



Frequently Asked Questions on character sets and languages in MT and MX free format fields

Version Final
17 January 2008



Preface

The Frequently Asked Questions (FAQs) on character sets and languages that follow apply **only** to free text fields.

Other fields, like codes, identifiers, amounts and dates, have rules that define the limited set of values and/or characters that can be used.





1 MT messages

- ***What character sets can I use in an MT message?***

All fields must use characters from the X, Y or Z character set (all three are provided in Appendix I), which are basically Latin letters, numbers from 0 to 9 and a small number of special characters and other symbols.
It is important, incidentally, to distinguish character sets from languages, as there is no unique relationship between them.
- ***What are the X, Y and Z character sets?***

The X-character set is based on the set of characters that can be transported over telex.
The Y-character set is equivalent to the EDIFACT level A character set.
The Z-character set is a combination of the X- and Y-character set, plus @ and #.
- ***What language can I use in an MT message?***

It is strongly recommended to use English in MTs.
A Closed User Group (CUG) – or another bilateral/multilateral agreement, typically for domestic traffic – can use a different language, but this should be agreed explicitly.
- ***How do I fill in a word or name using a different character set in an MT?***

A number of countries have created a tactical work-around in the form of transliteration tables between local character sets and the X, Y or Z character set. Cyrillic in Russia¹ and Ukraine and the “four corners character encoding” in China² are three examples.
A more general broadcast [ref. 4268] was sent out by SWIFT on 18 October 2000 regarding the transliteration of some missing EBCDIC characters.
Remember, this applies only to free text fields, not to coded or numeric fields.
- ***Is there any chance that MT messages will accept non-Latin characters in the future?***

SWIFT does not expect this to happen.
- ***What if I really need to use the actual foreign characters – assuming the back office systems on both sides can handle them?***

This cannot be done using MTs if the foreign characters are not in the X, Y or Z character set.
- ***What should I do with special characters that are not part of character sets X, Y or Z?***

To specify a character that is not part of the X, Y or Z character set in a SWIFT MT message, SWIFT recommends using the character’s hexadecimal EBCDIC code, preceded by two question marks (??) as escape sequence. Some examples are provided in Appendix III.

¹ Examples of usage can be found in the latest versions of the RUR and RUS market practices on the Russian SWIFT association’s website, www.swift.ru.

² Known to be used by various Chinese users of MT messages. A description of the encoding method can be found [here](#), and the transliteration tables themselves can be found [here](#).



2 MX messages

– **What character sets can I use in an MX message?**

The default character set for MX is Basic Latin (see Appendix II), *except if* a Closed User Group (CUG) – or another bilateral/multilateral agreement - is set up in which other principles are agreed.

If the service description of a CUG does not specify a character set, it means only Basic Latin can be used within that CUG.

If other characters need to be transported (e.g., Cyrillic), then this must be explicitly mentioned within the Service Description of that CUG.

– **But I thought UNICODE / UTF-8 was the ISO 20022 (UNIFI) standard's character set?**

The UNICODE character set, encoded in UTF-8, is indeed the official ISO 20022 character set. Appendix IV provides some technical details about UTF 8 encoding and UNICODE. However, SWIFT added a rule to restrict the set of allowed characters for free text XML elements to Basic Latin (see above).

This will protect customer applications from receiving unwanted characters and character sets.

The SWIFT network will not validate this rule, making it possible to have a bilateral, multilateral or CUG-wide agreement to ignore the rule and to use other characters and character sets in domestic and intra-company communications.

– **What language can I use in an MX message?**

English is strongly recommended to be used in MXs.

A Closed User Group (CUG) – or another bilateral/multilateral agreement, typically for domestic traffic – can use a different language, but this should be agreed explicitly.



MX and MT Character Set and Language FAQ - Final

APPENDIX I - MT messages: the X, Y, Z character sets

The table below lists the characters used in the three character sets that can be used in MTs.

| X | Y | Z | Character | Description |
|---|---|---|-----------|---|
| * | | * | "a" - "z" | 26 small characters of the Latin alphabet |
| * | * | * | "A" - "Z" | 26 capital characters of the Latin alphabet |
| * | * | * | "0" - "9" | 10 numeric characters |
| * | * | * | "/" | Solidus (slash) |
| * | * | * | "_" | Hyphen |
| * | * | * | "?" | Question mark |
| * | * | * | ":" | Colon |
| * | * | * | "{" | Opening parenthesis |
| * | * | * | "}" | Closing parenthesis |
| * | * | * | "." | Full stop |
| * | * | * | "," | Comma |
| * | * | * | "'" | Apostrophe |
| * | * | * | "+" | Plus |
| * | * | * | " " | Space |
| | * | * | "=" | Equal to |
| | * | * | "!" | Exclamation mark |
| | * | * | "\"" | Quotation mark |
| | * | * | "%" | Percentage |
| | * | * | "&" | Ampersand |
| | * | * | "*" | Asterisk |
| | * | * | "<" | Less than |
| | * | * | ">" | Greater than |
| | * | * | ":" | Semi-colon |
| | | * | "@" | At |
| | | * | "#" | Pound (hash) |
| * | | * | "{" | Opening curly bracket |
| * | | * | "}" | Closing curly bracket |
| * | | * | CR | Carriage return |
| * | | * | LF | Line feed |



MX and MT Character Set and Language FAQ - Final

APPENDIX II - MX messages: the Basic Latin character set

The Basic Latin code chart contains a number of characters not included in the X, Y or Z sets. The list below shows both the characters common to MT and MX message character sets and - in *italics* - the most relevant characters specific to the Basic Latin set.

| Basic Latin | Character | Description |
|-------------|-----------|---|
| * | "a" - "z" | 26 small characters of the Latin alphabet |
| * | "A" - "Z" | 26 capital characters of the Latin alphabet |
| * | "0" - "9" | 10 numeric characters |
| * | "/" | Solidus (slash) |
| * | "_" | Hyphen |
| * | "?" | Question mark |
| * | ":" | Colon |
| * | "(" | Opening parenthesis |
| * | ")" | Closing parenthesis |
| * | "." | Full stop |
| * | "," | Comma |
| * | "'" | Apostrophe |
| * | "+" | Plus |
| * | " " | Space |
| * | "=" | Equal to |
| * | "!" | Exclamation mark |
| * | "\"" | Quotation mark |
| * | "%" | Percentage |
| * | "&" | Ampersand |
| * | "*" | Asterisk |
| * | "<" | Less than |
| * | ">" | Greater than |
| * | ":" | Semi-colon |
| * | "@" | At |
| * | "#" | Pound (hash) |
| * | "\$" | <i>Dollar</i> |
| * | "{" | Opening curly bracket |
| * | "}" | Closing curly bracket |
| * | CR | Carriage return |
| * | LF | Line feed |



MX and MT Character Set and Language FAQ - Final

| Basic Latin | Character | Description |
|--------------------|------------------|------------------------------------|
| * | [| <i>Left square bracket</i> |
| * |] | <i>Right square bracket</i> |
| * | \ | <i>Reverse solidus (backslash)</i> |
| * | _ | <i>Low line (underscore)</i> |
| * | ^ | <i>Circumflex</i> |
| * | ` | <i>Grave accent</i> |
| * | | <i>Vertical line</i> |
| * | ~ | <i>Tilde</i> |
| * | | <i>A set of control characters</i> |



APPENDIX III - MT messages: Work-around for characters not included in X, Y or Z character set

To specify a special character that is not part of the X, Y or Z character set in a SWIFT MT message, SWIFT recommends using the character's hexadecimal EBCDIC code, preceded by two question marks (??) as escape sequence. Some examples are provided in the table below.

| Character (some examples) | Work-around (including escape sequence) |
|--------------------------------------|--|
| @ | ??7C |
| _ | ??6D |
| & | ??50 |
| % | ??6C |
| = | ??7E |
| * | ??5C |
| # | ??7B |
| \$ | ??5B |
| ! | ??5A |
| ^ | ??5F |
| ~ | ??A1 |

APPENDIX IV - The UNICODE character set

UNICODE and UCS (ISO10646) are the names of two character standards sets, i.e., tables that link characters to a number. They are in fact supersets of all commonly known character tables, like ASCII and EBCDIC.

As of 1993, UNICODE (version 1.1) and UCS (ISO10646-1:1993) have been kept compatible.

UNICODE 4.0 covers a set of 144 character code charts. Each of these code charts is fully specified and explained in UNICODE. A well-known example is the "Basic Latin" code chart.

UNICODE also offers code charts that cover Cyrillic, Hindi, Japanese, Chinese, Malay, Thai, Arabic, mathematical representations (e.g., letters with arrows on top), artistic languages (e.g., Tolkien), etc.

The UTF-8 encoding scheme is a technical implementation in XML schemas that allows specification of any character used anywhere. The scheme bases itself on UNICODE and Universal Character Set (ISO10646). The UTF-8 encoding scheme allows representation of any UNICODE character (or any character of the Universal Character Set) as a sequence of bytes.