

**A GUIDE TO QUARK HTML TEXT IMPORT
XTENSIONS SOFTWARE**

©1986–1999 by Quark Technology Partnership. All rights reserved.

Quark, QuarkXPress, QuarkXPress Passport, and XTensions are trademarks of Quark, Inc. and all applicable affiliated companies, Reg. U.S. Pat. & Tm. Off. and in many other countries. The Quark logo is a trademark of Quark, Inc. and all applicable affiliated companies.

All other trademarks are the properties of their respective owners.

 **Table of Contents**

Introduction	4
Using Quark HTML Text Import XTensions Software	5
Supported HTML Tags	6

 **Introduction**

Quark HTML Text Import XTensions™ software adds functionality to QuarkXPress™ and QuarkXPress Passport™ that lets you import HTML text into a QuarkXPress document. This makes it easy to use information from HTML files in QuarkXPress documents.

Using Quark HTML Text Import XTensions Software

Quark HTML Text Import XTensions software lets you import HTML text into a QuarkXPress document. It preserves the HTML document's formatting by interpreting HTML tags and associating them with type styles and style sheets in QuarkXPress.

To use Quark HTML Text Import XTensions software, follow these steps:

- 1 Launch QuarkXPress or QuarkXPress Passport.
 - 2 Create a new document or open an existing document, then select a text box.
 - 3 Choose **File** → **Get Text** to display the **Get Text** dialog box.
 - 4 Select the HTML text file you want to import and then click **Open**.
- !!!** *Windows only:* To import an HTML file with a file name that ends in ".html," choose **Display All Files (*.*)** from the **Files of type** pop-up menu in the **Get Text** dialog box.

Supported HTML Tags

HTML tags are mapped to type styles and style sheets according to the information in this section. *Tags that are not listed in this section are not supported.*



Quark HTML Text Import XTensions software does not interpret Cascading Style Sheets.

Basic formatting tags

The following table shows how text tagged with specific HTML tags is styled when it's imported into QuarkXPress or QuarkXPress Passport.

HTML	QuarkXPress
&NBSP;	Non-breaking space
, 	Bold
 	Soft return
<BLOCKQUOTE>	Paragraph with a left and right indent of 36 points
<CENTER>	Centered text
<HR>	Rule above/below
<I>, <ADDRESS>, , <CITE>	Italic
<P>	Hard return
<STRIKE>, <S>	Strikethrough
<SUB>	Subscript
<SUP>	Superscript

Lists

HTML tags that specify text in a list are interpreted as follows.

, , <DIR>, <MENU>

Text tagged with these tags is interpreted as a single paragraph. List items are separated with soft returns.

Text tagged with this tag is indented 24 points and is preceded by a bullet in an unordered list, or a number in an ordered list. Text in the list that does not follow an tag does not include a bullet or number and is indented 36 points. The tag is interpreted as a soft return. If a tag is not used, a soft return is inserted before the next tag.

Here is an example of HTML text for an unordered list and how it displays in QuarkXPress.

```
<UL>
  <LI>First item
  <LI>Second item
  <LI>Third item
</UL>
```

How it displays in QuarkXPress:

- First item
- Second item
- Third item

Here is an example of HTML text for an ordered list and how it displays in QuarkXPress.

```
<OL>
  <LI>First item
  <LI>Second item
  <LI>Third item
</OL>
```

How it displays in QuarkXPress:

1. First item
2. Second item
3. Third item

```
<DL>
```

Text tagged with the `<DL>` tag is not indented. The `</DL>` tag is interpreted as a hard return.

Font size

HTML `` tags with values that do not include a “+” or a “-” are interpreted in relation to the text size for the “Normal” style sheet in QuarkXPress. The `` tag is equal to the “Normal” text size (12 points by default). If the text size for “Normal” is changed, the interpretation of HTML `` tags also changes.

HTML `` tags with values that include a “+” or a “-” are interpreted in relation to the most recently specified font size.

The following table describes HTML `` tags and their corresponding text sizes in QuarkXPress. The point sizes in parentheses indicate the text sizes when the “Normal” text size setting is 12 points.

HTML	QuarkXPress
<code></code>	Normal + 12 points (24 points)
<code></code>	Normal + 6 points (18 points)
<code></code>	Normal + 4 points (16 points)
<code></code>	Normal + 2 points (14 points)
<code></code>	Normal (12 points)
<code></code>	Normal - 2 points (10 points)
<code></code>	Normal - 4 points (8 points)
<code></code>	Current font size + 2 points
<code></code>	Current font size + 4 points
<code></code>	Current font size + 6 points
<code></code>	Current font size - 2 points
<code></code>	Current font size - 4 points
<code></code>	Current font size - 6 points
<code><BIG></code>	Normal + 2 points (14 points)
<code><SMALL></code>	Normal - 2 points (10 points)

Headline styles

When headline tags are imported, they are matched to existing paragraph style sheets in QuarkXPress. If a match is not found, a new style sheet is created. For example, if `<H1>` is detected in the HTML file and an “h1” style sheet is not found, a new paragraph style sheet named “h1” is created. The new “h1” style sheet is set to “Normal” + 6 points and bold (see below).

The following table describes how text in HTML heading tags is styled in QuarkXPress.

HTML	QuarkXPress style sheet
<code><H1></code>	Normal + 6 points and bold
<code><H2></code>	Normal + 4 points and bold
<code><H3></code>	Normal + 2 points and bold
<code><H4></code>	Normal and bold
<code><H5></code>	Normal - 2 points and bold
<code><H6></code>	Normal - 4 points and bold

Font face

When a `` tag is encountered, QuarkXPress compares the font name specified in the tag to the list of installed fonts. If a matching font is available, it is used. If a matching font is not available, QuarkXPress uses the font for the “Normal” style sheet.

Font color

`` tags are honored as follows:

If	Then
Color name exists in QuarkXPress	QuarkXPress version is used
Color name exists in table below	Color is added using RGB value from table
Color name does not exist in QuarkXPress document and is not listed in table	Color is ignored
Color is specified in hexadecimal RGB format (for example, #FF0732)	Color is added and named after hexadecimal RGB values (for example, "#FF0732")

The following table lists the supported color names and their RGB values.

Color Name	RGB Value
Aqua	(#00FFFF)
Black	(#000000)
Blue	(#0000FF)
Fuchsia	(#FF00FF)
Gray	(#808080)
Green	(#008000)
Lime	(#00FF00)
Maroon	(#800000)
Navy	(#000080)
Olive	(#808000)
Purple	(#800080)
Red	(#FF0000)
Silver	(#C0C0C0)
Teal	(#008080)
White	(#FFFFFF)
Yellow	(#FFFF00)

Other tags

Additional tags are interpreted as follows.

<PRE>

White space and line breaks are preserved. Lines that are wider than the text box break according to QuarkXPress rules. The default font is Monaco on Mac OS, and Courier New on Windows.

<DIV>

Only the `ALIGN` attribute of this tag is recognized. For example, if the tag `<DIV ALIGN="CENTER">` is encountered in the HTML file, the text enclosed in the `DIV` tags is centered in the QuarkXPress text box.

Tables

Columns in tables are separated by tabs, and rows are separated by hard returns. The tabs do not attempt to recreate column width settings.

Forms

Form tags are ignored and content inside form tags is not imported.

Images

Image tags are ignored.

Frames

Frame tags are ignored, and content inside frame tags is not imported.

Nesting

To be properly imported, tags must be properly nested. For example, the following HTML is properly nested because the `` tag occurs before the `</P>` tag:

```
<P>This is a paragraph containing <B>bold</B> text.</P>
```

However, the following example is improperly nested, because the closing `` tag comes *after* the closing `</P>` tag:


```
<P>This is a paragraph containing <B>bold text.</P>
<P>The bold text continues into this paragraph.</B> Now some plain
text.</P>
```

In the above example, Quark HTML Text Import considers the first `</P>` tag to close the `` tag as well as the `<P>` tag, so the bold formatting is not applied after the end of the first paragraph. You can solve this problem by simply closing the `` tag before the end of the first paragraph, and then creating a new `` tag at the beginning of the next paragraph, like so:

```
<P>This is a paragraph containing <B>bold text.</B></P>
<P><B>The bold text continues into this paragraph.</B> Now some
plain text.</P>
```

Head content

Content between `<HEAD>` and `</HEAD>` tags is ignored.

 Quark HTML Text Import currently imports only ISO-LATIN 1 characters. It should not be used to import Cyrillic or two-byte text.

Improperly structured HTML

Unexpected formatting may result if an imported HTML file uses improperly structured HTML (for example, if the file uses opening tags without the corresponding closing tags).