NAME

mkw – freq 形式のファイルから WAM を作成する

SYNOPSIS

mkw *handle* [cvecbufsz=#] [hash-table-options] *[@]freqfile* [compression_type]

DESCRIPTION

mkw は *freqfile* から libwam(3) でアクセスできる形式の, 行名と列名の辞書 (CW) および圧縮された行列ファイル (XR)を作成する. その結果 *freqfile* から, 行の name を格納している行 CW と, 列の name を格納している列 CW , さらに *freqfile* を行方向に圧縮した行 XR と, 列方向に圧縮した列 XR の2つが作成される. *handle* には WAM のハンドルを渡す. **mkw** は, このハンドルから , wam_assoc_fname(3) により , 作成する CW および XR のパス名を決定する. *freqfile* の前に @ があれば, *freqfile* の各行が読むべきファイルのパス名であるとして扱われる. WAM の圧縮タイプは *compression_type* で指定することができる. 現在使用可能なのは NONE, RAW, NIB である. (デフォルト値は NIB)

hash-table-options で名前表作成時に使用するハッシュ表のパラメタおよび, 初期化用の単語辞書を指定することができる. 大抵の場合, デフォルト値で十分であるが, メモリ不足などが発生する場合は修正の必要がある. 詳細は mkw.c にコメントがあるのでそちらに当たられたい.

mkw は逆引き表を作成する際に作業用として FRAGSZDFL だけのメモリを確保する. この値は実行時引き数 cvecbufsz=# によりオーバーライドすることができる. なお, FRAGSZDFL は lib/wam/mkw.c で宣言されている. 確保されるメモリのサイズは **mkw** の実行時間に影響し, なるべく大きいほど良い (効果が期待できる最大は WAM をメモリに読み込むことができるだけのサイズである.) しかしながら, この値を大きくしすぎるとメモリの確保に失敗し, **mkw** が異常終了することになる. このメモリ確保は **mkw** が全処理を 2/3 ほど終えた所で行われるため, 失敗した場合のペナルティは大きい.

SEE ALSO

maw(1), freqfile(5)