

CHINESE-JAPANESE-KOREAN AUTOMATION NEWS

Unicode - a character set standard

A consortium of major computer companies have formally affiliated as Unicode Inc. to maintain and promote Unicode, a newly-developed multilingual text and character encoding standard. Board members of Unicode Inc. include vice presidents or directors of Apple, GO, IBM, Metaphor, Microsoft, NeXT, Novell, The Research Libraries Group and Sun Microsystems.

Unicode simplifies the development of international information systems through a single uniform character standard. Unicode encompasses all major languages and offers a strong foundation for international information systems. This standard will make multilingual software easier to write, information systems easier to manage, and international information exchange more practical.

An informal Unicode working group, the predecessor of the incorporated consortium, has completed a review draft of the Unicode standard containing more than 27,000 characters. This draft is currently being reviewed by hundreds of individuals, companies and institutions.

Unicode character codes are fixed-width 16-bit codes, unambiguous and sufficiently numerous to include all characters likely to be used in electronic information processing. The codes are grouped by linguistic and functional categories.

The group has worked to ensure mutual convertibility between Unicode and existing national, international and inter-company encoding standards.

More information can be obtained from: Tim Helms or Kelli Ramirez at The Benjamin Group Inc., 1965 Charleston Road, Mountain View, CA 94043, USA.

From: "Major computer companies join to promote worldwide standard character encoding."
Press release dated Feb. 19, 1991.

Reactions to Unicode

There is some incompatibility between Unicode and the draft international standard ISO 10646. ISO 10646 has been criticised for the fact that the Basic Multilingual Plane no longer contains ideographic characters, while Unicode has been criticised for "imposing" its own unified set, rather than accepting either national standards or an East Asian unified set. Unicode is so far away from the basically 7-bit structure of ISO/TC46 that it is unlikely to find direct support from that quarter, particularly as Unicode places graphic characters where all ISO standards place control characters...

Unicode has a fuller repertoire in a 16-bit code than ISO 10646 provides in its 32-bit code. However, the structure of Unicode allows no backward compatibility with older equipment. The Standards Council of Canada has now proposed ISO 10646U, which would integrate Unicode and ISO 10646.

Excerpts from "Unicode and ISO 10646."
Sesame Bulletin 4(2) Summer 1991:51-52

CJK system developments in Brisbane

There has been some recent significant progress regarding the development of a PC based CJK system centred in Brisbane. A feasibility study is about to commence to determine whether the Jiejing word processing system can be successfully incorporated into the Softlink library system (variously known as OASIS or ALARM). This feasibility study is a joint project between Griffith University, the University of Queensland, Softlink (Australia) Pty. Ltd and Jiejing Pty. Ltd. It will determine the feasibility of:

1. the acceptance of Jiejing CJK character inputs by the Softlink system and their storage in a MARC compatible database system.
2. the incorporation of the Jiejing stroke order indexing system within the Softlink system,
3. the screen display and report printing of both CJK and ASCII characters.

The study will also cost the development of a beta-tested functional CJK library system based on 386 microcomputer technology. This is to be costed against a preliminary specification.

The Jiejing system retrieves characters by input via a modified standard querty keyword of character stroke elements which, operating against a large character

database, predicts the likely character required. Using this system a CJK character can be identified and then input to text with an average of 2.5 keystrokes.

Encoding of CJK is in 16 bit bytes (2 standard bytes) and it is envisaged that this will conform with the Unicode standard. The Research Libraries Group (RLG) has advised that Unicode can be used for data interchange from the Research Libraries Information Network (RLIN) (and presumably also the OCLC Online Computer Library Center) database which currently employs the three byte East Asian Character Code (EACC) coding system for internal storage.

The Softlink PC system has currently about 500 installations. Both companies are located in the Brisbane area.

As well as developing a systems for QU and Griffith East Asian libraries, it is envisaged that a marketable product will result, at what is hoped would be moderate cost.

Bruce Edwards at Griffith University (07-875 7134) will be glad to answer any queries.

Bruce Edwards
Associate University Librarian
Griffith University

