

XHTML Format For WorldPad File Equivalence (Version 1.0)

*Sharon Correll,
SIL Non-Roman Script Initiative (NRSI)
2003-04-03*

1. Introduction	1
1.1. Comment on the approach taken to hard-formatting.....	2
2. Summary of supported tags and attributes.....	2
3. Examples and discussion.....	4
3.1. Top level structure	4
3.2. Paragraph contents	4
3.3. Ranges of text.....	5
3.4. Character properties	6
3.4.1. Font and size.....	6
3.4.2. Inverting properties	6
3.4.3. Color	7
3.4.4. Vertical offsets.....	8
3.4.5. Underlining	8
3.4.6. Font features	9
3.5. Paragraph formatting.....	9
3.5.1. Alignment and direction	9
3.5.2. Indentation.....	10
3.5.3. Paragraph spacing and leading.....	11
3.5.4. Borders	11
3.5.5. Bullets	12
3.5.6. Numbering.....	14
3.6. Named styles	15
4. Unsupported features.....	18

1. Introduction

This document describes a subset of the XHTML/CSS standard that provides the equivalent functionality to WorldPad's XML format. This will enable WorldPad to function as an editor for documents that can be accessed over the Internet and viewed in a web browser. A particular goal is to provide a process for authoring web documents containing text that can be displayed using Graphite, SIL's smart-font rendering technology. This technology is currently supported by WorldPad and will be supported by the Mozilla browser in the near future, so a path to convert documents from the former format to the latter will enable us to meet this goal.

The aim of this specification is simply to define a format for web data which will fairly closely approximate what is seen WorldPad. The purpose is not to define the most compact or human-readable format, or one that facilitates further editing of the document in an HTML editor. There are also a few kinds of formatting where XHTML produces some deviation from the exact appearance of the data in WorldPad. These cases are summarized in section 4.

Ultimately the visual result of XHTML data will depend on the implementation of the application in which it is displayed. An attempt has been made to verify that the Internet Explorer and Netscape browsers provide at least reasonable renderings of the formats discussed here, and because it is the Netscape-based browser technology that will eventually have Graphite support, that browser is of the higher priority. However, I cannot guarantee that these renderings match either each other or WorldPad's display exactly.

1.1. Comment on the approach taken to hard-formatting

In WorldPad, the mechanisms for hard-formatting (associating formatting information directly with a range of text) and the style mechanism are quite similar. The property names and values are identical, as well as the approach for "cascading" or inheriting the values. In XHTML, there are two distinct but overlapping approaches. The standard way to apply hard-formatting is within XHTML proper, with attributes assigned to tags. Styles, on the other hand, are handled within CSS. Not only is the exact syntax of the two approaches different (`attribute="value"` vs. `attribute: value;`), but in many cases the attribute names differ (`dir` vs. `direction`, `align` vs. `text-align`, etc.).

It is also possible to achieve the effect of hard-formatting by putting CSS style definitions within the value of the style property attached to a tag, e.g., `<p style="prop: value; prop: value">`. In this paper I am suggesting that we use this approach, not only because it allows a more elegant structure to the program (i.e., the routines will generate one consistent syntax rather than two), but also because it seems to more accurately reflect the nature of the WorldPad data.

It is possible that in future versions of XHTML, using the style property to create hard-formatting in this way may be deprecated, at least within some contexts. In that event, we may need to auto-generate styles to represent each instance of hard-formatting.

2. Summary of supported tags and attributes

Item	WorldPad XML tag or attribute	XHTML tag or attribute	Comments about XHTML	See section
Top level structure	<WpDoc>	html		3.1
Main content of document	<Body>	<body>		3.1
Style information	<Styles>	<head>, <style>	All styles are represented within a single <style> tag.	3.6
Language information	<Languages>, <LgWritingSystem>	-none-	Language information is not represented in the XHTML format, with the exception of the default font.	
Page set-up information	<PageSetup>, <PageInfo>, etc.	-none-	Not represented.	
Paragraph contents	<StTxtPara>, <Contents16>	<p>, , 		3.2
Document direction	docRightToLeft	dir		3.2
Range of text with specified properties	<Str>, <Run>			3.2
Style definition	<Style>	<style>	All styles are represented within a single <style> tag.	3.6
Style properties	<Rules17>, <Prop>			3.6
Encoding-specific style properties	<WsStyles9999>, <WsProp>	stylenameENC	Each encoding has a corresponding style class.	3.6
Paragraph style	<StStyle>	<style>		3.5
Paragraph and text direction	rightToLeft	direction	The dir attribute is not available within CSS styles, so for the sake of	3.5.1

XHTML Format For WorldPad File Equivalence (Version 1.0)

Language identification	enc, ws	class=styleNameENC	Encoding information is not represented in the XHTML, except that it is used to choose a relevant style.	3.6
Font	fontFamily	font-family		3.4.1
Font size	fontSize, fontSizeUnit	font-size		3.4.1
Italic	italic	font-style		3.4.2
Bold	bold	font-weight		3.4.2
Foreground color	forecolor	color		3.4.3
Background color	backcolor	background-color		3.4.3
Vertical positioning	offset, offsetUnit	vertical-align		3.4.4
Superscripts and subscripts	superscript	vertical-align: sub/super, font-size: 75%		3.4.4
Underlining	underline	text-decoration, text-decoration-style	Only the single (solid) underline style is supported in CSS2	3.4.5
Underline color	undercolor	text-decoration-color	Underline color is not supported in CSS2.	3.4.5
Font features	fontVariations	-sil-font-feature-ID	This is not part of the CSS standard.	3.4.6
Named styles	namedStyle	class		3.6
Paragraph alignment	align	text-align		3.5.1
Paragraph first-line indentation	firstIndent	text-indent, margin-left(/right)		3.5.2
Paragraph indentation	leadingIndent	margin-left(/right)		3.5.2
Paragraph indentation	trailingIndent	margin-right(/left)		3.5.2
Vertical space above paragraph	spaceBefore, marginTop	margin-top		3.5.3
Vertical space below paragraph	spaceAfter, marginBottom	margin-bottom		3.5.3
Paragraph leading	lineHeight, lineHeightUnit	line-height	"At least" line spacing is not supported.	3.5.3
Border width	borderTop/Bottom/Leading/Trailing, padLeading/Trailing	border- width, border-top/bottom/left/right-width, padding-left/right		3.5.4
Border color	borderColor	border-color		3.5.4
Labeling of paragraphs with bullets or numbers	bulNumScheme	, , list-style, list-style-type, list-style-position, list-style-image?	Values supported in CSS2 are disc, circle, square, decimal, lower/upper-roman, lower/upper-alpha, none	3.5.5, 3.5.6
Numbering	bulNumStartAt	start		3.5.5
Numbering formatting	bulNumTxtBef/Aft	before, :after		3.5.6

Numbering
formatting bulNumFontInfo

Supported tags and attributes

Not supported in
the XHTML format. 3.5.6

3. Examples and discussion

3.1. Top level structure

WorldPad XML:

```
<WpDoc ...>
  <Languages>
    ...
  </Languages>
  <Styles>
    ...
  </Styles>
  <Body ...>
    ...
  </Body>
  <PageSetup>
    ...
  </PageSetup>
</WpDoc>
```

XHTML:

```
<html>
  <head>
    <style type="text/css">
      ...
    </style>
  </head>
  <body>
  </body>
</html>
```

Note that there are no sections in the XHTML format to correspond to <Languages> or <PageSetup>.

3.2. Paragraph contents

WorldPad XML:

```
<Body docRightToLeft=true>
  <StTxtPara>
    <StyleRules15>
      <Prop .../>
    </StyleRules15>
    <Contents16>
      <Str>...</Str>
    </Contents16>
  </StTxtPara>
</Body>
```

XHTML:

```
<body dir=rtl>
  <p style = "...">
    ...
  </p>
</body>
```

Styles can also be specified within the <p> tag with a named style, as shown below:

WorldPad XML:

```
<Body docRightToLeft=false>
  <StTxtPara>
```

```

    <StyleRules15>
      <Prop namedStyle="heading"/>
    </StyleRules15>
    <Contents16>
      <Str>...</Str>
    </Contents16>
  </StTxtPara>
</Body>

```

XHTML:

```

<body dir=ltr>
  <p class=heading>
    ...
  </p>
</body>

```

3.3. Ranges of text

WorldPad XML:

```

<Str>
  <Run enc=XXX ws=## ...>run one</Run>
  <Run enc=XXX ws=## ...>run two</Run>
</Str>

```

XHTML:

```

<p>
  <span ...>run one</span>
  <span ...>run two</span>
</p>

```

If the writing system is right-to-left, the `` tag will include a `direction` attribute. This attribute is not required for left-to-right writing systems.

WorldPad XML:

```

<Str>
  <Run enc="HEB" ws="0" ...>Hebrew</Run>
  <Run enc="ENG" ws="0" ...>English</Run>
</Str>

```

XHTML:

```

<p>
  <span style="direction: rtl">Hebrew</span>
  <span>English</span>
</p>

```

Named character styles are handled by appending the encoding ID to the name of the style. (This is only necessary and ideally will only happen when the style definitions differ for the writing systems.)

See section 3.6 for more discussion.

WorldPad XML:

```

<Str>
  <Run enc="HEB" ws="0" namedStyle="highlight">Hebrew</Run>
  <Run enc="ENG" ws="0" namedStyle="highlight">English</Run>
</Str>

```

XHTML:

```

<p>
  <span class=highlightHEB style="direction: rtl">Hebrew</span>
  <span class=highlightENG>English</span>
</p>

```

3.4. Character properties

3.4.1. Font and size

WorldPad XML:

```
<Str>
  <Run ... fontFamily="Gentium">text</Run>
  <Run ... fontFamily="default serif">text</Run>
  <Run ... fontFamily="default sans serif">text</Run>
  <Run ... fontFamily="default monospace">text</Run>
  <Run ... fontsize="12000" fontsizeUnit="mpt">text</Run>
  <Run ... fontsize="7500" fontsizeUnit="mpt">text</Run>
</Str>
```

XHTML:

```
<p>
  <span style="font-family: Gentium">text</span>
  <span style="font-family: serif">text</span>
  <span style="font-family: sans-serif">text</span>
  <span style="font-family: monospace">text</span>
  <span style="font-size: 12pt">text</span>
  <span style="font-size: 7.5pt">text</span>
</p>
```

Note that currently the only value of `fontsizeUnit` that is generated by WorldPad is "mpt".

3.4.2. Inverting properties

In FW data, the bold and italic character properties can take the value "invert," which makes their meaning dependent on the context in which they occur. Most often "invert" mean "on", but within a run of text where bold is already on, "invert" means "off." Thus when the "Bold Chars" style is defined using `bold = "invert"` we get the following mapping:

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop namedStyle="Bold Paragraph"/>
  </StyleRules15>
  <Contents16>
    <Str>
      <Run ... >bold</Run>
      <Run ... bold="invert">not bold</Run>
      <Run ... bold="on">bold</Run>
      <Run ... bold="off">not bold</Run>
      <Run ... namedStyle="Bold Chars">not bold</Run>
    </Str>
  </Contents16>
</StTxtPara>
<StTxtPara>
  <StyleRules15>
    <Prop namedStyle="Non-Bold Paragraph"/>
  </StyleRules15>
  <Contents16>
    <Str>
      <Run ... >not bold</Run>
      <Run ... bold="invert">bold</Run>
      <Run ... bold="on">bold</Run>
      <Run ... bold="off">not bold</Run>
      <Run ... namedStyle="Bold Chars">bold</Run>
    </Str>
  </Contents16>
</StTxtPara>
```

XHTML:

```
<p class=boldParagraph>
  <span>bold</span>
  <span style="font-weight: normal">not bold</span>
  <span style="font-weight: bold">bold</span>
  <span style="font-weight: normal">not bold</span>
```

XHTML Format For WorldPad File Equivalence (Version 1.0)

```

<span class=boldChars style="font-weight: normal">not bold</span>
</p>
<p class=nonBoldParagraph>
  <span>not bold</span>
  <span style="font-weight: bold">bold</span>
  <span style="font-weight: bold">bold</span>
  <span style="font-weight: normal">not bold</span>
  <span class=boldChars>bold</span>
</p>

```

The fifth run of text in the first paragraph demonstrates a special problem. In WorldPad, the application of the "Bold Paragraph" style turns bolding on, then the application of the "Bold Chars" style inverts the property, turning it off. Since there is no concept of inverting in CSS, both styles will produce bold text. So we must explicitly turn bolding off using an inline style, to ensure the same results as the WorldPad format.

A similar approach is needed for the italic property.

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop namedStyle="Italic Paragraph"/>
  </StyleRules15>
  <Contents16>
    <Str>
      <Run ... >italic</Run>
      <Run ... italic="invert">not italic</Run>
      <Run ... italic="on">italic</Run>
      <Run ... italic="off">not italic</Run>
      <Run ... namedStyle="Italic Chars">not italic</Run>
    </Str>
  </Contents16>
</StTxtPara>

```

XHTML:

```

<p class=italicParagraph>
  <span>italic</span>
  <span style="font-style: normal">not italic</span>
  <span style="font-style: italic">italic</span>
  <span style="font-style: normal">not italic</span>
  <span class=italicChars style="font-style: normal">not italic</span>
</p>

```

3.4.3. Color**WorldPad XML:**

```

<Str>
  <Run ... forecolor="red">text</Run>
  <Run ... forecolor="cyan">text</Run>
  <Run ... bgcolor="yellow">text</Run>
  <Run ... forecolor="888888">text</Run>
</Str>

```

XHTML:

```

<p>
  <span style="color: red">text</span>
  <span style="color: aqua">text</span>
  <span style="background-color: yellow">text</span>
  <span style="color: #888888">text</span>
</p>

```

The following table shows the color equivalences between WorldPad XML and XHTML. All other colors will be described in term of an 6-digit hex number (RRGGBB).

WorldPad XML	XHTML	Value
white	white	FFFFFF
black	black	000000
red	red	FF0000

XHTML Format For WorldPad File Equivalence (Version 1.0)

green	green	008000
blue	blue	0000FF
magenta	fuchsia	FF00FF
cyan	aqua	00FFFF
yellow	yellow	FFFF00
transparent	transparent	

Color equivalences

The "magenta" and "cyan" color values seem to be recognized by some XHTML implementations, but given that they are not documented, it is preferable to use "fuchsia" and "aqua."

3.4.4. Vertical offsets

WorldPad XML:

```
<Str>
  <Run ... offset="3000" offsetUnit="mpt">text</Run>
  <Run ... offset="-3000" offsetUnit="mpt">text</Run>
  <Run ... offset="-6000" offsetUnit="mpt">text</Run>
  <Run ... superscript="super">text</Run>
  <Run ... superscript="sub">text</Run>
  <Run ... superscript="super" offset="2000" offsetUnit="mpt">text</Run>
</Str>
```

XHTML:

```
<p>
  <span style="vertical-align: super">text</span>
  <span style="vertical-align: sub">text</span>
  <span style="vertical-align: sub">text</span>
  <span style="vertical-align: super; font-size: 75%">text</span>
  <span style="vertical-align: sub; font-size: 75%">text</span>
  <span style="vertical-align: super; font-size: 75%">
    <span style="offset: 2pt">text</span>
  </span>
</p>
```

Notice that two levels of `` are required to handle a superscript or subscript with an adjusted offset.

Currently the only value of `offsetUnit` that is generated by WorldPad is "mpt".

3.4.5. Underlining

In CSS2, only a single, solid underline in a color matching the text is supported. However, in CSS3, various styles of colored underlining, such as double, dotted, and dashed, will be supported. For the sake of forward compatibility, it seems wise to implement the CSS3-compatible XHTML at the outset; within CSS2, the new attributes will simply be ignored, resulting in solid underlines.

WorldPad XML:

```
<Str>
  <Run ... underline="single">text</Run>
  <Run ... underline="double">text</Run>
  <Run ... underline="dotted">text</Run>
  <Run ... underline="dashed">text</Run>
  <Run ... underline="squiggle" undercolor="red">text</Run>
</Str>
```

XHTML:

```
<p>
  <span style="text-decoration: underline; text-decoration-style: solid">
    text</span>
  <span style="text-decoration: underline; text-decoration-style: double">
    text</span>
  <span style="text-decoration: underline; text-decoration-style: dotted">
    text</span>
  <span style="text-decoration: underline; text-decoration-style: dashed">
    text</span>
</p>
```



```

<span style=
  "    text-decoration: underline;
      text-underline-style: wave;
      text-underline-color: red">
  text</span>
</p>

```

3.4.6. Font features

Font features, such as those used for Graphite rendering, are not directly supported by CSS. An extension will be required to handle them, using the recommended syntax for proprietary identifiers:

WorldPad XML:

```

<Str>
  <Run ... font="MyGraphiteFont" fontVariations="100=2, 101=1">text</Run>
</Str>

```

XHTML:

```

<p>
  <span style=
    "    font-family: MyGraphiteFont;
        -sil-font-feature-100: 2;
        -sil-font-feature-101: 1"
  >text</span>
</p>

```

Because the `-sil-font-feature` property is proprietary, it is currently not supported by any browser, and we expect this to be the situation for the foreseeable future.

3.5. Paragraph formatting

3.5.1. Alignment and direction

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop align="right"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
<StTxtPara>
  <StyleRules15>
    <Prop align="center"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>

```

XHTML:

```

<p style="text-align: right">
  ...
</p>
<p style="text-align: center">
  ...
</p>

```

As with WorldPad, default alignment automatically follows the paragraph direction, and therefore does not need to be specified, as shown in the first paragraph below.

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop rightToLeft="1"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
<StTxtPara>

```

```

<StyleRules15>
  <Prop rightToLeft="1" alignment="trailing"/>
</StyleRules15>
<Contents16>...</Contents16>
</StTxtPara>

```

XHTML:

```

<p dir="rtl">
  ...
</p>
<p dir="rtl" style="text-align: left">
  ...
</p>

```

3.5.2. Indentation

The following represents a standard indentation:

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop firstIndent="36000"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>

```

XHTML:

```

<p style="text-indent: 36pt">
  ...
</p>

```

The following represents a hanging indentation. Notice that both the `text-indent` and `margin-left` attributes are needed in the XHTML format.

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop firstIndent="-36000"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>

```

XHTML:

```

<p style="text-indent: -36pt; margin-left: 36pt">
  ...
</p>

```

Specifically, the formula for `margin-left` is $(\text{leadingIndent} - \text{firstIndent})$, as shown by the following example:

WorldPad XML:

```

<StTxtPara>
  <StyleRules15>
    <Prop firstIndent="-16000" leadingIndent=20000/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>

```

XHTML:

```

<p style="text-indent: -16pt; margin-left: 36pt">
  ...
</p>

```

When dealing with right-to-left paragraphs, use `margin-right` for `leadingIndent` and `margin-left` for `trailingIndent`:

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop rightToLeft="1" leadingIndent="36000" trailingIndent="18000"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="direction: rtl; margin-left: 18pt; margin-right: 36pt">
  ...
</p>
```

3.5.3. Paragraph spacing and leading**WorldPad XML:**

```
<StTxtPara>
  <StyleRules15>
    <Prop spaceBefore="6000" spaceAfter="3000"
      lineHeight="2" lineHeightUnit="rel"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="margin-top: 6pt; margin-bottom: 3pt; line-height: 2.0">
  ...
</p>
```

"At least" line spacing is not supported in CSS, and will be treated as "exact."

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop lineHeight="-20000" lineHeightUnit="mpt"/>
  </StyleRules15>
  <Contents16>this text is supposed to use exact line spacing</Contents16>
</StTxtPara>
<StTxtPara>
  <StyleRules15>
    <Prop lineHeight="20000" lineHeightUnit="mpt"/>
  </StyleRules15>
  <Contents16>this text is supposed to use 'at least' spacing</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="line-height: 20pt">
  this text is supposed to use exact line spacing
</p>
<p style="line-height: 20pt">
  this text is supposed to use 'at least' spacing
</p>
```

3.5.4. Borders

When all four borders are visible, the border-width property can be used:

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop padLeading="4000" padTrailing="4000"
      borderLeading="2250" borderTrailing="2250"
      borderTop="2250" borderBottom="2250" borderColor="blue"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="padding-left: 4pt; padding-right: 4pt;
border: solid; border-color: blue; border-width: 2.25pt">
...
</p>
```

Note that in WorldPad all borders are solid, 4 points of padding to the right and left of the border are automatically included, and visible borders must be a uniform width and color.

When not all borders are visible, the missing ones must be specified in XHTML as width zero:

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop borderLeading="6000" borderTrailing="6000"
borderColor="00ff90c0"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="padding-left: 4pt; padding-right: 4pt;
border: solid; border-color: #ff90c0;
border-left-width: 6pt; border-right-width: 6pt;
border-top-width: 0; border-bottom-width: 0">
...
</p>
```

When working with a right-to-left paragraphs, `border/padLeading` correspond to `border/padding-right` and `border/padTrailing` correspond to `border/padding-left`, as shown below:

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop rightToLeft="1" padLeading="4000" borderLeading="1000"/>
  </StyleRules15>
  <Contents16>...</Contents16>
</StTxtPara>
```

XHTML:

```
<p style="border: solid; padding-right: 4pt border-right-width: 1pt;">
...
</p>
```

3.5.5. Bullets

In CSS2, there is a limited number of available bullet styles, smaller than those possible in WorldPad. It is possible to handle a wider number either by using web addresses to access picture files on the Internet or using the advanced `marker` construct. For the first version, however, I suggest using a simple mapping between WorldPad's bullets and the styles available in CSS2, as shown in the following table:

WorldPad	WorldPad bulNumScheme	CSS2
● •	101, 102	disc
○	103	circle
■ ■ □ □ □	104, 105, 106, 107, 108, 109	square
all others	100, 110-124	disc

Bullet equivalences

In CSS3, additional bullet styles will be supported, at which time the following correspondences will be appropriate:

WorldPad	WorldPad bulNumScheme	CSS3
● •	101, 102	disc
○	103	Circle

XHTML Format For WorldPad File Equivalence (Version 1.0)

◻ ◻ ◻ ◻	106, 107, 108, 109	square
▪ ■	104, 105	box
✓	124	check
◆ ◆ ❖	111, 112, 113	diamond
all others	100, 110, 114-123	disc

Bullet equivalences in CSS3

All paragraphs with identical bullet properties will be grouped together under the same tag.

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop firstIndent="-18000" bulNumScheme="106"
      bulNumStartAt="1"/>
  </StyleRules15>
  <Contents16><Str><Run>item one</Run></Str></Contents16>
</StTxtPara>
<StTxtPara>
  <StyleRules15>
    <Prop firstIndent="-18000" bulNumScheme="106"/>
  </StyleRules15>
  <Contents16><Str><Run>item two</Run></Str></Contents16>
</StTxtPara>
```

XHTML:

```
<ul style="margin-left: 18pt; list-style-type: square">
  <li>item one</li>
  <li>item two</li>
</ul>
```

(Any bulNumStartAt property in the WorldPad data is superfluous.)

The above represents a standard hanging indent as is commonly used for bulleted items. The sample below produces a bullet inside the flow of the paragraph, with no indentation.

WorldPad XML:

```
<StTxtPara>
  <StyleRules15>
    <Prop bulNumScheme="101"/>
  </StyleRules15>
  <Contents16><Str><Run>item one</Run></Str></Contents16>
</StTxtPara>
<StTxtPara>
  <StyleRules15>
    <Prop bulNumScheme="101"/>
  </StyleRules15>
  <Contents16><Str><Run>item two</Run></Str></Contents16>
</StTxtPara>
```

XHTML:

```
<ul style="list-style-type: disc; list-style-position: inside;
  margin-left: 0">
  <li>item one</li>
  <li>item two</li>
</ul>
```

As a general principle, whenever the WorldPad format indicates a hanging indent (firstIndent is negative), the CSS list-style-position should be "outside" (the default). In other situations list-style-position should be "inside." If the position is outside, the margin-left setting should not be less than about 6pt.

As always, if the paragraph is right-to-left, margin-left should be replaced by margin-right. Nothing else special is needed to ensure that bullets appear on the right side of the paragraph.

3.5.6. Numbering

According to the CSS2 on-line documentation, the following should generate paragraphs with automatic formatted numbering:

WorldPad XML:

```
<Styles>
  <StStyle>
    <Name17><Uni>Upper Bracket</Uni></Name17>
    <Rules17>
      <Prop firstIndent="-18000" bulNumScheme="13" "bulNumStartAt="3"
        bulNumtxtBef="[" bulNumtxtAft="]"/>
    </Rules17>
  </StStyle>
  <StStyle>
    <Name17><Uni>Red Bold</Uni></Name17>
    <Rules17>
      <Prop bold="invert" forecolor="red"/>
    </Rules17>
  </StStyle>
</Styles>
<Body>
  <StTxtPara>
    <StyleRules15>
      <Prop namedStyle="blueBorder"/>
    </StyleRules15>
    <Contents16>
      <Str><Run namedStyle="redBold">This is bold, red text.</Run></Str>
    </Contents16>
  </StTxtPara>
</Body>
```

XHTML:

```
<head>
  <style type="text/css">
    ol.upperBracket { counter-reset: item 3 }
    li.upperBracket:before {
      content: "[" counter(item, upper-alpha) " ] ";
      counter-increment: item }
  </style>
</head>
<body>
  <ol class=upperBracket>
    <li class=upperBracket >item one</li>
    <li class=upperBracket >item two</li>
    <li class=upperBracket >item three</li>
  </ol>
</body>
```

However, there does not seem to be good support in either the Netscape or IE browsers for the counter mechanism. So instead we will use the XHTML start attribute, and the formatting of the label using before- or after-text will not be handled. Thus the above example becomes:

XHTML:

```
<head>
  <style type="text/css">
    ol.upperBracket { list-style-type: upper-alpha }
  </style>
</head>
<body>
  <ol start=3 class=upperBracket>
    <li>item one</li>
    <li>item two</li>
    <li>item three</li>
  </ol>
</body>
```

The table below shows the supported numbering styles:

WorldPad	WorldPad bulNumScheme	CSS2
1, 2, 3 and 01, 02, 03	10, 15	decimal

I, II, III	11	upper-roman
i, ii, iii	12	lower-roman
A, B, C	13	upper-alpha
a, b, c	14	lower-alpha

Numbering equivalences

In WorldPad, all paragraphs that have the same `bulNumScheme` property are considered to be part of the same numbered list, and will be placed together within a single `` tag. Differences in the paragraph indentation or the formatting of the label should not be considered for this purpose.

3.6. Named styles

In theory, one might expect that WorldPad's character and paragraph styles could be handled in a straightforward manner using the concept of CSS style classes.

WorldPad XML:

```
<Styles>
  <StStyle>
    <Name17><Uni>Blue Border</Uni></Name17>
    <Rules17>
      <Prop borderColor="blue" borderTop="2000" borderBottom="2000"
        borderLeading="2000" borderTrailing="2000"/>
    </Rules17>
  </StStyle>
  <StStyle>
    <Name17><Uni>Red Bold</Uni></Name17>
    <Rules17>
      <Prop bold="invert" forecolor="red"/>
    </Rules17>
  </StStyle>
</Styles>
<Body>
  <StTxtPara>
    <StyleRules15>
      <Prop namedStyle="blueBorder"/>
    </StyleRules15>
    <Contents16>
      <Str><Run namedStyle="redBold">This is bold, red text.</Run></Str>
    </Contents16>
  </StTxtPara>
</Body>
```

XHTML:

```
<head>
  <style type="text/css">
    .blueBorder { border: solid; border-color: blue;
      border-width: 2pt }
    .redBold { color: red; font-weight: bold }
  </style>
</head>
<body>
  <p class=blueBorder>
    <span class=redBold>This is bold, red text.</span>
  </p>
</body>
```

In reality, however, the issue is complicated by WorldPad's handling of language-specific character formatting. There are two aspects to this issue.

In WorldPad, each style essentially incorporates a set of language-specific sub-styles that are used for formatting the character properties of text in each language. This capability exists for both paragraph and character styles.

Therefore, each WorldPad style may potentially result in a set of style classes in XHTML, with items in the set corresponding to the languages for which the character properties are specified. (Ideally, this will only happen when the language-specific character properties do in fact differ for the various languages.) The

XHTML Format For WorldPad File Equivalence (Version 1.0)

language-specific style class will be applied to spans of text that use that language and for which the style was in effect.

In addition, each writing system is assigned a default font. This information is stored in the <DefaultSerif24> tag of the writing system. (The <DefaultSansSerif24> and <DefaultMonospace24> tags are currently redundant and can be ignored.) When no other font is specified for a given style, the default font should be used.

Furthermore, each paragraph property will generate a corresponding non-language-specific style class containing the paragraph formatting (alignment, borders, etc.). This will be applied at the paragraph level, while the language-specific class is applied at the text span level. The following example shows a paragraph style with language-specific character formatting:

WorldPad XML:

```
<Languages>
  <LgWritingSystem id="ENG">
    ...
    <DefaultSerif24><Uni>Times New Roman</Uni></DefaultSerif24>
  </LgWritingSystem>
  <LgWritingSystem id="DEV">
    ...
    <DefaultSerif24><Uni>Code2000</Uni></DefaultSerif24>
  </LgWritingSystem>
  <LgWritingSystem id="SPN">
    ...
    <DefaultSerif24><Uni>Arial</Uni></DefaultSerif24>
  </LgWritingSystem>
</Languages>

</Languages>
<Styles>
  <StStyle>
    <Name17><Uni>Large Centered Title</Uni></Name17>
    <Rules17>
      <Prop align="center">
        <WsStyles9999>
          <WsProp ws="DEV" fontsize="22pt" fontsizeUnit="mpt" />
          <WsProp ws="ENG" fontsize="16pt" fontsizeUnit="mpt" />
          <WsProp ws="SPN" fontsize="16pt" fontsizeUnit="mpt"
            italic="invert" />
        </WsStyles9999>
      </Prop>
    </Rules17>
  </StStyle>
</Styles>
<Body>
  <StTxtPara>
    < StyleRules15>
      <Prop namedStyle="Large Centered Title" />
    </StyleRules15>
    <Contents16>
      <Str>
        <Run ws="ENG">The Use of </Run>
        <Run ws="SPN">Hola</Run>
        <Run ws="ENG"> in Mexican Spanish</Run>
      </Str>
    </Contents16>
  </StTxtPara>
</Body>
```

XHTML:

```
<head>
  <style type="text/css">
    .largeCenteredTitle { text-align: center }
    .largeCenteredTitle_DEV { font-family: Code2000; font-size: 22pt }
    .largeCenteredTitle_ENG { font-family: Times New Roman; font-size: 16pt }
    .largeCenteredTitle_SPN { font-family: Arial; font-size: 16pt; font-style:
  italic }
  </style>
</head>
<body>
  <p class=largeCenteredTitle>
```


XHTML Format For WorldPad File Equivalence (Version 1.0)

```

    <span class=largeCenteredTitle_ENG>The Use of </span>
    <span class=largeCenteredTitle_SPN>Hola</span>
    <span class=largeCenteredTitle_ENG> in Mexican Spanish</span>
  </p>
</body>

```

Note that in WorldPad, one style may be based on another, inheriting all its formatting properties. For the purposes of XHTML output, the equivalent style must include all the property settings inherited from all super-styles.

When there is a cascading effect of language-specific styles, two levels of are required, one for the paragraph style and one for the character style. Implementationally, it may be useful for the XHTML generation routines to carefully analyze when language-specific formatting exists and when it does not, in order to avoid superfluous tags. The following example shows the use of both a paragraph and a character style, each of which having language-specific character formatting:

WorldPad XML:

```

<Styles>
  <StStyle>
    <Name17><Uni>Centered Title</Uni></Name17>
    <Rules17>
      <Prop align="center" />
      <WsStyles9999>
        <WsProp enc="DEV" font-weight="bold" />
        <WsProp enc="ENG" />
        <WsProp enc="SPN" font-style="italic" />
      </WsStyles9999>
    </Prop>
  </Rules17>
</StStyle>
<StStyle>
  <Name17><Uni>Large Text</Uni></Name17>
  <Rules17>
    <Prop />
    <WsStyles9999>
      <WsProp enc="DEV" fontsize="22000" fontsizeUnit="mpt" />
      <WsProp enc="ENG" fontsize="18000" fontsizeUnit="mpt" />
      <WsProp enc="SPN" fontsize="18000" fontsizeUnit="mpt" />
    </WsStyles9999>
  </Prop>
</Rules17>
</StStyle>
<Styles>
<Body>
  <StTxtPara>
    < StyleRules15>
      <Prop namedStyle="Centered Title" />
    </StyleRules15>
    <Contents16>
      <Str>
        <Run enc="ENG" namedStyle="Large Text">The Use of </Run>
        <Run enc="SPN" namedStyle="Large Text">Hola</Run>
        <Run enc="ENG"
          namedStyle="Large Text"> in Mexican Spanish</Run>
      </Str>
    </Contents16>
  </StTxtPara>
</Body>

```

XHTML:

```

<head>
  <style type="text/css">
    .centeredTitle { text-align: center }
    .centeredTitle_BEN { font-weight: bold }
    .centeredTitle_SPN { font-style: italic }
    .largeText_BEN { font-size: 22pt }
    .largeText_ENG { font-size: 18pt }
    .largeText_SPN { font-size: 18pt }
  </style>
</head>
<body>
  <p class=centeredTitle>
    <span class=largeText_ENG>The Use of </span>

```

```
<span class=centeredTitle_SPN>
  <span class=largeText_SPN>Hola</span>
</span>
<span class=largeText_ENG> in Mexican Spanish</span>
</p>
</body>
```

4. Unsupported features

The following behaviors are supported in WorldPad but not in the corresponding XHTML format:

- Formatting of auto-generated labels with before- and after-text.
- The font specification for bullets and numbers in labeled paragraphs: the font used will be that for the paragraph itself.
- "At least" line spacing; will be treated as "exact."
- The approach to handling indentation of bulleted lists will produce approximately but not exactly the same results in XHTML as in WorldPad.

Features that will be supported in CSS3 include:

- Double, dashed, and dotted underlines.
- Underline color.
- Box, check, and diamond bullet styles.

This document was generated from
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=WorldPadXhtmlFormat
on Fri, 29 Aug 2003 12:22:25 -0500.

END