

VorteXML Designer

Easy conversion of structured text data into valid XML without programming.

Key Benefits

- Works with output from any host system
- No programming knowledge required
- Rapid learning curve for setting up and processing files
- Requires no changes to existing applications
- Allows datatype recognition and processing even with DTD output
- Speeds delivery of business critical data-sharing integrated applications
- Drastically lowers the price of entry for XML projects
- Offers significant, immediate return on investment
- Developed by a company with over 14 years proven expertise in parsing and recognition technology

WHAT IS VORTEXML DESIGNER?

VorteXML Designer is a powerful data transformation tool that visually converts structured text data to XML without programming. It is expressly designed to empower business users to quickly and easily transform their operational data into XML. Users only need to understand their source data – no programming experience is required.

The VortexXML solution suite is comprised of two powerful products that work together:

- **VorteXML Designer** provides users a visual interface that allows them to build and test re-useable profiles (see sidebar for profile definition) to extract, transform and map data from existing text documents into XML without programming. It may also be used for simple one-time conversions.
- **VorteXML Server** provides a scalable, high-volume product that automates the extraction and conversion of text documents into XML. It includes facilities that perform automated processing of profiles created by VortexXML Designer as well as a simple Web Service interface.

WHO SHOULD USE VORTEXML DESIGNER?

VortexXML uses standard text files for input, so anyone can use it. A simple point-and-click user interface lets users define the data they want to extract. This makes VortexXML a quick and easy to use tool for technical and non-technical users alike who want to web-enable operations such as purchasing, billing, supply chain operations, financial transactions, and customer records.

WHY USE DOCUMENTS AS THE SOURCE?

Documents are the primary medium of information interchange in business today. Whether on paper or in electronic form, they contain required transactional information in a concise, recognizable and understandable format. Using text data files eliminates the need to decipher complex database schema or program logic. Documents provide the most straightforward access to information and are the medium most readily available to business users.

SOME CHALLENGES TO USING DOCUMENTS

Many organizations have thought of using business documents as a source for data conversion. However, positional text documents such as invoices and purchase orders present significant challenges to most software programs.

1. They contain irregular field lengths, varying numbers of lines, and patterns such as addresses, which are hard for software to consistently recognize and parse

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<Invoice InvNum="536017">
  <ShipTo>
    <Address>
      <Addressee>Betty's Music Store</Addressee>
      <Addr1>Muscatine Plaza</Addr1>
      <Addr2>200 Lower Muscatine</Addr2>
      <Addr3>Cedar Falls, IA 50613</Addr3>
    </Address>
  </ShipTo>
  
```

VortexXML designer provides an easy way of dealing with the challenges of using text documents.

Minimum System Requirements

Hardware:

Pentium III

64 MB RAM

30 MB disk space

Software:

Microsoft Windows

98, ME, 2000,

NT4 (SP6+), XP

Internet Explorer 5.5
SP2 +

MSML 3.0

MSML 4.0 (for XSD
schema support)

2. Source data may not be in the correct format, i.e. the source has "John Q. Public" and the DTD needs:
<LastName> Public </LastName>
<FirstName> John </FirstName>
3. Conversion rules developed for one instance of a document may fail when a subsequent document has more data fields or wrapping lines

VORTEXML PROVIDES THE SOLUTION

VorteXML solves these problems by providing users with a tool that is uniquely suited to the task of converting and augmenting today's business documents to XML.

The core of VortexXML is a powerful data extraction engine that can pull data out of virtually any structured text output, from simple columnar reports to highly formatted documents with headers, footers and multiple sort levels. VortexXML relies on the inherent regularity of patterns within a document to recognize and extract fields of information. It uses proven parsing and pattern recognition technology from Datawatch's industry leading Monarch™ report mining software to identify and extract data from documents. The template designer simplifies the definition process and requires no programming skills.

HOW VORTEXML WORKS

To begin the process, a representative document is loaded and an extraction template is defined for each distinct data level within the input. VortexXML's recognition engine automatically identifies data types and formats for date, numeric and currency fields and can even take a "best guess" at defining and naming fields. Visual feedback mechanisms provide verification of line trapping, field selection and data typing.

Extracted data is stored as distinct datasets determined by the template level from which the data was extracted. The user clarifies the input data hierarchy by specifying the direction and relationship

of the extraction templates to create an internal framework from the source document while keeping the datasets as separate entities (VortexXML contains an auto-definition feature to help with this process). This allows for the generation of output structures that are completely independent of the input sequence.

VortexXML's robust calculation and transformation engine can then filter, convert or derive information from within the datasets using a library of logical, mathematical and string manipulation functions. The ability to transform data alleviates concerns associated with using DTDs by providing date and numeric capabilities prior to generating the XML. Filters can also be applied to the data, to limit generation of documents to a specific subsets from a broader input source.

VortexXML allows the user to load the DTD/Schema for mapping of data and the creation of the XML document. If no DTD/Schema is available, VortexXML will generate an internal one based on the input sequence. VortexXML displays the XML structure in a tree format, showing all the elements and attributes. The easy-to-use mapping engine provides a mechanism for matching data with the appropriate output and establishing element repetition settings. Users may optionally specify a XSL or CSS style sheet to be applied to the output, and a DTD/Schema to validate the XML output.

XML CREATION

Once the parameters have been defined, the XML file is generated. The generated document can be browsed within the product, validated against the chosen DTD/Schema, and saved as output files. When multiple output files are created, a single file prefix with successive numbering differentiates the output.

All of the input definition, data transformation and mapping parameters can be saved as a profile for reuse for future exports or for subsequent automation with VortexXML Server.



www.datawatch.com