BPI Web Scraping documentation

Note – I wasn’t sure if documentation meant simply showing XML syntax or describing the software in more detail.

Future improvements

* Future improvements would involve:
  + Implementing dependency injection to create more modular and cohesive services
  + Unit testing mock google html requests
  + Implementing MVP (Model-View-Presenter).
  + Moving raw string data to read-only strings!

Known bug:

* HTTP Web Requests do not always return all html request data, e.g. sometimes only 19 Google search results returned instead of 20 .
  + Most likely due to an asynchronous issue of not waiting for all request data to return however despite using a .Wait() to wait for asynchronous data the bug still persists.

XML software Documentation

<?xml version="1.0"?>

<doc>

</name>

</assembly>

<members>

<member name="M:WebScraperTest.WebScraperForm.InitializeComponent">

<summary>

Required method for Designer support - do not modify

the contents of this method with the code editor.

</summary>

</member>

<member name="T:WebScraperTest.Models.SearchResult">

<summary>

The model to contain search results.

</summary>

</member>

<member name="P:WebScraperTest.Models.SearchResult.Order">

<summary>

The rank of the search result.

</summary>

</member>

<member name="P:WebScraperTest.Models.SearchResult.Title">

<summary>

The title of the search result.

</summary>

</member>

<member name="P:WebScraperTest.Models.SearchResult.Url">

<summary>

The url of the search result.

</summary>

</member>

<member name="P:WebScraperTest.Models.SearchResult.SearchTerm">

<summary>

The search term of the search result.

</summary>

</member>

<member name="T:WebScraperTest.NetworkRequests.NetworkRequestsService">

<summary>

Service to deal with network queries.

</summary>

</member>

<member name="M:WebScraperTest.NetworkRequests.NetworkRequestsService.GetResponseData(System.String)">

<summary>

Takes in a url and returns the stream html data.

</summary>

<param name="url">The url to return data.</param>

<returns>The stream data of the html request</returns>

</member>

<member name="T:WebScraperTest.DataConversionService">

<summary>

A service to convert data to files.

</summary>

</member>

<member name="M:WebScraperTest.DataConversionService.SaveDataToCSV(System.Collections.Generic.List{WebScraperTest.Models.SearchResult},System.String)">

<summary>

Writes search results to a csv file.

</summary>

<param name="searchResults">The search results to write to.</param>

<param name="path">The location to save the file.</param>

</member>

<member name="T:WebScraperTest.WebScraperPresenter.WebScraperPresenter">

<summary>

Presenter class that uses all components to scrape data.

</summary>

</member>

<member name="M:WebScraperTest.WebScraperPresenter.WebScraperPresenter.InitializeServices">

<summary>

Initializes component services.

</summary>

</member>

<member name="M:WebScraperTest.WebScraperPresenter.WebScraperPresenter.InitializeScraping(System.String)">

<summary>

Returns specified data from a google search term.

</summary>

<param name="searchTerm">TThe search term to be queried.</param>

</member>

<member name="M:WebScraperTest.WebScraperPresenter.WebScraperPresenter.ConvertToCSV(System.Collections.Generic.List{WebScraperTest.Models.SearchResult},System.String)">

<summary>

Calls the data conversion method to convert to a csv.

</summary>

<param name="searchResults">The search results to be converted.</param>

<param name="fileLocation">The csv location to be saved to.</param>

</member>

<member name="T:WebScraperTest.WebScraper.WebScraperService">

<summary>

Service to scrape web data.

</summary>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.ScrapeData(System.String,System.IO.StreamReader)">

<summary>

Takes in raw stream data and returns the required list of SearchResult.

</summary>

<param name="searchTerm">The search term to be documented.</param>

<param name="htmlStream">The raw stream data to be processed.</param>

<returns>A list of SearchResult</returns>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.GetSearchResultsData(System.Collections.Generic.List{HtmlAgilityPack.HtmlNode},System.String)">

<summary>

Filters and returns specific data of the node list.

</summary>

<param name="orderedNodelist">The ordered node list to be filtered.</param>

<param name="searchTerm">The search term used in the method.</param>

<returns>List of SearchResult entities.</returns>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.DecodeString(System.String)">

<summary>

Decodes an encoded string.

</summary>

<param name="encodedString">The encoded string to be decoded.</param>

<returns></returns>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.ConvertStreamToHtmlDocument(System.IO.StreamReader)">

<summary>

Loads a HtmlDocument that takes in a stream.

</summary>

<param name="htmlStream">The stream that will be used.</param>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.SelectHtmlDocumentNodes(HtmlAgilityPack.HtmlDocument,System.String)">

<summary>

Selects nodes based on an xpath filter.

</summary>

<param name="htmlDoc">The HtmlDocument that will be filtered.</param>

<param name="xpathFilter">The xpath.</param>

<returns>A node collection based on the filter</returns>

</member>

<member name="M:WebScraperTest.WebScraper.WebScraperService.OrderNodesByPosition(HtmlAgilityPack.HtmlNodeCollection)">

<summary>

Orders the nodes based on their actual position on the html page.

</summary>

<param name="nodeCollection">The collection to be ordered.</param>

<returns>An ordered list of HtmlNode</returns>

</member>

</members>

</doc>