第 6 章

ライブラリ解説

本章では，システム辞書検索用のライブラリについて，その仕様と，検索関数，アクセス関数について説明する．

6.1 概要

辞書検索ライブラリは，検索用関数と辞書レコードアクセス関数とから構成される．詳細な関数の仕様については，6.2 に示す．

検索キーを用いた検索の手順は，以下の通りである．

- 検索キーを用いて，インデックスパートを二分探索し，該当する最初のエントリと最後のエントリを見つける．
- 各エントリの項目番号から固定長辞書パートでのレコードのオフセットを求め，固定長レコードにアクセスする．
- レコードが検索条件に合致していれば，検索結果として保存し，相違していれば，棄却する．保存の際，可変長フィールドがある場合には，オフセットから可変長パートにアクセスし，情報を得る．

検索モードとしては，「完全一致」「部分一致」「前方一致」の3種類を提供する．各々の検索モードと検索結果は，以下の通りである．

検索キー	検索モード	検索結果
東京	完全一致	「東京」のみ
	部分一致	「東」「東京」
	前方一致	「東京」「東京都」「東京音頭」など

ただし，概念識別子など，検索キーが可変長でないもの場合には，検索モードとして完全一致のみが有効であり，その他のモードが指定された場合にも，完全一致により検索を行う．

検索結果は，以下に示す辞書レコードのリストとなる．リストの末端では，nextPtrには NULL が格納される．

```
typedef struct EDRecList {
        struct EDRecList    *nextPtr;     //次のレコードへのポインタ
        unsigned int        dicType :4;   //辞書レコードの種類
        unsigned int        selected:1;   //選択済みフラグ
        unsigned int        marked  :1;   //マーク付きフラグ
        unsigned char       *dictRec;     //辞書レコードへのポインタ
} EDRecList;
```

※ 選択済みフラグ，マーク付きフラグは，EDBroW でのみ利用されるものであり，通常は必要ない．

辞書ファイルのオープン / クローズ

各辞書ファイルは構造を持っている．このため，辞書構造体（構造体名 DictPtr ）を用意し，辞書固有の情報をここに格納するとともに，検索のための辞書指定には，この辞書構造体を利用する．

辞書構造体に格納されるデータは，システム辞書の先頭に格納されているものを読み込んだものである．また，辞書構造体にデータをセットするための，辞書ファイルオープン関数，および辞書クローズ関数が，辞書検索ライブラリの一部として提供される．

複数キー / フィルタキー指定による検索

インデックスに対応する複数のキーと，インデックスには対応しないが，フィルタとして利用できるキーとを与え，必要な情報のみを取り出す関数を用意する．この関数は，どれか一つのキーを利用して辞書を検索し，結果をその他のキーによりフィルタリングすることにより実現される．複数キー指定の場合の，検索モードとフィルタリング方法の対応は以下の通りである．

キーの種類	フィルタリング方法
インデックスに対応する検索キー	
可変長キーの場合	与えられた検索モードに依存
固定長キーの場合	完全一致するもののみ
インデックスに対応しないフィルタキー	完全一致するもののみ

いくつかの特殊な検索は，検索モードと複数 / フィルタキーの組み合わせにより実現できる．

例えば，単語辞書検索に利用される「語尾付き単語見出し検索」は，単語見出しによる部分一致検索に，「語尾付き単語」をフィルタキーとして与えることにより実現可能である．

フィルタキーとして利用できるのは，以下の１０種類である

FLT_LANG	0	言語識別（単語 / 対訳 / 共起）
FLT_KIND	1	レコード種別（単語 / 対訳）
FLT_POS	2	品詞（単語 / 対訳）
FLT_CRARC	3	関係子（概念記述）
FLT_CRTF	4	真偽値（概念記述）
FLT_CRTYPE	5	記述区分（概念記述）
FLT_WDHDW	6	語尾付き見出し（単語）
FLT_CCARC	7	共起関係子（共起）
FLT_CCCID1	8	受け側概念識別子（共起）
FLT_CCCID2	9	係り側概念識別子（共起）

なお，フィルタキーは，以下に示すフィルタキー構造体（フィルタキー作成関数により作成される）を，複数ある場合にはリストにして関数に渡す．リストの末端では，nextKeyに NULL を格納する．

```
typedef     struct FilterKey {
        struct FilterKey *nextKey; //次のフィルタキーへのポインタ
        int     filterType;        //フィルタキー種別
        char    *filterKey;        //フィルタキー
} FilterKey;
```

6.2 関数一覧

6.2.1 辞書のオープン / クローズ用関数

DictPtr *DictOpen (char *dictName)

機能： システム辞書をオープンして，辞書構造体を作成する．
引数： dictName システム辞書ファイルのパス名
返値： 正常 ... 辞書構造体の先頭アドレス
異常 ... NULL

void DictClose (DictPtr *dictPtr)

機能： システム辞書をクローズする．
引数： dictPtr 辞書構造体の先頭アドレス
返値： なし

6.2.2 辞書検索関数

EDRecList *EDsearch (DictPtr *dictPtr, unsigned char *keyPtr, int idxNum, int srchMode)

機能： システム辞書を検索する
引数： dictPtr 辞書構造体の先頭アドレス
keyPtr 検索キー
idxNum インデックス番号
srchMode 検索モード
返値： 正常 ... 辞書レコード構造体の先頭アドレス
異常 ... NULL（検索キーを含むレコードが無かった場合を含む）
備考： 概念識別子の場合には，unsigned long の概念識別子へのポインタを，unsigned char にキャストして検索キーとして渡す．

※インデックス番号，検索モードは，以下の数値により指定する．

インデックス番号

IDX_1	0	単語見出し（単語・対訳），概念識別子 1（概念記述・概念体系），受け側単語見出し（共起），概念識別子（概念見出し）
IDX_2	1	読み（単語），概念識別子（対訳），係り側単語見出し（共起），概念識別子 2（概念記述，概念体系）
IDX_3	2	概念識別子（単語）

検索モード

SRCH_FULL	0	完全一致
SRCH_PART	1	部分一致
SRCH_HEAD	2	前方一致

EDRecList *EDmltKeySearch (DictPtr *dictPtr, char *keyPtr1, char *keyPtr2, char *keyPtr3, FilterKey *fltKeyPtr, int srchMode)

機能： システム辞書を複数キーで検索する
引数： dictPtr 辞書構造体の先頭アドレス
keyPtr1~3 検索キー（インデックス #1 ~ #3に対応）
fltKeyPtr フィルタキーリストの先頭アドレス
srchMode 検索モード

返値： 正常 ... 辞書レコード構造体の先頭アドレス
異常 ... NULL（検索キーを含むレコードが無かった場合を含む）

備考： 概念識別子の場合には，unsigned longの概念識別子へのポインタを，unsigned charにキャストして検索キーとして渡す．
指定する必要のない検索キーには，長さ 0 の文字列を渡す．
フィルタキーは，makeFltKey()で作成されたものを，複数ある場合はリストにして渡す．また，フィルタキーの指定が必要ない場合には，NULL を指定する．

6.3.3 フィルタキー用関数

FilterKey *makeFltKey (int filterType, char *filterKey)

機能： フィルタキーを作成する

引数： filterType フィルタキーの種類
filterKey フィルタキー文字列
※真偽値，言語識別など 0/1の値を持つフィールドをフィルタキーとする場合には，"0"，"1" の文字列を渡す

返値： 正常 ... フィルタキーの先頭アドレス
異常 ... NULL

void *freeFltKey (FilterKey *filterKey)

機能： フィルタキー格納領域を解放する

引数： filterKey フィルタキー構造体リストの先頭アドレス

返値： なし

6.3.4 辞書レコードアクセス関数

辞書レコードアクセス関数は，そのほとんどがほぼ同様の方法で利用できる．特に断わらない限り，以下の手順を基本とする

- アクセス関数には，辞書レコード構造体の，レコード本体の先頭アドレスを引数として与える（以下，変数名 recPtrで統一）
- アクセス異常の場合 -1 を返す
- 固定長の情報（概念識別子，各種フラグなど）はレコード本体の先頭アドレスのみを引数として取り，関数の返り値として，データの値を返す．
- 可変長の情報の場合，関数の第 2 引数に，可変長文字列を格納する領域の先頭アドレスを渡し，関数は可変長文字列の長さを返す．ただし，文字列を格納する領域は，呼び出し側で確保する．

概念辞書アクセス関数

unsigned long CIgetCID (unsigned char *recPtr)

機能： 概念見出しレコードの概念識別子を返す

備考： エラー時は 0 を返す

int CIgetECExpl (unsigned char *recPtr, char *strPtr)

機能： 概念見出しレコードの英語概念説明を得る

int CIgetECWord (unsigned char *recPtr, char *strPtr)
機能： 概念見出しレコードの英語概念見出しを得る

int CIgetJCExpl (unsigned char *recPtr, char *strPtr)
機能： 概念見出しレコードの日本語概念説明を得る

int CIgetJCWord (unsigned char *recPtr, char *strPtr)
機能： 概念見出しレコードの日本語概念見出しを得る

unsigned long CDgetCID1 (unsigned char *recPtr)
機能： 概念体系・記述レコードの概念識別子を返す
備考： エラー時は 0 を返す

unsigned long CDgetCID2 (unsigned char *recPtr)
機能： 概念体系・記述レコードの概念識別子を返す
備考： エラー時は 0 を返す

int CDgetArc (unsigned char *recPtr, char *relPtr)
機能： 概念記述レコードの関係子を得る

int CDgetTFVal (unsigned char *recPtr)
機能： 概念記述レコードの真偽値を返す

int CDgetType (unsigned char *recPtr)
機能： 概念記述レコードの記述区分を返す
備考： I 記述の場合 1 を，E記述の場合 0 を返す

単語辞書アクセス関数

unsigned long WDgetCID (unsigned char *recPtr)
機能： 単語辞書レコードの概念識別子を返す
備考： エラー時は 0 を返す

int WDgetHWord (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの単語見出しを得る

int WDgetRead (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの読みを得る

int WDgetLAdj (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの左連接情報を得る

int WDgetRAdj (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの右連接情報を得る

int WDgetLang (unsigned char *recPtr)
機能： 単語辞書レコードの言語識別を返す
備考： 日本語の場合 0 を，英語の場合 1 を返す

int WDgetIdmFlg (unsigned char *recPtr)
機能： 単語辞書レコードの慣用句フラグを返す
備考： 慣用句の場合 1 を，そうでない場合 0 を返す

int WDgetTechFlg (unsigned char *recPtr)
機能： 単語辞書レコードのレコード種別を返す
備考： 専門用語の場合 1 を，そうでない場合 0 を返す

int WDgetPOS (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの品詞を得る

int WDgetGram (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの文法情報を得る

int WDgetPron (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの発音情報 / 構成語情報を得る
備考： 構成語情報 / 発音情報は格納しているフィールドが同じであるため，
アクセスには同じ関数を利用する．

int WDgetSTree (unsigned char *recPtr, char *strPtr)
機能： 単語辞書レコードの構文木を得る

int WDgetWEnd (unsigned char *recPtr, char *wendPtr)
機能： 単語辞書レコードの単語見出し語尾を得る

int WDgetREnd (unsigned char *recPtr, char *wendPtr)
機能： 単語辞書レコードの読み語尾を得る
備考： 日本語単語辞書の場合のみ

対訳辞書アクセス関数

unsigned long BIgetCID (unsigned char *recPtr)
機能： 対訳辞書レコードの概念識別子を返す
備考： エラー時は 0 を返す

int BIgetHWord (unsigned char *recPtr, char *strPtr)
機能： 対訳辞書レコードの単語見出しを得る

int BIgetRead (unsigned char *recPtr, char *strPtr)
機能： 対訳辞書レコードの読みを得る

int BIgetLang (unsigned char *recPtr)
機能： 対訳辞書レコードの言語対識別を返す
備考： 日英の場合 0 を，英日の場合 1 を返す

int BIgetTechFlg (unsigned char *recPtr)
機能： 対訳辞書レコードのレコード種別を返す
備考： 専門用語の場合 1 を，そうでない場合 0 を返す

int BIgetPOS (unsigned char *recPtr, char *strPtr)
機能： 対訳辞書レコードの品詞を得る

int BIgetTrans (unsigned char *recPtr, char *strPtr)
機能： 対訳辞書レコードの対訳情報を得る

共起辞書アクセス関数

int CCgetCCHWord (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの句見出しを得る
備考： 句見出しは，単語見出しと関係単語表現から合成する

int CCgetHWord1 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの受け側単語見出しを得る

int CCgetHWord2 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの係り側単語見出しを得る

int CCgetHwInfo1 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの受け側表記情報を得る

int CCgetHwInfo2 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの係り側表記情報を得る

int CCgetCCArc (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの共起関係子を得る

int CCgetArcHWord (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの関係単語表記を得る

int CCgetText (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの例文を得る

unsigned long CCgetCID1 (unsigned char *recPtr)
機能： 共起辞書レコードの受け側概念識別子を返す
備考： エラー時は 0 を返す

unsigned long CCgetCID2 (unsigned char *recPtr)
機能： 共起辞書レコードの係り側概念識別子を返す
備考： エラー時は 0 を返す

int CCgetArc (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの概念関係子を得る

int CCgetPOS1 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの受け側品詞を得る

int CCgetPOS2 (unsigned char *recPtr, char *strPtr)
機能： 共起辞書レコードの係り側品詞を得る

int CCgetLang (unsigned char *recPtr)
機能： 共起辞書レコードの言語識別を返す
備考： 日本語の場合 0 を，英語の場合 1 を返す

int CCgetHwOrder (unsigned char *recPtr)
機能： 共起辞書レコードの表記順を返す
備考： 係り側→受け側の順の時 0 を，逆の場合 1 を返す

int CCgetArcDir (unsigned char *recPtr)
機能： 共起辞書レコードの関係子方向を返す
備考： 受け側 = 始点側の時 0 を，そうでない場合 1 を返す

int CCgetIdmFlg1 (unsigned char *recPtr)
機能： 共起辞書レコードの受け側慣用句フラグを返す

int CCgetIdmFlg2 (unsigned char *recPtr)
機能： 共起辞書レコードの係り側慣用句フラグを返す

int CCgetSfFreq (unsigned char *recPtr)
機能： 共起辞書レコードの表層共起頻度を返す

int CCgetDpFreq (unsigned char *recPtr)
機能： 共起辞書レコードの共起項目頻度を返す

int CCgetHwFreq1 (unsigned char *recPtr)
機能： 共起辞書レコードの受け側共起要素頻度を返す

int CCgetHwFreq2 (unsigned char *recPtr)
機能： 共起辞書レコードの受け側共起要素頻度を返す