Welcome to
Nokia's WAP 2.0
XHTML browser
for small devices.

# Advantages of XHTML for Wireless Data

NOKIA
CONNECTING PEOPLE

# Contents

# Introduction
# WAP 2.0 is XHTML

For mobile devices, the future of WAP lies in its close alignment with widely accepted Internet standards. The WAP Forum and the World Wide Web Consortium (W3C) have successfully defined mobile Internet standards over the past several years  Most recently, the WAP Forum has moved to adopt the XHTML Basic™ standard from the W3C as the basis for WAP 2.0. This transition to XHTML Basic will strengthen the mobile browser's position within the Internet mainstream and allow for a far greater range of presentation design and formatting possibilities than previously possible. Consequently, consumers will enjoy a wider array of services, more intuitive user interfaces, and a generally more useful experience. At the same time, it will also enable carriers to exercise more control over the look and feel of the services they provide through their wireless portal.

According to the W3C specification, XHTML Basic defines a document type that is rich enough to be used for content authoring and precise document layout, yet can be shared across different classes of devices—desktop, PDA, TV, and mobile handset. XHTML Basic is the mobile adaptation of XHTML 1.0, and includes everything in XHTML 1.0 except those capabilities, such as frames, that are not appropriate for devices with small screens. XHTML itself, according to the W3C, is the first major change to HTML since HTML 4.0 was released in 1997. In fact, the latest version of HTML (v. 4.1) forms the basis for XHTML: all tag definitions and syntax are the same. XHTML simply adds modularity and enforces strict adherence to language rules. As a result, XHTML brings clean structure to Web pages, which is especially important given the small screens and limited power of mobile devices. W3C is recommending XHTML for all future Web development for desktops as well as all other devices, including mobile handsets.

XHTML is the keystone in W3C's effort to create standards that provide richer Web content on an ever-increasing range of platforms. Using XHTML, content providers will find it easier to produce content for a broad set of platforms and with better assurance of how the content is rendered. In contrast, the initial WAP markup language (WML 1.x) does not ensure consistent layout across different devices. This is of particular concern to carriers that are managing mobile handsets from a number of providers and with different user interfaces from handset model to handset model.

With XHTML Basic, a document can be presented on the maximum number of Web clients, including a variety of mobile handsets featuring different display formats and presentation capabilities. And since XHTML Basic is sanctioned by both the W3C and the WAP Forum, it is an integral part of the overall Internet standards set, which will ensure its widespread acceptance now and continued viability over the long term.

This briefing paper describes the key features of XHTML Basic for mobile browsers and its benefits for content developers, carriers and end users. It will also present examples demonstrating the display opportunities and presentation flexibility enabled by XHTML Basic.

# XHTML Basic: Key Features and Capabilities

Support of XHTML Basic will bring a number of positive changes to the mobile browser. To start, it is a stricter progression of an established markup language since it is based on the same HTML tags used in every Web page today. This means that existing Web design and presentation tools can work with it. Similarly, a developer can view an XHTML Basic screen in a standard desktop Web browser and the layout will be consistent. But there are significant differences too.

Figure 1: The same XHTML page, viewed using the Nokia WAP Browser (left) and Microsoft Internet Explorer (right)

## Well-Formed XML

From a technical standpoint, XHTML Basic strictly enforces the rules for well-formed, valid XML (eXtensible Markup Language). XML is a W3C standard for specifying well-structured documents that can be processed by computers, e.g. a browser. A well-formed XML document conforms to all the rules for XML syntax. For example, every document must have an identifying XML header that is written properly, each tag (instruction) is matched by a closing tag and correctly nested, and attribute values are handled correctly. By following all the XML syntax rules XHTML Basic enables layout that is consistently displayed on different browsers. By comparison, HTML need not be well formed, but the resulting display might contain elements that do not appear as intended on some browsers.

### HTML

```
<p> This is a paragraph.
<P> This is a another.
. . .
```

### XHTML

```
<p> This is a paragraph. </p>
<p> This is a another. </p>
. . .
```
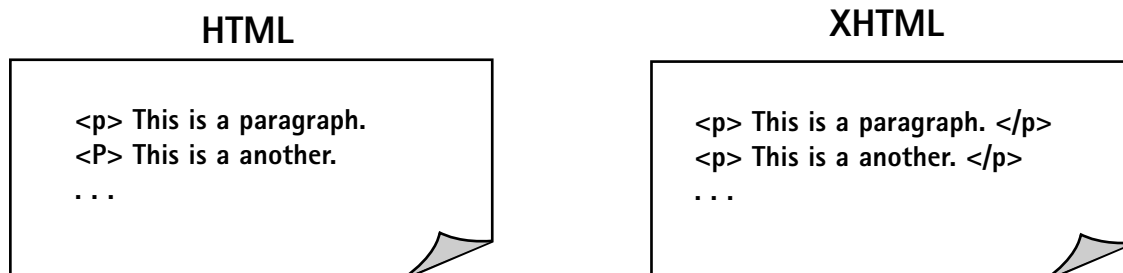
Figure 2: HTML tags are not always closed, e.g. the <p> tags are not closed by a </p> tag, but XHTML tags must be closed.

## Document.xhtml

```
<?xml version="1.0" ?>
<!DOCTYPE wml PUBLIC
"-//WAPFORUM//DTD WML 2.0//EN"
"http://www.wapforum.org/DTD/wml_2.0/xml">
<head>
```

Figure 3: An example DTD specifier (not actual approved DTD)

## Valid XML and the DTD

Valid XML (as noted above, all XHTML documents must contain only valid XML code) simply means that the document correctly uses the language—that it includes only defined tags, and each tag is used in the correct context. For example, in XHTML a table row tag is only valid inside a table tag, but a table row would be senseless outside of a table. The rules for XHTML (or any valid XML language) are defined by its Document Type Definition (DTD). The DTD (or the address at which to access the DTD) appears in each XHTML document's header. The header forms the beginning section of a Web document and contains information that is not part of the actual page content. It also contains controls for the overall document.

According to the W3C, the DTD facilitates hassle-free production and presentation. The DTD identifies what each tag means and specifies how it should be treated. In short, it provides the rules used to validate the online document while it ensures that the document can be displayed on any XHTML browser. This is particularly important in the mobile world where devices offer a wide range of display formats and capabilities.

By strictly following XHTML's rules, the memory size of mobile browsers can be reduced, allowing for a smaller memory footprint and greater performance efficiency, which is essential in today's mobile devices. If you look at the size of standard desktop Web browsers, much of the complexity comes from handling the ambiguities of HTML. As the Web moves towards XHTML, applications designed for mobile devices will lead the way.

## Cascading Style Sheets (CSS)

Central to XHTML Basic is its support for cascading style sheets (CSS). CSS describes how documents are presented on screen in the browser. The W3C has actively promoted the use of CSS on the Web for use with all browsers, desktop and mobile. Through the use of CSS, document creators can control the presentation of documents without sacrificing device independence or adding new markup language tags, as was done with WML 1.x.  The use of well-known standard HTML tags will reduce content development costs by eliminating the need for developers to learn new tags, to store multiple versions of content, or to master different tools. In addition, servers won't experience the added overhead that resulted from the expensive transcoding required to prepare content for an array of different markup languages.

## Document.xhtml

```
<head>
<link rel="stylesheet"
href="Style.css"
type="text/css"/>
<head>
. . .
```

## Style.css

```
li
    {list-style-type:disc;
    margin-bottom: 4px;
    margin-left:2px;}
. . .
```
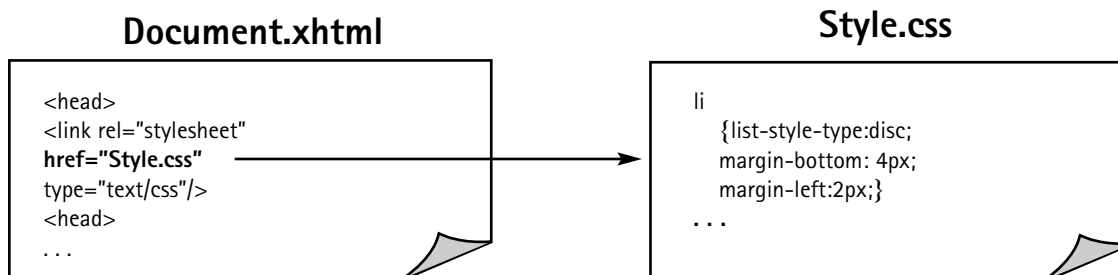
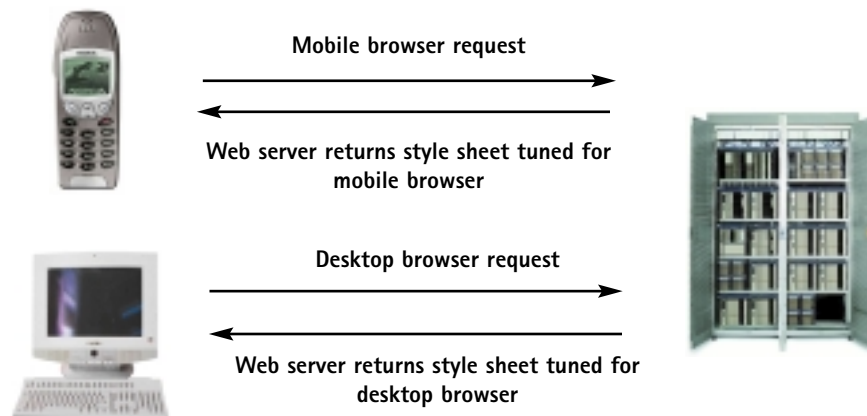Figure 4: An XHTML document referencing a style sheet.

Figure 5: All pages in the web site can be the same for any browser; only the style sheet is tuned for different needs

Through CSS a document creator can specify the presentation of a Web application in one place, the style sheet. If the presentation needs to be changed at any time, the change is made once in the style sheet and the modification is dynamically reflected throughout all the pages in the site.

CSS separates the content of the document or application from the presentation. This allows developers to easily create browser-specific versions of the same content simply by creating the appropriate style sheet. For example, when a user requests a Web page, the content server identifies the requesting device and returns the content with a link to the appropriate style sheet. The style sheet is downloaded once and cached by the browser for use with subsequent pages, which speeds the rendering of all pages in the site.

The power of CSS lies in the precise control it offers document creators and the ease with which they can optimize content for presentation on any device. Every aspect of the document—positioning, fonts, text attributes, borders, margin alignment, flow—can be defined in the style sheet. A change to any aspect of the document needs to be made only once.

CSS also gives carriers greater control over the look and feel of the services they provide through their wireless portal. Through XHTML Basic, the operator can define a default style sheet for all devices it supports, which will ensure a basic look and feel consistent for all devices, even those from different manufacturers. For example, a carrier might specify that all links are underlined or all headings are bold and centered. Of course, content developers can choose to modify that style through their own use of CSS if different behavior is required, but the default style sheet determines the overall look of all applications for a particular browser.

## CSS Browser Examples

The precise control provided by CSS makes it easy to reformat content for different mobile device formats.

In Figure 1 above, you can see the same XHTML Basic pages displayed for different browsers. Through the power of XHTML Basic the presentation, style, and navigation are easily reformatted for each browser. The following figures show other examples of the same page with different formatting applied from different CSS style sheets.

Note how XHTML Basic's support for CSS allows padding and paragraph indentation to achieve precise positioning. Even sophisticated typographic treatments such as drop-caps (see Figure 6 right) can be used. Flexible alignment (center, left, right), borders of varying thickness, and margins can also be used to create advanced effects such as floating paragraphs. (See Figure 7)
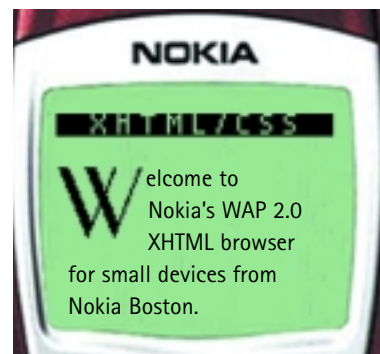


Figure 6: Precise layout control using CSS and XHTML

5

Figure 7: Identical files, but different CSS style sheet changes appearance. Font size, alignment, border, and link appearance are a few properties that are set by CSS.



Figure 8: The different bullet styles are not defined in the XHTML page. CSS can specify many bullet styles, from Roman to Latin numerals, squares and circles.

## Transformation

As an XML-based standard, XHTML Basic allows for the automatic parsing and transcoding of content through the use of XSLT (eXtensible Stylesheet Language: Transformations), part of XSL, the style sheet language for XML. A W3C standard, XSLT provides a language for transforming XML documents into other XML documents.

Using XSLT, a system can automatically transform the same XML content into multiple markup languages depending on the browser. Through such transformations, content can be created for one device and automatically transformed to appear on another device. For example, a news service could create its news feed once in XML and use XSLT to dynamically convert it to XHTML Basic and HTML for presentation on mobile handsets or other devices.

Better still, as XHTML becomes the standard for all Web browsers, a Web site could use XHTML for all its content. In this way, developers would create content once for display on different devices, eliminating the need to write different versions of the same content for different devices.

# Comparison with Compact HTML

XHTML Basic is similar to Compact HTML (cHTML), the language used by DoCoMo's i-mode wireless service in Japan, with a few exceptions. The two biggest limitations of cHTML are that it is not a W3C standard and it does not support cascading style sheets. By comparison, XHTML Basic is a W3C and WAP Forum standard widely accepted throughout the industry.  Furthermore, cHTML need not be well-formed, which means that it can produce inconsistent results from browser to browser, and it lacks the XML header and DTD specifier, which gives XHTML Basic much of its capabilities [see Appendix B: Comparison of cHTML and XHTML].

Mobile browsers that support only cHTML cannot take advantage of CSS. As a result, content providers have to expend considerably more effort to present documents on different devices, and carriers will not have the control CSS gives them.

Figure 9:  cHTML screen on a mobile device vs. layout with XHTML using the Nokia WAP Browser.

# Benefits of XHTML Basic Browser

XHTML Basic is leading the way to the next generation of WAP technology.  It has been widely embraced by the industry, and its modular architecture makes it easy to adapt to any changes that result from the evolution of new technologies. In addition, the XHTML Basic browser delivers a number of specific benefits for all parties involved in mobile computing. These are described in the following table.

| Group | Benefit |
|---|---|
| Device manufacturer | • Ensures device complies with the latest WAP Forum and W3C standard, accepted industry-wide<br>• Results in a more attractive browser because of enhanced layout control<br>• Makes the browser content easier to read due to the use of sophisticated typographic elements<br>• Reduces the memory footprint and increases performance due to the smaller size of XHTML Basic and its strict validation rules |
| Carrier | • Control the look and feel of content delivered through its wireless portal to all supported devices<br>• Enables carrier to define default CSS style sheet for all supported devices, creating consistent look across all manufacturers' browsers<br>• Standards-based browser makes it easier to attract content providers/developers<br>• Automated, dynamic transformation capabilities ensure access to more content |
| Content developer/provider | • Create content once<br>• Use existing tools<br>• Develop for an industry standard<br>• Precise control of content layout and display<br>• Style sheets allow for the easy application of design changes to whole site in one place |
| Mobile device user | • Browser that is more aesthetically pleasing<br>• Easier to use and navigate<br>• Enables access to more content |

# XHTML is the Future

The adoption of XHTML Basic within the mobile industry as the standard browser language brings wireless communications into alignment with the latest advances in Internet content development and presentation. Supported by the WAP Forum and the W3C, XHTML Basic establishes a global standard for all browsers. Even the i-mode service (which uses proprietary markup languages), appears to be moving toward XHTML compliance.

By adopting XHTML Basic as their browser standard, mobile device manufacturers, carriers, developers and end-users alike can get a head start on experiencing the many benefits its offers. Foremost among those benefits is compliance with the global standard for mobile Internet computing going forward.

# Major XHTML terms and abbreviations

| Terms and abbreviations | Description |
|---|---|
| **Attribute** | Modifier of an element (tag) that sets some special value for this particular tag, e.g. <table border="1"> is a table element with a border attribute, setting the border width to 1 pixel. |
| **Compact HTML (cHTML)** | A subset of HTML that is the basis for i-mode HTML, the language used in the i-mode service available from DoCoMo in Japan. Does not include CSS. Does not require strictly valid XML syntax. |
| **Cascading Style Sheets (CSS)** | A W3C specification for setting the display attributes for HTML and XHTML elements. CSS can control font and other text attributes, spacing, borders, margins, flow and floating of elements, and many other precise layout features. |
| **DoCoMo** | The wireless division of the Japanese telecommunications company NTT. |
| **Document Type Definition (DTD)** | A file that specifies the XML rules and definitions for the tags used in the associated document. |
| **Element** | A part of an XML or HTML document enclosed by angle brackets <> that specifies how a portion of a document should be displayed, e.g. <em>display this text with emphasis</em>. Some elements are "empty", and contain no content (text), but may break the text flow, e.g. <br/>, or insert a line, e.g. <hr/>. |
| **Header** | The lines at the beginning of a document that contain elements that are not displayed as part of the content of the document, but control its overall structure, e.g. links to style sheets, document title. |
| **Hyper Text Markup Language (HTML)** | The language for specifying contents and layout of WWW documents. The precursor of XHTML, it does not require strictly valid XML syntax. |
| **i-mode** | The name of the wireless service provided by DoCoMo. |
| **i-mode HTML** | A proprietary form of HTML based on Compact HTML (cHTML) |
| **Tag** | Informal name for element. In effect, it is a programmatic instruction. |
| **Valid** | XML code that conforms with the language as specified in the DTD. |
| **World Wide Web Consortium (W3C)** | An industry consortium that promotes standards for the evolution of the Web and interoperability between products by producing specifications and reference software. The W3C is vendor-neutral and its products are freely available to all. |
| **Wireless Application Protocol (WAP)** | The specification for a set of communication protocols to standardize the way wireless devices access and use the Internet. With WAP 2.0, the WAP protocol has become XHTML Basic. |
| **WAP Forum** | A consortium of organizations dedicated to fostering the advancement of wireless communication. It created WAP and has directed the transition of WAP into XHTML. |
| **Well-Formed** | Code that strictly conforms to the syntax and rules of the language. |
| **Wireless Markup Language (WML)** | A tag language that allows the text portions of Web pages to be presented on mobile handsets, personal digital assistants (PDA), and other mobile devices via wireless communication. |

| | |
|---|---|
| **XHTML** | The latest browser markup language, which replaces HTML as the W3C standard browser language. |
| **XHTML Basic** | A version of XHTML optimized for mobile computing. |
| **XML** | The W3C markup language used to describe content in terms of the meaning of its data.  XML content can be processed as data or displayed. |
| **XSL** (eXtensible style sheet) | The W3C tag language used to create CSS. |
| **XSLT** (eXtensible style sheet: transformation) | A version of XSL used to specify the conversion of a document from one XML-based language to another. |

# Appendix A: Comparison of HTML and XHTML Code

HTML does not require strict adherence to the coding rules of the tag language. This may result in an inconsistent display of the content on some browsers. XHTML's strict enforcement of the coding rules ensures a consistent presentation of the content a wide range of browsers and devices.

**Table 1: Example of HTML vs. XHTML tags**

```
HTML:
        <UL>
                <li><b><I>item</b> one</I>
                <LI>item two</li>
        </ul>
        <BR>
        <A href=document.html>a link</a>
```

```
XHTML:
            <ul>
                    <li><b><i>item</i> one</b></li>
                    <li>item two</li>
            </ul>
                    <br/>
    <a href="document.xhtml">a link</a>
```

The above code shows some examples of the kinds of lack of adherence to coding rules that is acceptable in HTML but not in XHTML.

1. HTML allows upper case in tag names.

2. The first <li> tag is missing its </li> tag.

3. The <b> and <I> tags are improperly nested in the HTML code.

4. The href attribute's value is not quoted in the HTML code. Attribute values must always be quoted in XHTML.

5. The break tag (<br/>) in XHTML must be self-closing, i.e. it has a closing / inside the tag to indicate that there is no corresponding </br> tag.  HTML browsers must manage this ambiguity. XHTML is unambiguous.

# Appendix B: Comparison of Compact HTML and XHTML

The main difference between cHTML and XHTML is cHTML's lack of XML support. There is no XML header and DTD specification in cHTML, and cHTML also will accept non-valid, non-well formed markup language. Similarly, cHTML does not support CSS, which makes all the benefits of CSS inaccessible to the cHTML environment.

cHTML has eight tags lacking in XHTML but these tags can be easily rewritten by the low-level browser as DIV with specific built-in styles. Also note that the <BLINK> tag is proposed as an extension to XHTML Basic for WML 2.0 (XHTML Basic).

| **i-mode/cHTML:** | **XHTML** |
|---|---|
| **Elements in cHTML but not in XHTML Basic** | **Elements in XHTML Basic but not in cHTML** |
| **cHTML 1.0 (all cHTML browsers)**<br>**<CENTER>**<br>    centers a block of text and other elements. In XHTML, use styles.<br><br>**<DIR>**<br>    a list of menus or directories, each preceded by <LI> tag.<br><br>**<HR>**<br>    horizontal rule (line). Has been added to WML 2.0 extensions to XHTML Basic.<br><br>**<MENU>**<br>    a menu list, each preceded by <LI> tag.<br><br>**<PLAINTEXT>**<br>    Displays text exactly as entered (but will be wrapped if wider than screen).<br><br>**Tel: URL**<br>    Specifies a "Phone-To" URL containing a phone number. In WML 2.0 there will be equivalent syntax for dialing a phone call, as well as more advanced telephony features.<br><br>**Picture Symbols (icons)**<br>    i-mode terminals have about 160 pre-defined picture symbols that can be displayed with Shift-JIS code or escaped 6-byte decimal character codes, in the form "&#decimal code;", e.g. "&#63919;". | **CSS (Style Sheets)**<br>**Tags**<br>    <link rel="stylesheet"...> is used to specify an external style sheet.<br><br>**<style>...</style>**<br>    defines a style block within a document head.<br><br>**Attributes**<br>    Style="..." is not officially supported by the XHTML specification [XHTMLREC] but we do support it in the Nokia WAP Browser.<br><br>    Name, id, class attributes are used to identify elements for applying styles.<br><br>**XHTML Basic elements not in cHTML 1.0 (but in cHTML 2.0)**<br>**<li>**<br>    lists<br><br>**<ol>**<br>    lists<br><br>**<select> & <option>**<br>    "pulldown" lists of options.<br><br>**<body>**<br>    in cHTML 2.0 use BODY to designate colors of background, text, links. In XHTML can also specify script to execute on events (e.g. onload). |

**Accesskey attribute**

> In <A> and <INPUT> tags the accesskey attribute assigns a phone key to a link. * and # keys not supported on some terminals.

## CHTML 2.0

> All of cHTML 1.0 [above] plus:

**i<BLINK>**

> blinks a character string. Has been added to WML 2.0.

**<FONT>**

> designates font color. In XHTML, use styles. Does not support other font attributes besides color.

**<MARQUEE>**

> scrolls a character string horizontally. Has been added to WML 2.0.

**istyle="input mode"**

> designates input character types in <INPUT> field.Assumes two-byte Shift-JIS character set.

**<meta>**

> to include non-displayed values, used as hints by browsers or search engines.

## Text Module

> cHTML does not support multiple fonts, so does not support the following tags.

**<abbr>**

**<acronym>**

**<address>**

**<cite>**

**<code>**

**<dfn>**

**<em>**

**<kbd>**

**<q>**

**<samp>**

**<span>**

**<strong>**

**<var>**

## Basic Tables Module

> cHTML does not support tables.

**<caption>**

**<table>**

**<td>**

**<th>**

**<tr>**

## Link Module

**<link>**

> Used for external CSS style sheets

Nokia Mobile Phones
Browser Entity
5 Wayside Road
Mail Stop 3042
Burlington, MA 01803 USA

For additional information:
URL:    www.nokia.com/corporate/wap
Email:  info.wapbrowser@nokia.com

# NOKIA
## CONNECTING PEOPLE