

# Unicode Breaks The ASCII Barrier

ASCII looked great in the '60s, but it's just an ugly American in the global '90s. Now Aldus, Apple, IBM, Microsoft and Xerox are trying to replace ASCII with Unicode—the code of a thousand character sets.

BY JOHN R. VACCA

**S**ometimes it's the little things that create the biggest problems. Take simple, little old ASCII.

Today, ASCII is the most common and most widely accepted computer code in the world. Born in 1967 as the American Standard Code for Information Interchange, ASCII was supposed to save the world from being locked into IBM's then-proprietary Extended Binary-Coded Decimal Interchange Code (EBCDIC).

At the ripe old age of 24, ASCII is positively ancient compared with most computer technologies. It has survived for so long because it is both simple and predictable: an unambiguous, 256 character, 8-bit fixed-length code with a well-organized and apparently numerous character set.

In recent years, however, ASCII has been showing its age and its limitations. Global computing becomes a whole lot more difficult when limited to a mere 256 characters. And that old 8-bit standard is holding back several hardware and software vendors that are looking for ways to enter major markets in Asia, where text comprises thousands of characters—not just the ABCs.

In fact, ASCII's biggest drawback is embodied in its name: American. ASCII can't begin to provide all the characters needed for information processing in Western Europe, not to mention Africa, Asia, Eastern Europe and the Soviet Union.

This month, a Mountain View, Calif.-based consortium of major computer companies, formally known as Unicode Inc., are mounting a challenge to ASCII. They plan to announce Unicode, a new and much more powerful multilingual text and character encoding standard, which will provide an international computer code for storage and transmission of text.

Think of Unicode as double-width global ASCII.

Unicode will establish an unambiguous, fixed 16-bit codeset. Since Unicode knows no country, no 7-to-8 bit shifting or code-type communication announcement mechanism will be required when transmitting or receiving coded text.

Major contributors to Unicode's development include Apple Computer Inc., Go Corp., IBM, Metaphor Computer Systems Inc., Next Inc., The Research Libraries Group Inc., Sun Microsystems Inc. and Xerox Corp. Recently, Aldus Corp., Lotus Development Corp. and Novell Inc. have taken active roles in the development of Unicode.

## The Multilingual Codeset

Unicode will simplify application development at international MIS shops by providing a single, uniform character standard for 99% of the languages known. Plans are in place to include historical languages, as well. "The Unicode standard evolved from the industry's need for a 16-bit version of ASCII," explains Joe Becker, a principal scientist for Stamford, Conn.-based Xerox Corp. and a technical vice president for Unicode Inc. "Information professionals gathered together to take the best from existing standards and apply their multilingual software experience to designing a simple, complete character encoding.

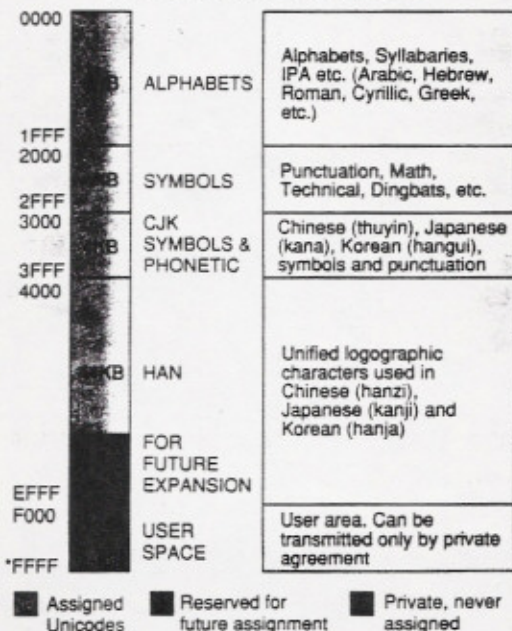
"In other words," contin-

ues Becker, "Unicode encompasses all major languages and offers a strong foundation for international MIS shops. This standard will make multilingual software easier to write, MIS shops easier to manage and international information exchange more practical."

The Unicode project was initiated in early 1989 by an informal working group

## Load 16 Bits And What Do You Get?

By replacing standard ASCII 256 character code space with the new Unicode 64-kilobyte standard, every application can speak in tongues. Here's how Unicode allocates the code space for all those new characters



\*FFFF is permanently reserved as an application-specific sentinel value (e.g. missing character)

Source: Apple Computer Inc.

## SOFTWARE STANDARDS

of linguists, engineers, managers and information professionals from organizations that were using or developing international information systems. The group determined the most effective way to achieve an international standard with the same level of simplicity and efficiency as ASCII would be to establish an inter-company, international character standard.

### UNICODE EVOLVED FROM THE INDUSTRY'S NEED FOR A 16-BIT VERSION OF ASCII.

Under the guidance of research librarians, Unicode staff reviewed bibliographic standards and requirements for modern and ancient texts and incorporated these whenever possible.

#### **A Catchy Code?**

Is Unicode part of a computer architecture or is it just a communications standard? That's up to the users to decide. Some MIS shops may end up using

Unicode only as an exchange medium to facilitate communication between different national systems with incompatible codesets. Other MIS shops may implement entirely new computer hardware architectures to support 16-bit Unicode text directly (examples would be to create extended keyboard memory or to make actual hardware changes on the keyboard itself).

The review draft of Unicode contains about 25,000 characters alone, covering all the world's major scripts, including some 18,000 Han—ideographic characters—incorporating national and de facto industry standards in China, Japan, Korea and Taiwan. Unicode has room for a total of 65,000 characters within its 16 bit-wide memory space (see table, "Load 16 Bits And What Do You Get?"). Obsolete scripts such as cuneiform, hieroglyphs and ancient Han characters (used in specialized research) will be added to future releases as required.

Of course, Unicode will only become a useful standard if widely adopted. Microsoft Corp. says it will incorporate Unicode in the TrueType font technology

of its upcoming Windows 3.1 environment, as well as in its Windows-32 version. Apple Computer plans to include it in its newest character set technology. Even IBM is looking at Unicode as the standard for future hardware.

Although Unicode may never replace ASCII altogether, a large number of MIS hardware and software vendors have shown support for Unicode.

### UNICODE WILL MAKE MULTILINGUAL SOFTWARE EASIER TO WRITE.

The final review draft of Unicode version 1.0 came out in December of 1990. The review period closed in mid-February. Since then, the Unicode consortium has been working to incorporate members' comments into the soon-to-be released Unicode version 1.0.

*John R. Vacca is a freelance technical data-processing and air & space contract writer based in Sterling, Va.*

## HIREL® STEP TO AN OPEN WORLD



*Exceptional technological concepts can help to gain a more global view of an evolving enterprise. IBM users can take a big step forward from a Hierarchical to an open RELational world by using the HIREL knowledge based re-engineering technology.*

*It supports the user on his way to a truly relational world, by performing a highly automated and reliable migration of data and programs from IMS/DB or DL/I to DB2. A wide range of specialized migration products based on the same technology is available.*

WE MAKE INNOVATION HAPPEN  
**SWS**  
Software Services

Worldwide - SWS  
Software Services GmbH  
☎ + 49 (72 31) 3 78 80

Germany - FWU/SWS  
Software Services GmbH  
☎ + 49 (71 41) 40 77 20

USA - SWS  
Software Services Inc.  
☎ + 1 (4 07) 2 77 03 01

Italy - AD Advanced Data  
☎ + 39 (2) 92 10 56 20

In addition, IBM is  
an international  
SWS marketing partner.

HIREL is a trademark of SWS Software Services. IBM, DB2, DL/I, IMS/DB are trademarks of International Business Machines Corp. HIREL/VREL are announced IBM logo products.