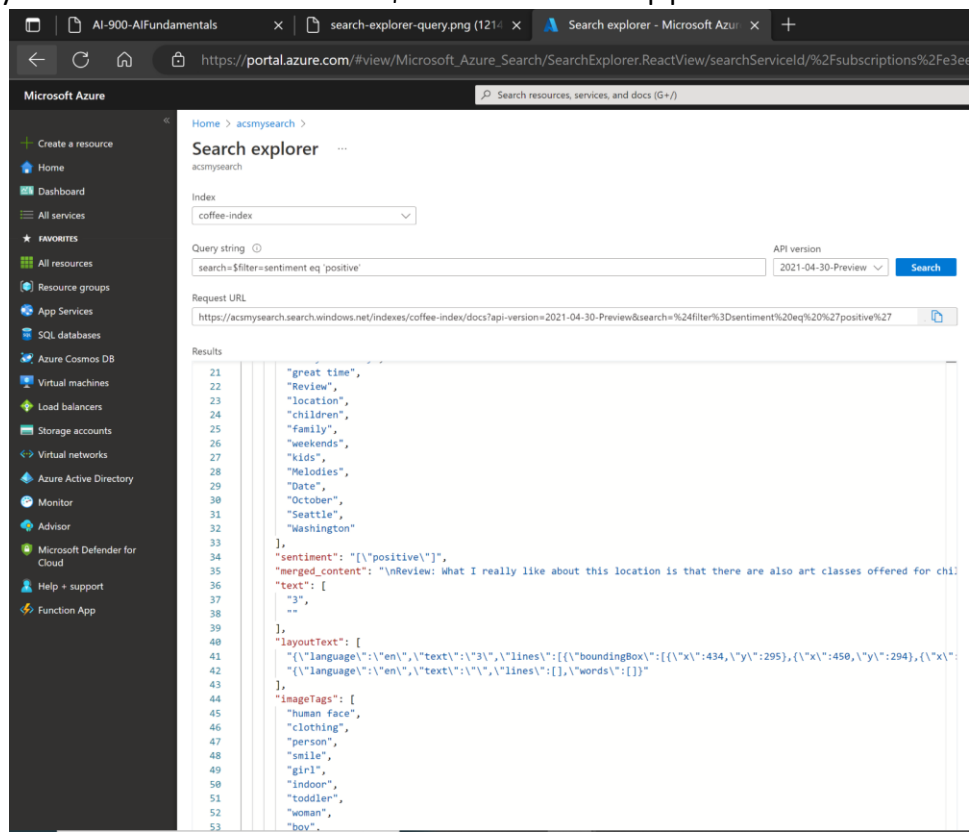
- Query - search=*&$count=true

- 

  - Filter by location - search=$filter=locations eq 'Chicago'
- 

  - filter by sentiment - search=$filter=sentiment eq 'negative'

- 

  - filter by sentiment Positive - search=$filter=sentiment eq 'positive'
- 

Conclusion

Azure Cognitive Search uses current modern text retrieval methodologies to provide customers very broader abilities to search on data especially unstructured data. It can be used to identifies, categorizes, and redacts sensitive information, language detection, sentiment analysis, opinion mining, summarization, key phrase extraction, entity linking, text analytics for health, custom text classification, build custom AI models to extract custom entity categories, conversational language understanding (CLU), question answering, etc.

In additional to Azure Cognitive Search, Azure also provide features worth to explore for data mining that includes Azure Synapse Analytics, Azure HDInsight, Azure Databricks.

References

[1] Sebastião Pais, João Cordeiro & M. Luqman Jamil. NLP-based platform as a service: a brief review.

https://link.springer.com/article/10.1186/s40537-022-00603-5

[2] Microsoft. What is Azure Cognitive Service for Language?

https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/overview

[3] Microsoft. Text analytics
https://azure.microsoft.com/en-us/products/cognitive-services/text-analytics/#documentation

[4] Microsoft. Natural language processing technology

https://learn.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/natural-language-processing

[5] Microsoft. Introduction to Azure Cognitive Search

https://learn.microsoft.com/en-us/training/modules/intro-to-azure-search/

[6] Microsoft. Full text search in Azure Cognitive Search

https://learn.microsoft.com/en-us/azure/search/search-lucene-query-architecture