

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N1037

Date: 2004-5-20

Doc. Type:	Member body contribution
Title:	Revised China's Basic Subset Proposal
Source:	China
Status:	Input to IRG
Action:	
Distribution:	IRG Members and Ideographic Experts
Reference:	
No. of pages:	1
Medium:	Electronic

This proposal proposes 5,080 characters to the IICore.

In the previous proposal (IRG N991, IRG#21), China proposed 7,184 candidates (category A 4,911, category B 918, category C 1,355). After rechecking these characters against **70 Million Contemporary Han Language Balance Corpus** (National Language Committee, Ministry of Education 教育部國家語委) and the **500 Million Hanzi Dynamically Circulating Corpus** (Beijing Language and Culture University 北京語言大學), with coverage over 99.9% as rationale, and with artificial adjusting, 5,080 characters are selected. Category C characters are almost completely withdrawn, 2,100 characters from category A and B are withdrawn.

See attached list.