

Київський національний університет ім. Тараса Шевченка

Кафедра мережевих та інтернет технологій

Лабораторна робота № 4

Дисципліна: Хмарні технології

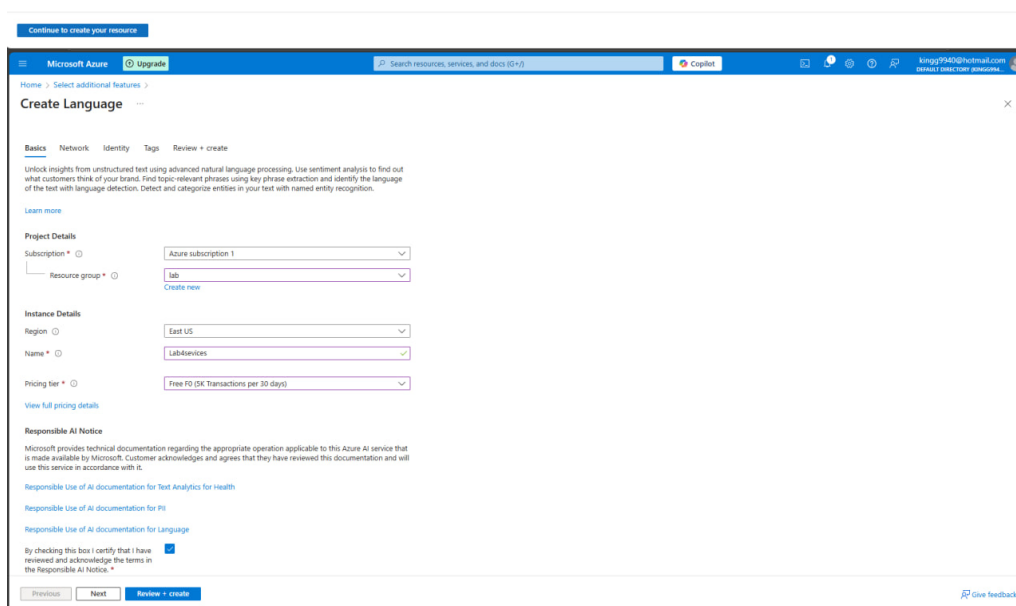
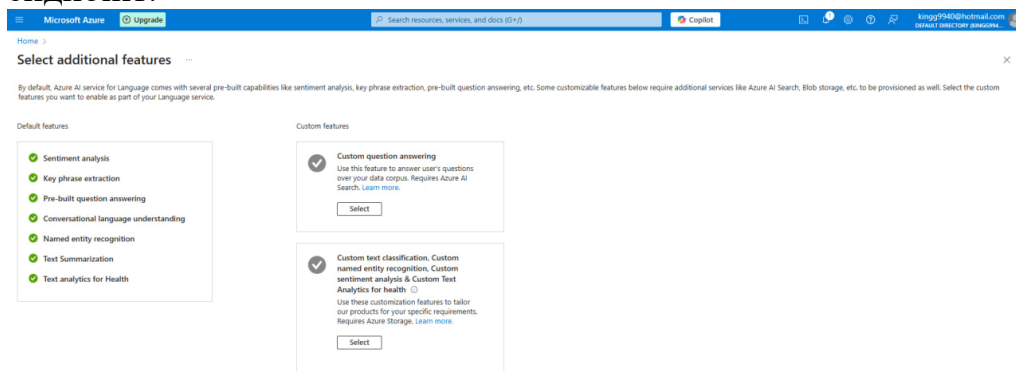
Тема: Створення класифікатора документів із застосуванням Azure AI

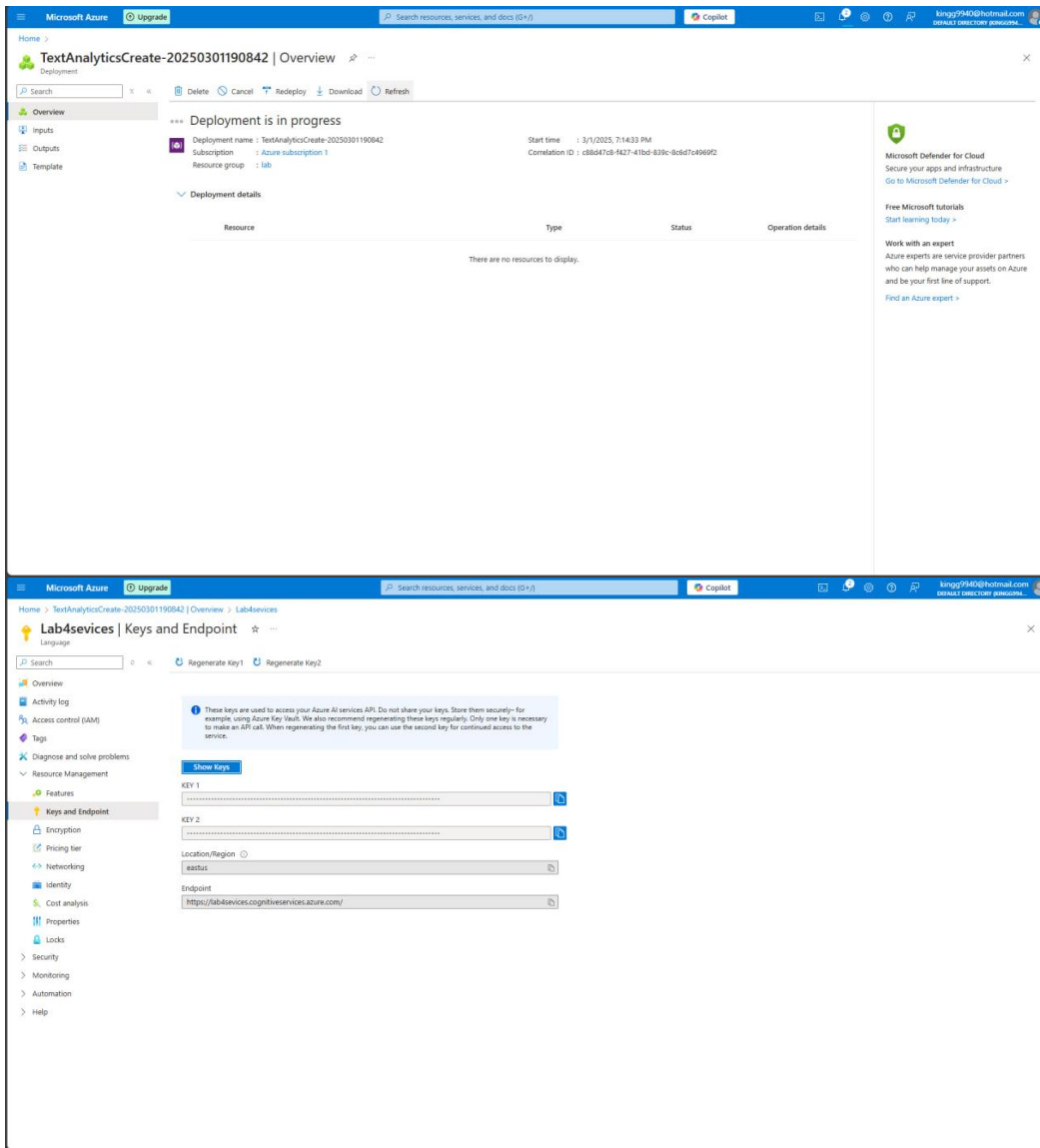
Виконав: Студент групи МІТ-31

Полюхович Андрій

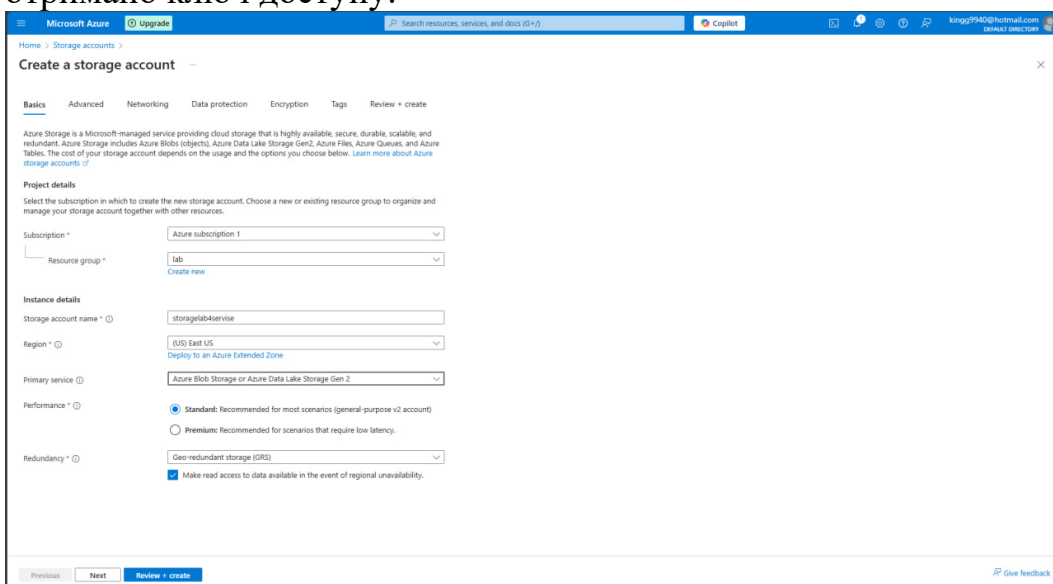
Хід роботи

- Для початку роботи було створено новий Azure AI Language сервіс з необхідними налаштуваннями, після чого отримано його ключ доступу та ендпоінт.





- Далі було створено новий Storage Account, в якому створено контейнери «polyukhovichsource» та «polyukhovichdestination». Після цього було також отримано ключ доступу.



Microsoft Azure | Upgrade | Search resources, services, and docs (0/1) | Copilot | kingg7940@hotmail.com | DEFAULT DIRECTORY

Home > storagelab4service_1740850153115 | Overview >

storagelab4service

Storage account

Search

Upload | Open in Explorer | Delete | Move | Refresh | Open in mobile | CLI / PS | Feedback

Essentials

Resource group	: lab	Performance	: Standard
Location	: eastus	Replication	: Read-access geo-redundant storage (RA-GRS)
Primary/Secondary Location	: Primary: East US, Secondary: West US	Account kind	: StorageV2 (general purpose v2)
Subscription	: Azure subscription 1	Provisioning state	: Succeeded
Subscription ID	: d2f102c-786f-4dd7-a088-e9429c8b54f	Created	: 01.03.2025, 19:29:26
Disk state	: Primary: Available, Secondary: Available		
Tags	: Add tags		

Properties | Monitoring | Capabilities (7) | Recommendations (0) | Tutorials | Tools + SDKs

Blob service

Hierarchical namespace	: Disabled
Default access tier	: Hot
Blob anonymous access	: Disabled
Blob soft delete	: Enabled (7 days)
Container soft delete	: Enabled (7 days)
Versioning	: Disabled
Change feed	: Disabled
NPS v3	: Disabled
Allow cross-tenant replication	: Disabled
Storage tasks assignments	: None

File service

Large file share	: Enabled
Identity-based access	: Not configured
Default share-level permissions	: Disabled
Soft delete	: Enabled (7 days)

Queue service

Security

Require secure transfer for REST API operations	: Enabled
Storage account key access	: Enabled
Minimum TLS version	: Version 1.2
Infrastructure encryption	: Disabled

Networking

Allow access from	: All networks
Private endpoint connections	: 0
Network routing	: Microsoft network routing
Access for trusted Microsoft services	: Yes
Endpoint type	: Standard

Microsoft Azure | Upgrade | Search resources, services, and docs (0/1) | Copilot | kingg7940@hotmail.com | DEFAULT DIRECTORY

Home > storagelab4service_1740850153115 | Overview > storagelab4service

storagelab4service | Access keys

Storage account

Search

Set rotation reminder | Refresh | Give feedback

Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location like Azure Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other.

Remember to update the keys with any Azure resources and apps that use this storage account.
[Learn more about managing storage account access keys.](#)

Storage account name: storagelab4service

key1 Rotate key
Last rotated: 01.03.2025 (0 days ago)

Key: [Redacted] Show

Connection string: [Redacted] Show

key2 Rotate key
Last rotated: 01.03.2025 (0 days ago)

Key: [Redacted] Show

Connection string: [Redacted] Show

Microsoft Azure | Upgrade | Search resources, services, and docs (0/1) | Copilot | kingg7940@hotmail.com | DEFAULT DIRECTORY

Home > storagelab4service

storagelab4service | Containers

Storage account

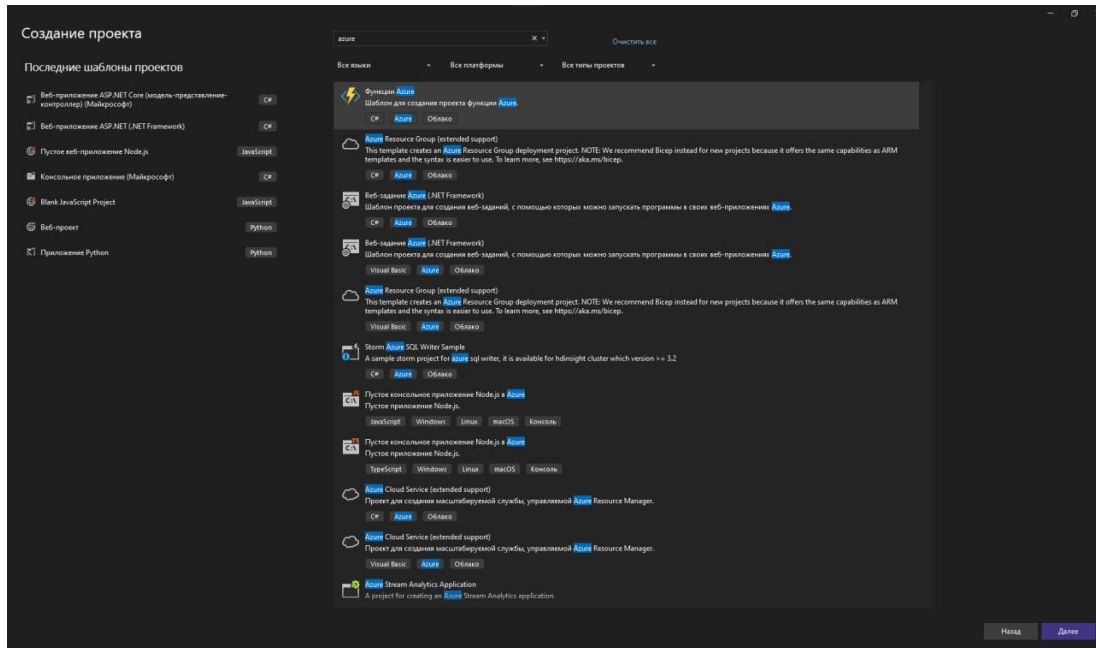
Search

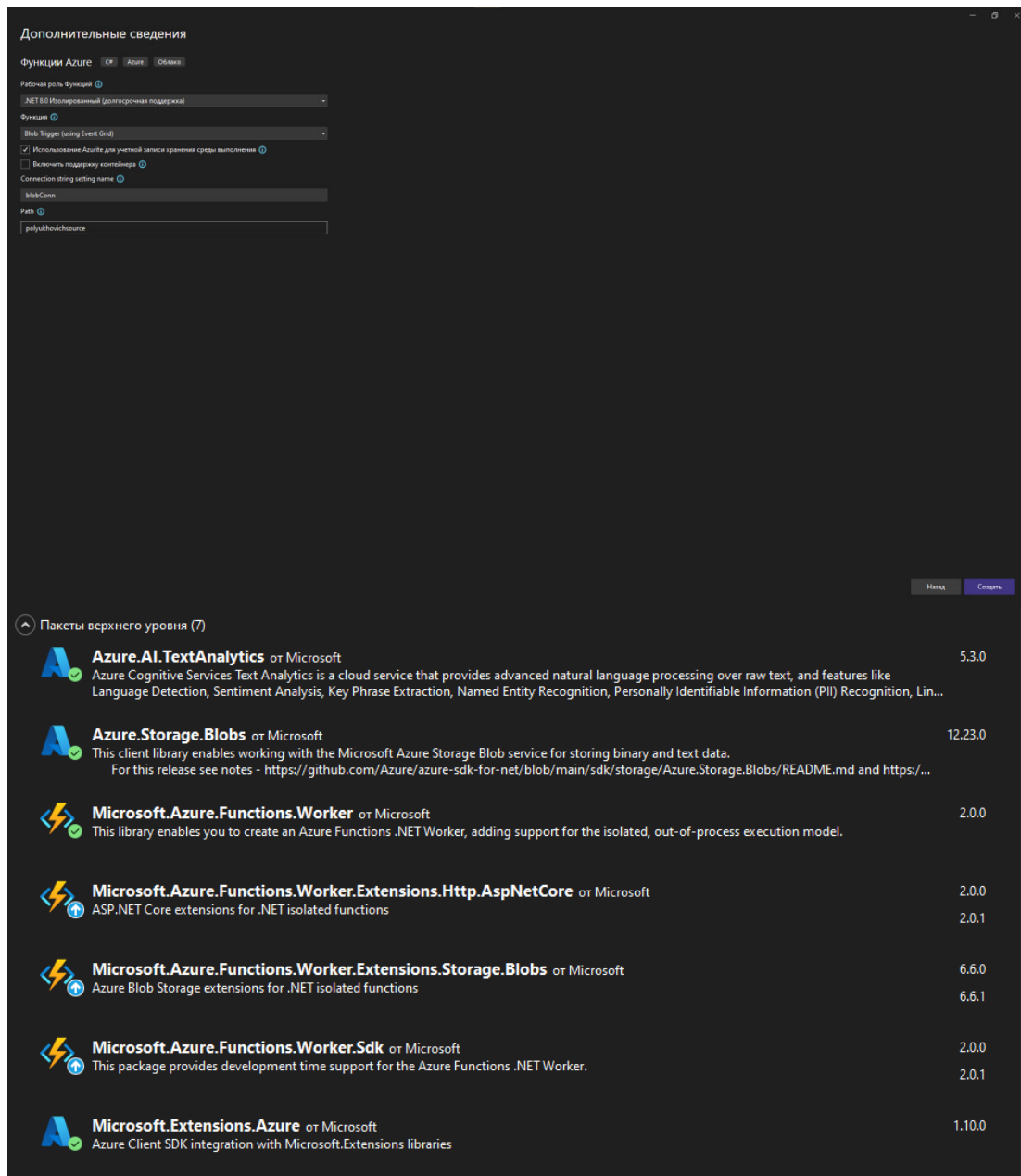
+ Container | Change access level | Restore containers | Refresh | Delete | Give feedback

Search containers by prefix: [Empty] Show deleted containers

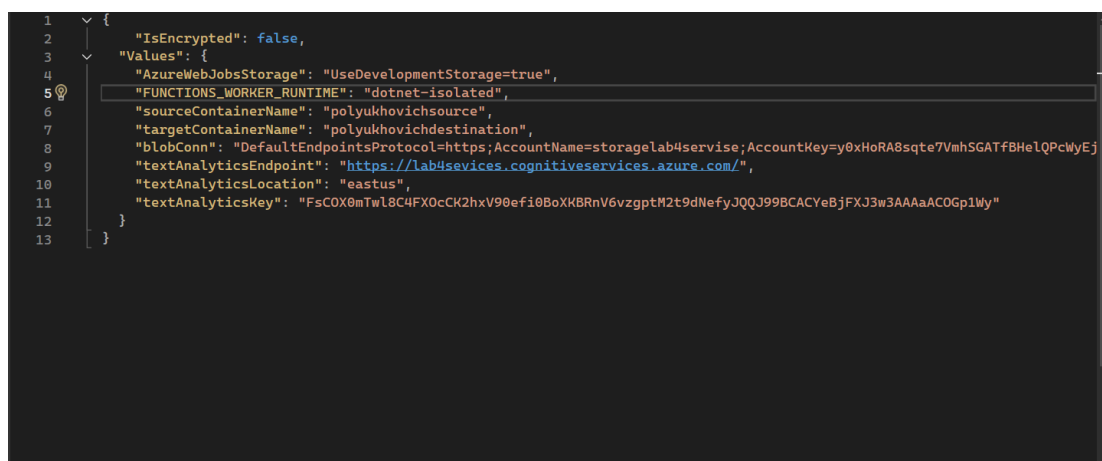
Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> \$logs	3/1/2025, 7:29:47 PM	Private	Available ...
<input type="checkbox"/> polyukhovichdestination	3/3/2025, 11:32:09 AM	Private	Available ...
<input type="checkbox"/> polyukhovichsource	3/3/2025, 11:32:41 AM	Private	Available ...

- Наступним кроком стало створення нового проекту у Visual Studio 2022 за шаблоном Azure Functions, після чого до проекту було додано необхідні пакети NuGet.





- Далі було додано ключі та строку підключення до local.settings.json, після чого було змінено код Program.cs та Function1.cs, де було створено Blob Service клієнт та імплементовано роботу з файлами з «polyukhovichsource» за допомогою AI Text Analytics.



```

1  using Azure;
2  using Microsoft.Azure.Functions.Worker;
3  using Microsoft.Azure.Functions.Worker.Builder;
4  using Microsoft.Extensions.Azure;
5  using Microsoft.Extensions.DependencyInjection;
6  using Microsoft.Extensions.Hosting;
7
8
9  var builder = FunctionsApplication.CreateBuilder(args);
10
11 builder.ConfigureFunctionsWebApplication();
12
13 builder.Services.
14     AddAzureClients(b =>
15     {
16         b.AddBlobServiceClient(Environment.GetEnvironmentVariable("blobConn"));
17         var endpoint = new Uri(Environment.
18             GetEnvironmentVariable("textAnalyticsEndpoint"));
19         var credential = new AzureKeyCredential(Environment.
20             GetEnvironmentVariable("textAnalyticsKey"));
21         b.AddTextAnalyticsClient(endpoint, credential);
22     });
23
24 builder.Build().Run();


```

```

1  using Azure.AI.TextAnalytics;
2  using Azure.Storage.Blobs;
3  using Microsoft.Azure.Functions.Worker;
4  using Microsoft.Extensions.Logging;
5  using System.Text;
6
7  namespace LanguageBasedDocumentClassifier
8  {
9      [Route("")]
10     public class LangBasedDocumentClassifier
11     {
12         private readonly ILogger<LangBasedDocumentClassifier> _logger;
13         private readonly TextAnalyticsClient _textAnalyticsClient;
14         private readonly BlobServiceClient _blobServiceClient;
15         private readonly BlobContainerClient _blobSourceContainerClient;
16         private readonly BlobContainerClient _blobDestinationContainerClient;
17         private readonly BlobContainerClient _blobTargetContainerClient;
18         public LangBasedDocumentClassifier(ILogger<LangBasedDocumentClassifier> logger,
19             TextAnalyticsClient textAnalyticsClient,
20             BlobServiceClient blobServiceClient)
21         {
22             _logger = logger;
23             _textAnalyticsClient = textAnalyticsClient;
24             _blobServiceClient = blobServiceClient;
25             _blobSourceContainerClient = _blobServiceClient.GetBlobContainerClient(Environment.
26                 GetEnvironmentVariable("sourceContainerName"));
27             _blobDestinationContainerClient = _blobServiceClient.GetBlobContainerClient(Environment.
28                 GetEnvironmentVariable("targetContainerName"));
29             _blobTargetContainerClient = _blobServiceClient.GetBlobContainerClient(Environment.
30                 GetEnvironmentVariable("targetContainerName"));
31         }
32         [Function(nameof(LangBasedDocumentClassifier))] // BlobTriggerSource.LogsAndContainerScan
33         public async Task Run([BlobTrigger("polyukhovichsource/{name}", Source =
34             BlobTriggerSource.LogsAndContainerScan, Connection = "blobConn")] Stream stream, string name)
35         {
36             using var blobStreamReader = new StreamReader(stream);
37             var content = await blobStreamReader.ReadToEndAsync();
38             _logger.LogInformation($"CS Blob Trigger (using Event Grid) processed blob\n Name: {name}" +
39                 "\n Data: { content}");
40             var detectedLanguage = await _textAnalyticsClient.DetectLanguageAsync(content);
41             var languageName = detectedLanguage.Value.Name;
42             _logger.LogInformation($"Detected language: {languageName}");
43             string targetBlobName = $"{languageName}/{name}";
44             BlobClient blobClient = _blobDestinationContainerClient.GetBlobClient(targetBlobName);
45             byte[] byteArray = Encoding.UTF8.GetBytes(content);
46             await blobClient.UploadAsync(new MemoryStream(byteArray));
47             _logger.LogInformation($"Upload blob - {targetBlobName} in the " +
48                 $"{Environment.GetEnvironmentVariable("targetContainerName")} container.");
49             await _blobSourceContainerClient.DeleteIfExistsAsync(name);
50             _logger.LogInformation($"Deleted blob - {name} from " +
51                 $"{Environment.GetEnvironmentVariable("sourceContainerName")} container.");
52         }
53     }
54 }

```

- Після цього було додано невеликі файли різними мовами (Варіант 2: Ukrainian, French, Italian, Polish, Spanish) до «polyukhovichsource» та запущено програму. В результаті її роботи було створено 5 папок у «polyukhovichdestination», що відповідають 5тьом обраним мовам, в яких містяться класифіковані файли з «polyukhovichsource». Класифікація відбулась без помилок.



polyukhovichsource

Container

Search

Upload

Change access level

Refresh

De

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra us

Location: polyukhovichsource

Search blobs by prefix (case-sensitive)

Add filter

Name

☐ sample1.txt

☐ sample2.txt

☐ sample3.txt

☐ sample4.txt

☐ sample5.txt

```
C:\Program Files\dotnet\dotnet.exe

Azure Functions Core Tools
Core Tools Version:      4.0.6821 Commit hash: N/A +c09a2033faa7ecf51b3773308283af0ca9a99f83 (64-bit)
Function Runtime Version: 4.1036.1.23224

[2025-03-06T19:12:19.482Z] Found D:\university\3 курс\Хмарн? технорн?1\lab4\FunctionApp1\FunctionApp1\FunctionApp1.cspr
oj. Using for user secrets file configuration.
[2025-03-06T19:12:29.003Z] Azure Functions .NET Worker (PID: 5952) initialized in debug mode. Waiting for debugger to at
tach...
[2025-03-06T19:12:29.104Z] Worker process started and initialized.

Functions:

    LangBasedDocumentClassifier: blobTrigger

For detailed output, run func with --verbose flag.
[2025-03-06T19:12:32.908Z] Executing 'Functions.LangBasedDocumentClassifier' (Reason='New blob detected(LogsAndContainer
Scan): polyukhovichsource/sample1.txt', Id=0cdd476e-9db6-468b-8c6d-040e9b072257)
[2025-03-06T19:12:32.913Z] Trigger Details: MessageId: 1ece6f98-8677-4229-9693-b7d90ac8d68f, DequeueCount: 1, InsertedOn
: 2025-03-06T19:12:31.000+00:00, BlobCreated: 2025-03-03T09:33:08.000+00:00, BlobLastModified: 2025-03-03T09:33:08.000+0
0:00
[2025-03-06T19:12:32.993Z] Executing 'Functions.LangBasedDocumentClassifier' (Reason='New blob detected(LogsAndContainer
Scan): polyukhovichsource/sample2.txt', Id=595233c6-9be7-494a-946f-653c93f9463b)
[2025-03-06T19:12:32.998Z] Trigger Details: MessageId: b8932282-c687-4016-83f6-1dceac7ae4d0, DequeueCount: 1, InsertedOn
: 2025-03-06T19:12:32.000+00:00, BlobCreated: 2025-03-03T09:33:08.000+00:00, BlobLastModified: 2025-03-03T09:33:08.000+0
0:00
[2025-03-06T19:12:33.154Z] Executing 'Functions.LangBasedDocumentClassifier' (Reason='New blob detected(LogsAndContainer
Scan): polyukhovichsource/sample3.txt', Id=62c1cd85-b872-4609-830c-d1b587cc6999)
[2025-03-06T19:12:33.163Z] Trigger Details: MessageId: 0fc0bc2b-7d49-4d3e-8c24-546c8e8ff66f, DequeueCount: 1, InsertedOn
: 2025-03-06T19:12:32.000+00:00, BlobCreated: 2025-03-03T09:33:08.000+00:00, BlobLastModified: 2025-03-03T09:33:08.000+0
0:00
```

Name	Modified	Access tier	Archive status
<input type="checkbox"/> French			
<input type="checkbox"/> Italian			
<input type="checkbox"/> Polish			
<input type="checkbox"/> Spanish			
<input type="checkbox"/> Ukrainian			

Висновок: У цій лабораторній роботі було розглянуто процес створення класифікатора файлів за мовою з використанням Azure AI Language Service та контейнерів Azure.