

1. Host Schema Evaluation

This evaluates: *Word ML*

I found some of the Microsoft documentation including the Overview of WordprocessingML and "New XML Features of the Microsoft Office Word 2003 Object Model" by Peter Vogel.

I also prepared a simple WordML XML file and loaded it into Microsoft Word successfully:

```
<w:wordDocument xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml" xmlns:wx="http://schemas.microsoft.com/office/word/2003/wordml/wx" >  
<w:body><wx:sect><w:p><w:r><w:t>aaa</w:t></w:r></w:p></wx:sect></w:body>  
</w:wordDocument>
```

By experimentation, I determined that the two xml namespace definitions were necessary.

1.1. Object Representation

Q: Does the host schema use a generic structural markup model?

A: The basic WordML document consists of just five tags:

- wordDocument
- body
- p
- r
- t

p is a paragraph, r allow for grouping text into runs to allow them to be formatted. And the text itself is inside t It also supports the section.

Q: Does the host schema define a "clause" object?

A: No, but a section or sub-section could be used for this purpose. (However, Neville Holmes and myself have found problems with how style sheets are applied to sections in Microsoft Word.

Q: Does the host schema define a paragraph level object that represents a structural or grammatical paragraph?

A: Yes, p.

Q: Using the host schema, can the clause equivalent object be inserted at arbitrary levels in the document hierarchy without transformation?

A: Unclear, as paragraphs don't contain sections.

Q: In the host schema, are element names and the structure sufficiently flexible that the clause and paragraph level objects can be used for other legal and business documents?

A: section and p elements are obviously used by Microsoft Word for many purposes.

1.2. Metadata

Q: Does the host schema provide a mechanism to add semantic information about:

- whole documents
- distinct objects, such as clauses, within documents?

A:

Q: If so, is the metadata model for the host schema sufficient for contracts or will it be necessary to extend it?

A: The DocumentProperties element includes information on title, version, and author. It is unclear to me how to extend this.

Q: Does the host schema allow embedded values to be represented and semantic information to be added to these values?

A: Microsoft Word XML supports field processing. It seems unnecessarily complex with fldData elements and several fldChar interspersed among r and t elements in a precise but difficult-to-use manner.

1.3. Processing Technologies

Q: Does the host schema require use of a particular processing technology?

A: Obviously, WordML is designed to be processed by Microsoft Word. However, one could easily process a simplified subset of this using other technologies such as XSLT. (I have seen web articles on creating WordML documents using XSLT.)

Q: Does the design of the host schema preclude use of particular currently available processing technologies?

A: See above.

1.4. Number of Content Objects

Q: Does the host schema permit the numbering of clauses, paragraphs, lists and other objects to be represented in the markup?

A: Lists can be created by including a `listPr` in a paragraph. This allows the user to override the numbering for a particular list item or define a style for a list.

Q: Does the host schema provide a mechanism to define the numbering schema applied to the document so that two applications could apply the same numbering, if desired?

A: See answer for previous question.

1.5. Complete Document Representation

Q: Using the host schema, will it be possible for the contract author to explicitly represent all parts of the narrative contact terms or will it be necessary to imply some parts?

A: The `subsection` and `section` elements could be used for this purpose.

Q: Does the host schema represent the relationship between all significant components in a way that allow high quality print and web rendition of of contact documents?

A: Obviously, Microsoft Word does provide quality style documents. Extensive style elements and style sheets are available in WordML.

1.6. Variables Definition

Q: Does the host schema include a mechanism for defining variables for embedded data values?

A: See above.

Q: If the host schema does not include such a mechanism, is there any obstacle to adding it?

A: See above.

1.7. Ease of use for authors

Q: Based on the following factors is the host schema easy for contract authors to use:

- Does it require authors to know only a small number of elements (positive factor)?
- Does it require authors make unnecessary or subtle distinctions that will be applied inconsistently (negative factor?)

- Does it have a clear logical structure that can be quickly explained to new users (positive factor)?
- Does it allow authors to re-locate content objects within a document hierarchy with minimal or no need for transformation of markup (positive factor)?

A: The use of sub-section, section, paragraph. Lists are supported, but as an add-on to paragraphs, they are relatively hard to use.

As, there is a definite hierarchy, moving text around from hierarchy to hierarchy does not seem feasible.

1.8. Schema Syntax

Q: Is the host schema a DTD only or can it also be expressed as an XML Schema or other schema type?

A: The schema is an XML Schema with multiple name spaces.

1.9. Adaptability to contracts

Q: Does the host schema provide for the complete representation for the distinct structures commonly found in contracts?

A: See answers above.

Q: If not, does the host schema explicitly allow additional distinct structures to be added?

A: I did not see any extension mechanisms on Microsoft's Web site.

Q: Does the host schema allow elements not considered necessary for contracts markup to be removed without contract documents being incompatible in a disadvantageous way with other documents using the host schema?

A: It is unclear how to remove elements from Microsoft's defined schema. However, only five elements are needed for Microsoft Word to read a document in and display it nicely.

Q: If distinct contract structures are added to the host schema, will this result in contracts documents being incompatible in a disadvantageous way with other documents using the host schema?

A: I added an element from an unknown schema as a test to the WordML XML. When I loaded the document into Microsoft Word, the element displayed with a purple highlighter.

1.10. Vendor and Developer Support

Q: Is the host schema already in widespread or general use for markup of narrative documents?

A: This is the representation for Microsoft Word documents.

Q: Are there already developed applications that will make it easy for organizations to implement the TC's specification based around the host schema?

A: This is the representation for Microsoft Word documents.

Q: Is there any reason to expect that the host schema will prove any particular advantages in gaining market support.

A: As mentioned in the Requirements Document, most Contracts are prepared using Microsoft Word.

1.11. Other Factors

Q: Does the host schema provide any other advantages for use in the TC's specification?

A:

Q: Does the host schema have any other disadvantages that make it undesirable for use in the TC's specification?

A: Nevill Holmes article in the November 2001 issue of *IEEE Computer* explained problems in formatting Microsoft Word documents, particularly when cutting and pasting from one document into another. In reading the specification for the XML representation for these documents, it is obvious why Microsoft Word has the problems that it does.