TIBCO Foresight® Products

Unicode at TIBCO Foresight

August 2017

Two-Second Advantage®

Important Information

SOME TIBCO SOFTWARE EMBEDS OR BUNDLES OTHER TIBCO SOFTWARE. USE OF SUCH EMBEDDED OR BUNDLED TIBCO SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED TIBCO SOFTWARE. THE EMBEDDED OR BUNDLED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER TIBCO SOFTWARE OR FOR ANY OTHER PURPOSE.

USE OF TIBCO SOFTWARE AND THIS DOCUMENT IS SUBJECT TO THE TERMS AND CONDITIONS OF A LICENSE AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED SOFTWARE LICENSE AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER LICENSE AGREEMENT WHICH IS DISPLAYED DURING DOWNLOAD OR INSTALLATION OF THE SOFTWARE (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH SOFTWARE LICENSE AGREEMENT OR CLICKWRAP END USER LICENSE AGREEMENT, THE LICENSE(S) LOCATED IN THE "LICENSE" FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.

This document contains confidential information that is subject to U.S. and international copyright laws and treaties. No part of this document may be reproduced in any form without the written authorization of TIBCO Software Inc.

TIBCO, Two-Second Advantage, TIBCO Foresight Archive and Retrieval System, TIBCO Foresight Community Manager, TIBCO Foresight HIPAA Validator Desktop, TIBCO Foresight Instream, TIBCO Foresight Transaction Insight, and TIBCO Foresight Translator are either registered trademarks or trademarks of TIBCO Software Inc. in the United States and/or other countries.

Enterprise Java Beans (EJB), Java Platform Enterprise Edition (Java EE), Java 2 Platform Enterprise Edition (J2EE), and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle Corporation in the U.S. and other countries.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only.

The United States Postal Service holds the copyright in the USPS City State Zip Codes. (c) United States Postal Service 2017.

THIS SOFTWARE MAY BE AVAILABLE ON MULTIPLE OPERATING SYSTEMS. HOWEVER, NOT ALL OPERATING SYSTEM PLATFORMS FOR A SPECIFIC SOFTWARE VERSION ARE RELEASED AT THE SAME TIME. SEE THE README FILE FOR THE AVAILABILITY OF THIS SOFTWARE VERSION ON A SPECIFIC OPERATING SYSTEM PLATFORM.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. TIBCO SOFTWARE INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "READ ME" FILES.

Copyright © 2010-2017 TIBCO Software Inc. ALL RIGHTS RESERVED.

General Contact Information
TIBCO Software Inc.
3303 Hillview Avenue
Palo Alto, CA 94304 USA
Tel: +1 650 846 1000
Fax: +1 650 846 1005

Technical Support
E-mail: support@tibco.com
Web: https://support.tibco.com
(Note: Entry to this site requires a username and password. If you do not have one, you can request one. You must have a valid maintenance or support contract to use this site.)

Contents

Document Purpose

This document is for users of TIBCO Foresight products who are working with Unicode data.

Unicode is a standard for handling text expressed in various languages.

Examples of text that can be expressed in Unicode:

English	Chinese	Hebrew
HOSPITAL	醫院	בית חולים

TIBCO Foresight products shown below handle the most common encodings: UTF-8 and UTF-16.

Unicode and TIBCO Foresight Products

Product	Level of Unicode support
TIBCO Foresight® Community Manager®	Validates Unicode
TIBCO Foresight® HIPAA Validator® Desktop	Validates Unicode
TIBCO Foresight® Instream®	Validates Unicode Response Generator reads and writes Unicode
TIBCO Foresight® Translator	Translates to and from Unicode
TIBCO Foresight® Transaction Insight®	Unicode is not supported
TIBCO Foresight® Archive and Retrieval System	Unicode is not supported

Unicode Capabilities in Instream®

Validating

When validating, Instream passes any Unicode through to the detail results file, which it writes as a UTF8 file.

GEN record 10224 describes the encoding in the input file:

```
   GEN          010224 1 0encoding=[UTF-8]
```

This information is used by Response Generator.

Controlling Response Generator Encoding in Output

Default behavior for Response Generator:

- Input ASCII creates output ASCII
- Input UTF8 or UTF16 creates UTF8 output with BOM

If you do not want to use the default behavior when creating response documents, use Response Generator's -e command-line parameter to control the encoding of its output, as shown in the following table.

Input data	RespGen output data	RespGen -e parameter	RespGen 999 IK4-04
Extended ASCII	ASCII	-eTEXT or none	Populated with encoded data
	UTF8 BOM	-eUTF8	Populated with encoded data
	UTF16 BOM	-eUTF16	Populated with encoded data
UTF8	UTF8 BOM	-UTF8	Populated with encoded data
	UTF8 no BOM	-eTEXT or none	Populated with encoded data
	UTF16 BOM	-UTF16	Populated with encoded data
UTF16	UTF8 with BOM	-UTF8	Populated with encoded data
	UTF8 no BOM	-TEXT or none	Populated with encoded data
	UTF16 with BOM	-UTF16	Populated with encoded data

Encoding Examples in Instream's Scripts directory

V_RG_837P_5010_ExtendedASCII.bat
V_RG_837P_5010_UTF8.bat
V_RG_837P_5010_UTF16.bat

Unicode Capabilities in Translator

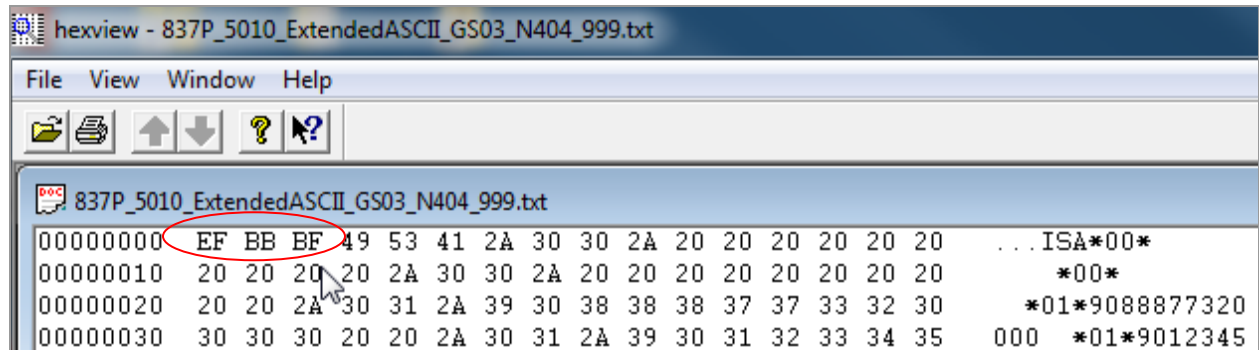Please see **TIB_translator_<n.n>_usersguide.pdf**.

Seeing Encoding in the Output

Seeing the BOM

Using a hex editor, you can look at the first characters in the output to see the BOM.
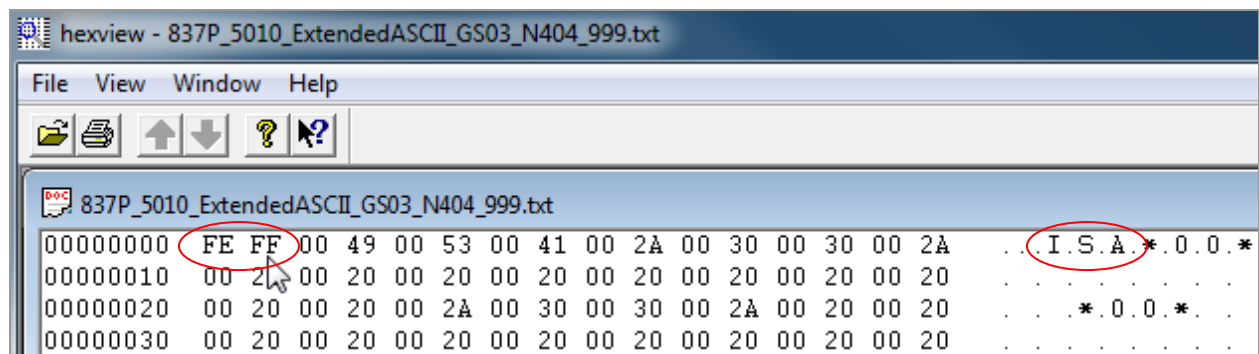
UTF8 BOM is three bytes:

```
hexview - 837P_5010_ExtendedASCII_GS03_N404_999.txt
File   View   Window   Help

837P_5010_ExtendedASCII_GS03_N404_999.txt
00000000  EF BB BF 49 53 41 2A 30 30 2A 20 20 20 20 20 20   ...ISA*00*
00000010  20 20 20 20 2A 30 30 2A 20 20 20 20 20 20 20 20       *00*
00000020  20 20 2A 30 31 2A 39 30 38 38 38 37 37 33 32 30     *01*9088877320
00000030  30 30 30 20 20 2A 30 31 2A 39 30 31 32 33 34 35   000   *01*9012345
```
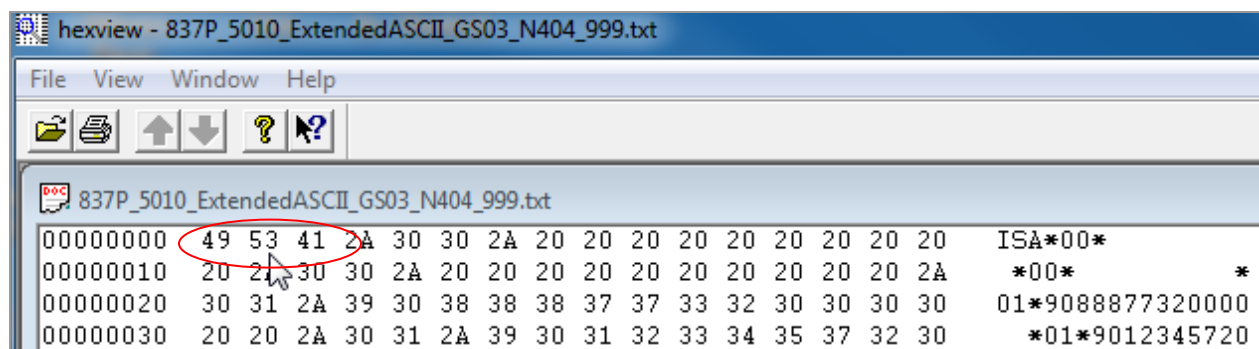
UTF16 BOM is two bytes, and every character in a UTF16 file is two bytes:

```
hexview - 837P_5010_ExtendedASCII_GS03_N404_999.txt
File   View   Window   Help

837P_5010_ExtendedASCII_GS03_N404_999.txt
00000000  FE FF 00 49 00 53 00 41 00 2A 00 30 00 30 00 2A   ..I.S.A.*.0.0.*
00000010  00 2A 00 20 00 20 00 20 00 20 00 20 00 20 00 20   . . . . . . . .
00000020  00 20 00 20 00 2A 00 30 00 30 00 2A 00 20 00 20   . . .*.0.0.*. .
00000030  00 20 00 20 00 20 00 20 00 20 00 20 00 20 00 20   . . . . . . . .
```
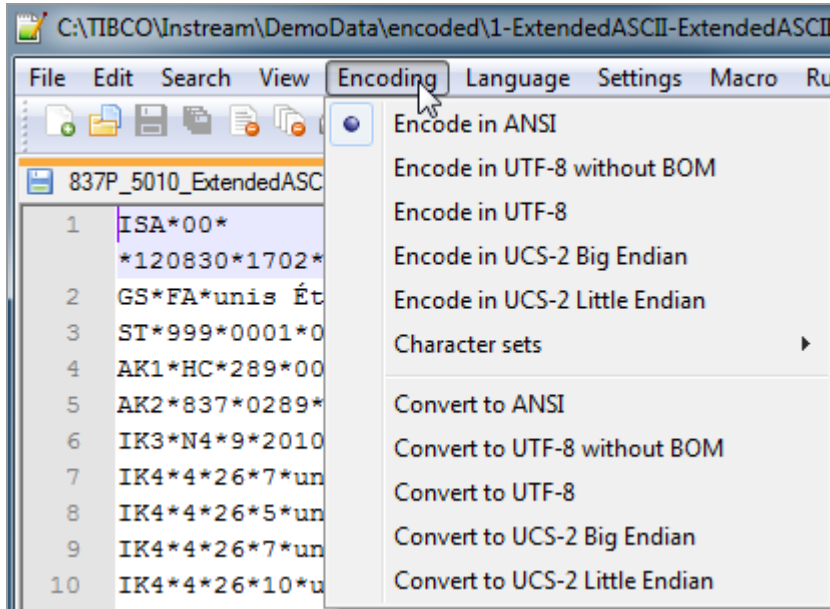
No BOM:

```
hexview - 837P_5010_ExtendedASCII_GS03_N404_999.txt
File   View   Window   Help

837P_5010_ExtendedASCII_GS03_N404_999.txt
00000000  49 53 41 2A 30 30 2A 20 20 20 20 20 20 20 20 20   ISA*00*
00000010  20 2A 30 30 2A 20 20 20 20 20 20 20 20 20 20 2A    *00*          *
00000020  30 31 2A 39 30 38 38 38 37 37 33 32 30 30 30 30   01*9088877320000
00000030  20 20 2A 30 31 2A 39 30 31 32 33 34 35 37 32 30    *01*9012345720
```
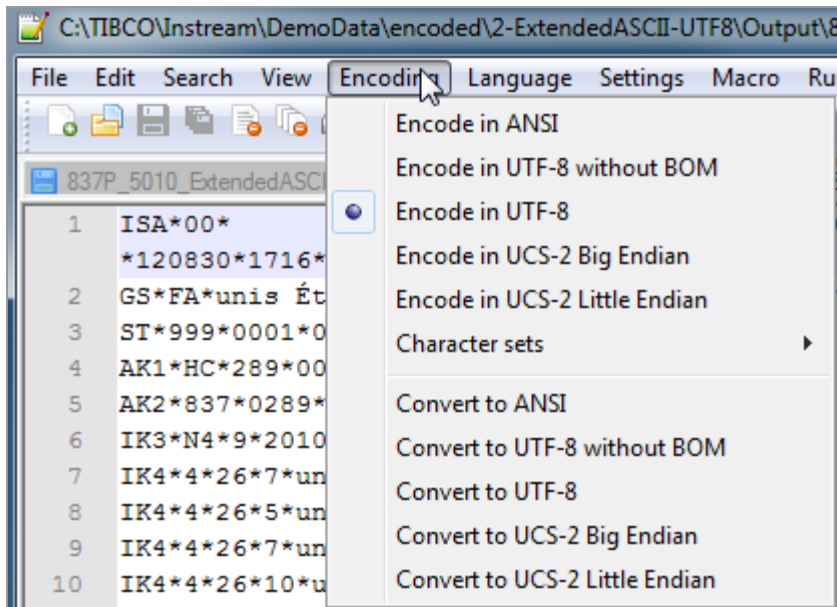
Seeing the Encoding

Using an editor that supports encoding, you can check the Response Generator output.

ASCII

UTF8

UTF16