

Unicode ê Pit-kì

Lâu Kiatt-gák

1. *Unicode* ê lèk-sú

1991 V1.0 28,302 jī.

1996 V2.0 38,885 jī.

2000 V3.0 49,194 jī.



Iáu-kú teh cheng-ka sin ê jī jip--khì, m̄-kú í-keng tī lâi-té..ê tō bē koh kái.

2. *Unicode* ê hōan-ûi

Óah ê (iáu teh sú-iōng..ê) gí-giân.

Phó-thong ê bûn-jī (plain text).

Unicode chhú-lí jī ê piau-sī hong-hoat. Tàk ê jī se"-chòe siá"-khóan sī lêng-gōa 1 chân tã-chì.

3. *Unicode* ê bók-phiau

Hō chōan sè-kài bûn-jī ê tiān-chú chu-liâu ē-tàng liū-thong kau-ōa".

Í -chá ê siat-kè lóng-sī 1-e/1-kóa gí-giân ēng 1 thò phiau-chún (kiò-chòe *codepage*). Kap pát-chióng gí-giân ê tiān-chú chu-liâu bô hoat-tō kau-ōa". Lí nā ū su-iàu ēng 2 chióng chha khah chē ê gí-giân (phì-jû kóng Hoat-gí kap Jit-gí), tō thâu thià"--ah. Ài sú-iōng tèk-piát ê íng-thé; chu-liâu kap pát-lâng mā bô it-tēng ē-tàng kau-ōa".

Unicode ê siat-kè hō bô káng gí-giân ê chu-liâu ē-tàng khng chò-hóe, kau-ōa" mā bô bûn-tôe--ah. Nā khòa" bô, tō chng

jī-hêng; nā beh siá, tō chng hit-chióng gí-giân ê su-jip-hoat tō hó.

Chia ū 1 ê chhù-bī ê koan-tiám. *Unicode* sui-bóng khòa"--khí-lâi sī kóng "thóng-it-bé", m̄-kú i ê siat-kè sī beh hō jiòk-sè gí-giân mā ē-tàng kap kiông-sè gí-giân ū káng-khóan ê tē-ûi, lóng ē-tàng ēng. Tâi-ôan cheng-hú tēng ê phiau-chún tian-tò-péng: "Lí nā ē-sái ēng chit thò, pát-hāng lóng bē-sái ēng--tit."

4. *Unicode* ê hiān-hóng

Windows NT/2000 lóng sī ēng *Unicode* lâi siat-kè..ê.

Word 2000 mā sī. (*Word 97* tùi *Unicode* ê chi-ôan bô ôan-chéng.)

HTML 4.0 ê phiau-chún kui-tēng ēng *Unicode*. Chit-má IE 5.x, *Netscape 6.x* lóng sī ēng *Unicode*.

Java tit-chiap ēng *Unicode*.

Kî-tha chi-ôan *Unicode* ê íng-thé iáu chiâ" chē (ē jú-lâi jú-chē). M̄-kú bák-chēng tōa pō-hūn ê íng-thé bô chi-ôan *Unicode*.

5. *Unicode* ê pian-bé

Unicode ēng 16 bit lâi piáu-sī 1 jī (*character*). Só-í ē-tàng piáu-sī 65536 jī ($2^{16}=65536$). M̄-kú i koh siat-kè ēng D800-DBFF(1024 jī) kap DC00-DFFF

(1024 jī) kap DC00-DFFF(1024 jī) 2 jī lâi piáu-sī chin-chià" ê 1 jī, só-í koh ke-chhut 1,048,576 jī (1024x1024=1,048,576). Lóng-chóng ē-tàng piáu-sī 1,114,112 jī (65536 + 1048576=1,114,112) (0-1,100,000). 0-127 ê pō-hūn hām ASCII sio-siāng.

Chin-chià" tiān-chú chu-liāu ê piáu-sī hong-hoat ū UTF-16, UTF-16LE, UTF-16BE, UTF-8, téng-téng. Siōng chiáp ēng..ê sī UTF-8. UTF-8 ē-sái kap tōa pō-hūn kū ê (chi-ôan ASCII..ê) ñng-thé sio-thong.

Unicode pau-koat tōa pō-hūn hiān-iú ê pian-bé phiau-chún. Chiū-sī kóng lí ē-tàng kā hiān-iú bô käng jí-giân ê chu-liāu lóng chóan-chòe *Unicode*, chiah-ê chu-liāu tō ē-tàng kau-ōa", bô khi chhiong--tiòh--ah.

Unicode ê pian-bé ū i ka-kī ê kui-chek, bô it-téng hām lán phó-thong kóng ê 1 ê jī sio-siāng. Phì-jû kóng 1 ê Arabic ê jī (grapheme) khó-lêng ài 4 ê *Unicode* ê jī lâi piáu-sī. Phì-jû kóng "á" ē-sái ēng *Unicode* 00E1 1 ê jī lâi piáu-sī (precomposed character), mā ē-sái ēng *Unicode* "a" kap *Unicode* 0301 2 ê jī lâi piáu-sī (combining character sequence).

6. *Unicode* kap ISO 10646

Kán-tan kóng 2 ê ē-sái kóng sī käng-khóa"..ê.

Pian-bé hong-bīn, *Unicode* ùi V1.1 í-āu tō lóng kap ISO 10646 í-chhi it-tì.

M̄-kú, *Unicode* ū ke 1-kóa jī ê tēng-gī. Só-í ē-sái kóng hū-háp ISO 10646..ê tō ē hū-háp *Unicode*.

7. *Unicode* kap jī-hêng

Unicode chhú-lí pian-bé kap jī ê tēng-gī, jī-hêng chhú-lí beh chái"-iū" hián-sī, che sī 2 chān tāi-chì. Bô käng ê hē-thóng sú-iōng bô käng ê hián-sī hong-hoat.

Beh hián-sī *Unicode* sió-khóa mâ-hoan. Thong-siōng sú-iōng .OTF ê jī-hêng tóng-àn. Nng-thé hong-bīn, *Microsoft* ū thê-kiong 1 ê *Uniscribe* hō ñng-thé siat-kè-chiá hong-piān sú-iōng. Tī *Mac* thia" kóng ū 1 ê ATSUI. Lēng-gōa koh ū 1 ê *Freetype* thia" kóng *Windows/Mac/Linux* lóng thong-iōng.

Góa siat-kè ê jī-hêng sī .TTF ê tóng-àn nā-tiā", iáu bô hoat-tō chi-ôan *Unicode*.

8. *Unicode* kap La-teng-jī

Lán it-poa" tiā" khòa"--tiòh--ê ê La-teng-jī, *Unicode* lóng ū chi-ôan. Phì-jû kóng Basic Latin (0020-007F: Eng-gí), Latin-1 (00A0-00FF: Hoat-gí, Tek-gí téng), Latin Extended-A (0100-017F: Latvian, Romanian, Polish, Lithuanian, Croatian, Esperanto, Maltese, Irish, Czech téng),

Latin Extended-B (0180-0233: Zulu, Pan-Nigerian, Zhuang, Acrican, Ewe, Pinyin téng),
 Latin Extended Additional (1EF0-1EF9: Livonian, Vietnamese téng).

9. *Unicode* kap Pêh-ôe-jī (POJ)

Ē-kha chit ê pió chiū-sī *Unicode* í-keng tit-chiap tēng-gī ê jī (precomposed).

Á	á	À	à	Â	â	Ã	ã	Ä	ä				
Ď	ď	Đ	đ										
É	é	È	è	Ê	ê	Ë	ë						
Í	í	Ì	ì	Î	î	Ï	ï	Ĭ	ĭ				
Ĺ	ĺ												
Ŋ	ŋ												
Ń	ń	Ň	ň	ⁿ									
Ó	ó	Ò	ò	Ô	ô	Õ	õ						
Ŕ	ŕ												
Ť	ť												
Ƴ	Ƴ	ú	Û	ù	Ũ	û	Ū	ū	Ŭ	ŭ	Ű	ű	Ų

Ē-kha chit ê pió sī *Unicode* bô tit-chiap tēng-gī ê jī, khó-lêng ē-tàng ēng kúi-ê-á *Unicode* ê jī lâi chō-háp. Phì-jû kóng “á” ē-sái ēng *Unicode* 'a' kap *Unicode* 030D lâi chō-háp.

À	á												
È	é												
Ì	í												
Ŋ	ŋ	Ŋ	ŋ	Ŋ	ŋ	Ŋ	ŋ						
Ŋ	ň	Ŋ	ň	Ŋ	ň								
Ò	ó												
Ƴ	Ƴ	ú	Û	ù	Ũ	û	Ū	ū	Ŭ	ŭ	Ű	ű	Ų

Ē-kha chit ê pió sī *Unicode* bô tit-chiap tēng-gī mā bē-tàng ēng kúi-ê-á *Unicode* ê jī lâi chō-háp. Su-iàu sin-chhéng sin ê *Unicode*.

Ō	ơ	Ó	ó	Ò	ò	Ô	ô	Ō	ó
---	---	---	---	---	---	---	---	---	---

Kám ē-sái hiòng *Unicode* chō-chit sin-chhéng hiah-ê su-iàu chō-háp ê jī chiâⁿ-chòe *precomposed* ê jī? Chiàu in ê gôan-chek, nã ē-tàng ēng chō-háp..ê, tô khah bô ài chiap-siū.

10. *Unicode* ê bāng-chí

Unicode chō-chit: <http://www.Unicode.org>

Khai-su: <http://www.egt.ie/standards/la/taioan.html>

Proposal to add Latin characters required by Latinized Taiwanese languages to ISO/IEC 10646: <http://www.taioan.com/Unicode/Unicode.html>

POJ *Unicode*:A-ki:<http://www.taioan.com/Unicode/Unicode.html>

<http://www.taioan.com/Unicode/kihosamp.html>

11. Kiat-lūn

Tâi-oan beh Pún-thó-hòa tiòh-ài sú-iōng 100 gōa tang ê pún-thó bûn-jī, POJ. Im-phiau ēng kok-chè im-phiau IPA (International Phonetic Alphabet) tō ē-sái--tit.

Kok-chè-hòa tiòh-ài kap bô-kāng bûn-jī, bûn-hòa ê kok-ka kau-liū. Che ēng POJ kap IPA siōng sek-háp.

Chu-sin-hòa ēng *Unicode* m̄-nā ē-sái chú-lí POJ, koh ē-sái kap chōan sè-kài kau-ōaⁿ tiān-chú chu-liāu, pún-thó-hòa kiam kok-chè-hòa it-kiam-jī-kò.

Sui-bóng ū-ê POJ *Unicode* bô tit-chiap tēng-gī, m̄-kú chiah-ê ki-sút-sèng ê bûn-tôe lóng ē-tàng kái-koat. Kap tōa pō-hūn ê gí-giân pí--khí--lâi bē khah khùn-lân.



Chhài-chhoah-bīn ê Kò-sū

Tēⁿ Si-chong

A-koan..a chiāng-sī ū khah bô-ēng, hák-hāu ê kong-khò pek-liáu chin ân, khó-chhì tùi chá-sī khó kàu àm koh khó bē-soah, pò-kò chit-phiⁿ koh kòe chit phiⁿ siá lóng bē-liáu, kóaⁿ kah lāng kiông beh khí-siáu. Ták-mī àm-tng chiáh-liáu chiū khuk tī í-á téng chin khò-pún, kóaⁿ pò-kò, it-tit piàⁿ kàu hoan-tiám, toh-á téng iū-gōan koh-sī sè pún chheh thiap tōa pún chheh, chit tui ná soaⁿ. Ták-jit khùn bô kàu kúi tiám-cheng, sim kông khi gèk. Phah-sng sī án-ne, bīn--nih ê thiâu-á ná kām-chúi puh-íⁿ ê má-lêng-chî-gê, seⁿ kah móaⁿ-móaⁿ sī.

A-koan..a khòaⁿ-tiòh hiah-ê thak

bē-liáu ê chheh, koh khòaⁿ-tiòh ka-kī móa-bīn chōan tâu-hoe, put-sī chiò-kiàⁿ chek thiâu-á, jek hún-chhì, chhòng kah kui-ê bīn āng-kòng-kòng, thiâu-á-lāng lōan-lōan poe.

In lāu-bú khòaⁿ i tiāⁿ-tiāⁿ teh siòng hit-ê kiàⁿ, chai-íáⁿ i teh hōan hiah-ê thiâu-á-chí, kā kóng, "Koan--ah, lí mā khah tú-á hó--leh! Chheh to teh thak bē-liáu--ah, koh ū sim-chiáⁿ teh chiò-kiàⁿ siòng-bīn!

A-koan..a in Lāu-bú ê chhui, "Má--ah, m̄-sī góa ài án-ne. Góa tùi 12 hòe sió-hák