

mendex：索引整形ツール

アスキー・メディアワークス

日本語 T_EX 開発コミュニティ

2016 年 11 月 12 日

概要

mendex は文書の索引を作成するコマンドラインツールです。L^AT_EX により抽出された索引リストファイル (.idx) を並べ替え、実際の索引のソースファイルの形に整形します。makeindex と互換性があり、さらに「読み」の扱いの手間を減らすように特殊化されています。出力される索引の形式は、スタイルファイルに従って決定されます。また、辞書ファイルを与えることにより、索引中の漢字の読みが登録されます。索引の階層は 3 段階まで作成することができます。

1 使用法

はじめに mendex の使用法を示します。

```
mendex [-ilqrcgfEJSU] [-s <sty>] [-d <dic>] [-o <ind>] [-t <log>] [-p <no>] [-I <enc>]  
[--] [<idx0> <idx1> <idx2> ...]
```

2 オプション

mendex で利用可能なオプションは以下の通りです。

- i 索引リストファイルが指定されている場合でも、標準入力を索引リストとして使用します。
- l 索引のソートを文字順で行います。指定されなければ単語順のソートになります（ソート方法については後述）。
- q 静粛モードです。エラーおよび警告以外は標準エラー出力に出力しません。
- r ページ範囲表現を無効にします。指定しないと、連続して出てくる索引については“1–5”のようにページ範囲で表現されます。
- c スペースやタブといったブランクを短縮して、すべて 1 個の半角スペースにします。また、前後のブランクは削除されます。
- g 日本語の頭文字の区切りを「あかさた……わ」にします。指定しないと「あいうえ……わをん」になります。
- f 辞書ファイルにない漢字も強制的に出力するモードです。
- s <sty> ファイル <sty> をスタイルファイルと見なします。スタイルファイルを指定しなければ、デフォルトの索引形式で作成します。
- d <dic> ファイル <dic> を辞書ファイルと見なします。辞書ファイルは日本語の“<漢字> <読み>”のリストで構成されます。

- o *<ind>* ファイル *<ind>* を出力ファイルと見なします。指定がない場合は最初の入力ファイルの拡張子を *.ind* としたもの、入力ファイルが標準入力のみであれば標準出力に出力します。
- t *<log>* ファイルをログファイルと見なします。指定がない場合は最初の入力ファイルの拡張子を *.ilg* としたもの、入力ファイルが標準入力のみであれば標準エラー出力のみに出力されます。
- p *<no>* *<no>* を索引ページの先頭ページとして指定します。また、 $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ のログファイル (*.log*) を参照することにより *any* (最終ページの次のページから)、*odd* (最終ページの次の奇数ページから)、*even* (最終ページの次の偶数ページから) といった指定の仕方も可能です。
- E エンコーディングを EUC-JP に指定します。入力ファイル、出力ファイルとも EUC-JP として扱います。
- J エンコーディングを JIS (ISO-2022-JP) に指定します。入力ファイル、出力ファイルとも JIS として扱います。
- S エンコーディングを Shift_JIS に指定します。入力ファイル、出力ファイルとも Shift_JIS として扱います。
- U エンコーディングを UTF-8 に指定します。入力ファイル、出力ファイルとも UTF-8 として扱います。
- I *<enc>* 内部バッファのエンコーディングを *<enc>* に指定します。*<enc>* には *euc* (EUC-JP) または *utf8* (UTF-8) が指定可能です。このオプションが指定されていない場合のデフォルト値は *euc* ですが、オプション *-U* が明示的に指定されている場合は *utf8* となります。
- 以降はオプション文字列と解釈しません。

3 スタイルファイル

スタイルファイルは *makeindex* のものと上位互換です。形式は “(スタイルパラメータ) (引数)” のリストで構成されます。パラメータの記述順序は自由です。また ‘%’ 以降はコメントと見されます。

以下にスタイルパラメータとそのデフォルト値の一覧を示します。

■入力ファイルスタイルパラメータ

keyword	<i><文字列></i> 規定値: “ <i>\indexentry</i> ” 索引エントリを引数として持つコマンド。
arg_open	<i><文字></i> 規定値: ‘{’ 索引エントリ文字列開始を表す文字。
arg_close	<i><文字></i> 規定値: ‘}’ 索引エントリ文字列終了を表す文字。
range_open	<i><文字></i> 規定値: ‘(’ ページ範囲の開始を示す文字。
range_close	<i><文字></i> 規定値: ‘)’ ページ範囲の終了を示す文字。
level	<i><文字></i> 規定値: ‘!’ 従属レベルであることを示す文字。

actual	<p>〈文字〉 規定値：‘@’</p> <p>このシンボルに続く文字列が実際の索引文字列として出力ファイルに書かれる。</p>
encap	<p>〈文字〉 規定値：‘ ’</p> <p>このシンボルに続く文字列が、ページ番号に付くコマンド名として使われる。</p>
page_compositor	<p>〈文字列〉 規定値：“-”</p> <p>階層化されたページ番号における階層間の区切り文字。</p>
page_precedence	<p>〈文字列〉 規定値：“rnaRA”</p> <p>ページ番号の記法の優先順位。‘R’ および ‘r’ はローマ数字、‘n’ はアラビア数字、‘A’ および ‘a’ はアルファベットによる記法を表す。</p>
quote	<p>〈文字〉 規定値：“”</p> <p>mendex のパラメータ文字に対するエスケープキャラクタ。</p>
escape	<p>〈文字〉 規定値：“\”</p> <p>一般的な文字に対するエスケープキャラクタ。</p>

■出力ファイルスタイルパラメタ

preamble	<p>〈文字列〉 規定値：“\begin{theindex}\n”</p> <p>出力ファイル先頭の文字列。</p>
postamble	<p>〈文字列〉 規定値：“\n\n\end{theindex}\n”</p> <p>出力ファイル末尾の文字列。</p>
setpage_prefix	<p>〈文字列〉 規定値：“n_{\setcounter{page}{”</p> <p>開始ページを設定するときの、ページ番号の前に付ける文字列。</p>
setpage_suffix	<p>〈文字列〉 規定値：“}\n”</p> <p>開始ページを設定するときの、ページ番号の後に付ける文字列。</p>
group_skip	<p>〈文字列〉 規定値：“\n_{\indexspace}\n”</p> <p>新項目（頭文字）の前に挿入する縦スペースを表す文字列。</p>
lethead_prefix	<p>〈文字列〉 規定値：“”</p> <p>頭文字の前に付けるコマンド文字列。</p>
heading_prefix	<p>〈文字列〉 規定値：“”</p> <p>lethead_prefix と同じ。</p>
lethead_suffix	<p>〈文字列〉 規定値：“”</p> <p>頭文字の後に付けるコマンド文字列。</p>
heading_suffix	<p>〈文字列〉 規定値：“”</p> <p>lethead_suffix と同じ。</p>
lethead_flag	<p>〈数値〉 規定値：0</p> <p>頭文字の出力のフラグ。0 のとき出力しない。0 より大きいときは英字を大文字で、0 より小さいときは小文字で出力する。</p>
heading_flag	<p>〈数値〉 規定値：0</p> <p>lethead_flag と同じ。</p>
item_0	<p>〈文字列〉 規定値：“n_{\item_”</p>

	主エントリ間に挿入するコマンド.
item_1	〈文字列〉 規定値: “\n _{UUUU} \\subitem _U ” サブエントリ間に挿入するコマンド.
item_2	〈文字列〉 規定値: “\n _{UUUUUU} \\subsubitem _U ” サブサブエントリ間に挿入するコマンド.
item_01	〈文字列〉 規定値: “\n _{UUUU} \\subitem _U ” 主-サブエントリ間に挿入するコマンド.
item_x1	〈文字列〉 規定値: “\n _{UUUU} \\subitem _U ” 主-サブエントリ間に挿入するコマンド (主エントリにページ番号がないとき).
item_12	〈文字列〉 規定値: “\n _{UUUU} \\subsubitem _U ” サブ-サブサブエントリ間に挿入するコマンド.
item_x2	〈文字列〉 規定値: “\n _{UUUU} \\subsubitem _U ” サブ-サブサブエントリ間に挿入するコマンド (サブエントリにページ番号がないとき).
delim_0	〈文字列〉 規定値: “, _U ” 主エントリと最初のページ番号の間の区切り文字列.
delim_1	〈文字列〉 規定値: “, _U ” サブエントリと最初のページ番号の間の区切り文字列.
delim_2	〈文字列〉 規定値: “, _U ” サブサブエントリと最初のページ番号の間の区切り文字列.
delim_n	〈文字列〉 規定値: “, _U ” ページ番号間の区切り文字列. どのエントリレベルにも共通.
delim_r	〈文字列〉 規定値: “--” ページ範囲を示すときの, ページ番号間の区切り文字列.
delim_t	〈文字列〉 規定値: “” ページ番号のリストの終端に出力する文字列.
suffix_2p	〈文字列〉 規定値: “” ページ番号が2ページ連続する場合に, delim_n と2ページ目の番号の代わりに付加する文字列. 文字列が定義されている場合にのみ有効.
suffix_3p	〈文字列〉 規定値: “” ページ番号が3ページ連続する場合に, delim_r と3ページ目の番号の代わりに付加する文字列. suffix_mp より優先される. 文字列が定義されている場合にのみ有効.
suffix_mp	〈文字列〉 規定値: “” ページ番号が3ページ連続する場合に, delim_r と3ページ目の番号の代わりに付加する文字列. 文字列が定義されている場合にのみ有効.
encap_prefix	〈文字列〉 規定値: “\” ページ番号にコマンドを付けるときの, コマンド名の前に付ける文字列.
encap_infix	〈文字列〉 規定値: “{” ページ番号にコマンドを付けるときの, ページ番号の前に付ける文字列.
encap_suffix	〈文字列〉 規定値: “}”

	ページ番号にコマンドを付けるときの、ページ番号の後に付ける文字列.
line_max	〈数値〉 規定値：72 1 行の最大文字数. それを超えると折り返す.
indent_space	〈文字列〉 規定値：“” 折り返した行の頭に挿入するスペース.
indent_length	〈数値〉 規定値：16 折り返した行の頭に挿入されるスペースの長さ.
symhead_positive	〈文字列〉 規定値：“Symbols” lethead_flag または heading_flag が正数の場合に数字・記号の頭文字として出力する文字列.
symhead_negative	〈文字列〉 規定値：“symbols” lethead_flag または heading_flag が負数の場合に数字・記号の頭文字として出力する文字列.
symbol	〈文字列〉 規定値：“” symbol_flag が 0 でない場合に、数字・記号の頭文字として出力する文字列. 文字列が定義されていれば、symhead_positive および symhead_negative より優先される (mendex 専用).
symbol_flag	〈数値〉 規定値：1 数字・記号の頭文字の出力フラグ. 0 のとき出力しない (mendex 専用).
letter_head	〈数値〉 規定値：1 日本語の頭文字の出力のフラグ. 1 のときカタカナ, 2 のときひらがなで出力する (mendex 専用).
priority	〈数値〉 規定値：0 英字と日本語との混在した索引語のソート方法についてのフラグ. 0 でなければ英字と日本語との間に半角スペースを入れた状態でソートする (mendex 専用).
character_order	〈文字列〉 規定値：“SEJ” 記号, 英字, 日本語の優先順位. ‘S’ は記号, ‘E’ は英字, ‘J’ は日本語を表す (mendex 専用).

4 日本語の扱いについて

mendex は日本語の索引をできるだけ楽に扱えるようになっています. makeindex では日本語の索引が正しく辞書順にソートするためにはひらがなまたはカタカナに揃え, 拗音, 撥音, 濁点を除いた読みを付けなければなりません (自動的に揃えるバージョンもある). mendex ではカナについてはすべて自動的に揃え, また漢字については辞書ファイルを設定することにより各索引語ごとに読みを付ける作業をかなり解消できます. 以下に内部でのカナの変換例を示します.

かぶしきがいしゃ	かふしきかいしゃ
マッキントッシュ	まつきんとつしゅ
ワープロ	わあふろ

辞書ファイルは“〈漢字〉〈読み〉”のリストで構成されます。〈漢字〉と〈読み〉の区切りはタブまたはスペースです。以下に辞書の例を示します。

漢字	かんじ
読み	よみ
環境	かんきょう
α	アルファ

辞書に登録する漢字については、読み方が1通りになるよう送り仮名を付けてください。「表」、「性質」のように送り仮名によらず2通りの読み方ができる語についてはどちらか1つしか登録できません。他の読み方については各索引語へ読みを付けることで対応してください。また、環境変数 `INDEXDEFAULTDICTIONARY` に辞書ファイルを登録することにより、自動的に辞書を参照します。環境変数に登録した辞書は `-d` で指定した辞書と併用できます。

5 ソート方法について

`mendex` は通常は入力された索引語をそのままソートします。`-l` オプションが付けられた場合、複数の単語で構成される索引語については、ソートするときに単語と単語の間のスペースを詰めてソートします。ここでは前者を単語順ソート、後者を文字順ソートと呼ぶことにします。文字順ソートの場合、実際に出力される文字列はスペースを含んだ状態のもので、索引語自体が変化することはありません。以下に例を示します。

単語順ソート	文字順ソート
X Window	Xlib
Xlib	XView
XView	X Window

また、日本語－英字間でも似たようなソート方法があります。スタイルファイルで `priority` を0以外に指定した場合、隣接した日本語と英字の間にスペースを入れてソートします。以下に例を示します。

<code>priority 0</code>	<code>priority 1</code>
<code>index sort</code>	<code>ind ファイル</code>
<code>ind ファイル</code>	<code>index sort</code>

6 環境変数について

`mendex` では以下のような環境変数を使用しています。

<code>INDEXSTYLE</code>	索引スタイルファイルがあるディレクトリ。
<code>INDEXDEFAULTSTYLE</code>	デフォルトで参照する索引スタイルファイル。
<code>INDEXDICTIONARY</code>	辞書があるディレクトリ。
<code>INDEXDEFAULTDICTIONARY</code>	常に参照する辞書ファイル。

7 詳細について

その他，詳細については `makeindex` に準拠.

8 既知の問題

複数のページ記法を使用する場合，ページ順に索引リストファイル (`.idx`) を与えないとページ番号を誤認することがあります.