

kmswx32辞書フォーマット解説書

■対応ソフト (kmsdiccat.exeが出力する辞書ファイルを使用)

kmswx32.exe v3.0, v3.1 (kmswx32.bin)

kmswx.exe v3.0, v3.1 (kmswx.bin)

*kmsdiccatはkdiccnvが出力するファイルを統合する

■はじめにお読みください

このフォーマット解説書はkmswx32.binおよびkmswx.binの使用してる辞書の構造を説明しています。辞書エディタやツールを作成される方は参考にしてください。このフォーマットを流用する場合は特に使用許諾を必要としていませんので、使用者の責任において使用してください。

■kdiccnvが生成するbinのフォーマット (v0.00.02 255ver)

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	'D'	'I'	'C'	暗号	文字1の開始位置			文字2の開始位置			文字3の開始位置					
	文字4				5			6			7					
	8				9			10			11					
中略																
+400	252				253			254			255					
	文字データ (最大32bit分:4G Byte - 0x400)															

文字1の開始位置はkmsdiccnvが出力するファイルの先頭からの意味

仮に文字1の開始位置が 0x00000400 (リトルエンディアン) で入っていたら 'DIC' 文字の 'D' から +400 の番地から格納している。暗号は現在使用していないので 0 を入れる (辞書の暗号化を使用する場合はこの数値で種類を判別する※kmswx32v3.0, v3.1では未使用)

各文字データの始めには開始記号と文字数およびブロック数が入っている。

例えば、2文字の単語が10ブロック分(10単語)あるとしたら、(00 11 22 33 44 55 66 77 88 99)

0xff 0x02 0x0a 0x00 という開始記号の後に

'00112233445566778899' が続く。終端記号は無い。

0xff はスタートの記号

0x02 は文字数

0x0a 0x00 は16bitの数値(unsigned, リトルエンディアン)

文字??位置に 0 (32bit)が入っていたらその文字数に相当する単語は無い。

ただし、その文字数の単語が無いだけで次の文字数の単語には位置が入っていることもある。

文字1から文字255までの位置情報を調べて0が入っているところと無いところを検査する。

また、総単語数を求めるには各文字数の位置情報で0以外が入っておりそこから位置を調べ0xff開始記号の次にあるブロック数を加算して求める。

kmsdiccnvが出力するデータを単独で利用できるが、次の項目で説明するkmsdiccatで複数の辞書を連結して使用する場合は主となる。

■kmsdiccatが生成するbinのフォーマット(v0.00)

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	'K'	'M'	'S'	'D'	'I'	'C'	'O'	'O'	'O'	0x1a	0	0	0	0	0	0
	辞書1位置				辞書2				3				4			
	5				6				7				8			
中略																
+ f0	57				58				59				60			
+100	辞書連結データ								辞書1							
中略																
+n	EXOR															

+0から+9までは 'KMSDIC000' , 0x1a

+aから+fまでは 0x00

辞書1位置は +0からの位置 (リトルエンディアン)

+nは全辞書連結後 +0 から +n-1 までを EXORした値(1 Byte)を入れる(辞書破損検出用)

KMSDIC000辞書は最大60まで連結して使用できるフォーマットだがkmswx32v3.0, v3.1では1から8までしか使用していない(1から4まで英文、5から8まで半角カナ文)

+100からは連結の日時および連結した辞書のファイル名を入れる。辞書1はこのあとにつける。念のため辞書の開始はアドレスの16で割り切れるところからにするためおしりの余った空白には0をいれておく(PAD)。kmsdiccatでは以下の連結日時フォーマットとファイル名を入れている。

C:2012/07/26|002130,1@file1.bin;2@file2.bin;3@;4@;5@;6@;7@;8@

上記の例では 2012年7月26日0時21分30秒作成で、辞書1はfile1.bin、辞書2はfile2.bin、辞書3から8までは空という意味になる。8@の後には'¥0'を入れておき、のこりは 0を入れておく。

仮に '8@', '¥0' の¥0がアドレスの 0x12dであったら0x12eと0x12fに 0を入れる。また、'8@', '¥0' の¥0が0x130となったら 0x131から0x13fまで 0を入れ、データは0x140から格納にする。

辞書1や辞書2はkmsdiccnvの出力データをそのままファイルのおしりに追加していただくだけである。また各辞書の終わりが16で割り切れるアドレスで終わるように足りない部分には0を入れて次に連結する辞書はアドレスの+0から結合できるようにしておく。たとえば辞書1が0x805で終了した場合は0x806から0x80fまで0を入れ、辞書2は0x810から結合する。

実際の構造はバイナリエディタでkmswx32.binまたはkmswx.binを開いてみると確認できる。

(c)2012, 2013, 2021 wnc develop

Web頁(wnc develop 記録)

<https://wncdevelop.livedoor.blog/>