

ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC 1/SC 2/WG 2
Universal Multiple-Octet Coded Character Set
(U C S)

ISO/IEC JTC1/SC2/WG2 **N2497**
Date: 2002-05-22

Title: Clarification of relations between UCS-4 and different transformation formats

Source: China, Singapore, US national bodies

Status: Submission

Action: For consideration by JTC1/SC2/WG2

As raised by the Chinese national body and the Singapore national body during the WG2 meeting in document N2485, there are two areas where additional information in the standard would significantly reduce confusion about the UCS-4 and its relation to different transformation formats. We suggest the following additions of text to Amendment 2 to ISO/IEC 10646-1.

Addition to Clause 5 (after paragraph 7)

NOTE - The use of the term "canonical" for this form does not imply any restriction or preference for this form over transformation formats that a conforming implementation may choose for the representation of UCS characters.

Addition to Clause 13.2: Add "(UCS-4)" in the header, and add the following note at the end

NOTE - When confined to the code positions in Planes 0 to 10 (U+0000 to U+10FFFF), UCS-4 is also referred to as UCS Transformation Format 32 (UTF-32). The Unicode Standard defines the following forms of UTF-32.

- UTF-32: the ordering of octets (6.3) is not defined and the signatures (Annex H) may appear;
- UTF-32BE: in the ordering of octets the more significant octets precede the less significant octets, as specified in 6.2, and no signatures appear;

- UTF-32LE: in the ordering of octets the less significant octets precede the more significant octets, and no signatures appear.