

Information on Unihan Database Proposed Additions for the UTC

(No Action Required)

John H. Jenkins

Apple Computer, Inc.

L2/05-123

There has been some discussion on the Unihan mailing list regarding upgrading the phonetic data available in the Unihan database.

I've entered all the pinyin (Mandarin) readings from the *Hanyu Da Zidian* and Richard Cook is working with Magda Danish to proof them. When this is complete, we will ask for UTC approval for inclusion. This will improve the ability of clients (like IUC) to generate Mandarin readings for characters, as it provides data from an authoritative source and covers thousands of characters for which readings are not currently available.

Tom Emerson of Basis Technology is reviewing the Korean readings. We hope that he will be able to provide a revised and extended set of readings for Korean. He has suggested that we switch to using *hangul* instead of romanization to represent Korean readings, and the consensus of the mailing list is in agreement with him.

The Japanese readings also need extensive revision, but there have been as yet no volunteers to do the work. The suggestion here has been that we consolidate the two Japanese fields into one, kJapanese, using hiragana for the *kun* readings and katakana for the *on* readings. There is no strong consensus on the mailing list as to whether or not this is the best way to proceed.