Panasonic
ideas for life

Mac Hanin Software

Hanin Dictionary Maker Manaul

Version: 0.02

Date: 2008/2/4

TM	PM	Prepared
Youko Ohiwa	Wu Jia-Kwan	Wu Jia-Kwan

PTL

Panasonic Taiwan Laboratories Co., Ltd. (PTL)

[Change History]

Date	Ver.	Category	Section	Changes	Changed by
2008/02/01	0.01	-	-	First edition	Wu Jia-Kwan
2008/02/04	0.02	D		remove ISM stamp on cover page	Wu Jia-Kwan

Category [A: Added, U: Updated, D: Deleted]

[Table of Contents]

1. Introduction

1.1 Abstract

HDicMaker is a utility to build the binary dictionary using text files of Chinese character and word lists for Hanin input algorithm. *HDicMaker* is a Win32 console application developed using Microsoft Visual C++ v6.0. It can be executed on a console terminal on Windows environment.

This document introduces the usage of *HDicMaker* to demonstrate how to build a Hanin executable dictionary. This document also describes the format of Hanin dictionary both of text file and binary file.

1.2 References

The following documents are inputs of this specification. When changes are made to the following documents, these changes will impact this specification.

```
[1] Liu Guei-Ji, Data Structure Modification for Hanin Dictionary,
    Technical Report PTL8804, Panasonic Taiwan Laboratories Co.,
    Ltd., Jun 21, 1999
[2] Liu Guei-Ji, Tools for Hanin Dictionary Maintenance,
    Technical Report MITT8737, Matsushita Electric Institute of
    Technology (Taipei) Co., Ltd., Mar 4, 1999
[3] Ouyang Yen-I & Lin Chin-Hsuen, Tools for Hanin System,
    Technical Report MTT8638, Matsushita Electric Institute of
    Technology (Taipei) Co., Ltd., Mar 31, 1998
```

2. Usage of HDicMaker

2.1 Build a Hanin Executable Dictionary

Syntax

HDicMaker *<character file> <word file> <target file> <label name>*

where, *<character file>* is a pre-prepared fixed format of character dictionary. Currently, this character dictionary is **char.v4.cmp**.

<word file> is the filename of text format of words

<target file> is the filename of binary dictionary to be created

<label name> is the label string of this binary dictionary

Example

HDicMaker *char.v4.cmp hv5.big HaninDic* 系統

2.2 Build an Aimei Dictionary

Syntax

HDicMaker *<aimei text file> <aimei binary file>*

where, *<aimei text file>* is the text format of aimei zhuyin

<aimei binary file> is the filename of binary aimei dictionary to be created

Example

HDicMaker *aimei.big AimeiDic*

3. Format of Hanin Text Dictionary

In Hanin text dictionary, each line contains zhuyin fuhao, Chinese characters code, and word frequency regarding word.   It can't put any space between these items to be a separator.

Example

ㄙㄨㄥˋㄒㄧㄚˋㄉㄧㄢˋㄑㄧˋ松下電器 **55**

If the first character in a line is '|' (0x7C), this line is regarded as a comment line.   For example:

```
|   漢音 5.0 版詞庫格式說明
|   頻率：
|   2 字詞:        ㄅㄞˊㄅㄢˇ白板 42
|
```

Regarding aimei dictionary, its format is different than other Hanin dictionaries.   In the aimei text file, each line contains a zhuyin fuhao and following zhuyin fuhaos which are aimei-yin of the first zhuyin fuhao.

Example

ㄅㄚ1ㄆㄚ1ㄅㄚ2ㄅㄚ3ㄅㄚ4ㄅㄚ5ㄆㄚ5

It means ㄆㄚ1,ㄅㄚ2,ㄅㄚ3,ㄅㄚ4,ㄅㄚ5,ㄆㄚ5 are aimei-yin of ㄅㄚ1.

4. Format of Hanin Binary Dictionary

The architecture of the Hanin binary dictionary is shown as below.

0x0000 | IDX1 (size = sizeof(IDX1) * MD_MAX_IDX)

`0x0200`	Hanin Version (12 bytes), Label Name (12 bytes)
`0x0400`	IDX2 (size = sizeof(IDX2) * 3550[1]) [1] The number of records of IDX2 will be increased if the number of words in dictionary is increased.
`0x5C00`	Chinese Character Cache Area
`0xAC00`	Chinese Character & Chinese Word (Chinese character & Index3 area for leading two compress code)