archana.mandape@verizon.com

DX- A language useful to transform delimited data to XML

In many information technology systems, data is usually stored in database and millions/trillions of database transactions are done everyday.

In order to view this data in a readable format, there is need of converting data from databases into XML form. Same applies to data stored in excel sheets. If this data is to be transferred over the internet it is beneficial if it is some predefined format.

Why XML?
In today's world XML(extensible markup language) is being used on a wide variety of platforms. XML is a structured form of data primarily designed to use on the web but it is beneficial to diverse applications. There are many XML parsers available today.

There is need for converting data from databases or from excel sheets into XML format.

The DX language that I plan to design and develop will give the user the ability to convert files exported from a database or from a excel sheet into XML format.

It can be extended to have procedures to perform operations on the xml data as well.

For eg, address records in the format
First_name;Last_name;city;email;phone_no
Tom ;Smith; Tampa; tomsmith@yahoo.com; 8133832844
Jerry ;Smith; New York; jerrysmith@google.com; 9149873456

will be converted to the following XML format

```
<address_book>
        <address_book_entry>
                <first_name>Tom</first_name>
                <last_name>Smith</last_name>
                <City>Tampa</City>
                <email></email>
                <phone_no>8133832844</phone_no>
        </address_book_entry>
        <address_book_entry>
                <first_name>Jerry</first_name>
                <last_name>Smith</last_name>
                <City>New York</City>
                <email>jerrysmith@google.com</email>
```

```
        <phone_no>9149873456</phone_no>
    </address_book_entry>
</address_book>
```

Although there may be some tools available to do the same, DX's goal is allow users to use this conversion from within code.

Some of the keywords that would be part of the language.

- openFile
- readRecord
- readHeaderRecord - would read header which defines the attribute names
- addRecToXML - add the records to XML
- delimiter - would be set to the delimiter in the input file
- RootTag - root element of the generated output xml
- RecordTag - record name in the generated output xml
- AttributeTag – record attribute in the generated output xml
- AttributeVal - record attribute value in the generated output xml.

To keep the language simple, there would be only one type of variable.

DX is a simple language with not many things to be done by a programmer. The programmer needs to specify input file, output file, delimiter used in the input file, the tags to be used in the output XML.

A sample program would look like :

```
Start Program;

    Set InputFile =  “infile.txt”;
    Set OutputFile = “outfile.txt”;
    Set delimiter = “,” ;
    Set RootTag = “address_book”
    Set RecordTag = “address_book_entry”
    Set AttributeTag = ReadHeaderRecord()
                        //Read first record

    While EOF
    {
        Set AttributeVal = AddRectoXML()
        //read rest of the records in the input file
    }
    Print(“outfile.txt”)
End Program;
```