

Internationalization

The SOA Gateway uses IBM's International Components for Unicode to support internationalization (i18n). This supports text data conversion between almost any codepage.

- Setting codepages
 - Which codepage do I use?
 - SOAP versus REST differences
 - Troubleshooting
-

Setting codepages

The Single Byte Character Set (SBCS) or Multi Byte Character Set (MBCS) codepage can be set on the driver, for more information, see here. The codepages can also be set on the on the web service itself, by simply left-clicking the web service and entering the codepage in the appropriate section of the web service Properties.

Important:

The codepage set on the web service overrides the one set on the driver.

Which codepage do I use?

This depends on what sort of information your service is going to return. Generally the ASCII codepage is sufficient for the English language. The ISO-8859-1 (often called latin1) codepage should suffice for most languages of Western Europe. The windows-1251 codepage supports Cyrillic languages such as Russian and Bulgarian. The ISO-8859-8 codepage can be used for Hebrew script.

The ICU home page has provided a useful web page which displays the ICU internal name, and a list of the aliases that the SOA Gateway will recognise. This page will also display the codepage map, which will allow you to choose the codepage best suited to your service.

SOAP versus REST differences

Generally when using WSDL and SOAP, once the correct codepage has been set, the payload should be recognised or returned correctly.

When using REST requests, things are slightly different. Non-ASCII characters entered on a URL bar of a browser will be escaped into their native hex value, of the form %XX. This native hex value differs depending on what codepage the browser recognises the character as. For example, a browser running in the latin1 codepage will recognise Á as %C1, but a browser running in the Cyrillic codepage will recognise Б as %C1.

For this reason the SOA Gateway allows users to provide an extra field on the REST request. This field is called __encoding. Thus users can indicate what codepage their browser is running in.

Important:

By default, the SOA Gateway assumes the escaped values are in the ISO-8859-1. The `__encoding` field is not required in this case.

Example 1

The browser escapes the Russian `Б` into `%C1`. You need to tell the SOA Gateway that this the Cyrillic encoding.

The URL should be

```
http://host:port/WebService?LIST&key=%C1&__encoding=windows-1251.
```

Example 2

The browser escapes the Hebrew Shin (`ש`) into `%F9`. You need to tell the SOA Gateway that this is the Hebrew encoding

The URL should be

```
http://host:port/WebService?LIST&key=%F9&__encoding=iso-8859-8.
```

Troubleshooting

When the SOA Gateway cannot display a character in the requested codepage, it writes a message to the error log, and continues to attempt to process the rest of the payload. If you find your responses are missing some characters, check the `error_log / error.log / XMIDCARD` on *nix, Windows and z/OS respectively.

The error message to check should be something like this :

```
Unicode char 0xF1 is not representable in encoding ASCII.
```