**Natural Language Requirements Specification Document**

**Objective**

Implement a Natural Language Question Answering (QA) model using local data and preprocessing HTML strings.

**1. Environment Setup**

* Use Visual Studio 2022 Community Edition (64-bit) - Preview, Version 17.11.0 Preview 2.1.
* Ensure .NET 6.0 SDK is installed.

**2. Project Structure**

* Create a new Console Application project in Visual Studio.

**3. Dependencies**

* **ML.NET Installation:**
  + Open Visual Studio.
  + Go to Tools > NuGet Package Manager > Manage NuGet Packages for Solution.
  + Search for Microsoft.ML and install the latest stable version.
* **HtmlAgilityPack Installation:**
  + Similarly, install HtmlAgilityPack using NuGet Package Manager.

**4. Data Preparation**

* Prepare a tab-delimited text file languageData.txt containing:
  + Column to predict
  + Context (HTML string)
  + Question
  + Answer Index

**5. Preprocessing**

* Implement HTML preprocessing in C# to clean and extract text from HTML strings.

**6. Model Training**

* Tokenize and pad data as required for model training.
* Configure ML.NET pipeline:
  + Featurize text columns (Context and Question).
  + Concatenate features and prepare labels (Answer Index).
  + Train the model using a suitable ML.NET trainer.

**7. Execution Flow (Structured Console Application)**

* Implement a structured console application without using the Main method style:
  + Use a top-level statement approach to structure your application.
  + Example structure:

csharp

Copy code

// Define your classes and methods here

DataPreparation.PreprocessHtmlData();

ModelTraining.TrainQAModel();

Validation.ValidateModel();

// Other application flow as required

**8. Testing and Validation**

* Validate the trained model against sample inputs.
* Ensure accuracy and performance meet acceptable levels.

**9. Documentation**

* Document code with comments and necessary explanations.
* Prepare a requirements document outlining setup, dependencies, data format, and execution flow.

**10. Future Considerations**

* Explore deployment options.
* Incorporate additional features or enhancements based on feedback.