

A Guide to Quark HTML Text Import



©2001 Quark Technology Partnership as to the content and arrangement of this material.
All rights reserved.

©1986–2001 Quark Technology Partnership and its licensors as to the technology. All rights reserved.

The information in this document may not be modified, distributed, or disseminated in paper or electronic form without prior written consent of Quark Technology Partnership or its licensee, Quark, Inc.

Quark, QuarkXTensions, QuarkXPress, and QuarkXPress Passport 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.

Introduction	3
Minimum System Requirements	3
Installing Quark HTML Text Import on Mac OS	3
Installing Quark HTML Text Import on Windows	3
Using Quark HTML Text Import	3
Supported HTML Tags	4
Contacting Quark	10

INTRODUCTION

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

MINIMUM SYSTEM REQUIREMENTS

QuarkXPress or QuarkXPress Passport 4.0x

INSTALLING QUARK HTML TEXT IMPORT ON MAC OS

- 1 Quit QuarkXPress.
- 2 Drag-copy the “HTML Text Import” file from the QuarkXPress CD-ROM to the “XTension” folder within your QuarkXPress application folder.
- 3 Launch QuarkXPress to access the features of Quark HTML Text Import.

INSTALLING QUARK HTML TEXT IMPORT ON WINDOWS

- 1 Exit QuarkXPress.
- 2 Drag-copy the “HTML Text Import.xnt” file from the QuarkXPress CD-ROM to the “XTension” folder within your QuarkXPress application folder.
- 3 Launch QuarkXPress to access the features of Quark HTML Text Import.

USING QUARK HTML TEXT IMPORT

Quark HTML Text Import QuarkXTensions 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, follow these steps:

- 1 Launch QuarkXPress.
- 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**.



When you import HTML following these steps, Quark HTML Text Import interprets the HTML tags, and the text imports fully formatted. To import the HTML tags themselves, press ⌘ (Mac OS) or Ctrl (Windows) while you click **Open** in the **Get Text** dialog box.



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 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 left and right indent of 36 points
<CENTER>	Centered text
<HR>	Rule below
<I>, <ADDRESS>, , <CITE>	Italic
<P>	Hard return
<STRIKE>, <S>	Strike through
<SUB>	Subscript
<SUP>	Superscript

LIST TAGS

Text tagged as an unordered list ``, ordered list ``, and definition list `<DL>` are interpreted as a single paragraph and display as follows:



QuarkXPress reads `<DIR>` and `<MENU>` tags. Text tagged with these tags is interpreted as a single paragraph and the list items are separated with soft returns.

``

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:

``

`Item`

`Item`

`Item`

``

How it displays in QuarkXPress:

•Item

•Item

•Item

``

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

``

`First item`

`Second item`

`Third item`

``

How it displays in QuarkXPress:

1. First item

2. Second item

3. Third item

`<DL>`

The `<DL>` tag is interpreted as the beginning of a definition list. The `</DL>` tag is interpreted as a hard return. You can enter the term and its definition within the `<DL>` tags.

Text entered after the `<DT>` tag is the term being defined. Text entered after the `<DD>` tag is the term's definition. Most browsers enter a soft return after the `<DT>` tag and indent the term's definition on the next line. Soft returns also separate each term/definition pair. You can add formatting such as bold `` or strong `` to help differentiate the term from the definition.

Here is an example of HTML text for a definition list and how it displays in QuarkXPress:

```
<B><UL>List of Terms</UL></B>
<DL>
    <DT><STRONG>Term A</STRONG>
    <DD>Definition A
    <DT><STRONG>Term B</STRONG>
    <DD>Definition B
    <DT><STRONG>Term C</STRONG>
    <DD>Definition C
</DL>
```

How it displays in QuarkXPress:

List of Terms

Term A

Definition A

Term B

Definition B

Term C

Definition C

FONT SIZE

HTML `` tags with values that do not include a plus sign (+) or a minus sign (-) 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 the **Normal** style sheet is changed, the interpretation of HTML `` tags also changes.

HTML `` tags with values that include a plus sign (+) or a minus sign (-) 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
	Normal + 12 points (24 points)
	Normal + 6 points (18 points)
	Normal + 4 points (16 points)
	Normal + 2 points (14 points)
	Normal (12 points)
	Normal – 2 points (10 points)
	Normal – 4 points (8 points)
	Current font size + 2 points
	Current font size + 4 points
	Current font size + 6 points
	Current font size – 2 points
	Current font size – 4 points
	Current font size – 6 points
<BIG>	Normal + 2 points (14 points)
<SMALL>	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
<H1>	Normal + 6 points and bold
<H2>	Normal + 4 points and bold
<H3>	Normal + 2 points and bold
<H4>	Normal and bold
<H5>	Normal – 2 points and bold
<H6>	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 (see following table)	Color is added using RGB value (see following 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 hexadecimal RGB value (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 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).

CONTACTING QUARK

For information about how to contact Quark, see the “Contacting Quark.pdf” file on the QuarkXPress CD-ROM.

IN THE AMERICAS

If you are unable to access the “Contacting Quark.pdf” file, contact Quark Technical Support by phone at (303) 894-8899 between 7:00 a.m. and 6:00 p.m. Mountain Time, Monday through Friday, or send an e-mail to Quark Technical Support at MacTech@quark.com or WinTech@quark.com.

OUTSIDE THE AMERICAS

If you are unable to access the “Contacting Quark.pdf” file, contact your local Quark office or Quark Full Service Distributor.

For an up-to-date list of Quark offices and Quark Full Service Distributors, click the “Contact Quark” link on the Quark Web site at www.quark.co.uk.

For e-mail support of QuarkXPress, send a message to techsupport@uk.quark.ch, MacTech@quark.com, or WinTech@quark.com.