

組み込み向け かな漢字変換モジュール
Embedded ASK Ver. 3.1 機能仕様書

株式会社 アクセス
ACCESS®

本操作仕様書におけるすべての文書、図面、プログラムなどの著作権は株式会社アクセスが所有します。

本操作仕様書の全部又は一部を株式会社アクセスの許諾なく複製し、頒布その他の行為を行うことはできません。

また、本操作仕様書の内容・構成を株式会社アクセスの許諾なく改変すること、改変したものを複製し、頒布その他の行為を行うこともできません。

Copyright © 1998 株式会社アクセス All rights reserved.

**組み込み向け かな漢字変換モジュール
Embedded ASK Ver. 3.1 機能仕様書**

平成 10 年 12 月 1 日作成

株式会社アクセス

〒101-0064 東京都千代田区神田神保町 1-64
神保町協和ビル 7F

目次

1	はじめに.....	1
2	ASK の機能仕様.....	2
2.1	メモリサイズ	2
2.2	変換機能.....	2
3	ASK 辞書	4
3.1	辞書サイズ.....	4
3.2	品詞情報.....	4
3.3	接続情報.....	5
3.4	辞書ユーティリティ.....	5
4	簡易型かな漢字変換 API ライブラリ	6
4.1	初期化.....	6
4.2	かな漢字変換	6
4.3	次候補.....	7
4.4	前候補.....	7
4.5	候補取得仕様.....	8
付録	ASK のローマ字変換対応表	9

1 はじめに

情報家電・組み込み向けのコンパクトなかな漢字変換エンジン ASK では、パソコン FEP 並みの高度な連文節の変換機能を提供します。Embedded ASK Ver. 3.1 では、組み込み用途に適した以下のような機能をもっています。

- (1) コンパクトサイズ
- (2) 高度な変換機能
- (3) ROM 化可能
- (4) アプリケーションから利用するための API
- (5) 辞書ユーティリティで辞書をカスタマイズ

ASK の変換アルゴリズムは、連文節最長一致をベースに、品詞、活用、接頭語・接尾語などの変換効率の良い接続ルールを抽出して一般化しています。変換エンジンでは、候補となる連続文節間の接続性を評価し、接続ポイントの最も高い候補を選択する処理を行っています。Embedded ASK のエンジンはライブラリとして提供されるので、ソフトウェア・キーボードから利用したり、インライン入力から利用するといったことが可能です。

アクセスの情報家電向けインターネットソフトウェア NetFront ブラウザ、メール、Java 仕様準拠モジュール JV-Lite などとともに、すでに多数製品化されています。搭載製品例としては、インターネットTV、セットトップボックス、ゲーム機、PDA、イントラネット端末、インターネット FAX、WebPhone などがあります。

2 ASK の機能仕様

2.1 メモリサイズ

Embedded ASK は、メモリ容量の厳しい環境の組み込み機器向けに、省メモリで動作する設計になっています。以下に、メモリ要求の概要は以下の通りです。

コード	約28kバイト	
データ	辞書	約600kバイト(ラージ)
		約330kバイト(スモール)
	初期化あり変数	約 6kバイト
	初期化なし変数	約56kバイト

※ コードは、CPU、コンパイラによって変わります。
上記は、日立SHプロセッサの場合の数字です。

2.2 変換機能

ASK のかな漢字変換エンジンは、連文節最長一致をベースに、各種接続情報を加味した評価方式をとっています。変換ルールの効率的な一般ルールを抽出することより、コンパクトなソフトウェアでありながら、高変換効率を実現しています。

例:

きょうはよいてんきです。
→ 今日は | よい | 天気です。

かんじのこうほからせんたくしてください。
→ 漢字の | 候補から | 選択して | ください。

※変換結果文字列の | は文節の切れ目を意味しています。

活用語の活用形とそれにつづく単語の品詞の接続性はもちろんのこと、各助詞の接続性や接頭語・接尾語の接続性を考慮した評価になっています。以下では、ASK 変換エンジンの変換ルールの例を示します。

- 例1 接頭語と次の単語との相性
- | | |
|---------|------------|
| 第 1の | (だい1の) |
| 大 感謝祭 | (だいかんしゃさい) |
| 最 有力 | (さいゆうりよく) |
| 再 変換 | (さいへんかん) |

- 例2 接尾語と前の単語との相性
- | | |
|-----------------|----------|
| 100 円 (100えん) | |
| 林檎 園 | (りんごえん) |
| 独占 権 | (どくせんけん) |
| 12 件 | (12けん) |
| 横須賀 港 | (よこすかこう) |
| 11 項 | (11こう) |
| 信長 公 | (のぶながこう) |

	33 冊	(33さつ)
	千円 札	(せんえんさつ)
例3	体言修飾と次の単語の相性	
	重い 荷物	(おもいにもつ)
	思い 願う	(おもいねがう)
	広い 家	(ひろいいえ)
	広井 家	(ひろいけ)
例4	用言修飾と次の単語の相性	
	深く 感動する	(ふかくかんだうする)
	不覚 不惑	(ふかくふわく)
	鞠を 蹴る	(まりをける)
	真利 ちゃん	(まりちゃん)
例5	助詞と次の単語の相性	
	味噌の 汁	(みそのしる)
	味噌を する	(みそをしる)
例6	文頭の助詞を判別	
	と 呼ぶと	(とよぶ)
	へ 向かった	(へむかった)
例7	例7 括弧などのうしろの助詞の判別	
	「 はい 」と 答えた	(「はい」とこたえた)
	(いいえ)の 返事	((いいえ)のへんじ)
例8	例8 英文字列をサ変名詞扱い	
	NTTの 株	(NTTのかぶ)
	PACKする 袋	(PACKするふくろ)

※以上の文例の変換は、ラージ辞書を使用した場合です。

3 ASK 辞書

3.1 辞書サイズ

かな漢字変換モジュール ASK 辞書では、標準辞書として、以下のようにラージ版とスモール版を用意しています。

■ラージ辞書

約5万語

約600kバイト

■スモール辞書

約3万語

約330kバイト

3.2 品詞情報

ASK では、単語に以下のような品詞情報をつけています。新規に単語を辞書に登録する場合には、以下の品詞情報を付けて登録することが必要になります。

品詞番号

1. 動詞カ行五段(飽・く)
2. 動詞ガ行五段(泳・ぐ)
3. 動詞サ行五段(愛・す)
4. 動詞タ行五段(打・つ)
5. 動詞ナ行五段(死・ぬ)
6. 動詞バ行五段(忍・ぶ)
7. 動詞マ行五段(読・む)
8. 動詞ラ行五段(走・る)
9. 動詞ワ行五段(使・う)
10. 動詞サ行変格(する)
11. 動詞ザ行変格(信・ずる)
12. 動詞カ行変格(来る)
13. 動詞一段(教え・る、見・る)
14. 形容詞(多・い)
15. 形容詞「げ」「がる」が付く(弱々し・い)
16. 形容詞「げ」「がる」「み」が付く(楽し・い)
17. 形容詞「げ」「がる」「め」が付く(熱・い)
18. 形容詞「げ」「がる」「み」「め」が付く(強・い)
19. 形容動詞(静か・な)
20. 形動名詞(唐突・な)
21. タルト名詞(延々・たる)
22. サ変名詞(補佐・する)
23. ザ変名詞(案・ずる)
24. 名詞(車)
25. 拡張名詞(ユーザ登録では使わない)
26. 地名(東京)
27. 人名(姓)(鈴木)
28. 人名(名)(太郎)
29. 団体名(自民党)
30. 物の名称(その他の固有名詞)
31. 数字・数表現(一)
32. 感動詞(あれ)

33. 接続詞(しかし)
34. 副詞(もはや)
35. 副詞「と」「する」が付く(はらはら・と)
36. 副詞「に」が付く(じき・に)
37. 副詞「の」が付く(もっぱら・の)
38. 連体詞(この)
39. 慣用句・挨拶表現(ありがとう)
40. 単漢字
41. 接頭語
42. 接尾語(名詞)
43. 接尾語(サ変名詞) 化、切り
44. 接尾語(形動名詞) 的、質、並み
45. 接尾語(形容詞) がまし・い、がたな・い
46. 接尾語(カ行五段動詞) め・く、づ・く
47. 接尾語(ラ行五段動詞) やが・る、がが・る、ば・る
48. 接尾語(数字・数表現) 十、百、千

3.3 接続情報

ASK では、接頭語(45)、接尾語(50～56)については、接続情報も単語登録時に指定する必要があります。以下の接続情報のうちの1つを指定します。

- | | | |
|---|------------------------|--------------|
| 0 | …名詞(20～25、50)に付く | |
| 1 | …地名(26)に付く | 「北」「橋」「口」 |
| 2 | …人名(27、28)に付く | 「故」「様」「さん」 |
| 3 | …数字・数表現(31、56)に付く | 「第」「約」「円」「匹」 |
| 4 | …サ変名詞(22)に付く | 「未」「再」 |
| 5 | …動詞の連用形に付く | 「付け・る」 |
| 6 | …数字・数表現、名詞に付く | |
| 7 | …団体名(29)に付く | 「銀行」「大学」 |
| 8 | …接頭語「お」だけ(ユーザ登録では使わない) | |

品詞情報や接続情報は、漢字文変換処理で利用されるので、適切な情報を登録することが必要です。

3.4 辞書ユーティリティ

ASK 辞書に単語登録あるいは削除などの編集を行うためのツールとして、Windows 上の辞書ユーティリティ(ASKDICM.EXE)が用意されています。

辞書ユーティリティの機能としては、以下のものがあります。

- ・ テキスト→辞書バイナリ変換
- ・ 辞書バイナリ→テキスト変換
- ・ 新規単語登録
- ・ 単語の編集
- ・ 単語の削除
- ・ 辞書のマージ
- ・ 辞書の再編集(空き領域の設定)

4 簡易型かな漢字変換 API ライブラリ

Embedded ASK では、簡単にアプリケーションが利用できる API として、単文節変換の簡易型 API を用意しています。以下では、簡易型 API の関数仕様について説明します。

4.1 初期化

【関数書式】

```
int      comkan_init(void)
```

【引数】

なし

【戻り値】

エラーコード

=0 : 正常

<0 : エラー

【解説】

- かな漢字変換モジュールの初期化を行う。
 - ・ROM 辞書のオープン
 - ・ヘッダ読み込み
 - ・インデックス読み込み
 - ・各種検索情報の初期化
- かな漢字変換モジュールの他の関数を使用する前に、この関数を一度呼んでおく必要がある。
(各関数の呼出し毎に必要ではない。)
- 対応する終了処理は必要としない。

4.2 かな漢字変換

【関数書式】

```
int      comkan_henkan(
                unsigned char  *outbuf,
                int             outsize,
                unsigned char  *inbuf,
                int             insize,
                int             direction
            )
```

【引数】

outbuf 変換結果の全角ひらがな変換結果の全角ひらがな文字列の最初の候補が格納される。

outsize outbuf の領域のバイト数。

inbuf 変換もとの文字列が格納される。

insize 変換もとの文字列のバイト数。

direction 0: 最初の変換候補を outbuf に格納

1: 最後の変換候補を outbuf に格納

【戻り値】

> : 正常 (候補数)

=0 : 候補なし

<0 : エラー

【解説】

- かな漢字変換を行う。
- 全角ひらがなをかな漢字変換する。
- 候補には、漢字変換済み候補、ひらがなの読み、カタカナの読みの3種類があり、漢字→ひらがな→カタカナの順に並んでいる。
例) にほん(読み)
1) 日本 2) 二本 3) にほん 4) ニホン
- 変換できないものは、そのままコピーされる。
この時の候補数は1が返る。
- 正常に変換できた場合、返り値として候補数が返る。
- **outbuf** には、1候補格納される。
- **outsize** \geq **insize** でなければならない。
- かな漢字変換された候補は、次回 **comkan_init()**, **comkan_henkan()** が呼ばれるまで保持される。

4.3 次候補

【関数書式】

```
int comkan_nextkouho(
    unsigned char *outbuf,
    int outsize
)
```

【戻り値】

>0 : 正常(候補番号)
<=0 : エラー

【解説】

- 次候補を取得し、その候補番号が返る。(候補番号とは、1からインクリメントされた番号。)
- 次候補が無い場合には、先頭候補(候補番号1)が返る。従って候補なしはあり得ない。

4.4 前候補

【関数書式】

```
int comkan_prevkouho(
    unsigned char *outbuf,
    int outsize
)
```

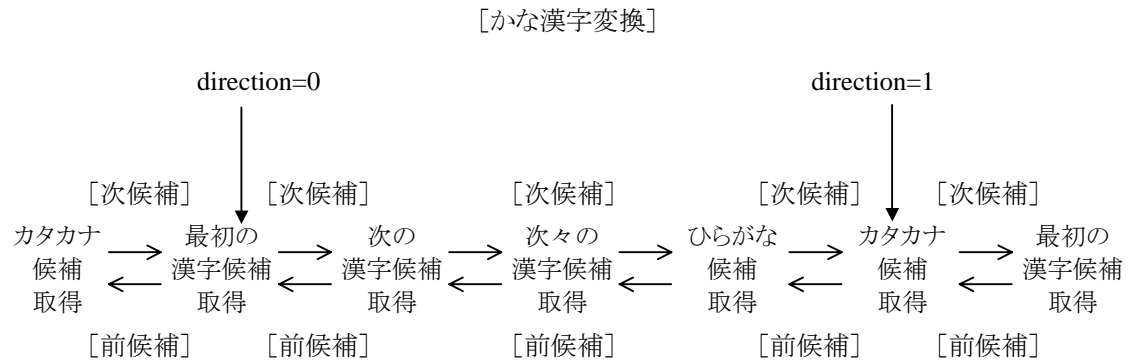
【戻り値】

>0 : 正常(候補番号)
<= : エラー

【解説】

- 前候補を取得し、その候補番号が返る。(候補番号とは、1からインクリメントされた番号。)
- 前候補が無い場合には、最終候補(候補番号最大)が返る。従って候補なしはあり得ない。

4.5 候補取得仕様



付録 ASK のローマ字変換対応表

あ	い	う	え	お
A	I (YI)	U (WU)	E	O
か	き	く	け	こ
KA (CA)	KI	KU (CU) (KWU) (QU)	KE	KO (CO)
さ	し	す	せ	そ
SA	SI (SHI) (CI)	SU	SE (CE)	SO
た	ち	つ	て	と
TA	TI (CHI)	TU (TSU)	TE	TO
な	に	ぬ	ね	の
NA	NI	NU	NE	NO
は	ひ	ふ	へ	ほ
HA	HI	HU (FU)	HE	HO
ま	み	む	め	も
MA	MI	MU	ME	MO
や		ゆ	いえ	よ
YA		YU	YE	YO
ら	り	る	れ	ろ
RA	RI	RU	RE	RO
わ	うい		うえ	を
WA	WI		WE	WO
ん				
Nの次に '(アポストロフィー), NN, Nの次に子音				
が	ぎ	ぐ	げ	ご
GA	GI	GU (GWU)	GE	GO
ざ	じ	ず	ぜ	ぞ
ZA	ZI (JI)	ZU	ZE	ZO
だ	ち	づ	で	ど
DA	DI	DU	DE	DO
ば	び	ぶ	べ	ぼ
BA	BI	BU	BE	BO
ぱ	ぴ	ぷ	ぺ	ぽ
PA	PI	PU	PE	PO
っ				
LTU, LTSU, XTU, XTSU, 子音を重ねて母音				

きゃ	きい	きゅ	きえ	きよ
KYA	KYI	KYU	KYE	KYO
しゃ	しい	しゅ	しえ	しよ
SHA (SYA)	SYI	SHU (SYU)	SHE (SYE)	SHO (SYO)
ちゃ	ちい	ちゅ	ちえ	ちよ
CHA (CYA) (TYA)	CYI (TYI)	CHU (CYU) (TYU)	CHE (CYE) (TYE)	CHO (CYO) (TYO)
にゃ	にい	にゅ	にえ	によ
NYA	NYI	NYU	NYE	NYO
ひゃ	ひい	ひゅ	ひえ	ひよ
HYA	HYI	HYU	HYE	HYO
みゃ	みい	みゅ	みえ	みよ
MYA	MYI	MYU	MYE	MYO
りゃ	りい	りゅ	りえ	りよ
RYA	RYI	RYU	RYE	RYO
ぎゃ	ぎい	ぎゅ	ぎえ	ぎよ
GYA	GYI	GYU	GYE	GYO
じゃ	じい	じゅ	じえ	じよ
JA (JYA) (ZYA)	JYI (ZYI)	JU (JYU) (ZJU)	JE (JYE) (ZYE)	JO (JYO) (ZJO)
ぢゃ	ぢい	ぢゅ	ぢえ	ぢよ
DYA	DYI	DYU	DYE	DYO
びゃ	びい	びゅ	びえ	びよ
BYA	BYI	BYU	BYE	BYO
ぴゃ	ぴい	ぴゅ	ぴえ	ぴよ
PYA	PYI	PYU	PYE	PYO
くわ	くい		くえ	くお
KWA	KWI (QI) (QYI)		KWE (QE) (QYE)	KWO (QO)
くあ				
QA				
くや		くゅ		くよ
QYA		QYU		QYO
つあ	つい		つえ	つお
TSA	TSI		TSE	TSO
ふあ	ふい		ふえ	ふお
FA	FI (FYI)		FE (FYE)	FO

あ LA (XA)	い LI (XI) (LYI) (XYI)	う LU (XU)	え LE (XE) (LYE) (XYE)	お LO (XO)
や LYA (XYA)		ゆ LYU (XYU)		よ LYO (XYO)
わ LWA (XWA)				

ふや FYA		ふゆ FYU		ふよ FYO
ぐわ GWA	ぐい GWI		ぐえ GWE	ぐお GWO
てや THA	てい THI	てゆ THU	てえ THE	てよ THO
でや DHA	でい DHI	でゆ DHU	でえ DHE	でよ DHO
ヴあ VA	ヴい VI	ヴ VU	ヴえ VE	ヴお VO

- * NetFront、AVE は、株式会社アクセスの日本国における登録商標です。
- * JV-Lite は、株式会社アクセスの商標です。
- * WAVE(Window-based Abstract Virtual Environment／簡易ウインドウマネージャ)は、株式会社アクセスの商標です。
- * Java 及びすべての Java 関連の商標及びロゴは、米国及びその他の国における米国 Sun Microsystems Inc.の商標又は登録商標です。
- * Windows は、米国 Microsoft Corporation の米国及びその他の国における商標又は登録商標です。
- * Netscape は、米国 Netscape Communications Corporation の米国及びその他の国における商標又は登録商標です。
- * 文中の会社名、商品名は各社の商標又は登録商標です。