HTML protocol handler

The HTML protocol handler parses HTML documents (websites) and extracts attribute values using XPath expressions. This protocol handler can even parse unbalanced HTML documents using the CyberNeko HTML parser.

Options

- **delay**: Delay of reading in milliseconds (default 0).
- **nanodelay**: Delay of reading in nanoseconds (default 0).
- *Attribute* : XPath for the attribute in the document

Example

PQL

HTML Protocol Handler

```
input = ACCESS({source='Web', wrapper='GenericPush',
transport='File',protocol='HTML',
  dataHandler='Tuple',options=[['symbol','//td[last()-1]'],['points','//td[last()]']],
  schema=[
    ['symbol','String'],
    ['points','Double'] ]
})
```