Solving International Label Printing Challenges with Unicode™

A ZEBRA **BLACK&WHITE** PAPER

Zebra

Many label printers can print this:

ABC 123456

Can these same printers also print this?

حلول الطباعة العالمية من Zebra متوافقة مع
Unicode™ من أجل طباعة حرارية إبداعية سريعة
بعدة لغات حسب الطلب. فقط من Zebra!

Zebra Technologies now makes it easy to do both.

The Arabic characters represented above are extremely challenging for many thermal printers to output, because character shapes have variable forms, the language is printed from right to left, and commonly used fonts can't express the characters. These are just a few of the many printing challenges that non-Western languages can present.

With companies transacting business around the world, it's more important than ever to be able to identify parts, products, and pallets of goods—plus their ports of call—in different languages on the same label.

Organizations that need to print shipping labels in these languages have been faced with implementing costly custom output solutions, or treating characters as graphics, which results in very slow label printing.

The Unicode™ Standard was created to address multiple language printing problems. Until recently, Unicode™ has not been available in bar code label printers. Zebra now offers Unicode™ support so users can output almost all the world's major languages on their Zebra® printers straight out of the box. This white paper provides an overview of international language printing challenges, describes Zebra's Global Printing Solution, and explains the alternatives for international language output on label printers.

International Character Printing Basics

Many organizations that do not have Unicode™-enabled printing capabilities print international characters from their business applications as graphics, because equipment and applications do not properly support the Unicode™ Standard for printing. The graphics approach simplifies the international character printing effort but drastically reduces printer performance and first label out time. Printers cannot download or process graphics as quickly as they can process text strings. Using graphics instead of fonts reduces the throughput of the printing application because it requires more processing time for each label. It also leads to delays between labels instead of continuous printing.

If users need to output even a single character not included in their codepage, they need to install additional codepages and supporting fonts, have a third-party business application provider develop language-specific solutions, or print non-supported characters as graphic images.

Output through native codepage support is the most convenient and cost-effective option because printing graphics is slow, and add-on codepage solutions become expensive and difficult to scale.

Printers can normally support multiple codepages. Organizations that need to print labels in multiple languages have traditionally upgraded their printers by developing language-specific codepages, licensing fonts and purchasing the additional memory required to accommodate them. These printing systems become complex and expensive to manage as business expands into new regions because of the font licensing, installation, and printer configuration required to support each new language.

Unicode

Unicode™ is an industry standard whose goal is to provide the means by which text of all forms and languages can be encoded for use by computers. ASCII and most other traditional codepage encoding systems support 256 characters or less. The Unicode™-supported character set covers almost 100,000 characters from all the world's major languages, including complex non-Western languages that can be challenging to print.
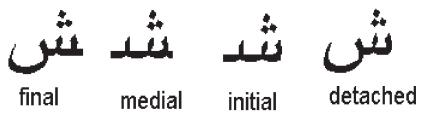
A Unicode™-enabled printer could seamlessly output any language, with no need for an operator to select the language, font, or codepage, or otherwise configure or adjust the printer. Many leading IT systems and enterprise software applications are now users of the Unicode™ Standard. Organizations can print international language labels directly from their applications by networking a Unicode™-enabled printer to these systems.

Additional Language Challenges

Some languages present additional challenges that are met different ways depending on the Unicode™-supported implementation. Asian, Middle Eastern, and Indian languages in particular require a robust Unicode™-solution to ensure proper printing.

Middle Eastern

Arabic and Hebrew are the two most common languages in the Middle East. They differ from most other languages because they are read and written from right to left. Another issue with Arabic is that characters are displayed cursively. Arabic characters change shape depending on the characters around them. For example, below are four different representations of the Arabic letter "Sheen."

ـش   ـشـ   شـ   ش

final   medial   initial   detached

Even though a character can have several different forms, it is assigned a single Unicode™ code point. Advanced Unicode™-enabled printing solutions can print the proper shape variants based on context. Other solutions simply print the Arabic characters as graphics.

Indian and Southeast Asian

Languages of India and Southeast Asia use scripts that can be difficult for printers to output. Languages have different ways of displaying the human-readable text, each using different scripts. English, for example, uses the Latin script to produce human-readable English text. A single script can be used for more than one language, and a language may use more than one script.

Countries in Southeast Asia including Thailand, India, Sri Lanka, the Philippines, and Bangladesh use scripts such as Thai, Devanagari, Telugu, Bengali, and Sinhala. The scripts feature headstrokes and combined characters. A headstroke is a horizontal line that runs across the top of each character. The character stems off from the headstroke. The characters combine and can change order depending on the characters around them. As with Arabic, even though a character can have several forms, it is only assigned one Unicode™ code point.
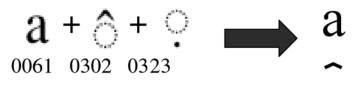
Other Asian Languages

The remaining Asian languages not covered by other regions are Japanese, Korean, Simplified Chinese, Traditional Chinese, and Vietnamese, commonly known as CJKV. The vast amount of characters each of these languages contains creates printer memory and output challenges. Although only around 2,000 to 3,000 characters are required for basic literacy in Japanese or Chinese, there are upwards of 80,000 characters listed in some dictionaries. Most of these characters are rarely used in everyday writing, but are commonly used in proper names—which means they are needed for shipping labels and other business documents. Most characters will have the same meaning in all CJKV languages, but may have a slightly different glyph (the visual representation of the character) in each. These languages also use multiple scripts. A sentence in Japanese, for example, may use up to four scripts. Therefore, the fonts that support CJKV languages are very large and memory intensive because of the amount of characters and representations they need to accommodate.

Storing large, memory-intensive fonts on a printer can reduce print speed. Most TrueType™ fonts that support CJKV languages are too large for the available printer memory. Font and memory problems can be solved by adding memory to the printer, or by having the font on a PC card that is inserted into the printer.

Fonts for CJKV languages still may not have all the required versions of characters. Vietnamese words must have a tone mark, which is a diacritical mark combined with a base character. Many of these characters do not have a presentation form. When using the presentation forms to render the characters, there is a potential problem if a presentation form is not available. This problem could arise for several reasons, including that the presentation form is not available in the selected font, or the user selected multiple combining diacritics that do not form a valid combination. The following illustration is an example of the output if there is no presentation form available in the font.

The character in this Vietnamese example is a lower case "a" with a dot below and a circumflex accent above. There is no presentation form available for this character.

The non-spacing diacritical marks, such as the combining dot below and accent above in this example, are required to print Pointed Hebrew and Vietnamese. The example above also shows that Vietnamese requires more vertical space to be displayed properly. To support these languages, the characters should be printed without using the presentation forms because the forms are not always available.

Zebra's Global Printing Solution

Zebra is now offering a Global Printing Solution that is currently not provided by its competitors. Zebra's Global Printing Solution allows most of the world's languages, including Arabic and Asian characters, to be printed without needing to develop the unique codepages for each language or to slow label processing. Maximum printing performance can be achieved without language-specific codepage development, label design, font licensing, or modification of business applications for different languages. It is now possible to design one label format, printing on one printer model, from one version of your business application software that can be used around the world. The solution is available for Zebra's *Xi*™ series, 105*SL*™, Z4Mplus™, and Z6Mplus™ printers, and *PAX*™ series print engines.

The preloaded TrueType font (Swiss™ 721) lets users print any European, Middle Eastern, or African (EMEA) language, including Arabic and Hebrew, right out of the box. Zebra preloads the Unicode™ codepage and supporting fonts into printers shipped to customers in Europe, the Middle East, and Africa (EMEA), and offers the solution as a free option anywhere else in the world.

To print Southeast Asian and CJKV languages, users only need to add a supporting font, which Zebra makes available with a factory-installed Flash memory upgrade or on a PC card. Zebra offers Worldtype® font for Asian languages (Andalé®), which also supports EMEA languages, so a single font can satisfy all printing needs and multiple fonts don't have to be stored on the printer. Zebra's solution supports multiple Unicode™ encoding methods, including UTF-8, UTF-16BE, and UTF-16LE. See Appendix A for a complete list of supported encodings, scripts, and languages.

The solution supports the OpenType® standard, which enables users to print the diacritic marks needed to properly print the Vietnamese and Pointed Hebrew languages. OpenType is a cross-platform font file format developed jointly by Adobe and Microsoft. The Andalé fonts have been revised to include OpenType tables for 10 Indian scripts, including Devanagari.

The solution includes the Microsoft® Windows® Private Character Editor, which gives users the ability to create their own logos and special characters for printing. The Private Character Editor allows the user to design a character that is assigned a code point in the Unicode™ private character space. This feature is very valuable for organizations that create Asian shipping labels, because many streets, company names, and other proper nouns are not included in Asian-language fonts.

Because Zebra's Global Printing Solution includes Unicode™ encoding, preloaded supporting fonts, and implementation of the OpenType standard, Zebra printers do not need to convert text to graphics for output. The long waits for the first label to print and delays between labels associated with graphics printing are thus avoided. Bidirectional printing is also supported, for fast output of Arabic, Thai, and other languages that require it. These productivity benefits are especially noticeable in mid- to high-volume printing operations.

Conclusion

International printing implementations have introduced new challenges for printer memory management, font compatibility, and codepage support. By including Unicode™ encoding and fonts in its printers, Zebra offers a simple, scaleable solution that requires minimal effort and support. With Zebra's Global Printing Solution, organizations can now develop a single printing application, business transaction, or label, and deploy it throughout the world without managing multiple configurations or requiring additional custom or codepage developments. The flexibility to print new languages without adding fonts and redeveloping labels provides a significant cost advantage for systems that will be deployed to support multiple languages as well as business transactions that could change after the initial implementation.

Appendix A: Encoding, Language, and Script Support

Zebra's global printing solution supports the following encodings, languages and scripts:

Supported Encodings

- Big5
- GB2312
- Shift JIS
- JIS
- Code Page 874

- GB18030-2000
- Big5 HKSCS
- UTF-8
- UCS-2
- Wansung

- UTF-16
- Johab
- Unified Hangul Code
- Code Page 850
- Code Page 1252

Supported Languages

- Albanian
- Arabic
- Azerbaijani
- Bulgarian
- Chinese (Traditional)
- Chinese (Simplified)
- Croatian
- Czech
- Danish
- Dutch
- English
- Estonian
- Farsi
- Finnish
- French

- German
- Greek
- Hebrew
- Hindi
- Hungarian
- Icelandic
- Indonesian
- Italian
- Japanese
- Kazakh
- Malay
- Moldavian
- Korean
- Norwegian
- Polish

- Portuguese
- Romanian
- Russian
- Serbian
- Slovak
- Slovene
- Spanish
- Swedish
- Tajik
- Thai
- Turkish
- Ukrainian
- Urdu
- Vietnamese52

Supported Scripts

- Arabic
- Bopomofo
- Cyrillic
- Devanagari

- Greek
- Han
- Hangul
- Hebrew

- Hiragana
- Katakana
- Latin
- Thai

Notes

Notes