

ViewpointTM Foreign Language Support and Detection

Application Version: 7.0

August 19, 2026

© 2021 Conduent, Inc. All rights reserved. Conduent and Conduent Agile Star are trademarks of Conduent, Inc. and/or its subsidiaries in the United States and/or other countries.

Other company trademarks are also acknowledged.

Document Version: 7.0 (August 2021).

Document Conventions

Convention	Explanation
Bold	For file names, commands, fields, menus, options, and window names.
<code>Courier New</code>	Commands as you should type them.
<code>Lucida Console</code>	Example output generated by the system.
<i>Italics</i>	For configuration variables, including variable portions of file names and URLs. Also indicates a document name.
Note / Blue Callout	The Blue Callout text indicates information that is of special interest or importance, an idea that could be useful or additional information about a product or a feature.
	The Caution icon along with the text indicates actions that can lead to problems in system operation or configuration settings if the instructions are not followed properly.

Document Revision History

This section tracks the initial creation of the document after each major version thereafter.

Ver:	Date	Description	Reviewed / Approved By
1.0	Feb 28, 2020	Initial Version	Team
1.1	August 23, 2021	Rebranded adhering to the latest Conduent brand central documentation standards/guidelines.	Technical Writer

Table of Contents

- 1. Foreign Language Support and Detection.....1-1
 - 1.1. Foreign Language Detection.....1-2

1. Foreign Language Support and Detection

Viewpoint supports the Unicode Standard. Viewpoint supports indexing, searching and displaying Unicode (and ASCII) documents including support in Near Duplicate (ND), Email Thread Analyzer (ETA) and Relationship Analyzer (RA). Viewpoint Concept Analyzer is not capable of processing Asian language documents at this time. The following is a partial list of the character sets supported by the Unicode standard:

Supported Unicode Character Sets		
Arabic	Hangul	Old Turkic
Aramaic, Imperial	Hanunóo	Osmanya
Armenian	Hebrew	Oriya
Avestan	Hiragana	Pahlavi, Inscriptional
Balinese	Javanese	Parthian, Inscriptional
Bamum	Kaithi	Phags-pa
Batak	Kannada	Phoenician
Bengali	Katakana	Rejang
Bopomofo	Kayah Li	Runic
Brahmi	Kharoshthi	Saurashtra
Buginese	Khmer	Samaritan
Buhid	Lao	Shavian
Canadian Syllabics	Latin	Sinhala
Carian	Lepcha (Rong)	Sumero-Akkadian Cuneiform
Cham	Limbu	Sundanese
Cherokee	Linear B	Syloti Nagri
Coptic	Lisu	Syriac
Cypriot	Lycian	Tagalog
Cyrillic	Lydian	Tagbanwa
Deseret	Malayalam	Tai Le
Devanagari (Hindi, etc.)	Mandaic	Tai Tham
Egyptian Hieroglyphs	Meetei Mayek	Tai Viet
English	Mongolian	Tamil
Ethiopic	Myanmar	Telugu
Georgian	New Tai Lue	Thaana

Glagolitic	N'Ko	Thai
Gothic	Ogham	Tibetan
Greek	Ol Chiki	Tifinagh (Berber)
Gujarati	Old Italic (Etruscan)	Ugaritic
Gurmukhi	Old Persian Cuneiform	Vai
Han (Japanese, Chinese, Korean)	Old South Arabian	Yi

1.1. Foreign Language Detection

The following languages / Unicode character sets are identifiable by Viewpoint's processing engine:

Viewpoint Identifiable Character Sets		
Albanian	Greek	Ol Chiki
Arabic	Hebrew	Philipino
Armenian	Hindi	Polish
Balinese	Hungarian	Portuguese
Batak	Icelandic	Rejang
Buginese	Indic	Romanian
Bulgarian	Indonesian	Russian
Canadian Aboriginal	Irish	Samaritan
Cham	Italian	Serbian
Cherokee	Japanese	Slovak
Chinese	Kayah Li	Slovenian
CJK	Khmer	Spanish
Croatian	Korean	Sundanese
Cyrillic	Lao	Swedish
Czech	Latvian	Tagalog
Danish	Lepcha	Tai Le
Dutch	Limbu	Tai Tham
English	Lithuanian	Thaana
Esperanto	Malay	Thai
Estonian	Mandaic	Tibetan
Ethiopic	Mongolian	Turkish
Finnish	Myanmar	Ukrainian

French	NKo	Vietnamese
German	Norwegian	

When the option to identify foreign languages is turned on, the processing engine will use one of two Language Models. The system defaults to use the **Full Language Set**; however, users can select the Short Language Set, which will only identify English, German, French, Italian, Portuguese, Spanish, Hebrew, Chinese, Japanese, Korean, Arabic and Hindi. Choosing the **Short Language Set** can help speed up the process and limit the number of potential false positives.