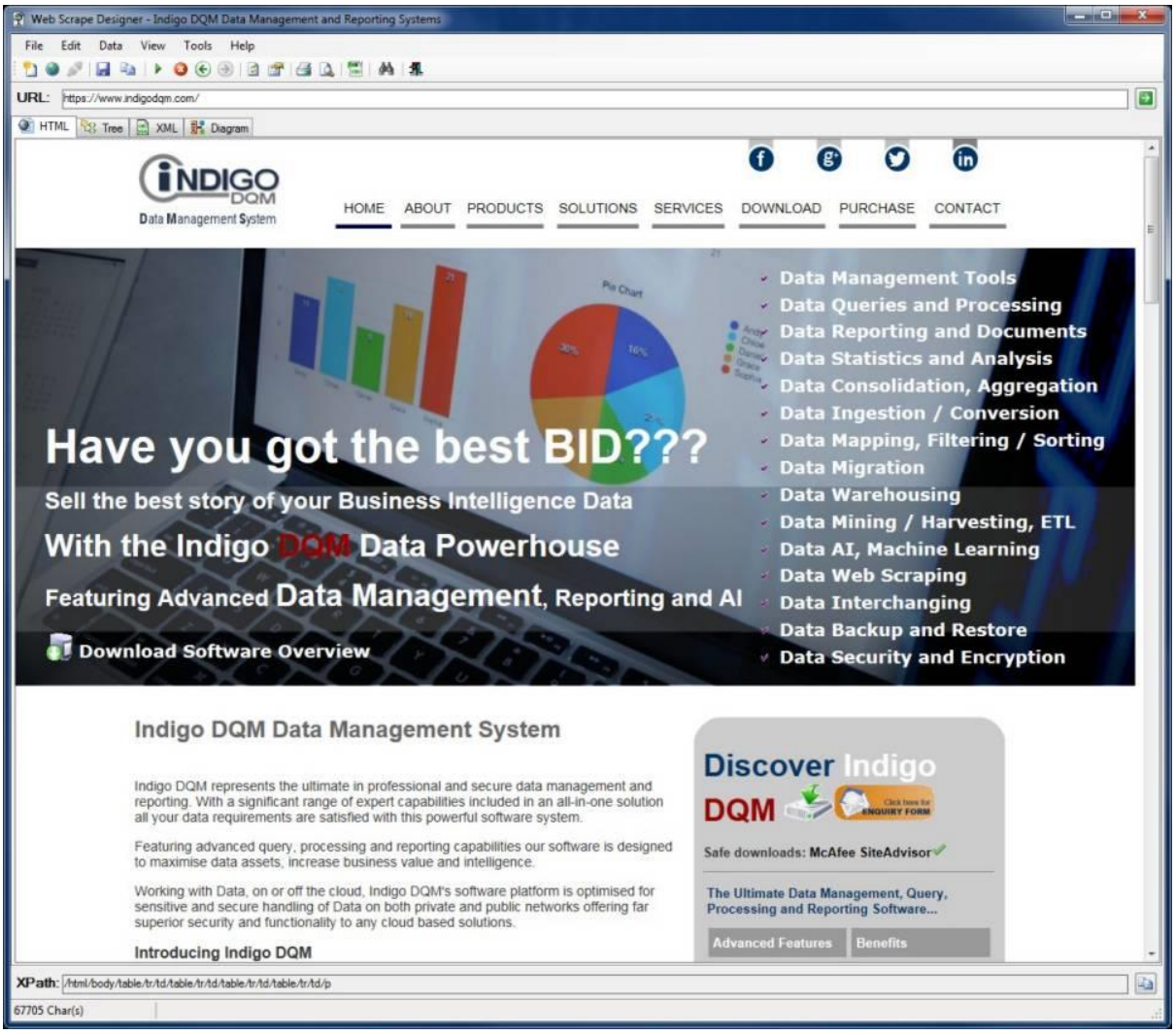
Indigo DQM Data Management System

Web Scrape Designer

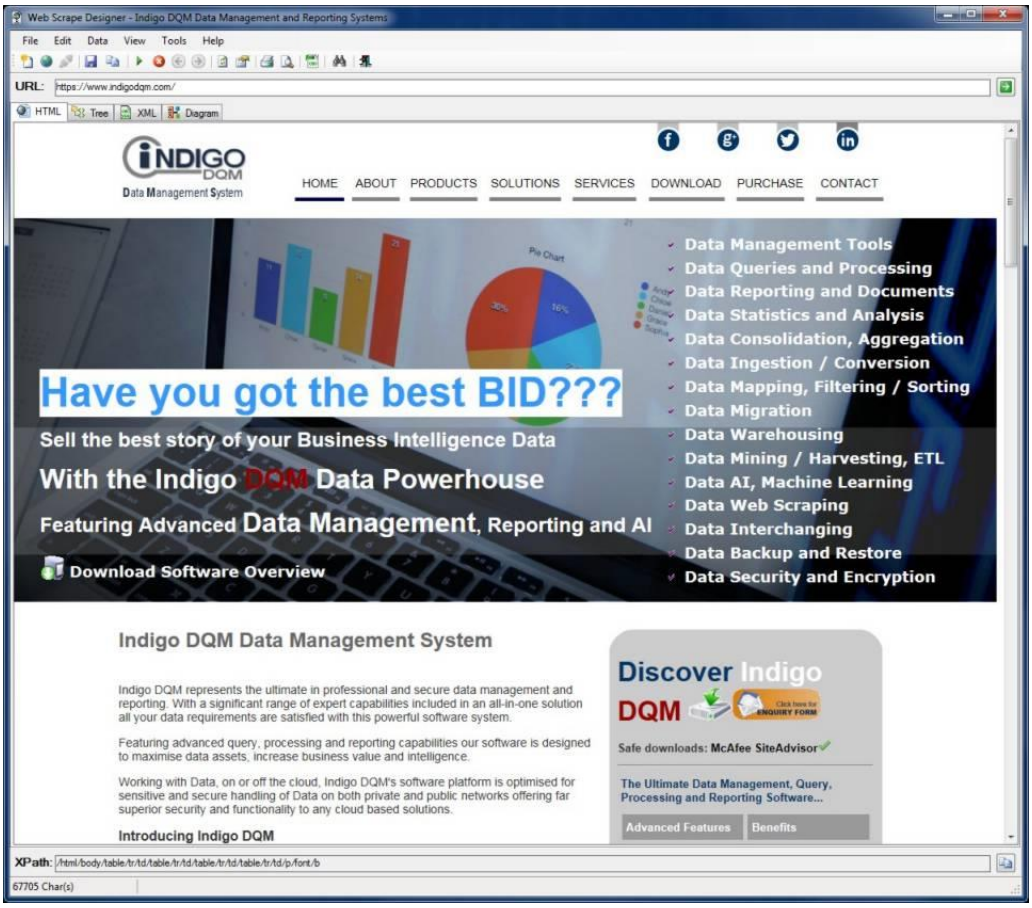
Indigo DQM Web Scrape Designer allows complex XQueries and XPath statements to be executed to extract or scrape elements from HTML Web Pages or Files.

XQuery is a query and functional programming language that is designed to query and transform collections of structured and unstructured Data, usually in the form of XML (Extensible Markup Language).

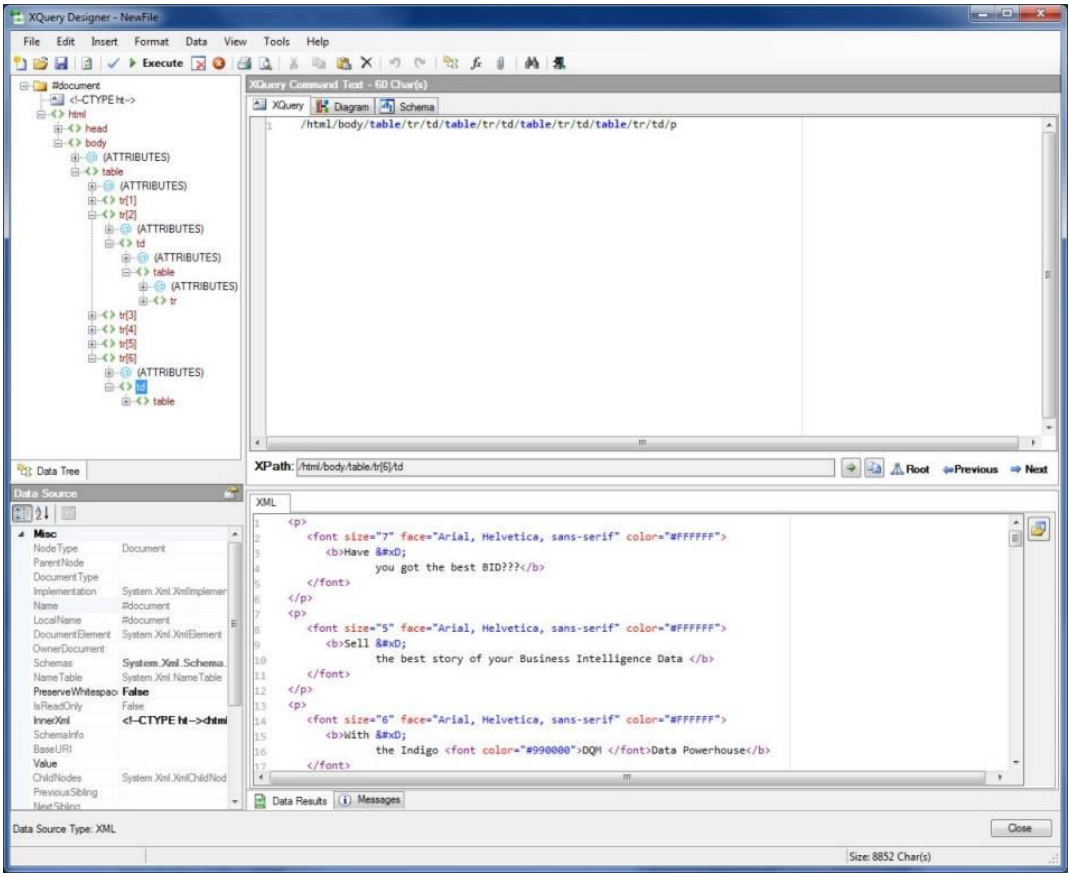


The language is based on the XQuery and XPath Data Model (XDM) which uses a tree-structured model of the information content of an XML document.

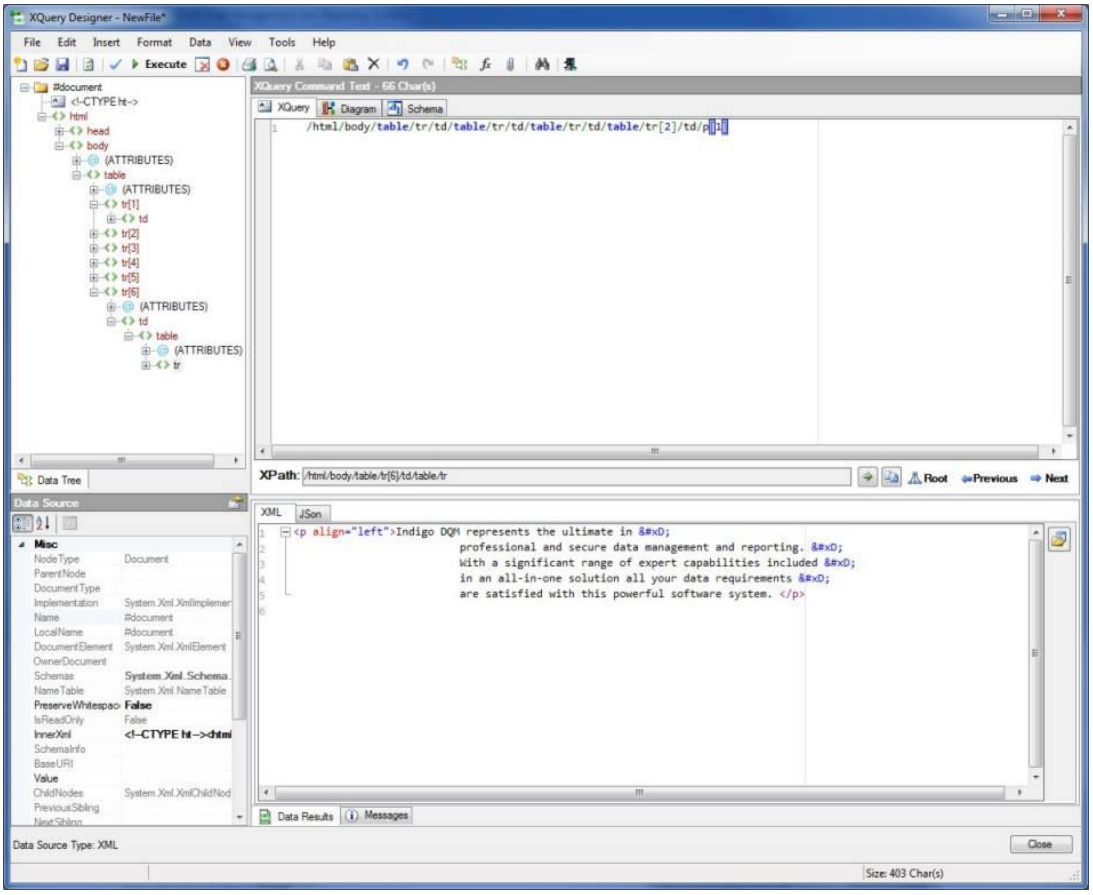
The Web Scrape Designer is a visual aid that allows web page elements to be clicked, selected or highlighted to automatically generate XPath statements.



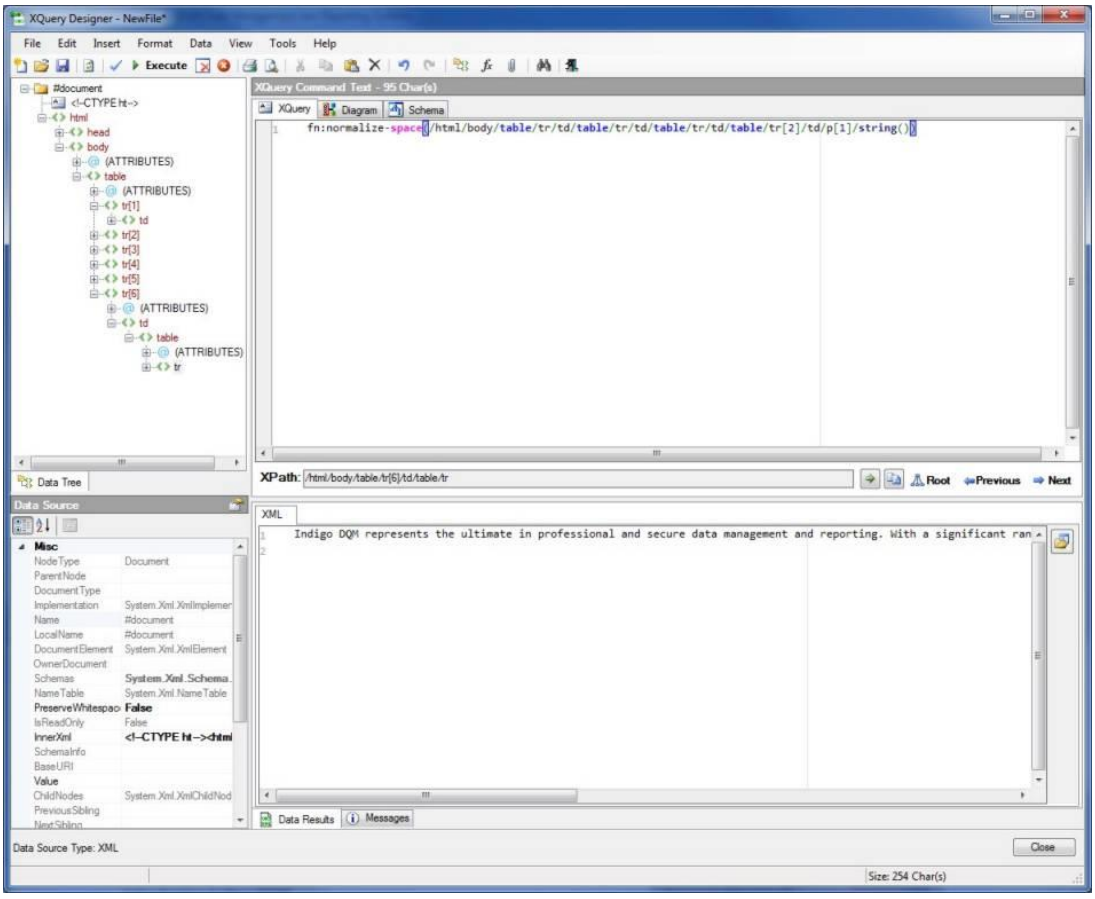
Web page data can be queried using XQuery Designer.



Extracting HTML elements from the Web Page using XQuery

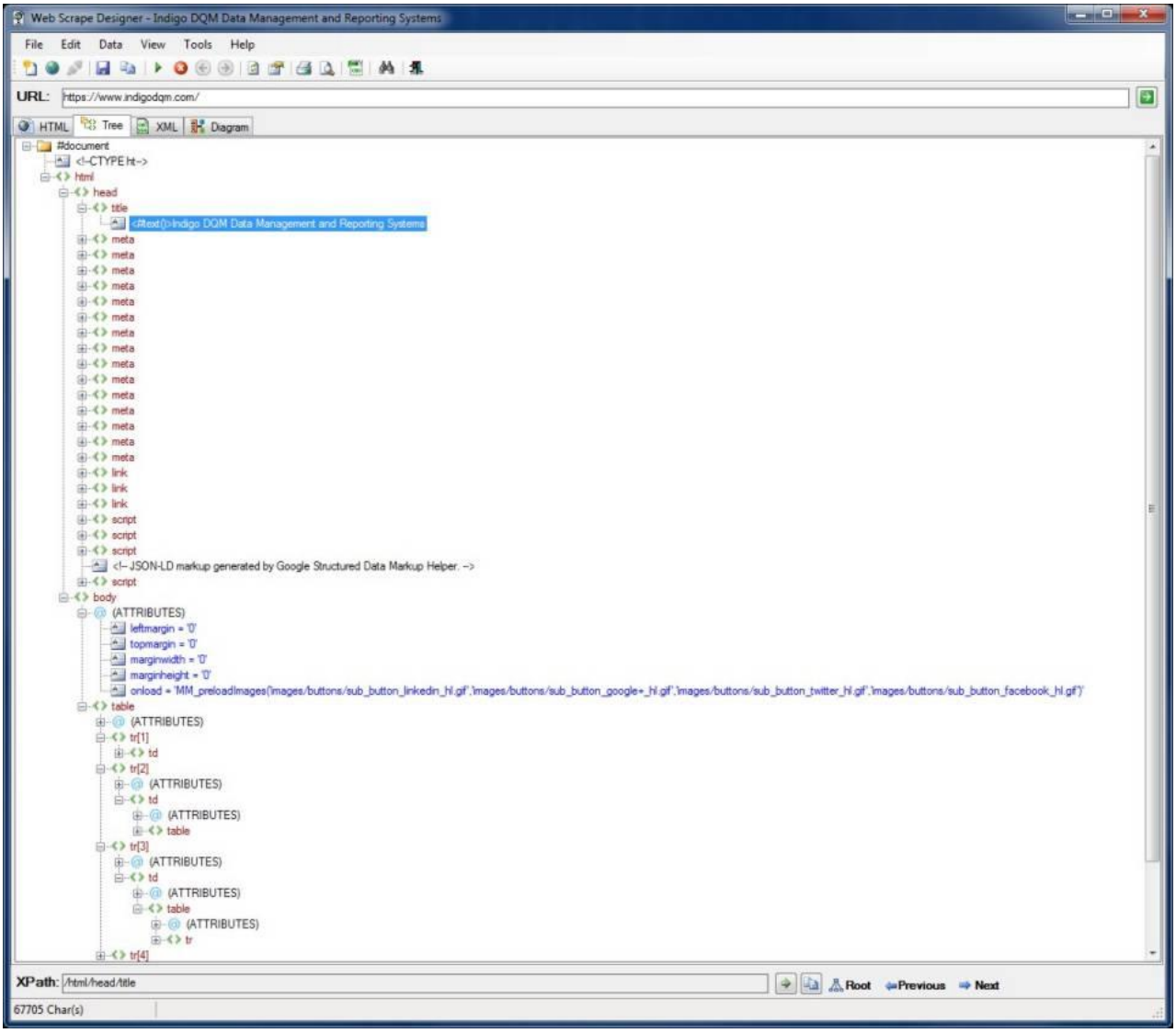


Using an XQuery function to normalize space and extract the plain text.



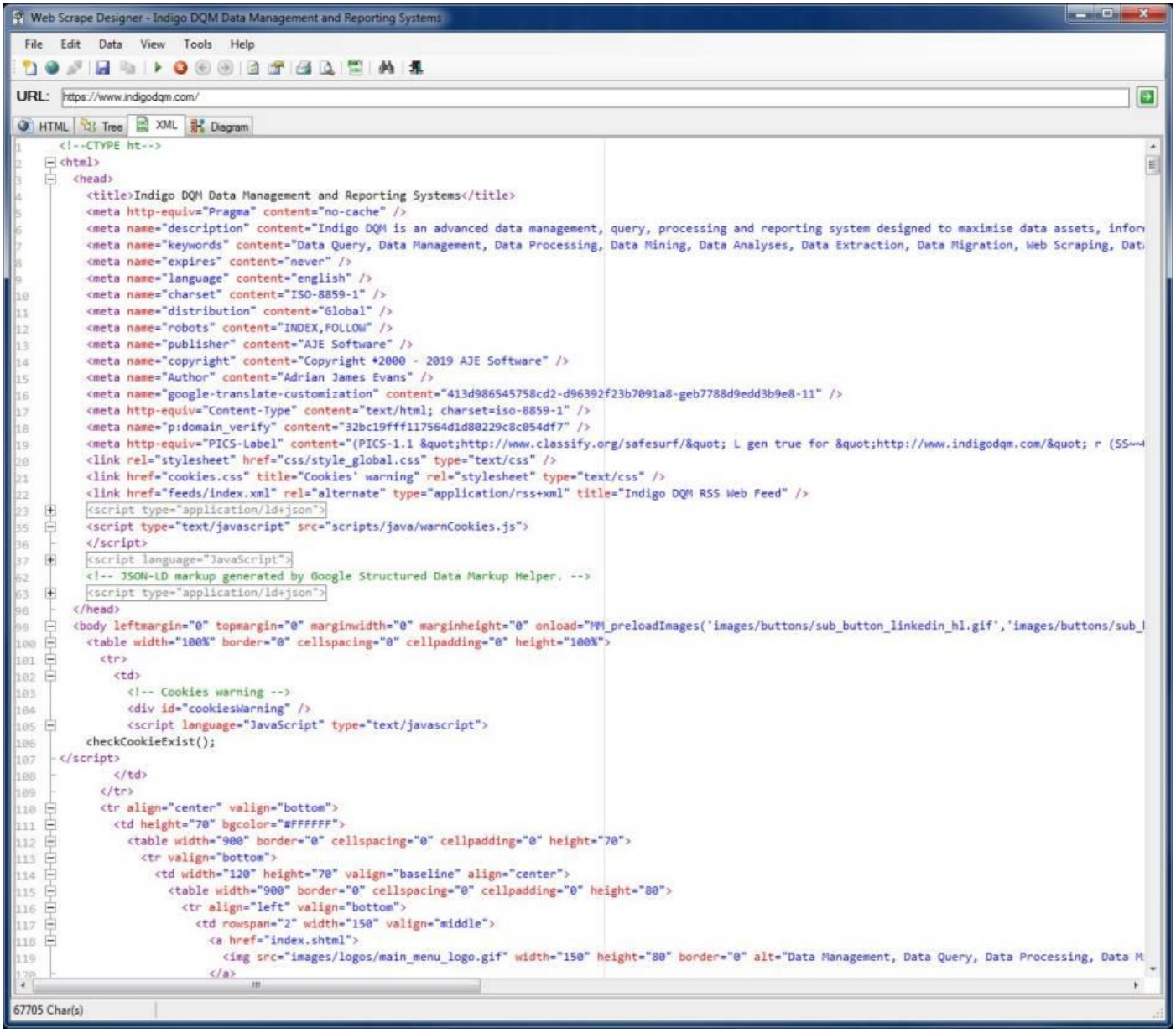
Data Tree View

Selecting a node in the Data Tree will update the current XPath for that node.



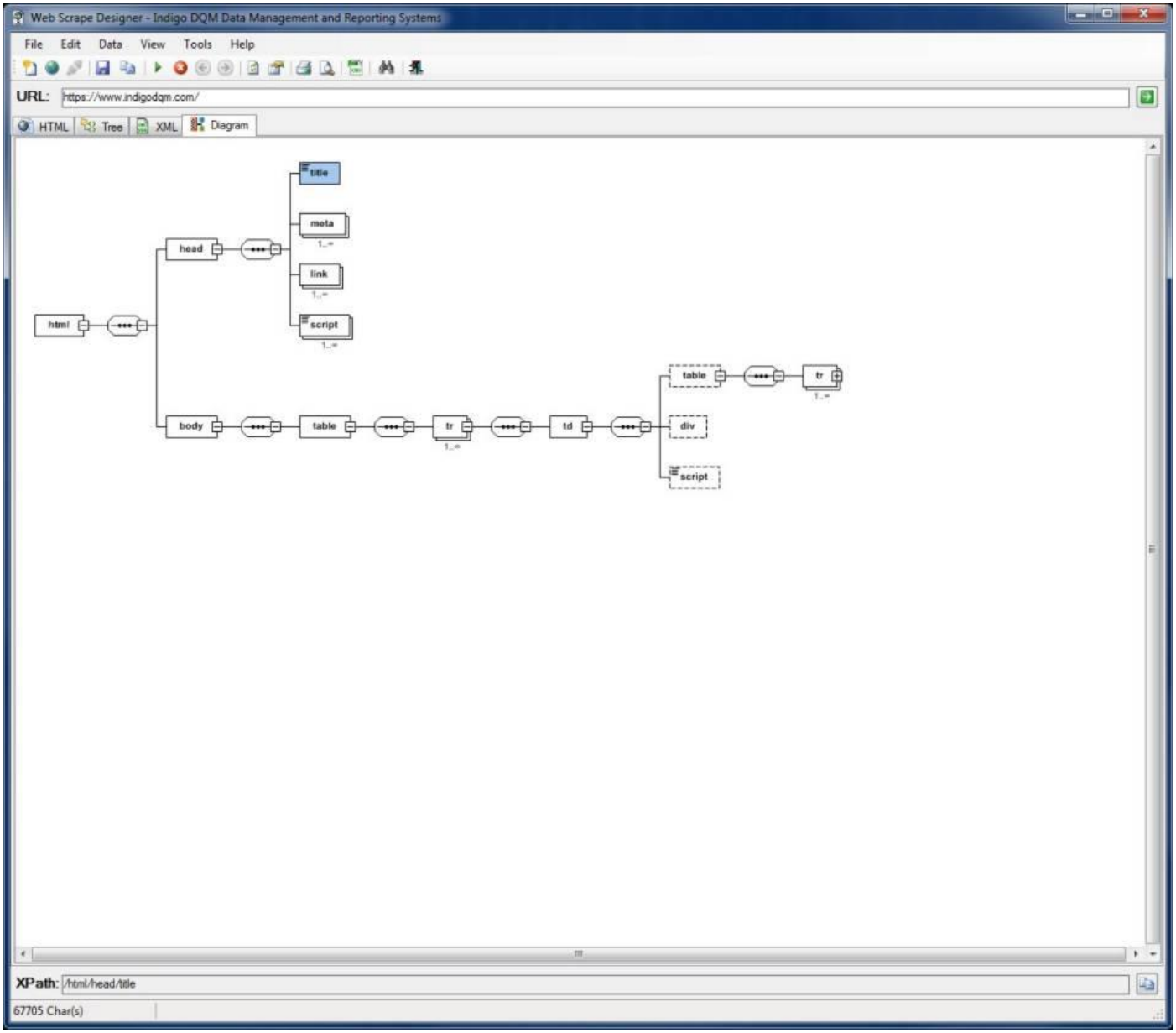
Web Page Source

Viewing the HTML to XML for the Web Page Source.



XSD Diagrams

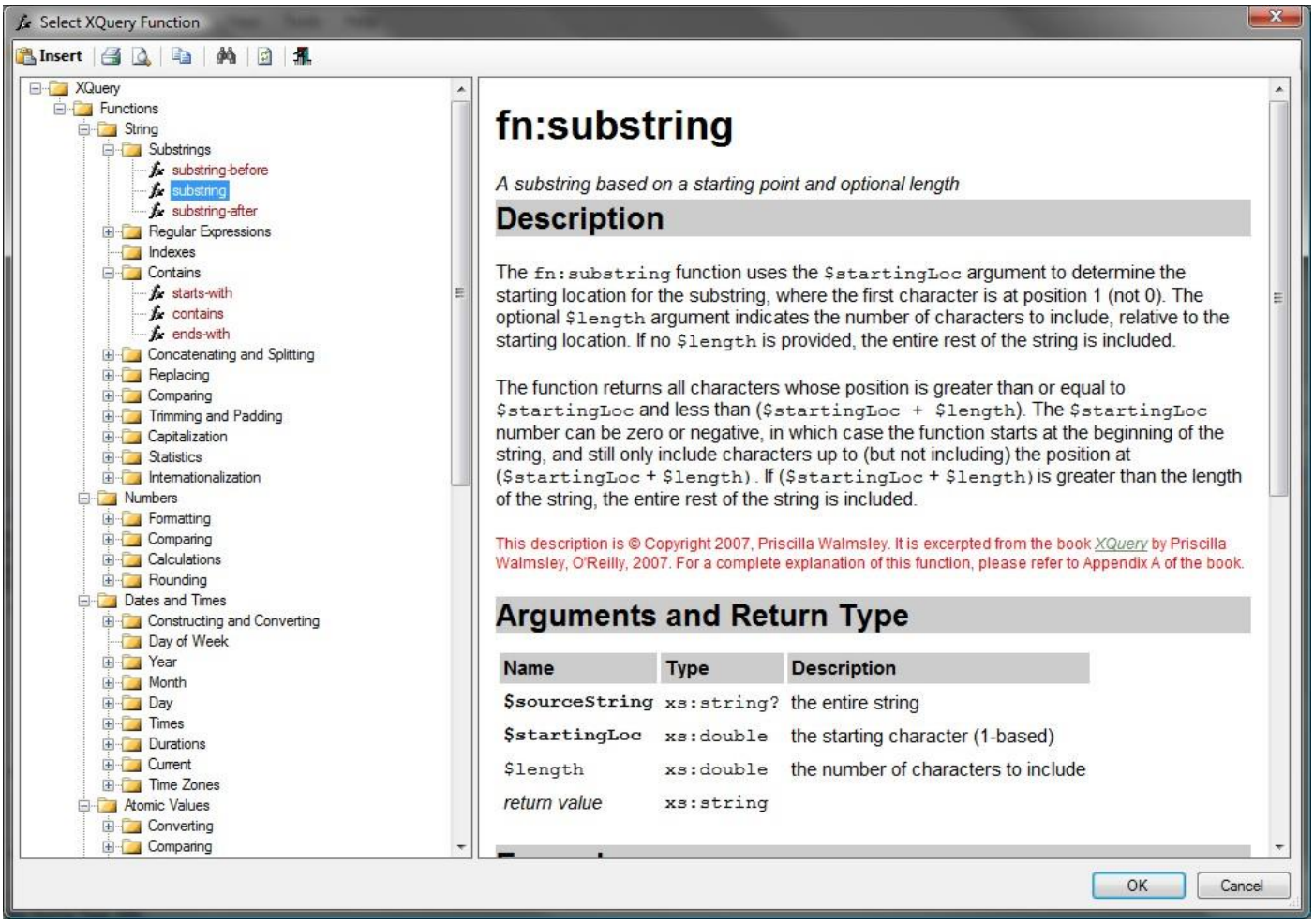
XSD Diagrams allow a visual representation of an XPath expression in the Data Schema. Click the Diagram tab and expand out the Diagram elements to show the structure of the Data Schema.



The XPath expression for the current element is shown in the XPath navigation bar. Elements can also be navigated using the navigation buttons.

Inserting an XQuery Function

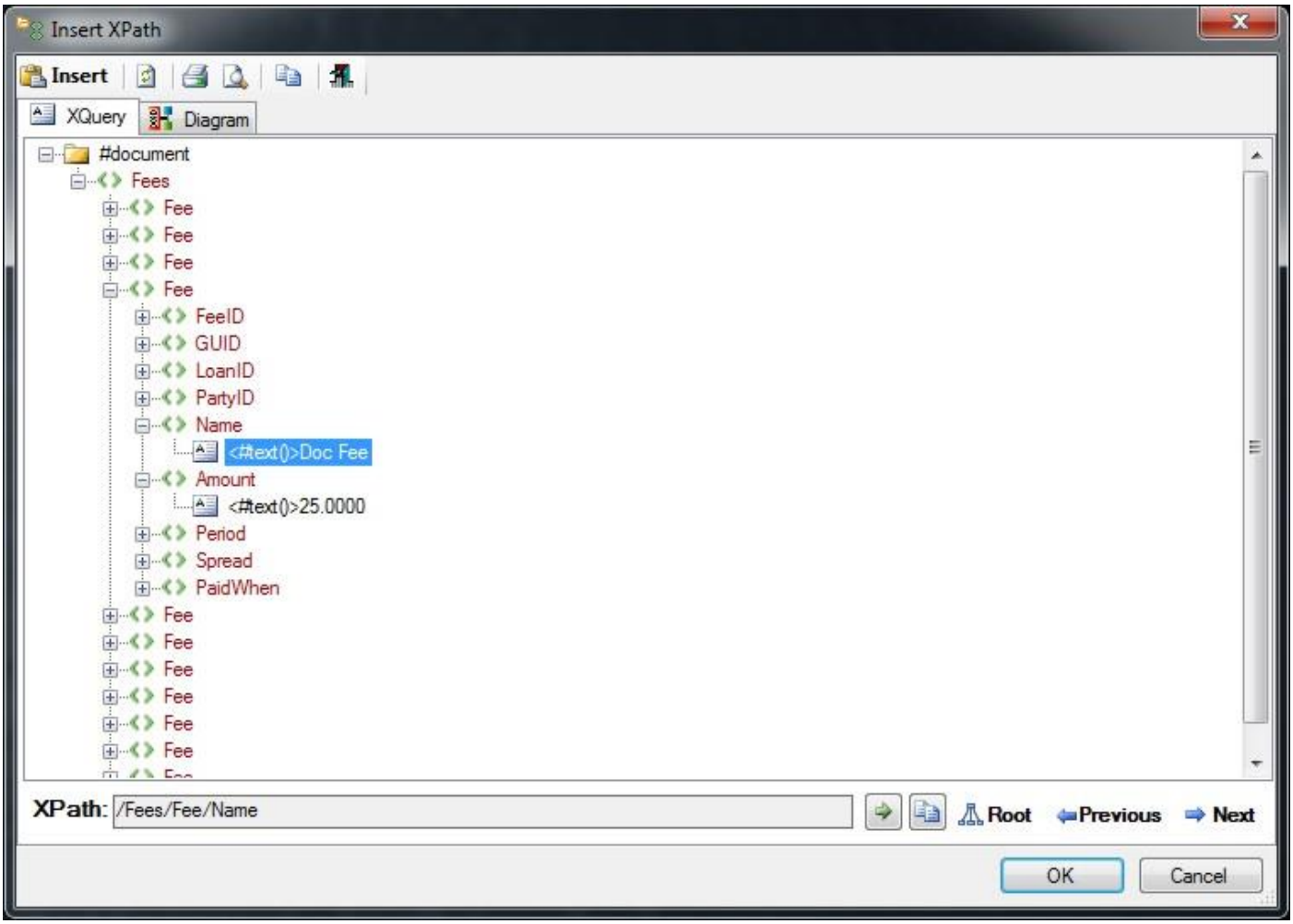
Predefined XQuery Functions can be inserted into the XQuery using the Function Tool.



XQuery contains a superset of XPath expression syntax to address specific parts of an XML document. The language is based on the XQuery and XPath Data Model (XDM) which uses a tree-structured model of the information content of an XML document.

Inserting an XPath

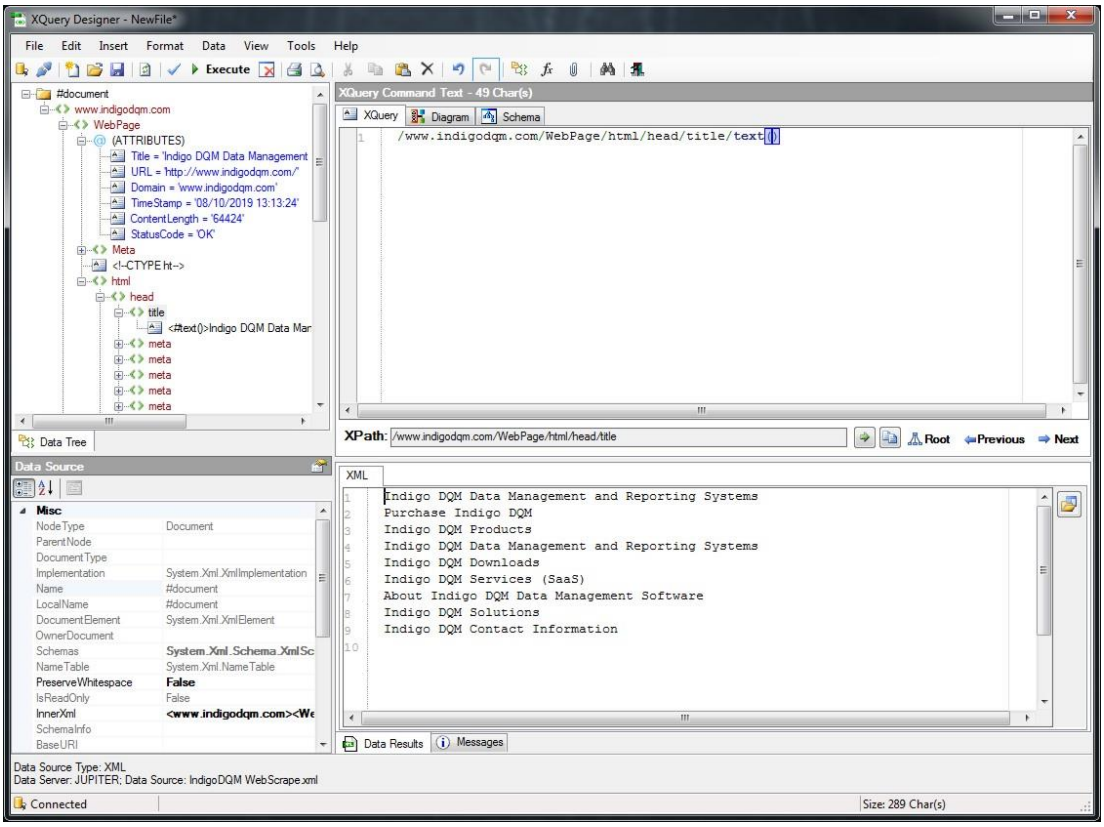
XPath can be used to navigate through elements and attributes in an XML document. XPath is a syntax for defining parts of an XML document and can be inserted by navigating the Data Tree or using the Insert Tool from the menu Insert | XPath.



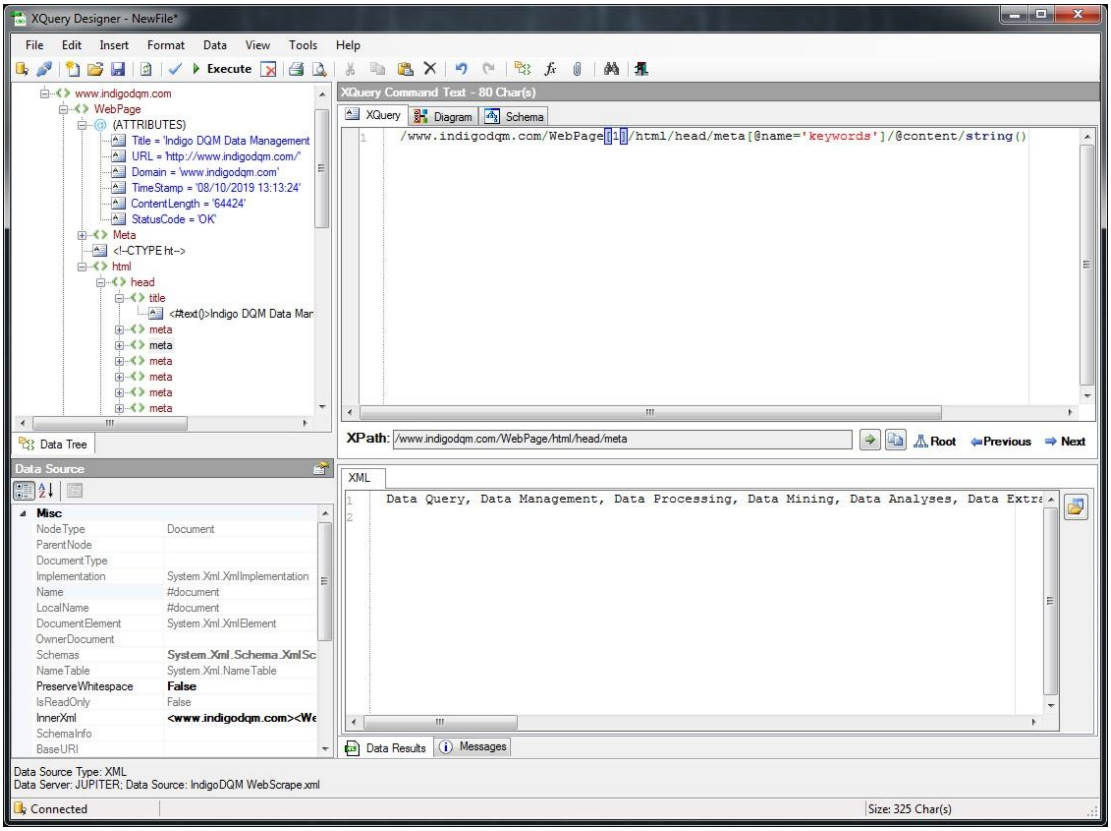
XPath uses path expressions to navigate in XML documents. Click Insert to add the current XPath expression to the XQuery Designer.

Extracting the Web Page Title using XQuery

Executing an XQuery statement for a Web Scrape to extract the Web Page Title.



Executing an XQuery statement for this Web Scrape to extract the Web Page Keywords.



XQuery contains a superset of XPath expression syntax to address specific parts of an XML document.