



by
Jeff
Eisen

Level: Intermediate
Works with: Domino 5.0
Updated: 07/01/99

Inside this article:

[An introduction to international MIME settings](#)

[Configuring the international MIME settings](#)

[Recommendations for configuring MIME](#)

Related links:

[Notes from Support: SMTP MTA International Character Sets](#)

[MIME Support in Notes 4.6](#)

[Character set detection flowchart](#)

Get the PDF:
Kb



Introduction

In today's global information age, your business will likely need to work across many languages. And language barriers will need to be a thing of the past. With the new MIME features in R5, you're one step closer to this goal.

The Multipurpose Internet Mail Extensions (MIME) standard opened up Internet messaging to include more than plain text. MIME was designed to enable mail to handle more complex messages with non-ASCII text, multimedia, images, and application-specific formats. The Domino and Notes support for MIME extended this capability for sending complex Internet messages to the Notes client.

This article describes the international settings for MIME in Domino and Notes R5. You'll first learn a little background about the history of Notes and MIME, and then learn how to configure the international MIME settings for both the Domino R5 server and the Notes R5 client. Finally, we'll look at some recommendations for setting up MIME for either a single language or multiple language environment.

A brief history of Notes and MIME

To understand better where we are with MIME in R5, let's take a look at where we've come from.

First came Notes...

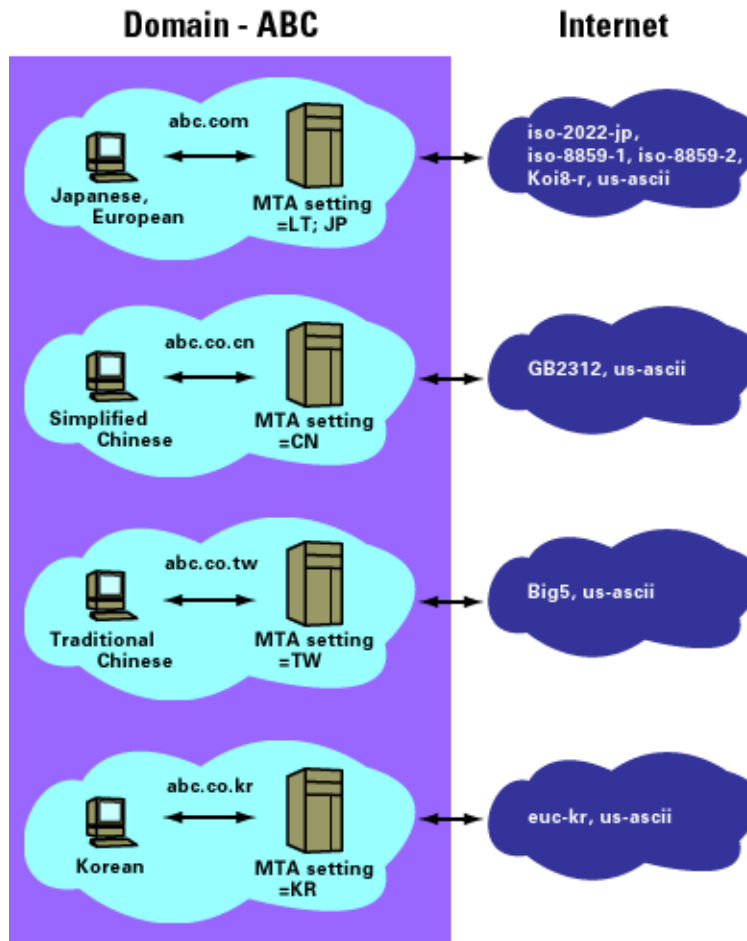
From the very beginning, international character set issues have been relatively straightforward in Notes and Domino. Notes stored textual data using its own character set called the Lotus Multi-Byte Character Set (LMBCS), which represents characters in virtually all of the world's languages.

As long as you stayed within Notes and Domino, there was no problem. The text remained in LMBCS in databases and was converted to and from the character set of your local operating system for display and input, respectively. Text file exporting and importing were treated similarly. Although there were some complex international issues (collation, word wrap, full text indexing, and so on), they were typically confined to Notes -- that is, they were not interoperability problems.

The primary exception came when you needed to send and receive Internet mail via the Simple Mail Transfer Protocol (SMTP). Then, the Lotus SMTP Message Transfer Agent (MTA) was responsible for inbound and outbound mail conversion between Notes and the Internet. The MTA had a very powerful, but complex configuration for international issues. However, the MTA had serious limitations that made many multilingual configurations impossible, particularly for Asian languages. You could not configure a single MTA to process multiple Asian languages, for example, both Korean and Japanese messages. Most organizations used "work-arounds," such as routing outgoing Japanese messages through "japanese.somecompany.com" and Korean messages through "korean.somecompany.com."

The following graphic shows a typical configuration using multiple R4.6 servers. Notice that a separate server is needed for outbound traffic for each of the four Asian "double byte" regions (Japanese, Simplified Chinese, Traditional Chinese, and Korean). For more information on the SMTP MTA and how it handles

international character sets in Domino R4.6, see [Notes from Support: SMTP MTA International Character Sets](#).



With the explosion of the Internet and R5's emphasis on being a good multinational citizen, these work-arounds are no longer enough! Now, in order for Notes/Domino to interpret and generate messages and Web pages, it needs to understand a multitude of different character sets *simultaneously* on a single client or server!

Then came MIME...

As mentioned earlier, the MIME standard was developed in the early '90s to expand the content of Internet messages. MIME allows for the representation and transmission of messages that contain data other than plain text, such as images, HTML, sound clips, spreadsheets, and so on. A MIME message contains parts, where each part is tagged with a content type, such as "text/html" or "image/gif."

Fortunately for MIME user agents (for example, mail reading and sending programs), textual content type parts -- typically "text/html" or "text/plain" -- can also be tagged with a character set name, such as "US-ASCII" or "EUC-KR" (Korean). This character set name indicates how to interpret the bytes in the text. For example, in "ISO-8859-1" (Western), the byte with decimal value 241 corresponds to the lower case "n" with a tilde, as in the Spanish word *Señor*; in "ISO-8859-7" (Greek), the same byte corresponds to the lower case letter rho.

So, continuing this example, if a Spanish message containing the word *Señor*

was incorrectly interpreted as Greek, then the word would appear as Σεπορ.

Unfortunately, many messages and Web pages do not contain the correct character set information. Problematic cases are:

- MIME messages that are tagged incorrectly (such as with "ISO-8859-1" when the content data is really "EUC-KR")
- MIME messages without a character set tag
- Non-MIME messages

For all of these cases, the user agent (in our case, Notes) must guess the character set for the textual data in a message or a Web page. For the first case, Notes/Domino must map the incorrect tag to the correct character set, possibly at the expense of messages correctly tagged with the problematic name ("ISO-8859-1" in the above example). For the second and third cases, Notes/Domino must try to guess the character set based on the bytes in the message.

Similarly, when Notes or Domino generates a message, it must decide which character set to use to encode the data. The easiest solution would be if Notes/Domino could just leave the text in LMBCS when formatting a message for the Internet, that is, when converting to MIME. However, this is not possible because LMBCS is a Lotus-defined, proprietary character set -- it is not an Internet standard.

Alternatively, Notes/Domino could convert all the text to the [Unicode character set](#), an Internet standard character set developed by the Unicode Consortium. Unicode, like LMBCS, is able to represent the characters in virtually all of the world's languages. Unlike LMBCS, it is an Internet standard and is therefore a "legal" character set for Internet messaging. Unfortunately -- here's the big catch -- not all user agents understand Unicode. This means that if you send a message to your friends who use "Joe's written-in-a-weekend mail reader," it is very possible that the message will appear as garbage to them. Of course, if your friends are using Notes, they won't have a problem because Notes and Domino can handle Unicode!

Then came MIME in Notes...

One of the great new features in Domino/Notes R5 is support for "native MIME." There was some MIME support in Notes/Domino before R5, but there was no integrated MIME storage pervading the products. In Notes/Domino R4.6, for example, MIME was supported in messaging by storing MIME attachments in mail and news documents, rather than by storing MIME data at the "item" level. (For an in-depth discussion of MIME in R4.6, see [MIME Support in Notes 4.6](#).) With the R5 support for native MIME, in most places, conversions between Notes and MIME are no longer necessary.

However, in the places where it is necessary to convert between the traditional Notes Composite Document (CD) format and MIME, the international MIME configuration information is available. For example, the conversions are still necessary when a Notes R4.6 client (which doesn't understand native MIME) needs to open a document that contains native MIME on an R5 server. In addition, there are other places (such as, Web browsing and reading MIME mail messages) where Notes doesn't truly convert from MIME to Notes CD format, but where it still needs to figure out MIME international character set and font information.

These conversions can occur on either the Notes client or the Domino server. Furthermore, the conversions can occur on the client during front-end processing (such as, the editor displaying HTML) or back-end processing (such as, LotusScript code opening a note using the NotesDocument class). Obviously,

the server only handles back-end processing.

Back-end conversions take place when converting MIME to Notes CD format. This may occur when a Domino R5 server replicates a native MIME note to a pre-R5 server. Or, it may occur when Domino delivers mail to a user (possibly using R5) whose Person document in the Domino Directory specifies to use the Notes Rich Text format for incoming mail.

An introduction to the international MIME settings

First, you should note that for many organizations and individual users, no configuration of the international MIME settings is necessary! The default values for the configuration settings should work in most circumstances (although, as you will see, the default values may work better for Asian, rather than European, locales).

For configuration purposes, Notes divides the world's languages and character sets into 16 groups:

- English
- Western (Danish, Dutch, Finnish, French, German, Italian, Portuguese, Norwegian, Spanish, Swedish)
- Central Europe (Czech, Hungarian, Polish, Romanian, Slovenian, Slovakian)
- Turkish
- Baltic Rim (Estonian, Latvian, Lithuanian)
- Greek
- Cyrillic (Bulgarian, Belorussian, Russian, Ukrainian)
- Arabic
- Hebrew
- Japanese
- Korean
- Simplified Chinese (Chinese used in People's Republic of China)
- Traditional Chinese (Chinese used in Taiwan)
- Thai
- Vietnamese
- Unicode (Universal character set)

Some of these groups (such as, Thai and Japanese) correspond to a single language, and others (such as, Western and Central Europe) correspond to a region where there are multiple languages that use mostly the same characters.

Note that the configuration described below is not currently used for all MIME handling within Notes and Domino. The biggest exception that comes to mind is the Domino HTTP server, which generates Web pages based on character set information contained in the database.

Primary and secondary character sets

The international MIME configuration settings first define a single primary character set group. Then, you can define multiple secondary character set groups. These primary and secondary choices control, among other things, how Notes and Domino perform autodetection of character sets.

For **inbound messages**, character set autodetection is required when the incoming MIME or non-MIME message does not contain character set information. Domino is able to distinguish with very good (but not perfect) accuracy among the various character sets used by "CJKT" languages, that is Simplified Chinese (used in the People's Republic of China), Japanese, Korean, and Traditional Chinese (used in Taiwan).

In order to perform this autodetection with the best accuracy possible, Domino needs to know what priority order to assign to the CJKT regions. For example, if a message appears to be either EUC-KR (a Korean character set) or GB2312 (a Simplified Chinese character set), Domino uses the priority order from the primary and secondary character set groups to determine which character set to

use. It chooses the primary group first, then the secondary (in an undefined order -- the order of your multiple secondary choices does not matter), and then the operating system group (for operating systems, such as Windows NT or Windows 98, where the locale can be queried).

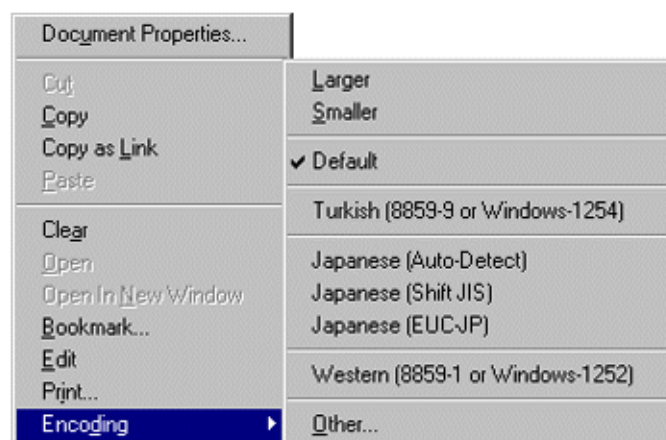
Note that for most European character sets, no inbound autodetection is possible without a complex linguistic analysis (such as, recognizing the words in different languages in the text). This is because the various European character sets tend to use the same bytes and byte ranges, while many of the CJKT character sets tend to use only certain byte ranges or use special uniquely identifiable control codes. These restricted byte ranges and control codes enable Domino to match patterns in the incoming message to a probable character set.



Important Note: Because of the differences between CJKT and European (and other) character sets, setting the primary or secondary character set groups to anything other than Simplified Chinese, Japanese, Korean, or Traditional Chinese does not help inbound autodetection. Setting these groups to non-CJKT values, though, can be useful for outbound autodetection, as explained below. You will also learn that you can use character set groups in the Notes client to affect the character set choices that are displayed for when you need to explicitly specify a MIME character set.

For **outbound messages**, Domino chooses a MIME character set based on the text of the message. For messages in some languages (such as Thai), it is usually pretty obvious which character set to choose. For other messages (such as those in some European languages), there is significantly more overlap in the character sets, and it is sometimes difficult to guess which MIME character set to use. Domino uses the primary, secondary, and operating system groups, in that order, to break ties in determining which character set to use. That means, for example, if a message contains all characters that could either be French or Turkish, then Domino uses the primary and secondary groups to determine which character set to use.

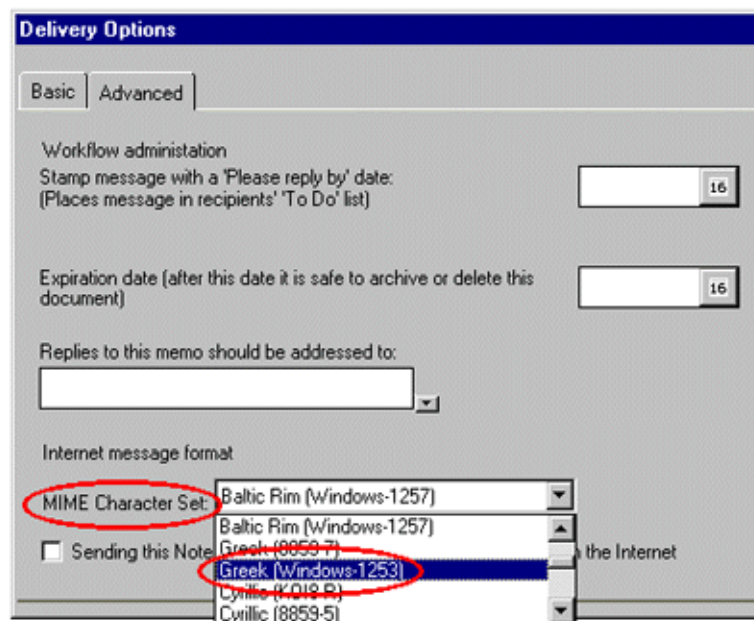
In addition, the client uses the primary and secondary character group settings to determine which character sets appear in the override Encoding dialogs. For example, if your primary character set group is Turkish, your secondary group is Japanese, and you are running on a Western or US configuration of Windows 98 (admittedly an unusual configuration), Notes displays the following dialog box when you right-click on an incoming MIME message:



Note the inclusion of Turkish, Japanese, and Western character sets in addition to the "Other..." choice, which will give you all choices when clicked.

This "override" dialog box allows you to override the default character set that Notes/Domino uses on incoming messages and on Web pages! This is one of the most important advantages of the native MIME storage, new in R5 -- even if Notes/Domino guesses wrong on an incoming message, you can specify the correct character set by explicitly overriding the character set choice.

Similarly, in the Notes client, you can explicitly specify the character set that Notes uses for outgoing mail, overriding the default Notes/Domino heuristics. You can specify this on a per-message basis on the Advanced tab of the Delivery Options dialog box, as shown below.



Per-character set group options

In addition to selecting the primary and secondary groups, you can configure options for each character set group. These options include both inbound and outbound message options. The following screen shows the configuration for the "Central Europe" group. In general, these per-group configurations should not be necessary unless you have special font preferences.

MIME settings by character set group: Central Europe

Settings for Central Europe

Inbound Message Options

Font Options

HTML Proportional: Default Serif

HTML Mono-spaced: Default Monospace

HTML Size: 12

Plain Text: Default Monospace

Plain Text Size: 10

Outbound Message Options

	Character Set	Encoding
Header:	ISO-8859-2	Quoted Printable
Body:	ISO-8859-2	Quoted Printable

Inbound options

The inbound options provide a way to control the default fonts that Notes uses, as well as their default sizes, when opening a MIME message (or a Web page, news article, and so on). You can configure these fonts on a per-character set group basis because you may require different fonts for different character sets.

For example, a font that can display Japanese needs many more characters or *glyphs* than a font that can display French (in the Western group). You can use the same fonts for each group (if you have a Unicode font that can display all of the character sets), but you may not want to. Instead, you may want to use a Unicode font for most of the groups, but a different non-Unicode font for Hebrew, for example, if you have a high-quality Hebrew font that you prefer.

Notice that you can configure "HTML" font information separately from "Plain Text" font information. Here are the default HTML font settings:

- The default HTML Proportional font is the Default Sans Serif font (FONT_FACE_SWISS for API programmers) for the Simplified Chinese, Japanese, Korean, and Traditional Chinese groups, and is Default Serif (FONT_FACE_ROMAN) for the others.
- The default HTML Mono-space font uses the Default Monospace font (FONT_FACE_TYPEWRITER). This font is used when the HTML contains a <TT>, <PRE>, <CODE>, or any of the other related tags.
- The default HTML font size is 12-point, which corresponds to the size "3" in an HTML tag. For HTML, these fonts are chosen unless there is an explicit FACE specified in a tag.

For Plain Text, the default configuration uses a 10-point Default Monospace font. (Note that all of these defaults may be different if you are using a localized version of Notes/Domino.)



Important Note: The Notes R5 client uses these font options for front-end processing -- such as, Web browsing and reading native MIME notes. Any back-end processing (typically performed on the server) currently uses the font size parameters, but ignores the HTML font face parameters. Also,

the back-end processing only obeys the Plain Text font face parameter if it is set to one of the following: Default Sans Serif, Default Serif, or Default Monospace.

Outbound options

The outbound options provide a way to control the character set that Notes/Domino uses when creating a MIME message. When creating a MIME message, Notes/Domino must convert the message from LMBCS to an Internet standard character set. Remember that in general, the character set should not be Unicode or UTF-8 (an encoding of Unicode), because many mail readers do not understand Unicode. An exception is made for multilingual messages where a single character set (other than Unicode) cannot represent all the characters in the message.

Internally, the procedure for choosing the character set is a two-step process. First, Domino/Notes analyzes the LMBCS data to determine the appropriate character set group to use. Then, it consults the outbound options to determine the specific character set within the chosen group to use. As shown in the previous screen, you can configure a header character set and a (potentially different) body character set. In general, the default value for the header and the body are the same for a given group. The exception is Korean, where the default header character set is EUC-KR and the default body character set is ISO-2022-KR. You can also choose the character set encoding for each character set choice. The possible encodings are None, Base64, and Quoted Printable. These encodings allow the transmission of 8-bit data over what is frequently a 7-bit medium.

The default values follow customary usage among the majority of mail programs and Web browsers. Using different values, in general, is a bad idea, unless you know exactly what the receiving mail reader(s) will understand or expect.

For each group, the choices of a character set are those character sets appropriate for that group. Directly above the configuration data for the screen, there is a checkbox labelled: For outbound message options below use all possible choices (Advanced Users). Selecting this checkbox (not recommended) allows you to configure nonstandard or even nonsensical character set options. For example, it allows you to send outbound Vietnamese messages using a Greek character set. If you do this, you should expect many characters in the message to use a "fallback" character (typically a question mark) because they cannot be represented in the chosen character set.

Advanced options

There are four advanced options related to international MIME settings. The procedure for configuring these options is slightly different on the client and on the server. The options are:

- Inbound Character set name aliases
- Outbound Character set name aliases
- Unknown (unable to Autodetect) Inbound 8-bit character set choice
- Multilingual outbound option

The alias options (both inbound and outbound) allow you to configure nonstandard mappings between character set names and character sets. For inbound aliases, this configuration may be necessary if you frequently receive messages (or browse Web pages) that incorrectly specify a character set. You can specify a mapping from an invalid name (such as, "US-ASCII" -- notice the incorrect spelling) to a valid character set (US-ASCII -- two I's). Or, you can specify a mapping from a valid name to a different character set. For example, it is common for many mailers to send mail improperly labelled as "ISO-8859-1." If you frequently receive such mail, you can create an alias mapping the name "ISO-8859-1" (Western) character set to the character set GB2312 (Simplified Chinese), for example, as shown in the screen below.

Note that there is one "built-in" alias that you get for free (though it doesn't automatically appear in the configuration form). This alias maps the name "ISO-8859-1" to the Windows-1252 character set. This mapping is necessary because it is very common for mailers to send mail in the Windows-1252 character set and label it as "ISO-8859-1." This built-in alias should not cause a problem because Windows-1252 is a superset of ISO-8859-1 -- that is, Windows-1252 is the same character set as ISO-8859-1, with the addition of some extra characters. You can disable this behavior by aliasing "ISO-8859-1" to ISO-8859-1, that is, aliasing it to itself.

For outbound aliases, you can create aliases that allow you to send messages labelled with an improper character set name. Using these aliases is not recommended because you will probably be sending messages that many mail readers will not understand!

Also, notice in the screen above the Unknown Inbound character set choice, labelled "For non-MIME messages or MIME messages with an unknown character set, 8-bit character set is assumed to be:". This choice allows you to control the inbound character set that Notes/Domino uses when it cannot autodetect the character set. The default value is Windows-1252.



Important Note: You must manually set the unknown inbound character set. That is, it is not automatically adjusted based on your choice of primary or secondary character sets, or based on your operation system locale. Due to CJKT autodetection of Asian character sets, if you do not change this value from the Windows-1252, things will probably work correctly for the Simplified Chinese, Japanese, Korean, and Traditional Chinese groups, but not for the other (such as the European) groups! Because of this problem, this setting may end up being the only one you will need to make, that is, where the defaults are inappropriate. This will be addressed in a future release.

Finally, the multilingual option specifies how to send outbound mail when Notes/Domino detects a multilingual message -- that is, a message containing characters that cannot be represented in a single character set other than a universal character set, such as Unicode.

You can send multilingual messages in Unicode (actually UTF-8, which is an 8-bit encoding of Unicode) or in the most representable character set, that is, the best match for the majority of characters in the message. When using the best match option, if the message is sent in plain text, the unrepresentable characters are essentially lost. (They are sent as a fallback character, which is typically a question mark.) If the message is sent in HTML and the best match option is chosen, the unrepresentable characters are sent as Unicode entities, that is,

their Unicode numeric values. A Unicode-enabled mail reader can decode such a message. The default value for this option is Unicode.

On the client, for both of these choices, you can also choose to be prompted for multilingual messages. Then, each time you send a multilingual message, Notes prompts you for how to send the message. (Choosing the "Use Unicode and Prompt" option, for example, causes Notes to prompt you with a dialog box where the default value is Unicode.)

Configuring the international MIME settings

You can configure the international settings for MIME processing in both the Personal Address Book and the Domino Directory (formerly called the Public Address Book). They are available in both places because, as mentioned above, both the client and server make MIME character set (and other international) decisions.

The configuration process is very similar for both types of address books. That is, most of the screens and choices are the same in both the Personal Address Book and the Domino Directory, although they are on different forms. The Domino Directory configuration has a few additional settings, primarily in the "Advanced" section, because the Domino Directory configuration screens include all MIME settings, not only international MIME settings.

Configuring the international MIME settings for Notes R5

To configure the international MIME settings for Notes R5:

1. In the Personal Address Book, choose Create - Advanced - International MIME Settings.
2. Specify the location for these settings to apply to, or keep the wildcard character "*" for all locations. (This means that if you change locations within a Notes session, the settings for the new location will take effect immediately.)
3. On the Basics tab, select the primary and secondary character set groups.
4. On the Settings by Character Set tab, specify the inbound and outbound options. (The default values should be OK, unless you have special font preferences.)
5. On the Advanced tab, enter any advanced options such as character set name aliases or the character set to use for unknown inbound character sets. (For many languages, you should manually set the unknown inbound character set.)
6. Save and Close the document
7. Create additional "International MIME Settings" documents for other locations, if desired.
8. To set multilingual message options, choose File - Preferences - User Preferences, click the Mail and News icon, and select an option in the Multilingual Internet Mail field. (The default is "Use Unicode and Prompt.")

Configuring the international MIME settings for Domino R5

In the Domino Directory, each Configuration Settings document applies to a specific server, or to a named group of servers, or to all servers (specified using a "*" wildcard). This configuration model, in general, lets you configure parameters at the granularity that makes sense for your organization. For international MIME settings, this hierarchical model makes sense (to avoid having to set the same parameter values for each server individually). However, unlike some other settings in the Configuration Settings document, the various parameters in the international MIME settings are interrelated in a way that makes mixing-and-matching from different documents (in the single server - to group of servers - to all servers hierarchy) infeasible.

For that reason, the international MIME settings for a server are taken from a single Configuration Settings document. Domino uses the most specific document for a server that has the "International MIME Settings for this document" field enabled. This allows you, for example, to set the same settings across your organization (by enabling this field in the "*" wildcard configuration document) and to only override the setting for a handful of servers (by enabling this field for those servers and not for other servers).

To configure the international MIME settings for Domino R5:

1. On the Configurations tab of the Domino Administrator, select the Server/Configurations view, select a Configuration Settings document, and click Edit Configuration.
2. On the Basics tab, select Enabled in the "International MIME settings for this document" field.
3. On the MIME/Basics tab, select the primary and secondary character set groups.
4. On the MIME/Settings by Character Set Groups tab, specify the inbound and outbound options. (The default values should be OK, unless you have special font preferences.)
5. On the MIME/Advanced tab, specify the inbound and outbound options. (For many languages, you should manually set the unknown inbound character set. The additional options on this tab are not related to international issues.)



Important Note: There are currently two restrictions in configuring international MIME settings in the Domino Directory that are still present as of the writing of this article (R5.0.1 timeframe). First, you can only configure international MIME settings at the "individual server" or at the "all servers" granularity. That is, configuring by groups (which works for other settings in the Configuration Settings document as well as for other non-international settings under the MIME tab) does not work for international MIME settings -- the settings will be ignored. Second, unlike in the Personal Address Book, the international MIME settings in the Domino Directory do not get reloaded during a server session. They are cached in memory and you must restart the server in order for changes to take effect. Both of these limitations should be fixed in a future release of the Domino server.

Recommendations for configuring MIME

In the previous sections, we addressed the basics of how to perform international MIME configuration. In this section, we will look at some common configurations and how to best set your configuration options for either a single language or multiple language environment.

Configuring MIME settings for a single language environment

If you only need to configure a server or client for one language group, use the recommended settings in the following table:

Language/ Region	Basics tab - Primary Character Set Group option	Advanced tab - Advanced Inbound Message Options - Unknown 8-bit Character Set	Advanced tab - Advanced Inbound Message Options - Character Set Name Aliases
Western	English (Default)	Windows-1252 (Default)	
Japanese	Japanese	Shift_JIS	
Simplified Chinese	Simplified Chinese	GB2312	ISO-8859-1 to GB2312
Traditional Chinese	Traditional Chinese	Big5	ISO-8859-1 to Big5
Korean	Korean	EUC-KR	ISO-8859-1 to ISO-2022-KR US-ASCII to ISO-2022-KR
Thai	Thai	Windows-874	ISO-8859-1 to Windows-874
Vietnamese	Vietnamese	Windows-1258	
Central European	Central Europe	ISO-8859-2	
Cyrillic	Cyrillic	KOI8-R	
Greek	Greek	Windows-1253	
Turkish	Turkish	Windows-1254	
Baltic Rim	Baltic Rim	Windows-1257	ISO-8859-1 to Windows-1257

You do not need to choose any secondary character set groups in your configuration because you are only configuring for one group. Since some groups contain multiple languages, this single group configuration can allow multiple related languages, such as French, Spanish, and German (all Western).

Note that the inbound aliases listed in the table above are based on erroneous character set information commonly sent by various mail packages. Of course, if you alias "ISO-8859-1" to a different character set and you receive a message that is truly in ISO-8859-1 (Western), the message will be interpreted incorrectly. You have to decide when configuring the settings whether more messages (or possibly more "important" messages) that are labelled "ISO-8859-1" are truly in ISO-8859-1, or whether more are in a different character set, such as Big5.

The "US-ASCII" to ISO-2022-KR mapping for the Korean group in the table above should not cause any problems even if incoming messages are truly in US-ASCII because US-ASCII is a subset of ISO-2022-KR. This is true for all other character sets available in the configuration, except for UTF-7 (an infrequently used Unicode encoding). That is, US-ASCII is a subset of all of the character sets available, excluding UTF-7.

Configuring MIME settings for a multiple language environment

Configuring for multiple language environments is a bit more complex. For the sake of this discussion, we classify the character set groups into the following three categories:

- English (English)
- Asian (Japanese, Korean, Simplified Chinese, Traditional Chinese)
- European (Western, Central Europe, Turkish, Baltic Rim, Greek, Cyrillic, Arabic, Hebrew, Thai, Vietnamese)

We chose the category names rather loosely, based on the majority of the languages in the category. So, for example, the European category contains Vietnamese, even though Vietnam is not in Europe. The Asian groups are what we referred to as the "CJKT" languages earlier. Note that the "Unicode" group is omitted here because it is not truly a language or world region, and because Unicode is treated specially by Notes and Domino for autodetection purposes.

To configure multiple language environments, use the following six "rules of the road." Note that these rules assume a sensible configuration. It is certainly possible (by using poor choices of Advanced options) to create a configuration that sends invalid messages and mishandles incoming messages.

RULE 1: All configurations support English.

The English language uses the US-ASCII character set, which includes the letters A-Z (upper and lowercase), the numbers 0-9, as well as punctuation characters. As mentioned previously, this character set is a subset of all other MIME character sets that Domino supports (with the exception of the infrequently used UTF-7). Because of this fact, even if a US-ASCII inbound message is misidentified as a different character set, such as ISO-2022-KR, Domino/Notes still handles the message correctly.

For outbound messages, Domino can always identify a US-ASCII message when converting to MIME -- there are no ambiguities. So, for example, if a message is ambiguously English (US-ASCII) or Greek (Windows-1253) in a server configured for Greek, Domino sends the message as US-ASCII because there is no loss of information, and because virtually all mail user agents can handle the US-ASCII message correctly.

There is one slight caveat to this rule, related to punctuation. US-ASCII contains all the letters and numbers used by a typical English language message or Web page, but it does not contain all the punctuation used by some messages or Web pages. For example, directional quotation marks are commonly used in English language Web pages, as in the following sentence:

She exclaimed, "I am!"

However, these quotation marks are not in the US-ASCII character set (they are in Windows-1252, though). So, if the Web page is incorrectly identified as US-ASCII rather than Windows-1252, the letters appear correctly, but the quotation marks do not.

RULE 2: You do not need to list "English" as a primary or secondary character set group.

Since all configurations support English, as stated in Rule 1, you do not need to explicitly list English as your primary or secondary character set group. However, if you do explicitly list English, it appears as one of the first tier (rather than "Other") options on the Encoding override dialog boxes in the Notes client. This listing does not affect autodetection, though.

RULE 3: You do not need a localized version of Notes/Domino or a localized version of an operating system to support a multiple language configuration.

You can configure all MIME settings for any version of Notes/Domino -- whether it's the US, International English, or a localized version such as Japanese. So, for example, you can configure a Japanese localized version of Domino to handle Greek as the primary character set group.

You should note the following three points for this rule:

1. Localized versions of Notes/Domino may contain different default values for font faces, font sizes, character sets, and so on. You can specify different

values -- these are only defaults.

2. As explained previously, the operating system locale provides some default values for character set group priorities. This means, for example, that while you do not need to explicitly list Japanese as a primary or secondary character set group to obtain Japanese-style autodetection on a Japanese locale server, you would need to list Japanese on a Western locale server.
3. On the Notes client, if you have proper fonts installed (typically Unicode fonts), you can display and create messages in any of the supported MIME languages. There are some exceptions, such as, you need to have an operating system that supports Arabic in order to perform some of the right-to-left character handling required by Arabic.

RULE 4: For outbound conversion (Notes-to-MIME), Notes/Domino supports all languages even in the absence of configuration information -- the configuration information is primarily used to resolve ambiguities.

When generating MIME, Notes/Domino must decide which character set to use to encode the text of a message. In most cases, this can be done on a single server for messages in multiple languages without any special configuration.

However, for some cases, you need to configure which language/region has the higher priority. For example, consider a message that, from strictly scanning the characters used, could be either French (Western) or Turkish. For a French company, clearly, Domino should send the ambiguous message using a Western character set. Similarly, for a Turkish company, it should use a Turkish character set.

For an organization that is neither French nor Turkish, the administrators have to decide which character set is more likely to be correct. For most, but not all, this will be Western (because Western services a much larger region than Turkish). Each administrator has to take into account, in this case, where the majority of their Internet traffic goes to and comes from.

Note that if the server is running on an operating system set for a French or Turkish locale, or if the client's ("general," not just MIME) international settings are set to French or Turkish, then no international MIME configuration is needed to resolve this ambiguity. Otherwise, you must select French or Turkish as a primary or secondary character set group.

RULE 5: For inbound conversion (MIME-to-Notes), Notes/Domino handles all MIME messages that contain the correct character set tagging information.

When we talk about handling multiple inbound languages, the "hard part" is guessing what to do when character set information is either missing or is incorrect. It is straightforward for Notes/Domino to interpret the text of a document assuming that a character set name is present and that it is a character set that Notes/Domino supports. This includes MIME messages as well as Web pages. In fact, for this common case, no special configuration is necessary. So, in some sense, in a "perfect world" where all messages are correctly tagged MIME, handling inbound messages is an easy problem.

Special configurations become necessary because in the "real world," there are many messages and Web pages that either have missing or incorrect information. You will see in the next rule what type of configurations are possible in R5.

It is interesting to note that in some ways, the extent of this guessing problem has been diminishing over the past few years, and in some ways it has gotten worse. The good news is that most of the leading e-mail programs and web browsers have improved their international support dramatically over the past few years -- they support Unicode for multilingual messages and they send MIME messages with the correct information. The bad news is that there is an

increasing number of free e-mail programs, such as the various Web-based, free e-mail accounts. These programs typically send messages either missing character set information or with incorrect information, such as tagging a message as "ISO-8859-1" (Western) when it is really ISO-2022-KR (Korean).

RULE 6: For inbound conversion (MIME-to-Notes), Notes/Domino handles (with varying degrees of accuracy) one European or Asian group, plus a high percentage of Japanese and Korean messages.

Finally, we discuss what is possible for inbound global MIME handling and how to accomplish it. For a graphical look at how Notes/Domino chooses what character set to use for incoming conversions, see the sidebar "[Character set detection flowchart](#)."

Domino's capabilities are based on the following three characteristics of the various character sets:

1. Remember Rule 1 -- that English essentially comes for free.
2. Notes/Domino R5 *cannot* distinguish the European character sets from each other. This is because the European character sets (which, as noted above, include some languages that are not truly European, such as Thai) are 8-bit character sets that use bytes in similar byte ranges.
3. Notes/Domino R5 *can* distinguish the various Asian character sets to some degree. This is because the Asian character sets are either 7-bit character sets that use different control codes distinguishing the character set, or are 8-bit character sets that use bytes only in a limited range of values.

The following table lists the Asian character sets that Notes/Domino handles as well as the characteristics of each character set:

Language/Region	Character Sets for Region	Characteristics of Character Set
Japanese	EUC-JP	8-bit character set using only certain byte ranges
	ISO-2022-JP (also known as JIS)	7-bit character set using recognizable distinct control codes
	Shift_JIS	8-bit character set using only certain byte ranges (mostly different ranges than EUC-JP)
Korean	EUC-KR	8-bit character set using only certain byte ranges
	ISO-2022-KR	7-bit character set using recognizable distinct control codes
Simplified Chinese	GB2312	8-bit character set using only certain byte ranges
Traditional Chinese	Big5	8-bit character set using only certain byte ranges
	EUC-TW	8-bit character set using only certain byte ranges

Notice in the above table that ISO-2022-JP and ISO-2022-KR are both 7-bit character sets that can each be recognized by the presence of distinct control codes. Fortunately, these two character sets are used for a significant percentage of Japanese and Korean messages, respectively.

It is therefore possible for us to recognize one character set group (European or Asian) in addition to ISO-2022-JP and ISO-2022-KR.

Overall tips for multiple configurations

Based on the preceding six rules, you should configure for multiple groups using the following guidelines:

1. Choose a primary group based on what group corresponds to the majority of your inbound and outbound internet traffic. This group can be either Asian or a European group. There is never a need to choose English -- most likely, Western is the best choice if your messages are in English in general.
2. Include any additional European groups that you require as your secondary groups (to aid in outbound detection). Note that the order of your secondary groups does not matter.
3. If your primary group is Asian and you need to handle other inbound Asian groups, add them to your secondary groups.
4. Set inbound character set name aliases based on your primary character set group as shown in the table in the "Configuring MIME settings for a single language environment" section above.
5. Set your unknown inbound 8-bit character set to the most prevalent character set for the primary group (again see the table in the single language environment section).



Important Note: Any Asian character set groups explicitly listed in the primary or secondary groups, or implicitly used by the "locale group" will trigger inbound autodetection of the Asian character sets. This can cause unintentional interference with the character set chosen as the "unknown inbound 8-bit character set" because the unknown character set is only used if the Asian autodetection fails (see the sidebar "[Character set detection flowchart](#)"). Because of this interference, you should only choose an Asian character set in the secondary group if your primary group is also Asian.

Note that automatic inbound detection of ISO-2022-JP and ISO-2022-KR, two of the most commonly used Japanese and Korean character sets, respectively, will be coming in R5.0.2. With this automatic detection, you will no longer need to specify Japanese or Korean as a primary or secondary group for detection of ISO-2022-JP and ISO-2022-KR. Thus, there will no longer be an interference problem for these two commonly used Asian character sets. You will still want to list Japanese or Korean if this is your primary language because that will enable EUC-JP versus ISO-2022-JP, versus Shift_JIS and EUC-KR, versus ISO-2022-KR autodetection.

The future of MIME in Notes/Domino

R5 is a major improvement over previous releases of Notes/Domino in terms of its MIME integration, and international and multilingual processing.

Perhaps the most important improvement for multinational customers -- which includes many Notes/Domino users in this global information age -- is that the limitation of the MTA is now resolved. Domino R5 provides global support for converting between Notes Rich Text and MIME. This is true for both outbound and inbound messages. The global outbound support is primarily achieved through character set autodetection. The global inbound support is achieved through a combination of character set autodetection and native MIME support, new in R5, which allows for explicit overriding of a character set in the Notes client when the autodetection guesses wrong.

Moving forward, the goals for MIME in Notes/Domino are really simplification and unification. For example, although you can do more MIME configuration in Notes R5 than you could in R4.6, we'd like the configuration to be more "in your face." That is, we hope to make the MIME settings available in menus, action

buttons, toolbars, and other shortcuts. In addition, we are working on better central administration of the configuration. For example, the configuration on the client in the Personal Address Book will be pushable via Dynamic Configuration.

Also, we are working on enhancing both our inbound and outbound autodetection technology, the enabling technology which allows you to have multilingual servers for MIME handling. This also will potentially mean taking advantage of the character set and language information present in multilingual databases (a new feature in R5). In addition to character set autodetection, font autodetection and substitution is an important priority.

Finally, we are working on providing better APIs for manipulating native MIME data found in notes. Clearly, these APIs will advance the power of our native MIME handling in general, not just for international issues. With all of these improvements, we'll truly be making language barriers a thing of the past.

ABOUT THE AUTHOR

Jeff Eisen joined Iris in July of 1995 to lead the effort of embedding Java applets into the Notes client. In fact, see [our interview with Jeff](#) and [an article Jeff wrote](#) to learn more about his work with Java applets. Before joining Iris, Jeff worked at Lotus where he focused on the platform-dependent layer of LotusScript. For R5, Jeff was responsible for the Notes browser as well as many other internet and MIME issues including the international work forming the basis for this article.

What do you
think about
this article?

Register
Here!

[Lotus Home](#) | [IBM Home](#) | [Iris Home](#) | [Feedback](#)
Copyright 1999 Iris Associates Inc.



Notes.net

Iris Today



[back to "[Worldwide messaging: Using international MIME in R5](#)"]

Character set detection flowchart (sidebar)

The following flowchart shows how Notes/Domino chooses the character set to use for inbound conversions (MIME-to-Notes). For Japanese and Korean 7-bit detection, this flowchart describes the process coming in R5.0.2, as noted by the asterisks at the end of the flowchart.

