



Maestría en Inteligencia Artificial Aplicada

Proyecto Integrador

Baseline

Alumno

Johanna Rodríguez Jaramillo

Katherine Vanegas Salamanca

Marcos Chávez Chávez

Matrícula

A01794010

A01794113

A01688507

Profesores Titulares

Dra. Grettel Barceló Alonso / Luis Eduardo Falcón Morales

Profesor Asistente

Horacio Martínez Alfaro

19 de mayo de 2024

Avance proyecto

Dentro del avance durante la última semana y posterior aprobación de la solicitud para el uso de Azure OpenAI, procedimos a realizar las siguientes actividades

1. Creación grupo de recursos

Un grupo de recursos es un contenedor que almacena los recursos relacionados con una solución de Azure. El grupo de recursos puede incluir todos los recursos de la solución o solo aquellos que se desean administrar como grupo, en nuestro caso es en este contenedor “Nuevo_Prueba” donde vamos a administrar todas las instancias.

Nombre	Tipo	Ubicación
<input type="checkbox"/> iagenerativenormas	Document intelligence	East US
<input type="checkbox"/> normascables	Azure OpenAI	East US 2
<input type="checkbox"/> normascables	Cuenta de almacenamiento	East US
<input type="checkbox"/> searchnormas	Servicio de búsqueda	West US 2
<input type="checkbox"/> searchnormas2	Servicio de búsqueda	West US 2

2. Creación de cuenta de almacenamiento o *blob storage*

La plataforma de Azure Storage es la solución de almacenamiento en la nube de Microsoft para los escenarios modernos de almacenamiento de datos. Azure Storage ofrece almacenamiento de alta disponibilidad, escalable de forma masiva, duradero y seguro para una gran variedad de objetos de datos en la nube. Es aca donde reposan los documentos de las normas para poder acceder a ellos desde cualquier servicio de Azure.

Microsoft Azure

normascables

Cuenta de almacenamiento

Buscar

Información general

Registro de actividad

Etiquetas

Diagnosticar y solucionar problemas

Control de acceso (IAM)

Migración de datos

Eventos

Explorador de almacenamiento

Storage Mover

Almacenamiento de datos

Contenedores

Recursos compartidos de archivos

Colas

Tablas

Essentials

Grupo de recursos (...): [Nuevo Prueba](#)

Ubicación: eastus

Ubicación principal/se...: Principal: East US, secundario: West US

Suscripción ([mover](#)): [Azure for Students](#)

Id. de suscripción: e8b4a95c-d41f-4fa7-bd34-a3ed5cc03e97

Estado del disco: Principal: Disponible, secundario: Disponible

Etiquetas ([editar](#)): ProjectType : aoi-your-data-service

Rendimiento: Estándar

Replicación: Almacenamiento con redundancia geográfica con acceso d...

Tipo de cuenta: StorageV2 (uso general v2)

Estado de aprovisiona...: Correcto

Creado: 4/5/2024, 17:59:13

Propiedades

Supervisión

Funcionalidades (7)

Recomendaciones (0)

Tutoriales

Herramientas y SDK

Blob service

Espacio de nombres jerárquico: [Deshabilitado](#)

Nivel de acceso predeterminado: [Hot](#)

Acceso anónimo al blob: [Deshabilitado](#)

Eliminación temporal de blobs: [Habilitado \(7 días\)](#)

Eliminación temporal de contenedores: [Habilitado \(7 días\)](#)

Seguridad

Requerir transferencia segura para las operaciones de API de REST: [Habilitado](#)

Acceso a la clave de la cuenta de almacenamiento: [Habilitado](#)

Versión de TLS mínima: [Versión 1.2](#)

Cifrado de infraestructura: [Deshabilitado](#)

Microsoft Azure

normascables1

Contenedor

Buscar

Información general

Diagnosticar y solucionar problemas

Control de acceso (IAM)

Configuración

Agregar filtro

Nombre	Modificado	Nivel de acceso	Estado del archivo	Tipo de blob	Tamaño	Estado de concesión
<input type="checkbox"/> NMX-B-001-SCFI-1988 METOD...	14/5/2024, 13:43:17	Frecuente (inferido)		Blob en bloques	1.36 MiB	Disponible ***
<input type="checkbox"/> NMX-B-172-CANACERO-2018...	14/5/2024, 13:43:00	Frecuente (inferido)		Blob en bloques	1.6 MiB	Disponible ***
<input type="checkbox"/> NMX-CC-19011-IMNC-2012 DL...	14/5/2024, 13:42:59	Frecuente (inferido)		Blob en bloques	1020.56 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-036-SCFI-1994 INDIC...	14/5/2024, 13:42:43	Frecuente (inferido)		Blob en bloques	256.11 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-070-SCFI-1993 TERM...	14/5/2024, 13:42:35	Frecuente (inferido)		Blob en bloques	114.3 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-099-IMNC-2005 MIC...	14/5/2024, 13:43:10	Frecuente (inferido)		Blob en bloques	922.89 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-110-1-SCFI-1993 INS...	14/5/2024, 13:43:06	Frecuente (inferido)		Blob en bloques	144.75 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-115-1-SCFI-1993 ALT...	14/5/2024, 13:43:12	Frecuente (inferido)		Blob en bloques	143.96 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-115-2-SCFI-1994.pdf	14/5/2024, 13:43:20	Frecuente (inferido)		Blob en bloques	274.89 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-131-1-SCFI-1993.pdf	14/5/2024, 13:43:21	Frecuente (inferido)		Blob en bloques	129.38 KiB	Disponible ***
<input type="checkbox"/> NMX-CH-7500-1-IMNC-2006.p...	14/5/2024, 13:43:18	Frecuente (inferido)		Blob en bloques	463.27 KiB	Disponible ***
<input type="checkbox"/> NMX-EE-161-SCFI-1983.pdf	14/5/2024, 13:43:19	Frecuente (inferido)		Blob en bloques	172.03 KiB	Disponible ***

3. Creación recursos Document Intelligence

Azure AI Document Intelligence (anteriormente Form Recognizer) es un servicio de Inteligencia artificial de Azure basado en la nube que usa modelos de aprendizaje automático para automatizar el procesamiento de datos en aplicaciones y flujos de trabajo.

The screenshot displays the Azure AI Document Intelligence Studio interface. The top section shows the service overview for 'iagenerativenormas' (Document Intelligence). Key details include:

- API Kind: FormRecognizer
- Plan de tarifa: Estándar
- Extremo: https://iagenerativenormas.cognitiveservices.azure.com/
- Administrar claves: Haga clic aquí para administrar claves.
- Escalaibilidad automática: Deshabilitado

The bottom section shows the 'General document' analysis interface. It includes a document preview and a sidebar showing detected text and tables. The document being analyzed is 'NMX-B-001... ES.pdf'. The sidebar shows the following text:

Text

Paragraph

Transferir a un vaso de 400 ml, por medio de una bureta de 50 ml aproximadamente 20 ml de solución de K3Fe (CN)6, Anotar la lectura de la bureta, Agregar 25 ml de agua, 10 ml de solución de citrato de amonio, y 25 ml de hidróxido de amonio. Enfríar entre 278 y 283 K (5 a 10 °C) y mantener esta temperatura durante la titulación. Transferir el vaso al aparato de titulación potenciométrica. Mientras se agita, titular el ferricianuro de potasio con la solución de cobalto (1 ml - 1.5 mg Co) usando una bureta de 50 ml. Titular a una velocidad tan rápida como sea posible hasta alcanzar el punto final, y entonces agregar el titulante gota a gota. Después de la adición de cada incremento, anotar la lectura de la bureta y diferencia de potencial, cuando se ha alcanzado el equilibrio, Estimar

En la anterior imagen se muestra como Document Intelligence detecta el texto de la norma, en caso de haber tablas también son detectadas.

4. Creación servicio de búsqueda (AI Search)

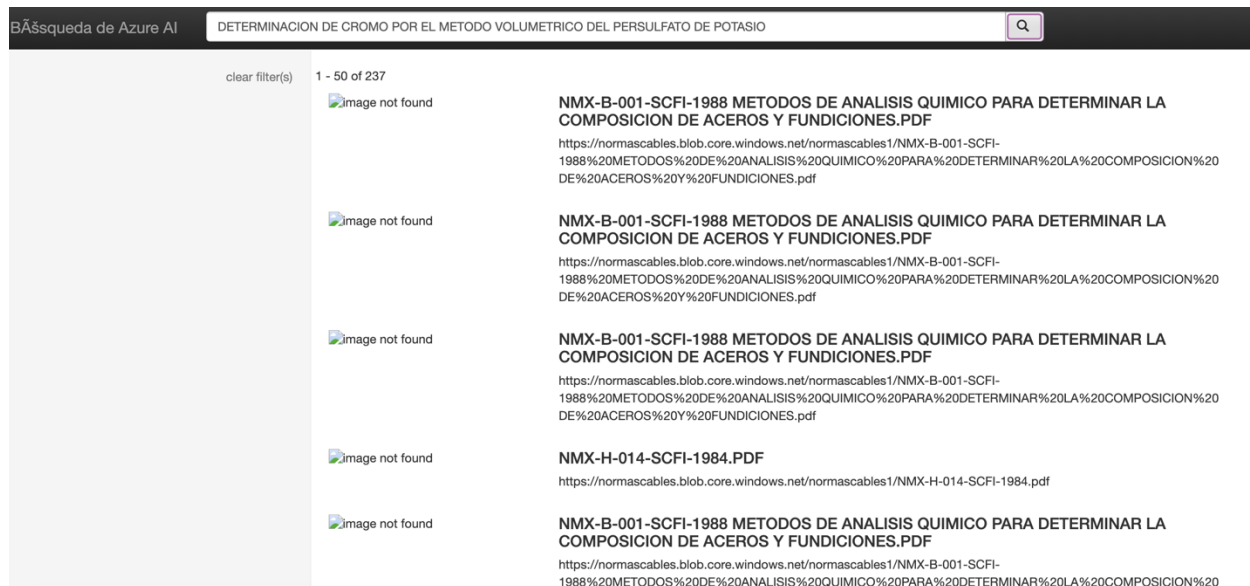
Búsqueda de Azure AI es una solución de recuperación de información de vectores y texto completo para la empresa y para escenarios de IA tradicionales y generativos. En nuestro caso se optó por seleccionar un plan básico, ya que el en el gratuito tiene limitaciones de aplicación.

The screenshot shows the Microsoft Azure portal interface for a search service named 'searchnormas2'. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (A01794113@tec.mx). The left sidebar contains a menu with options like 'Inicio', 'Nuevo_Prueba', 'Información general', 'Registro de actividad', 'Control de acceso (IAM)', 'Etiquetas', 'Diagnosticar y solucionar problemas', 'Administración de búsquedas', 'Configuración', 'Supervisión', 'Automation', and 'Ayuda'. The main content area displays the 'searchnormas2' service details, including its location (West US 2), subscription (Azure for Students), and various configuration options. A 'Comenzar' button is visible. Below the service details, there is a section titled 'Cree una experiencia de búsqueda de texto completo con IA y clasificador semántico' with a brief description.

Se realizó la vectorización de todos los documentos, en tal caso la aplicación permite buscar una palabra clave y encuentra la coincidencia en todos los documentos.

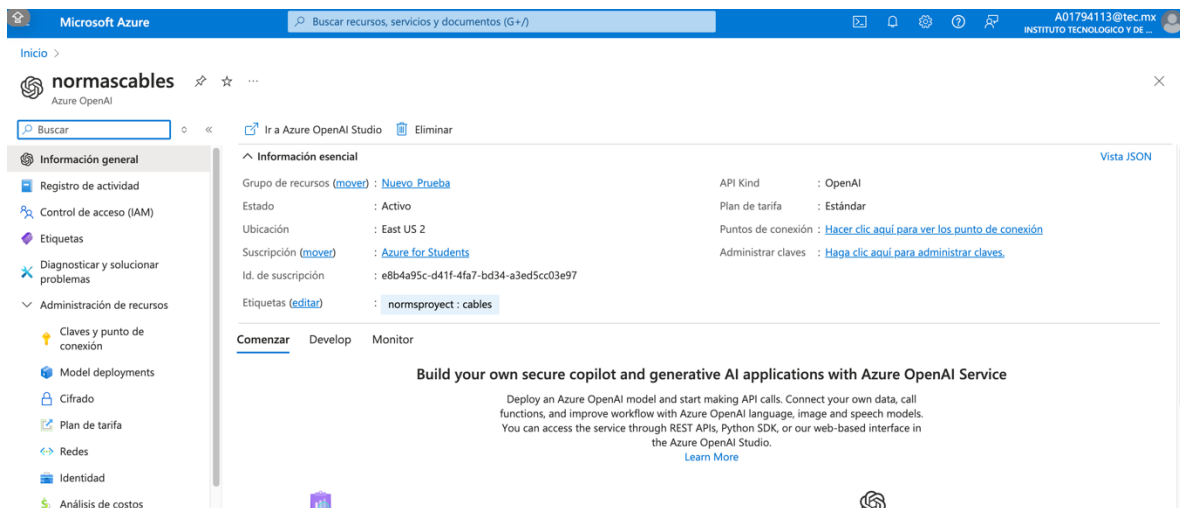
The screenshot shows the Microsoft Azure portal interface for the 'searchnormas2' service, displaying search results for the query 'manganeso'. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (A01794113@tec.mx). The left sidebar contains a menu with options like 'Inicio', 'Nuevo_Prueba', 'Información general', 'Registro de actividad', 'Control de acceso (IAM)', 'Etiquetas', 'Diagnosticar y solucionar problemas', 'Administración de búsquedas', 'Configuración', 'Supervisión', 'Automation', and 'Ayuda'. The main content area displays the search results, including the number of documents (239), total storage (1.33 MB), and vector index size (0 Bytes). The search results are displayed in a table format, showing the search score, content, file path, title, URL, ID, chunk ID, and last updated date. The search results are filtered by 'manganeso'.

Esta aplicación también permite la creación de una aplicación de demostración, donde se puede definir como mostrar los resultados, como parte de la exploración de la herramienta se realiza la prueba arrojando los siguientes resultados:

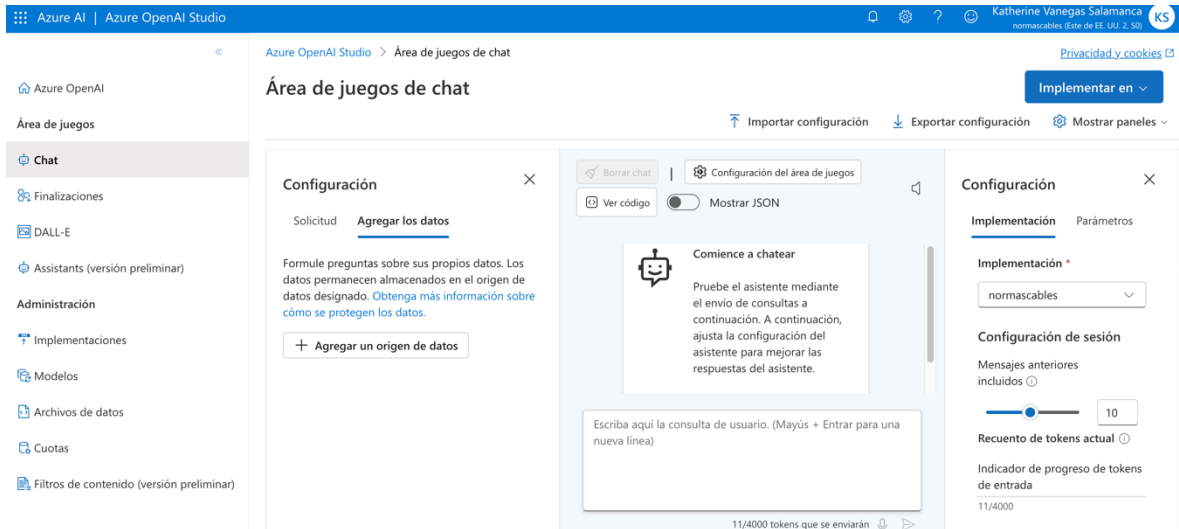


5. Creación servicio OpenAI

El servicio OpenAI de Azure ofrece herramientas para moderar el contenido generado y orientación para diseñar aplicaciones de forma segura.



Desde este menú se puede acceder a Azure OpenAI Studio, Azure OpenAI Services que ayuda a las organizaciones a generar información personalizada, contenido y búsquedas mediante sus orígenes de datos designados



Desde acá se puede alimentar el modelo con los datos propios, esta solución de acerca más a lo que estamos buscando, una especie de chatbot que responda a las consultas de forma fluida y concreta.

datos

Administración de datos

Revisar y finalizar

El origen de datos se usa para fundamentar los resultados generados con los datos. Seleccione un origen de datos existente o cree una nueva conexión de datos con Azure Blob Storage, bases de datos, búsqueda, URL o archivos locales como origen

[Obtenga más información sobre la privacidad y la seguridad de los datos en Azure AI.](#)

Azure Blob Storage (preview)

Suscripción *

Azure for Students

Selección del recurso de Azure Blob Storage *

normascales

Selección del contenedor de almacenamiento *

normascales1

Seleccione el recurso de búsqueda Azure AI *

searchnormas2

Escriba el nombre del índice *

normascales

El uso de Azure AI Search incurrirá en el uso de su cuenta. [Ver precios](#)

Programación del indexador *

Once

La programación del indexador se puede cambiar en la configuración del indexador para este recurso de búsqueda en el Azure Portal.

☒ Agregue la búsqueda de vectores a este recurso de búsqueda.

Insertando modelo

Seleccione un modelo de incrustación *

Azure OpenAI - normasvec

6. Modelo aplicado al proyecto

Basado en los objetivos del proyecto y el uso de la suite OpenAI de Microsoft, se aplicará el modelo de embeddings que es un tipo de modelo de aprendizaje automático que convierte datos en vectores de alta dimensionalidad, de manera que elementos similares estén más cerca unos de otros en este espacio vectorial. Estos modelos son ampliamente utilizados en procesamiento de lenguaje natural (NLP) y otras áreas de inteligencia artificial para representar palabras, frases, documentos y otros tipos de datos en un formato que pueda ser fácilmente manipulado por algoritmos de machine learning.

Este modelo funcionará de la siguiente manera:

Representación Vectorial:

Los modelos de embeddings asignan a cada palabra (o unidad de texto) un vector en un espacio de alta dimensionalidad.

Captura de Semántica:

La proximidad en este espacio vectorial refleja la similitud semántica entre palabras o textos.

Entrenamiento:

Los modelos de embeddings se entrenan en grandes corpus de texto. Durante el entrenamiento, el modelo aprende a colocar palabras que aparecen en contextos similares cerca unas de otras en el espacio vectorial.

Bibliografía

Azure OpenAI Service embeddings tutorial - Azure OpenAI. Microsoft Learn.

<https://learn.microsoft.com/en-us/azure/ai-services/openai/tutorials/embeddings?tabs=python-new%2Ccommand-line&pivots=programming-language-python>