1.Introduction / Project Purpose

This project, **SozcuNewsAppTask**, is a full-stack .NET 8-based web application designed to crawl real-time news articles from sozcu.com.tr, index them into **Elasticsearch**, and present them to users via a **Razor Pages** interface.

It demonstrates modern practices such as:

- Web scraping with HtmlAgilityPack.
- Full-text search with NEST and Elasticsearch.
- Clean separation of services, API, and UI.
- Interactive UI with search, detail, and original source redirection.

2. Technologies Used

- .NET 8 – Main backend framework
- ASP.NET Core Web API – To serve data
- Razor Pages (WebUI) – For frontend UI
- HtmlAgilityPack – For scraping HTML content
- NEST (Elasticsearch .NET client) – For indexing/searching documents
- Elasticsearch Cloud – Hosted search engine
- HttpClient – For fetching data
- Swagger / Postman – For testing APIs
- CSS / JavaScript / HTML – Razor frontend
- Regex – Text cleanup during scraping

3. Architecture: Layers and Classes

Application Layers:

- **NewsApp.Api** – Backend scraping + data indexing logic
- **NewsApp.WebUI** – Razor-based frontend interface

Class	Responsibility
NewsAppService.cs	Scrapes data using XPath via HtmlAgilityPack
ElasticNewsService.cs	Indexes crawled data into Elasticsearch using BulkAsync
ElasticsearchContext.cs	Configures Elasticsearch connection with cloudId, auth
Index.cshtml.cs	Lists news from Elasticsearch
Search.cshtml.cs	Executes multi_match, fuzzy, max_expansions queries
Detail.cshtml.cs	Displays single news with source redirection
Program.cs	DI setup for both API and UI

4. Sample Code Flows

NewsAppService.cs – Web Scraping

```
var document = web.Load(url);
var title = document.DocumentNode.SelectSingleNode("//h1").InnerText;
var content = document.DocumentNode.SelectSingleNode("//p").InnerText;
```

ElasticNewsService.cs – Indexing to Elasticsearch

```
var response = await _elasticClient.BulkAsync(b => b
    .Index("news-app-demo")
    .IndexMany(newsList));
```

Search.cshtml.cs – Full-Text Search

```
var response = await _client.SearchAsync<NewsAppDto>(s => s
    .Query(q => q
        .MultiMatch(m => m
            .Fields(f => f.Field(p => p.Title).Field(p => p.Content))
            .Query(SearchTerm)
            .Fuzziness(Fuzziness.Auto)
            .MaxExpansions(2))));
```

5. References

NewsAppService.cs

- *https://www.youtube.com/watch?v=m9zFq6KS94Y&ab_channel=ShaunHalverson*
- *https://github.com/zzzprojects/html-agility-pack*
- https://www.w3schools.com/xml/xpath_syntax.asp
- https://learn.microsoft.com/en-us/dotnet/api/system.net.http.httpclient
- https://stackoverflow.com/questions/15705092/do-httpclient-and-httpclienthandler-have-to-be-disposed
- https://learn.microsoft.com/en-us/dotnet/api/system.text.regularexpressions.regex.replace
- https://medium.com/%40ertekinozturgut/basic-web-scraping-via-selenium-in-c-for-google-news-c012c2b4939d
- https://www.scrapingbee.com/blog/html-agility-pack/
- https://www.zenrows.com/blog/html-agility-pack
- https://www.zenrows.com/blog/web-scraping-c-sharp

ElasticNewsService.cs

- https://www.youtube.com/watch?v=tw9svKWq6tg&t=634s&ab_channel=dotnetFlix
- https://www.elastic.co/blog/indexing-documents-with-the-nest-elasticsearch-net-client
- https://stackoverflow.com/questions/48819147/how-to-call-bulkall-method-in-elasticsearch-nest-asynchronously-in-a-windows-app
- https://stackoverflow.com/questions/68317127/how-can-i-write-multiple-updates-in-one-bulkall-method-in-elasticsearch-nest-7-1

Program.cs – NewsApp.Api

- https://www.youtube.com/watch?v=tw9svKWq6tg&ab_channel=dotnetFlix
- https://cloud.elastic.co/deployments/b5efd44d51b041638fcb92869db2c65f

ElasticsearchContext.cs

- https://blexin.com/en/blog-en/how-to-integrate-elasticsearch-in-asp-net-core/
- https://www.elastic.co/docs/reference/elasticsearch/clients/dotnet/connecting
- https://github.com/elastic/elasticsearch-net/issues/8184

Index.cshtml.cs

- https://www.elastic.co/docs/reference/elasticsearch/clients/dotnet/examples
- https://stackoverflow.com/questions/33834141/elasticsearch-and-nest-why-am-i-missing-the-id-field-on-a-query

Index.cshtml

- https://wpengine.com/resources/card-layout-css-grid-layout-how-to/
- https://www.w3schools.com/css/css_grid_container.asp

Detail.cshtml.cs

- https://www.elastic.co/docs/reference/elasticsearch/clients/dotnet/getting-started

Search.cshtml.cs

- https://www.elastic.co/docs/reference/query-languages/query-dsl/query-dsl-multi-match-query
- https://www.elastic.co/docs/reference/elasticsearch/clients/dotnet/query
- https://stackoverflow.com/questions/31086987/searching-for-an-input-keyword-in-all-fields-of-an-elasticsearch-document-using
- https://www.elastic.co/docs/reference/query-languages/query-dsl/query-dsl-fuzzy-query --> fuzzy query ve max expansions
- https://www.pipiho.com/es/7.7/en/query-dsl-multi-match-query.html
- https://stackoverflow.com/questions/74867381/elasticsearch-partial-search-with-fuzziness-on-multiple-fields

Program.cs – NewsApp.WebUI

- https://www.youtube.com/watch?v=8LXCxHzEIhc&ab_channel=TechWithPat