

## QuestionPoint's Character Set and Encoding Level

If we all communicated using the same written language there would be no character set issues. However, because QuestionPoint is a cooperative reference tool, at some point you will likely receive a message from a patron who is using a different language, or you will refer a question outside your institution, or a question will be routed to you from any where in the world.

QuestionPoint stores all data in the Unicode Character Set with the Encoding Level of UTF-8.

- **A character set** is a collection of characters that can be entered, displayed and stored by an application. Unicode supports every nation's alphabet, accented character, including the Chinese and Japanese ideographs.
- **UTF-8** is a multi-byte encoding. A character can be encoded into 1 to 4 bytes. The majority of the Western European languages require less than 2 bytes per character. For example, characters from Latin based scripts on the average require only 1.1 bytes, and the majority of Greek, Arabic, Hebrew, and Russian scripts require 1.7 bytes. The Japanese, Korean, and Chinese languages are mostly comprised of 3 bytes per character.

All QuestionPoint forms as well as the system-generated messages are coded as Unicode with UTF-8; these include email forms, basic and enhanced chat forms and the messages your patrons automatically receive. In the source code of each of these is the following HTML code:

```
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```

To ensure all data is correctly converted do not remove or edit this HTML string in any of your email or chat forms.

### How QuestionPoint Works

Incoming data is not automatically converted into Unicode / UTF-8 until the data is "submitted" into the system. Therefore text originating from a patron's email account may display incorrectly, until it has been saved / submitted to QuestionPoint.

Questions originating from an email or chat form or the QuestionPoint Patron Interface are controlled by the QuestionPoint's HTML code. However, the clarification or answer you send is viewed and/or responded to from the patron's email account, which is controlled by the patrons' application's settings. The application the patron uses to view a clarification, system-message or answer could be an application that came with their operating system, such as Microsoft Outlook, it may be an email service offered through their browser; such as Netscape Mail, or MSN HotMail, or it might be an e-mail account they created through another web site such as Yahoo, or Pegasus Mail. A patron may have several email accounts and depending on how they have them configured, they may respond back to you from one account one time and from another the next.

Currently there is no "official" international character set, and that is why there are so many character set and encoding choices. However, Unicode using UTF-8 has been proven to work successfully with most language sets. UTF-8 was chosen as the QuestionPoint standard because it is ASCII compatible, easily supported, compact and efficient for most scripts (languages), and is easily processed, unlike other multi-byte encoding.

As a QuestionPoint subscriber you must use Microsoft Internet Explorer (IE) 5.01 with Service Pack 2 or higher, **or Netscape 4.0 or higher**. Mozilla as well as other browsers are not supported. In order for your patrons to successfully communicate with you, they must use IE or Netscape 4.0 or higher or a MAC Web browser. Your patrons must use I.E, or Netscape 7 if using an International character set; Netscape 4 will not work with languages other than English.

Understanding why one email message might display diacritics / accents successfully when another one does not will help eliminate any frustration or confusion.

The following are scenarios that are caused by incompatible character sets and the suggested action you should take to resolve the problem.

Symptom	Cause and Resolution
<p>Librarian receives an <u>Email form</u> - the diacritics / accent characters are incorrectly displayed, or the message is garbled on a new message.</p>	<p>The source code on your email <u>form</u> has been altered and no longer contains this text string:</p> <pre data-bbox="706 478 1377 506">&lt;meta http-equiv="Content-Type" content="text/html; charset=UTF-8"&gt;</pre> <p>Have your administrator change the HTML in the email form's source code back to the default, as shown above.</p> <p>Exception: For Hotmail users see information below.</p>
<p>Librarian opens a chat session and the <u>Chat form</u> contains incorrect diacritics / accent characters, or the message is garbled.</p>	<p>The source code on your chat <u>form</u> has been altered and no longer contains this text string:</p> <pre data-bbox="706 720 1377 747">&lt;meta http-equiv="Content-Type" content="text/html; charset=UTF-8"&gt;</pre> <p>Have your administrator change the HTML in the chat form's source code back to the default, as shown above.</p> <p>Exception: For Hotmail users see information below.</p>
<p>The Patron responds back to the Librarian using his/her email account. When viewed in QuestionPoint, the diacritics / accent characters are incorrect or the message is garbled.</p> <p>Or</p> <p>The Patron replies back that the email clarification or answer you sent to him/her is garbled or contains incorrect diacritics / accent characters.</p>	<p>The patron's browser is set to an incompatible character-set. Anything you send to a patron will also be displayed incorrectly. If you will be replying back and forth with the patron ask him/her to set his/her browser's encoding to Unicode / UTF-8.</p> <p>In order to read specific text you may need to temporarily change your browser setting, but then change it back to Unicode – UTF8. There is no need to edit the patron's response before saving it, once you save / submit the record in QP, the system will convert it to UTF-8 and the patron's message will be properly converted for future viewing.</p> <p>Exception: For Hotmail users see information below.</p>
<p>The patron's email message looks fine when you view it from your email account but displays incorrectly within QuestionPoint.</p>	<p>Your email account's encoding level is compatible to the patron's. However, when you are in QuestionPoint the message does not look the same. The patron's email message has not yet been converted to UTF-8 so the encoding levels are incompatible. Once the email message has been saved or submitted to the knowledgebase it will be converted and will display correctly in QuestionPoint. Therefore there is no need to edit the message before saving / submitting it.</p> <p>Exception: For Hotmail users see information below.</p>
<p>If your patron uses Hotmail or any other email services that fixes encoding to match the language used, the display of characters in messages from these services will display improperly in QuestionPoint. Additionally, messages from QuestionPoint will display improperly to the patron.</p>	<p>Questionpoint is unable to successfully convert Hotmail messages to UTF-8, because Hotmail permanently fixes the encoding to match the language that is used to compose a message. There is no option in Hotmail to select another encoding type. For these users, you must ask them to change their browser's encoding level to Unicode UTF-8 and then have them resend the e-mail from within QuestionPoint.</p> <p><b>Note:</b> A fix is planned for the October2004 release of QuestionPoint that will automatically convert messages that are specifically from Hotmail accounts so that characters will display properly for the patron and the library.</p>

### Changing the Browser's Character set / Encoding level

Application	Procedure
Internet Explorer (IE)	Click on View / Encoding / Unicode – UTF-8
Netscape	Click on View / Character set encoding / Unicode (UTF-8)
Mozilla	Click on View / Character coding / Unicode (UTF-8)

If you have any questions or concerns regarding the character set and encoding level please contact the OCLC Customer Support Department at [support@oclc.org](mailto:support@oclc.org) or 1-800-848-5800.