Supported Character Sets

On this page:

Target audience:

CAST AI Administrators

Introduction

This document describes the character sets that can be used in the CAST Services (CAST Storage Service, Analysis Service, Dashboard Service and Management Service). For further information, see also **Collation compatibility between server hosting the Analysis Service and potential participating Servers** in Appendix - RDBMS requirements and configuration.

Supported Languages with Unicode

The supported languages with Unicode are:

- Albanian
- Belarusian
- Bosnian
- Bulgarian
- Catalan
- Chinese
- Croatian
- Czech
- Danish
- Dutch
- Estonian
- Faroese
- Finnish
- French
- German
- English
- Greek
- Greenlandic Inuktitut
- Hungarian
- Icelandic
- Japanese - note that we do not support:
    - EUC-JP
    - ISO_2022_JP_3
- Irish
- Italian
- Latvian
- Lithuanian
- Luxembourgish
- Macedonian
- Maltese
- Moldovan
- Norwegian
- Polish
- Portuguese
- Romanian
- Russian
- Serbian
- Slovak
- Slovenian
- Spanish
- Swedish
- Turkish

- Ukrainian

Any non-supported character that is present in the source code file encoded with one of the supported encodings, will be converted to an arbitrary supported character by CAST. The impact on the analysis result of this conversion depends on the situation in which the conversion occurs and on the character to which the conversion occurred. Therefore, the impact is unpredictable in a general way as for example:

- If the conversion occurs only in source code comments, there is no impact on the analysis result.
- If the conversion occurs inside an identifier and the converted identifier is no longer unique due to the conversion, resolution errors can occur.

Please also note that results of the analysis depend also on whether the Storage service (CAST Storage Service or commercial RDBMS) supports UNICODE (please see the details in the chapter below).

Code page equivalence for Unicode Character Sets

The language of the code page used in the Operating System hosting the CAST Analysis workstation (the machine on which the CAST Management Studio is run from) must be the same as the language used for the source code to be analyzed. For example on an OS in Turkish you must analyze source code that is Unicode encoded for the Turkish language.

Encoding support

Storage Services

CAST Storage Services

The CAST Storage Service (CSS) can be used to store analysis results of Unicode encoded source files provided the files use one of the below mentioned encodings:

- UTF-8 without BOM
- UTF-8 with BOM
- UTF-16 with BOM
- GB 18030 (standard Chinese character set)
- BIG5 (Chinese character set for Taiwan, Hong Kong and Macau)

Oracle Database Server

The following character sets are those corresponding to a single byte, ASCII with '€' coding. They can be used with CAST AIP products:

WE8PC858	IBM-PC Code Page 858 8-bit West European
EL8ISO8859P7	ISO 8859-7 Latin/Greek
WE8ISO8859P15	ISO 8859-15 West European
EE8MSWIN1250	MS Windows Code Page 1250 8-bit East European
CL8MSWIN1251	MS Windows Code Page 1251 8-bit Latin/Cyrillic
WE8MSWIN1252	MS Windows Code Page 1252 8-bit West European
EL8MSWIN1253	MS Windows Code Page 1253 8-bit Latin/Greek
TR8MSWIN1254	MS Windows Code Page 1254 8-bit Turkish
BLT8MSWIN1257	MS Windows Code Page 1257 8-bit Baltic

Please note that using an "Oracle Database Server" does not provide any support for any Unicode encoding.

Microsoft SQL Server

With Microsoft SQL Server, CAST recommends using Windows collations. Linked to the Windows locales, the code page '1252' is suitable for Western Europe, the Americas and Australia. In addition, the CS (Case sensitive) and AS (Accent Sensitive) attributes must be active.

Please note that using an "Microsoft SQL Server" does not provide any support for any Unicode encoding.

CAST AIP schemas

The **Dashboard Service**, **Analysis Service**, **Management Service** and **Measurement Service** schemas support the following encodings:

- UTF-8 without BOM
- UTF-8 with BOM
- UTF-16 with BOM
- GB 18030 (standard Chinese character set)
- BIG5 (Chinese character set for Taiwan, Hong Kong and Macau)

CAST AIP Analyzers

The following CAST analyzers:

- C/C++
- .NET > Please note that C/S links are not resolved when the T-SQL database collation is not the same as the server collation, and the machine collation is different then the server collation.
- ASP
- Visual Basic
- Universal Analyzer/Universal Importer
- JEE Analyzer extension (including EJB, Web Services and CAST Script)
- Mainframe
- ABAP
- PL/SQL
- T-SQL
- SQL Analyzer extension

support the following encodings:

- UTF-8 without BOM
- UTF-8 with BOM
- UTF-16 with BOM
- GB 18030 (standard Chinese character set)
- BIG5 (Chinese character set for Taiwan, Hong Kong and Macau)

CAST AIP features

The following CAST AIP components:

- CAST Server Manager
- CAST Management Studio, including:
    - Reference Patterns
    - Update Knowledge Base (Analysis Service) Assistant
    - XXL Table Quality Rule injection
    - Background Facts upload
    - Environment Profile Manager
- CAST Architecture Checker
- CAST Transaction Configuration Center
- Engineering Dashboard
- Health Dashboard
- Legacy CAST Engineering Dashboard
- Legacy CAST Discovery Portal
- CAST Report Generator
- CAST Logs
- CAST Delivery Manager Tool
- Command line (CAST Management Studio/CAST Delivery Manager Tool)
- Metrics Assistant

support the following encodings:

- UTF-8 without BOM
- UTF-8 with BOM
- UTF-16 with BOM
- GB 18030 (standard Chinese character set)
- BIG5 (Chinese character set for Taiwan, Hong Kong and Macau)

All other unmentioned CAST AIP components may provoke an arbitrary error when analyzing or working with a Unicode encoded source code file.

Note: BOM = Byte Order Mark, an indicator at the beginning of the Unicode encoded file that specifies in which order the bytes of a multi-byte character appear in the file ("Little Endian" vs. "Big Endian" encodings)