Hyper-Text Query Language COM Interface

2011/03/24

1.   About

Hyper-Text Query Language (HTQL) is a language for the query and extraction HTML data. This guide explains the use of HTQL COM interface for potential uses in JavaScript, Visual Basic, ASP, .NET, and C++ applications.   HTQL can be used to:

1) Extract content from HTML pages
2) Retrieve HTML page through HTTP protocol
3) Modify HTML pages

HTQL syntax can be found at http://htql.net/htql-manual.pdf.

2.   Installation

1) Download the HtqlCom.dll into a local directory, such as 'C:\htql\'.
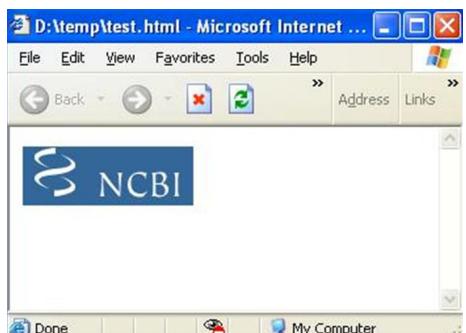2) Register the "HtqlCom.dll" by running:
   C:\htql\> regsvr32 HtqlCom.dll

3.   Simple Examples

The following example shows the use of HTQL in an HTML page with JavaScript.   The JavaScript code in this HTML page retrieves the first <a> tag from http://www.ncbi.nlm.nih.gov/ and show it in the HTML body.

```
<!--- test.html -->
<html> <base href="http://www.ncbi.nlm.nih.gov/">
<body>
<script language=JavaScript>
    var a= new ActiveXObject("HtqlCom.HtqlControl");
    a.setUrl("http://www.ncbi.nlm.nih.gov/");
    a.setQuery("<a>");
    document.write(a.getValueByIndex(1));
</script>
</body>
</html>
```

The resulting HTML page:

The following Visual Basic example does the same thing and shows the result in a message box:

```
' VB example
Dim a As Object
Set a = CreateObject("HtqlCom.HtqlControl")
i = a.setUrl("http://www.ncbi.nlm.nih.gov/")
i = a.setQuery("<a>")
MsgBox (a.getValueByIndex(1))
```

It will display a message box:

Project1

<A HREF="http://www.ncbi.nlm.nih.gov"><img src="corehtml/left.GIF" width="130" height="45" border="0" ALT="NCBI logo"></A>

OK

Below is a more complete example, which get HTQL results in a loop using moveFirst() and moveNext() functions until isEOF() returns true.

```
<!--- test2.html -->
<html> <base href="http://www.ncbi.nlm.nih.gov/">
<body>
<script language=JavaScript>
     var a= new ActiveXObject("HtqlCom.HtqlControl");
     a.setUrl("http://www.ncbi.nlm.nih.gov/");
     a.setQuery("<input>:name, type, value");
     for (a.moveFirst(); !a.isEOF(); a.moveNext()){
         for (i=1; i<a.getFieldsCount(); i++)
             document.write(a.getValueByIndex(i) + ", ");
         document.write("<br>\n");
     }
</script>
</body>
</html>
```

This example queries a page at http://www.ncbi.nlm.nih.gov/ with an HTQL expression:

   <input>:name, type, value

and displays the results in the HTML body as comma separated lines.

4. HTQL COM Interfaces

The COM interface includes functions in four categories: (1) to set the source page, (2) to set query expressions, (3) to set variables, and (4) to retrieve query results.   The following subsections describe these functions.

4.1 Set the Source Data

long setUrl(string url)

   Retrieve web page by url.

long setSourceData(string data, long length)
long setSourceUrl(string url)

   Set the web page and url directory.

long setUrlToPost(string url)
long setUrlParameter(string name, string value)
long setUrlCookie(string name, string value)
long postUrl()

   These functions simulate a form post.

4.2 Set HTQL queries

long setQuery(string query)
long moveFirst()
long moveNext()
long isEOF()

   Query the source data.

long dotQuery(string query)

   Continue a previous query

4.3 Use Global Variables

long setGlobalVariable(name, value)

   Global variables can be used in HTQL queries.

4.4 Get HTQL Query Results

string getValueByIndex(long index)
string getValueByName(string name)
long getTuplesCount()
long getFieldsCount()
string getFieldName(long index)

5. Reference

HTQL was designed and created by Dr. Liangyou Chen as part of his Ph.D. dissertation work:

Ad Hoc Integration and Querying of Heterogeneous Online Distributed Databases. 2004.
Mississippi State University.

Please cite the dissertation for references.