
libnmz

「libnmz API」
UNIX版 - 初級編 -

(Version 2.0.19 対応版)
第3版(Version 0.0.0.3)

アカネイロソフトウェア

発行: 2009/05

目 次

1.	はじめに.....	4
2.	初級編について.....	4
3.	libnmz のライセンス.....	4
4.	libnmz について.....	5
5.	インデックス	6
1.	NMZ.w.....	6
2.	NMZ.t.....	6
3.	NMZ.field.xxxx	6
6.	ヘッダーファイル	7
7.	ライブラリファイル.....	7
8.	コンパイル	8
9.	手順.....	9
10.	サンプルコード.....	10
11.	定数.....	13
1.	BUFSIZE.....	13
2.	nmz_stat.....	13
3.	nmz_sortmethod	14
4.	nmz_sortorder.....	15
12.	構造体	16
1.	nmz_data.....	16
2.	_NmzResult.....	17
13.	主要関数一覧	18
1.	libnamazu.h	18
2.	field.h	18
3.	hlist.h	18
4.	idxname.h	18
5.	i18n.h	18
6.	search.h.....	19
7.	codeconv.h	19
14.	関数.....	20
15.	「Namazu の開発日記」メルマガについて.....	31
15.1.	「Namazuの開発日記」メルマガについて	31
15.2.	メルマガ登録/解除	31

16.	免責.....	32
17.	著作權.....	32
18.	履歷.....	33

sample

1. はじめに

「libnmz API」は、日本語全文検索システム Namazu の `namazu` や `namazu.cgi` で利用している検索用ライブラリです。本書はこの libnmz の使用方法を簡単にまとめたものです。

本書は Solaris, Linux, FreeBSD 等の UNIX 系 OS での利用が前提となっています。Windows で利用される方は、「libnmz API Windows 版」をお読みください。

Namazu は常に最新版をインストールしてご利用ください。最新バージョンはセキュリティ上の問題を含め、多くのバグ修正、性能向上が行われています。

2. 初級編について

本書は初級編として、Perl モジュールの Search-Namazu が利用している libnmz の機能に限定して解説しています。

libnmz は Namazu の検索機能を切り出してライブラリ化したものですが、設計が悪くライブラリとして見た場合、汎用性に欠ける関数等も多く存在します。このため、全ての機能を説明するには、`namazu` や `namazu.cgi` の内部処理についてより具体的に解説する必要があります、非常に難しい内容となってしまいます。

Search-Namazu では検索に必要最小限の libnmz の機能を使って作られているため、これらに絞って解説することで必要最小限の libnmz の機能を比較的簡単に理解し、利用できるようになるものと思います。

3. libnmz のライセンス

libnmz のライセンスは LGPL ではなく、GPL ライセンスです。スタティックリンクであろうと、ダイナミックリンクであろうと、libnmz を使用するアプリケーションは GPL ライセンスと矛盾しないライセンスでなければなりません。

libnmz を使用したアプリケーションを配布する場合は、注意してください。

4. libnmz について

- libnmz を用いると Namazu 2.X 形式のインデックスファイルから検索を行うアプリケーションを比較的簡単に作成することができます。
- libnmz は単語検索、部分一致検索、正規表現検索、フィールド検索等、namazu や namazu.cgi がサポートする全ての検索式をサポートしています。
- 検索式は Shift_JIS, EUC-JP のどちらでも指定することができます。しかし、UTF-8 はサポートしていません。
- 検索結果はスコア順/日付順/フィールド順にソートして取り出すことができます。また、昇順/降順の指定も可能です。
- 各文書のフィールド値を取り出すことができます。ただし、漢字コードは EUC-JP です。
- URI は HTML エンコードされたまま取り出されます。必要な場合は、アプリケーション側でデコードしてください。
- URI の置換機能はサポートしていません。必要な場合は、アプリケーション側で対応してください。
- 検索語の強調表示機能はサポートしていません。必要な場合は、アプリケーション側で対応してください。
- その他、libnmz は namazurc(namazurc) を読み込みませんので、これらの設定ファイルで設定する項目は全てサポートしていません。
- 複数のインデックスに対応しています。複数のインデックスをまとめて一度に検索することができます。
- libnmz にはインデックスを作成する機能は含まれていません。インデックスは mknmz で作成したものを使用してください。

5. インデックス

1. NMZ.w

単語ファイル。

漢字コードは EUC-JP で記録されています。

2. NMZ.t

タイムスタンプが記録されているファイル。

-1 はその文書が削除されたことをしめします。

日付ソートのキーとして使用します。

3. NMZ.field.xxxx

1 文書当たり 1 行で、文書のフィールド情報が記録されているファイル。

フィールド名をファイル名に含みます。

漢字コードは EUC-JP で記録されています。

主なフィールド名

from	著者
date	日付
size	文書ファイルのサイズ
subject	サブジェクト、タイトル
summary	要約
message-id	メッセージ ID
newsgroup	ニュースグループ
to	宛先
uri	URI

6. ヘッダーファイル

libnmz を利用するアプリケーションでは、次のヘッダーをインクルードしてください。

```
#include "namazu/libnamazu.h"
#include "namazu/field.h"
#include "namazu/hlist.h"
#include "namazu/idxname.h"
#include "namazu/il8n.h"
#include "namazu/search.h"
#include "namazu/codeconv.h"
```

Namazu をソースからインストールした場合、ヘッダーファイルは標準では /usr/local/include にインストールされます。

7. ライブラリファイル

libnmz を利用するアプリケーションでは、libnmz をリンクしてください。

Namazu をソースからインストールした場合、ライブラリファイルは標準では /usr/local/lib にインストールされます。

インストールされている libnmz は、スタティックライブラリあるいはシェアードライブラリのいずれか、または両方の場合があります。

8. コンパイル

ヘッダーファイルが `/usr/local/include` ライブラリが `/usr/local/lib` にインストールされていると仮定すると次のように `-I`, `-L`, `-l` オプションを指定します。

```
$ cc program.c -o program -I /usr/local/include -L /usr/local/lib  
-lnmz
```

sample

9. 手順

簡単な libnmz の使用手順を以下に示します。

1. 準備

nmz_set_lang	言語設定
nmz_set_sortmethod	ソートの種類を設定
nmz_set_sortorder	昇順/降順の設定
nmz_set_sortfield	フィールドソート時の文字列
nmz_set_maxhit	最大ヒット数の設定
nmz_add_index	インデックスの追加
nmz_codeconv_query	検索式のコード変換

2. 検索

nmz_search	検索
------------	----

3. 結果

nmz_get_field_data	フィールド値の取得
--------------------	-----------

4. 終了処理

nmz_free_hlist	ヒットリスト解放
nmz_free_internal	内部メモリ解放

10. サンプルコード

```
#include "namazu/libnamazu.h"
#include "namazu/field.h"
#include "namazu/hlist.h"
#include "namazu/idxname.h"
#include "namazu/i18n.h"
#include "namazu/search.h"
#include "namazu/codeconv.h"

#include <stdio.h>
#include <stdlib.h>
#include <string.h>

int main()
{
    int i;
    char buff[BUFSIZE] = "";
    NmzResult hlist;

    putenv("LANGUAGE=");
    putenv("LC_ALL=");
    putenv("LC_MESSAGES=");
    putenv("LC_CTYPE=");
    putenv("LANG=");
    nmz_set_lang("ja_JP.eucJP");

    /* ソート順 */
    nmz_set_sortmethod(SORT_BY_DATE);

    /* インデックスの指定 */
    nmz_add_index("/usr/local/var/namazu/index");
```

11. 定数

libnmz で使用する主要な定数をしめします。
これらはヘッダーファイル内で定義されています。

1. BUFSIZE

BUFSIZE は、バッファサイズを示す定数です。

2. nmz_stat

nmz_stat は、ステータス定数です。

```
enum nmz_stat {
    FAILURE = -1,
    SUCCESS,
    ERR_FATAL,
    ERR_TOO_LONG_QUERY,
    ERR_INVALID_QUERY,
    ERR_TOO_MANY_TOKENS,
    ERR_TOO_MUCH_MATCH,
    ERR_TOO_MUCH_HIT,
    ERR_REGEX_SEARCH_FAILED,
    ERR_PHRASE_SEARCH_FAILED,
    ERR_FIELD_SEARCH_FAILED,
    ERR_CANNOT_OPEN_INDEX,
    ERR_NO_PERMISSION,
    ERR_CANNOT_OPEN_RESULT_FORMAT_FILE,
    ERR_INDEX_IS_LOCKED,
    ERR_OLD_INDEX_FORMAT
};
```

12. 構造体

libnmz で使用する主要な構造体をしめします。
これらはヘッダーファイル内で定義されています。

ここに示す構造体は、libnmz を利用するアプリケーションでは読み出し専用です。値を書き変えてはいけません。

1. nmz_data

検索結果 1 件(1 文書)単位の情報が格納されています。

```
struct nmz_data {  
    int score;  
    int docid; /* document ID */  
    int idxid; /* index ID */  
    int date; /* document's date */  
    int rank; /* ranking data for stable sorting */  
    char *field; /* field's contents for field-specified search */  
};
```

score	スコア
docid	ドキュメント ID
idxid	インデックス ID
date	日付(タイムスタンプ) 1970 年 1 月 1 日 00:00:00 UTC からの経過時間
rank	ランク
field	フィールドソート時に一時的に利用するもの (<u>読み書き禁止</u>)

13. 主要関数一覧

libnmz の主要関数の一覧です。

1. libnamazu.h

```
extern void nmz_set_sortmethod ( enum nmz_sortmethod method );  
extern void nmz_set_sortorder ( enum nmz_sortorder order );  
extern void nmz_set_maxhit ( int max );  
extern void nmz_free_internal( void );
```

2. field.h

```
extern void nmz_get_field_data ( int idxid, int docid, const char  
*field, char *data );
```

3. hlist.h

```
extern void nmz_free_hlist ( NmzResult hlist );  
extern void nmz_set_sortfield ( const char *field );
```

4. idxname.h

```
extern enum nmz_stat nmz_add_index( const char *idxname );
```

5. i18n.h

```
extern char *nmz_set_lang ( const char *lang );
```

14. 関数

<pre>extern void nmz_set_sortmethod (enum nmz_sortmethod method);</pre>		
引数	method	<p>SORT_BY_SCORE または SORT_BY_DATE または SORT_BY_FIELD を指定します。</p> <p>SORT_BY_SCORE ... スコア順 SORT_BY_DATE ... 日付(タイムスタンプ)順 SORT_BY_FIELD ... フィールド順</p>
戻り値		なし
説明		<ul style="list-style-type: none">● 検索結果のソート方法を指定します。● ソート順は nmz_set_sortorder で指定します。● SORT_BY_FIELD を指定した場合は nmz_set_sortfield でソートキーを指定する必要があります。

15. 「Namazu の開発日記」メルマガについて

15.1. 「Namazuの開発日記」メルマガについて

「Namazu の開発日記」メールマガジンは、Namazu に関する様々な情報をお知らせするメールマガジンです。Namazu の TIPS や、次世代 Namazu に関する機能案の紹介、Namazu for Windows で利用する ActivePerl 用の PPM に関する情報等も紹介しています。

多くは「ナマズのブログ」<http://namazu.asablo.jp/blog/> の内容を再編集した記事が元になっており、「ナマズのブログ」のダイジェストの役割を果たしています。

メルマガにはアフィリエイト広告が含まれています。広告収入は Web サイトの管理運営費の一部に充てています。

15.2. メルマガ登録/解除

「Namazu の開発日記」メールマガジンは、下記 URL から登録、解除が可能です。(無料)

<http://mg7.com/cgi-bin/hp.cgi?id=16084>

16. 免責

「libnmz API」は有用とは思いますが、頒布にあたっては、市場性及び特定目的適合性についての暗黙の保証を含めて、いかなる保証も行なっていません。

17. 著作権

「libnmz API」は著作物です。本書の内容のすべてまたは一部を許可なく配布することを禁じます。

Copyright (C) 2008-2009 TADAMASA TERANISHI

Copyright (C) 2008-2009 AKANEIRO SOFTWARE. All rights reserved.

sample

18. 履歴

- 2008/06/17 初版
- 2008/07/30 第2版
 - 「libnmz API Windows 版」について追記
 - サンプルに LC_CTYPE を追加
 - サンプルの解説を追加
- 2009/05/23 第3版
 - Namazu 2.0.19 対応

sample

索 引

—

_NmzResult 17

A

ASCENDING 15

B

BUFSIZE 13

D

DESCENDING 15

E

ERR_CANNOT_OPEN_INDEX 14

ERR_CANNOT_OPEN_RESULT_FORMAT_FILE 14

ERR_FATAL 14

ERR_FIELD_SEARCH_FAILED 14

ERR_INDEX_IS_LOCKED 14

ERR_INVALID_QUERY 14

ERR_NO_PERMISSION 14

ERR_OLD_INDEX_FORMAT 14

ERR_PHRASE_SEARCH_FAILED 14

ERR_REGEX_SEARCH_FAILED 14

ERR_TOO_LONG_QUERY 14

ERR_TOO_MANY_TOKENS 14

ERR_TOO_MUCH_HIT 14

ERR_TOO_MUCH_MATCH 14

F

FAILURE 14

N

NMZ.field.xxxx 6

NMZ.t 6

NMZ.w 6

nmz_add_index 27

nmz_codeconv_query 30

nmz_data 16

nmz_free_hlist 25

nmz_free_internal 23

nmz_get_field_data 24

nmz_search 29

nmz_set_lang 28

nmz_set_maxhit 22

nmz_set_sortfield 26

nmz_set_sortmethod 20

nmz_set_sortorder 21

nmz_sortmethod 14

nmz_sortorder 15

nmz_stat 13

NmzResult 17

S

SORT_BY_DATE 14

SORT_BY_FIELD 14

SORT_BY_SCORE 14

SUCCESS 14