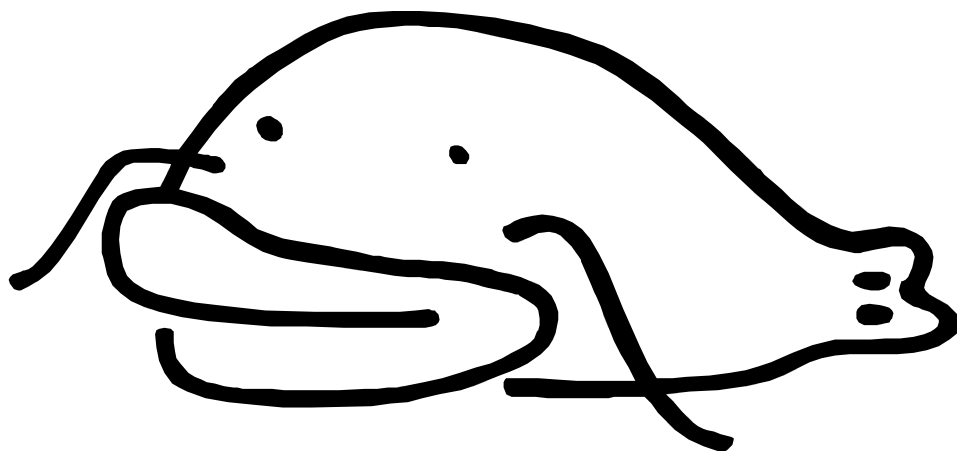


Namazu for Win32 メンテナンス物語  
オープンソースプロジェクトに参加したきっかけ



株式会社ドリーム・アーツ

竹迫 良範

<http://namazu.org/~takesako/>



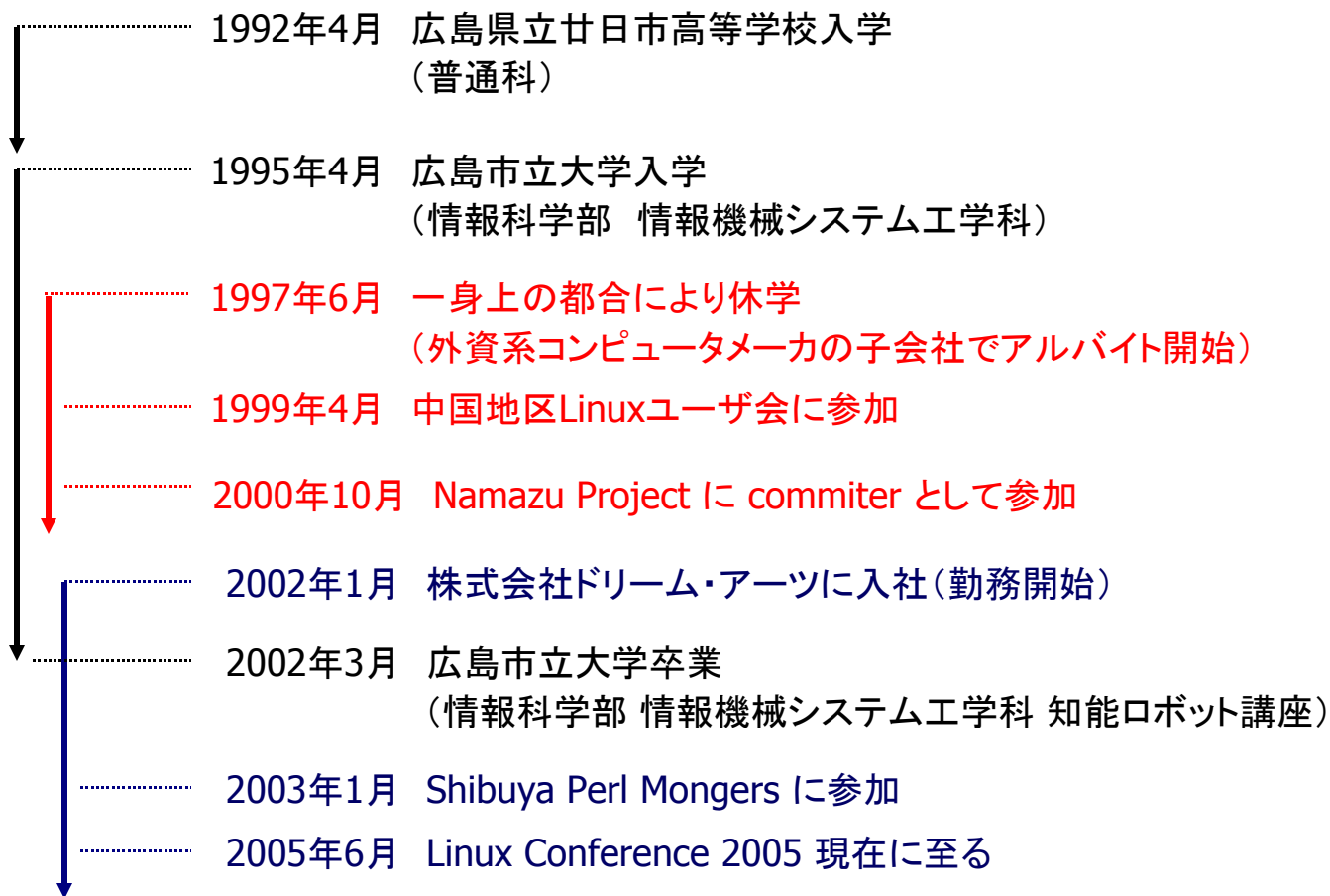
## 本日のアジェンダ

1. 開発者の自己紹介
2. そのシステムを作ろう／開発に参加しようとしたきっかけ
3. そのシステムの長所(と短所)
4. それがよくわかるようなデモ
5. 今後の展望や、今興味を持っているテーマなど





## 自己年表





## 全文検索システムの構築

検索キーワードの入力

### ■ 概要

- 社内のファイルサーバで共有している膨大な文書に対して高速な全文検索システムを導入することによって、社内の情報共有を促進
- オープンソースによって開発されているフリーの全文検索システム Namazu for Win32 を採用
- Text, Mail, HTML, Word, Excel, PowerPoint, PDF形式のオフィス文書
- 約4万ファイル(2.5GB)

検索キーワードの入力

検索キーワード: HP-UX 11.0 インストール手順

表示形式: 標準 表示件数: 20 ソート: 日付(新しい順)

検索対象:  
 file://\*:\*:\*:\*:\*:\*  
 file://\*:\*:\*:\*:\*:\*  
 file://\*:\*:\*:\*:\*:\*

検索結果

参考ヒット数:

- Central [HP-UX: 411][11.0: 163][インストール: 790][手順: 498]: 94][合計: 18]
- MOTES [HP-UX: 265][11.0: 105][インストール: 187][手順: 24]: 3][合計: 2]
- 検索結果 [HP-UX: 332][11.0: 161][インストール: 262][手順: 51]: 11][合計: 3]

検索式にマッチする 23 個の文書が見つかりました。

1. [b2355-97177.pdf](#) (スコア: 5,918)  
日付: Tue, 07 Aug 2001 16:13:24  
シス シス システム テム システム テム / ワーク ワーク ワーク グループ グループ グループ の 管 の 管 グループ の 管 理 理 の 管理 理 第 第 第 2 版 版 版 版 HP-UX 11.0 Manufacturing Part Number: B2355-97177 1998 年 年 年 年 7 月 月 月 月 U.S.A. Copyright 1  
[file:///C:/usr/local/Tools/HP-UX11.0/SE/OpenView/NNM/doc/HP-UX/1020/b2355-97177.pdf](#) (7,557,954 bytes)
2. [i1240-90055.pdf](#) (スコア: 638)  
日付: Wed, 04 Jul 2001 15:05:42  
HP OpenView ネット ネット ネット ワーク ワーク ワーク ノード ノード ノード マネージャージャージャ  
マネージャージャ データ分析 データ分析 データ分析 データ分析 / レポート レポート レポート レポート ガイド ガイド  
ド ガイド Win  
[file:///C:/usr/local/Tools/HP-UX11.0/SE/OpenView/NNM/doc/Nm62/i1240-90055.pdf](#) (635,195 bytes)
3. [HP-UX11iの概要-White Paper.pdf](#) (スコア: 346)  
日付: Thu, 31 May 2001 21:09:26  
HP HP-UX 11i の june 2000 White Paper: 概要 HP-UX 11i 目次 概要 4 インターネット・コンピューティングにおける要件とは? 4 コンピューティング環境 4 開発環境 4 HP-UX 11i の機能 4 HP-UX 11i の主な新機能 5 Linux/Windows と  
[file:///C:/usr/local/Tools/HP-UX11.0/Tools/HP-UX11i\(HP-UX11.11\)情報/HP-UX11iの概要-White Paper.pdf](#) (187,396 bytes)
4. [バックアップ/リカバリ計画書](#) (スコア: 181)  
日付: Fri, 30 Mar 2001 17:32:33  
[file:///C:/usr/local/Tools/HP-UX11.0/Tools/HP-UX11i\(HP-UX11.11\)情報/HP-UX11iの概要-White Paper.pdf](#)

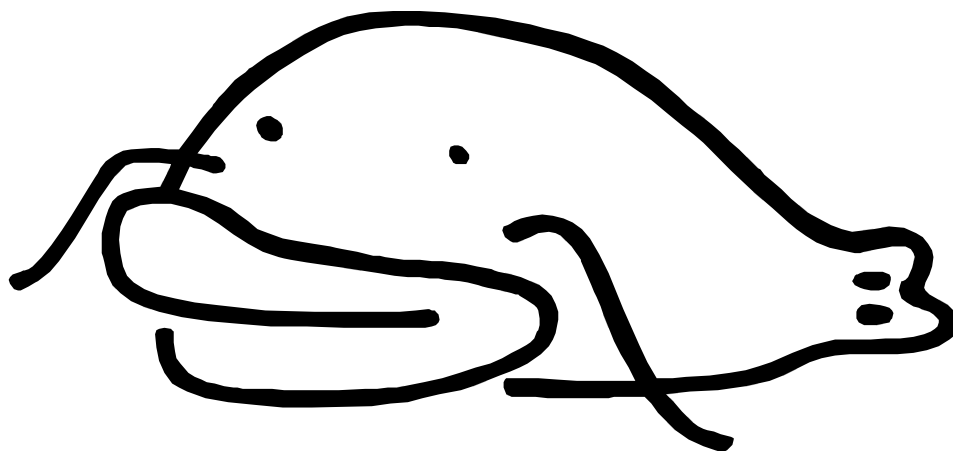


## Namazuz for Win32 を使ってみたが…

- 半角カタカナを使用したファイル名の処理に不具合
  - 半角カナ対応パッチの作成(社内での ad hoc 対応)
  - ライセンスが GPL だったので、開発元にフィードバック
- OLE文書フィルタのパフォーマンス問題
  - Excelファイルのインデックスに時間がかかる
    - OLE文書フィルタの性能改善
- Windows環境でのコンパイル
  - Cygwin + WinGW32 + MSVC++ + ActivePerl + nmake
  - メーリングリスト上でのユーザサポートなどなど

パッチを書いて送っていたら  
committer としてお誘いを受けた

## Namazu Project の紹介



日本語全文検索ソフトウェア



## Namazu の特徴

- 検索プログラムはCで記述 - `namazu`, `namazu.cgi`
  - 高速性が求められる用途は低級言語で実装 (チューニング)
  - `libnmz` として C API を提供 (拡張性)
- インデクサはPerlで記述 - `mknmz`
  - 検索ほど高速性を求められていない
  - 速くは作れないが早く作れる(スクリプト言語の利点)
    - 開発期間の短縮
  - 文書フィルタの拡張性
    - plain text, HTML, Mail/News, MHonArc, TeX, man, PDF, Word, Excel, PowerPoint, 一太郎, DocuWorks...





## 関連ソフトウェアが豊富

- **1998-01-31:** namazu.el-19980131.01 が開発される  
by やまだあきらさん, まつもとゆきひろさん, 馬場肇さん
- **1998-02-06:** bnamazu by 馬場肇さん
- **1998-03-04:** NamazuJAVA v0.2 by まつむらのぞみさん
- **1998-03-04:** pnamazu by 古川令さん
- **1998-03-25:** tknamazu v1.00 by 広瀬健一さん
- **1998-03-25:** search-s v0.7 by 山下誠二さん
- **1999-02-24:** mew-nmz.el by 白井秀行さん
- **1999-06-11:** Search Engine of Namazu on Java2  
by 坂口昭男さん

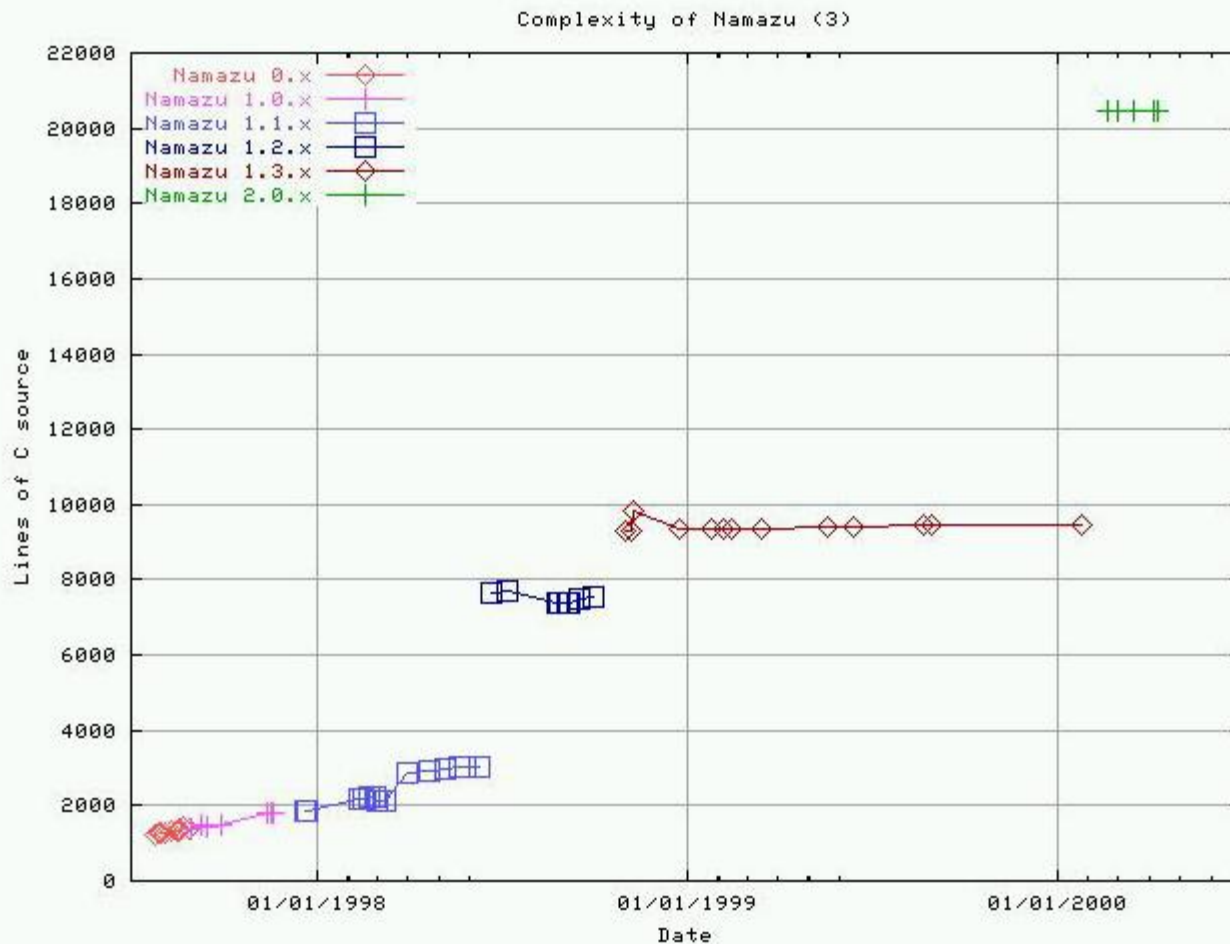


## Namazu Project の特徴

- 開発者が多い
  - 最初は高林さん一人が開発
  - 現在は CVS で共同開発（active commiter数名／18名）
  - メーリングリストでオープンな議論
- 幅広い対応プラットフォーム
  - UNIX系OS, Windows, OS/2, Javaクライアントも
- 歴史がある
  - 技術が枯れている、情報が多い、書籍もある



## ソースコードの規模の推移 (C)

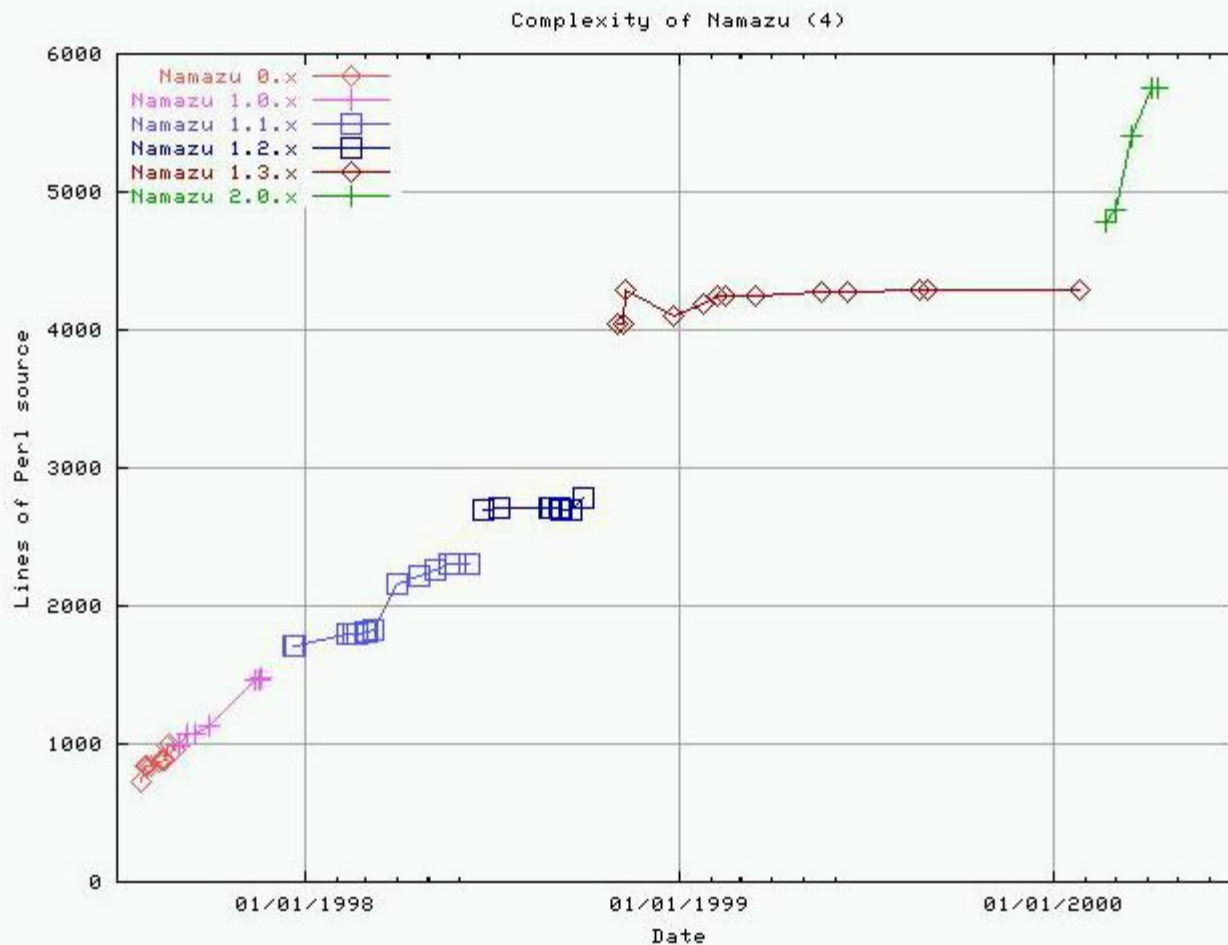


Namazuz is a full-text search engine. Namazu has an index maker, mknmz command, and a text searcher, namazu command. For searching a great amount of documents quickly, Namazu makes an index in advance. The concept of index is just similar to an index of book.

# Namazuz for Win32



## ソースコードの規模の推移 (Perl)

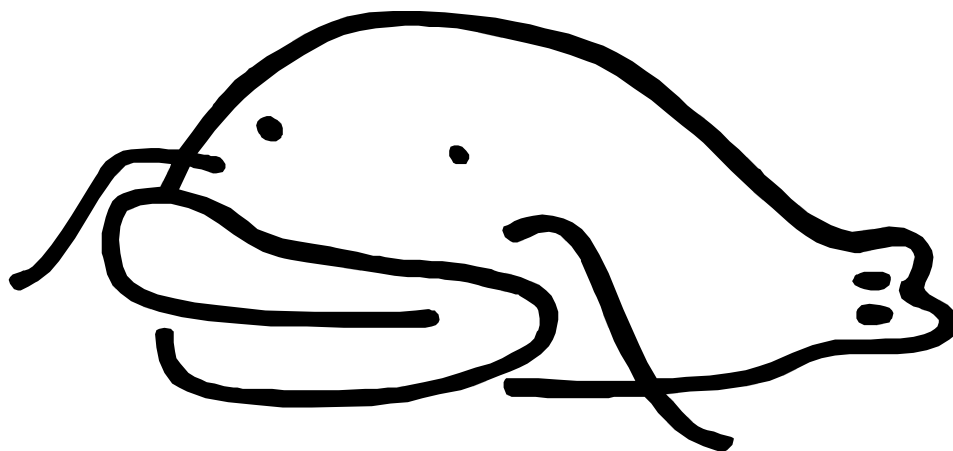




## Namazu の欠点

- 良くも悪くも分かち書き(形態素解析)方式
  - Kakasi, Chasen, MeCab...
  - tf·idfのスコア計算は、やりやすいが...
  - 日本語のフレーズ検索の精度が低い
- 設計が古い感は否めない
  - namazu.cgi ってどうよ?
  - libnmz はスレッドセーフではない
- コンパイルがちょっと大変なことがある
  - GNU autotools 呪縛
    - automake, autoconf, libtools, gettext...
  - Perl モジュールのインストール File::MMagic...

## Namzu for Win32 のメンテナンス



開発環境を用意するのが少し面倒…



## namazu.exe のビルド

### ■ make-namazu.sh (MinGW + MSYS の場合)

```
export MINGW32=yes
export CYGWIN=yes

export ac_cv_path_NKF=module_nkf
export ac_cv_path_KAKASI=module_kakasi
export ac_cv_path_CHASEN=chasen
export ac_cv_prog_LN_S=ln
export ac_cv_path_PERL=perl

CFLAGS=-O2 LIBS=-lintl ./configure --enable-nls ¥
--prefix=/namazu --datadir=/namazu/share ¥
--build=i386-pc-mingw32 --host= i386-pc-mingw32 ¥
--enable-tknamazu --disable-file-mmagic-checking ¥
--without-libiconv-prefix --without-libintl-prefix

make
make install
```



## libnmz.dll のビルド

### ■ make-nmz-dll.sh (MS Visual C++6.0 + Cygwin の場合)

```
export MINGW32=yes
export CYGWIN=yes

export ac_cv_type_off_t=no
export ac_cv_header_dlfcn_h=no
export ac_cv_header_unistd_h=no
export ac_cv_header_nl_types_h=no
export ac_cv_header_sys_param_h=no

CC="cl" CFLAGS="-nologo -Gf -W3 -O1 -MD -DNO_ALLOCA -DMSVC" LD="lib -nologo" ¥
./configure --prefix=/namazu --host=i386-pc-windows --build=i386-pc-windows ¥
--enable-shared=no --disable-nls --disable-tknamazu --disable-file-mmagic-checking ¥
--without-libiconv-prefix --without-libintl-prefix

cd lib; make all; OBJS=`echo *.obj`; lib -nologo -OUT:libnmzut.lib $OBJS; cd ..
Cd nmz; make all; OBJS=`echo *.obj`; lib -nologo -OUT:libnmz.lib $OBJS
echo "LIBRARY ¥"nmz¥"" > nmz.def; echo "EXPORTS" >> nmz.def
for F in $OBJS; do
    nm $F | grep 'T _' | sed 's/.* T _/ /' >> nmz.def
done
cl -nologo -Gf -W3 -O1 -MD -LD -Fnmz.dll $OBJS libnmzut.lib -link -def:nmz.def
```

※ 広瀬健一さんからノウハウを引き継ぎ





## (v)snprintf 問題

### ■ AC-Archive の AC\_FUNC\_SNPRINTF (m4マクロ) より

```
#include <stdio.h>
#include <stdarg.h>

int my_vsnprintf (char *buf, const char *tmpl, ...)
{
    int i;
    va_list args;
    va_start (args, tmpl);
    i = vsnprintf (buf, 2, tmpl, args);
    va_end (args);
    return i;
}

int main(void)
{
    char bufs[5] = { 'x', 'x', '\x00', '\x00' };
    char bufd[5] = { 'x', 'x', '\x00', '\x00' };
    int i;
    i = my_vsnprintf (bufs, "%s", "111");
    if (strcmp (bufs, "111")) exit (1);
    if (i != 3) exit (1);
    i = my_vsnprintf (bufd, "%d", 111);
    if (strcmp (bufd, "111")) exit (1);
    if (i != 3) exit (1);
    exit(0);
}
```

<http://www.t3.rim.or.jp/~raeva/diary/?200305b#200305144>

リンクする標準ライブラリによっては、  
戻り値が ISO C99 に準拠していない、  
あるいは NUL terminate しない実装がある

(例) MSVCRT.DLL の \_vsnprintf

```
#define vsnprintf _vsnprintf
```

Cygwin は独自の vsnprintf の  
実装を持っているので大丈夫だが  
Win32 native の環境では要注意



## PL2BAT で mknmz.BAT を作成

- ActivePerl が提供する PL2BAT (バッチファイルに変換)

### ▼ Hello.pl

```
#!/perl -w
#
use strict;
print "Hello world!¥n";
1;
```

Perlスクリプト

PL2BAT Hello.pl  
コマンドを実行

BATファイルを作成

### ▼ Hello.BAT

```
@rem = '---*Perl---'
@echo off
if "%OS%" == "Windows_NT" goto WinNT
perl -x -S "%0" %1 %2 %3 %4 %5 %6 %7 %8 %9
goto endofperl
:WinNT
perl -x -S %0 %*
if NOT "%COMSPEC%" == "%SystemRoot%¥system32¥cmd.exe" goto endofperl
if %errorlevel% == 9009 echo You do not have Perl in your PATH.
if errorlevel 1 goto script_failed_so_exit_with_non_zero_val 2>nul
goto endofperl
@rem ';
#!perl -w
#line 15
#
use strict;
print "Hello world!¥n";
1;
__END__
:endofperl
```



## Namazuz for Win32 のインストーラ

### ■ Namazuz 2.0.13(β版)から Inno Setupでインストーラを作成

- 1 インストーラの国際化  
日本語、英語の両言語メッセージを用意
- 2 環境変数の自動設定  
LANG, MKNMZRC, NAMAZURC, KANWADICTPATH, ITAIJIDICTPATH, NAMAZULOCALDIR, pkgdatadir
- 3 Windows NT4.0/2000/XPでは、システムの再起動なしに環境変数の設定を自動反映(Win95/98/Meは未対応)
- 4 Perl 5.8 対応 (5.8.3以上を推奨)  
ActivePerl Build 500系、600系のバイナリの他に  
ActivePerl 5.8.3 Build 809 の PPM バイナリを追加
- 5 同梱Perlモジュール(PPMバイナリ)のバージョンアップ  
File-MMagic-1.22  
NKF-2.0.4  
Text-Kakasi-2.04 (Kakasi 2.3.4)  
Search-Namazuz-0.92 (Namazuz 2.0.3RC13)
- 6 アンインストールが可能に  
Namazuz for Win32 のアンインストールプログラムが付属





## Inno Setup とは？

- Windowsインストーラ作成ソフト
  - .iss ファイルから .exe を作成
- 動作環境
  - Windows 95/98/Me
  - Windows NT4.0/2000/XP/2003
- オープンソースで開発が継続
  - Jordan Russell (1997~)
  - Martijn Laan (2000~)
- 開発環境
  - Borland Delphi 2.0~5.0
- 多言語インストール対応
  - 英語・日本語もOK

```
[Setup]
AppName=BAD Know-How
AppVerName=BAD Know-How 5.14
DefaultDirName={pf}¥BAD
Compression=lzma
SolidCompression=yes
ShowLanguageDialog=yes
LanguageDetectionMethod=UILanguage

[Languages]
Name: "en"; MessagesFile: "compiler:Default.isl";
Name: "ja"; MessagesFile: "compiler:Languages#Japanese.isl";

[Files]
Source: "myprog.exe"; DestDir: "{app}"
Source: "myprog.hlp"; DestDir: "{app}¥doc"
Source: "readme.txt"; DestDir: "{app}¥doc"; Flags: isreadme
```

<http://www.jrsoftware.org/>



## 設定ファイルからコンパイル(例)

### ■ bad.iss → bad.exe

```
[Setup]
AppName=BAD Know-How
AppVerName=BAD Know-How 5.14
DefaultDirName={pf}¥BAD
Compression=lzma
SolidCompression=yes
ShowLanguageDialog=yes
LanguageDetectionMethod=UILanguage

[Languages]
Name: "en"; MessagesFile: "compiler:Default.isl";
Name: "ja"; MessagesFile: "compiler:Languages¥Japanese.isl";

[Files]
Source: "myprog.exe"; DestDir: "{app}"
Source: "myprog.hlp"; DestDir: "{app}¥doc"
Source: "readme.txt"; DestDir: "{app}¥doc"; Flags: isreadme
```





## Inno Setup をお勧めする理由

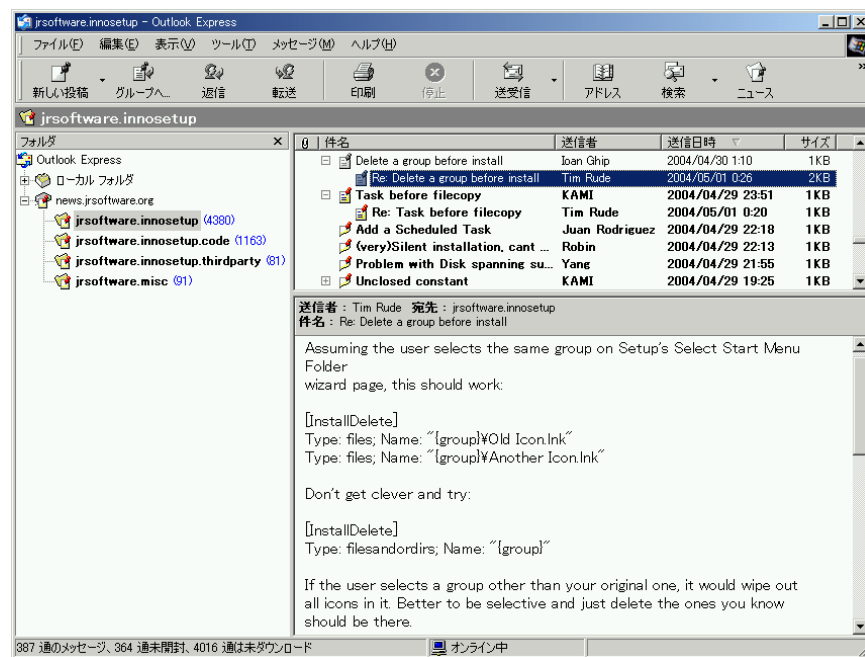
- オープンソースで開発が継続 (1997~)
  - 現在のバージョンは 4.2.2 (2004-04-27)
  - 全ソースコードが公開されているので安心
- 商用・非商用を問わず無償で利用可能
  - Papal.com / kagi.com で寄付受付中
- My Inno Setup Extensions (Inno Setup 4 から同梱)
  - 独自の Pascal Script でカスタマイズが可能
- 多言語対応 (現在45ヶ国語のカatalogが公開)

Arabic, Asturian, Belarus, Bulgarian, Catalan, Chinese (Simplified), Chinese (Traditional), Croatian, Czech, Danish, Dutch, Estonian, Finnish, French, Galician, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Lithuanian, Malaysian, Norwegian, Macedonian, Norwegian (nynorsk), Polish, Portuguese (Brazilian), Portuguese (Standard), Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Spanish (Argentina), Spanish (Colombia), Spanish (Mexico), Spanish (Standard), Swedish, Thai, Turkish, Ukrainian



## Inno Setup のサポート

- なんとニュースグループが存在
  - fj 世代の oldtype にも安心
  - newtype には新鮮な NNTP
  - 過去記事の検索も可能
- メーリングリストも存在するが…
  - バージョンアップのアナウンスのみ
  - メールの屑に埋もれる必要なし



## Inno Setup に関するニュースグループ

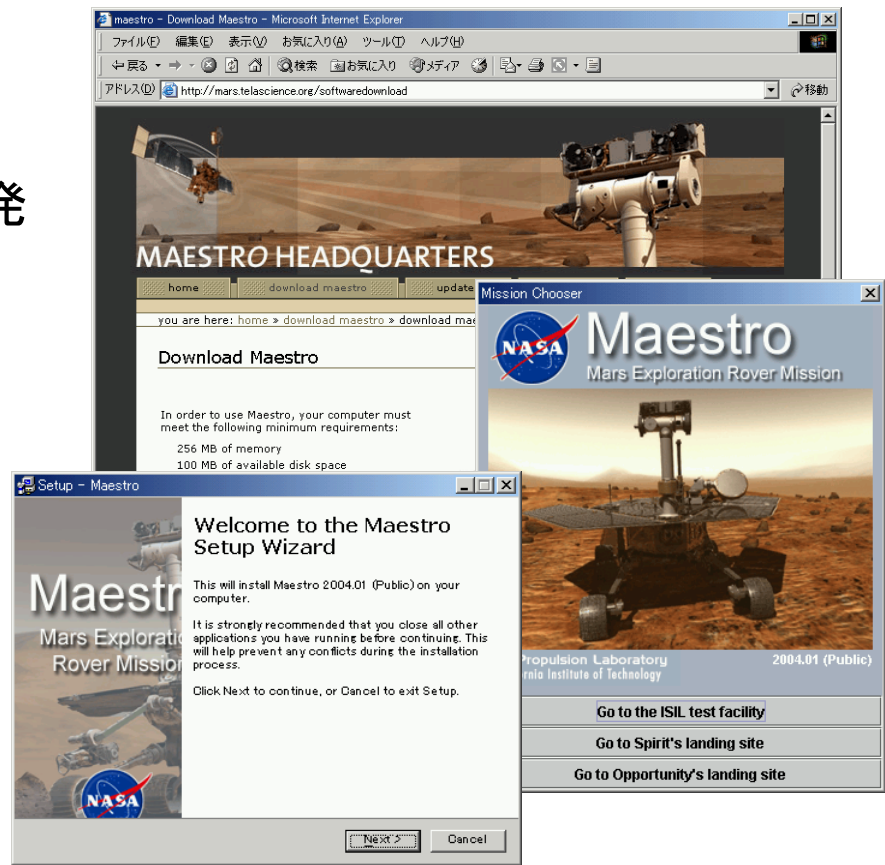
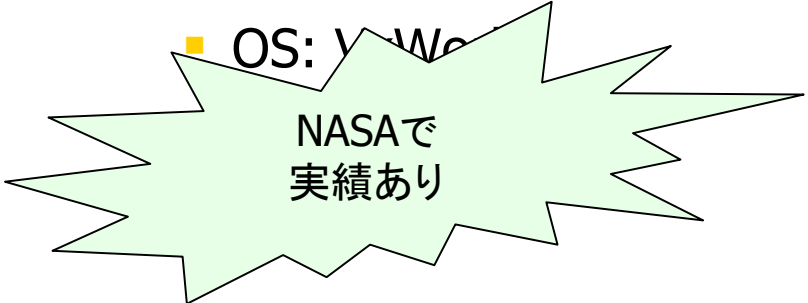
news://news.jrsoftware.org/jrsoftware.innosetup  
news://news.jrsoftware.org/jrsoftware.innosetup.code  
news://news.jrsoftware.org/jrsoftware.innosetup.thirdparty



## あの NASA が Inno Setup を採用!

### ■ Maestro

- Maestro-Win.exe
  - NASAジェット推進研究所が開発
  - Javaフリーソフト (Windows版)
- 火星探査車スピリット、オポチュニティーの遠隔制御・画像収集・解析を行う「地球側」の管制プログラム
  - 探査車のハードウェア仕様
    - MPU: RAD6000
    - OS: yWo



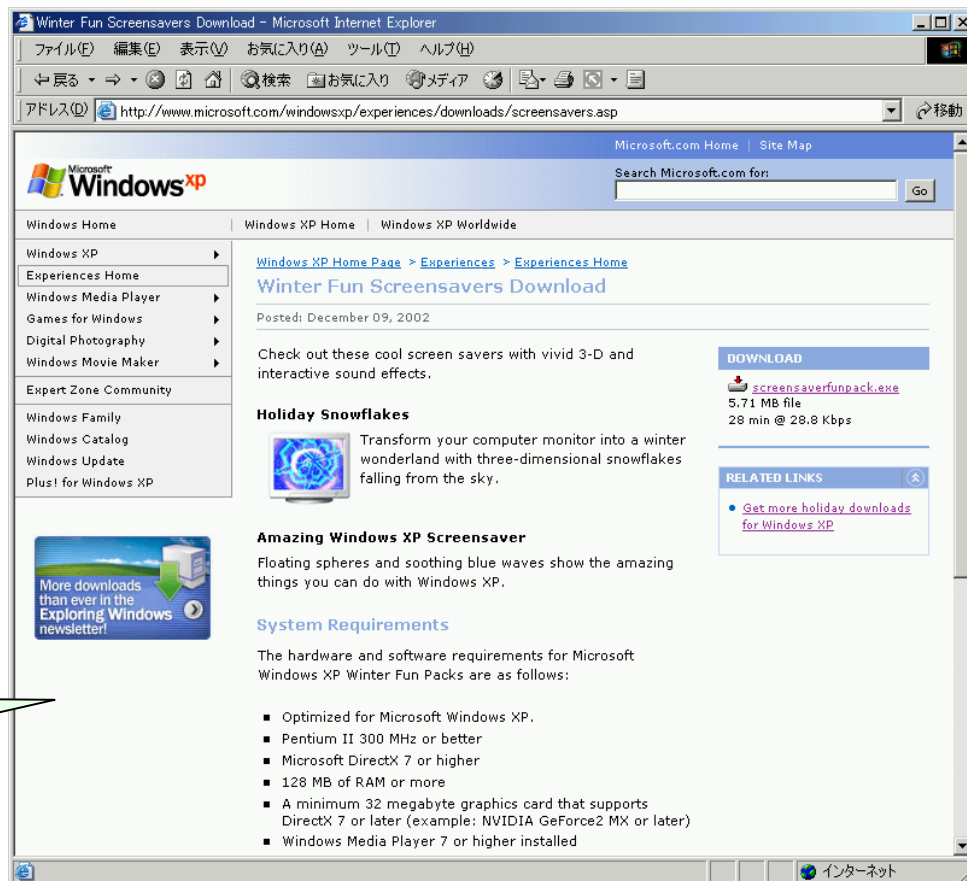
<http://mars.telascience.org/home>





## 実はMicrosoftも利用していた!

- Winter Fun Screensavers
  - screensaverfunpack.exe
    - Windows XP
    - スクリーンセーバー集
  - InstallShield と Inno Setup を組み合わせて利用



Microsoft も  
使っていた!

<http://www.microsoft.com/windowsxp/experiences/downloads/screensavers.asp>

ザモを

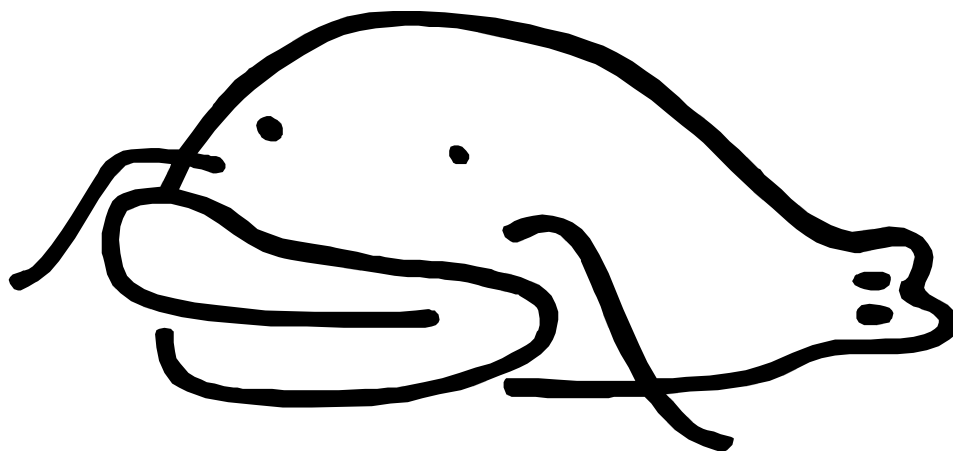
どろげ



## 今後の Namazu (拡張アイデア)

- 検索インデックスの UTF-8 対応 (現状EUC-JP)
  - Namazu 2.1.x 系
- A9.com OpenSearch 対応
  - 誰かやりませんか？
- インデクサの C API 化 (野首さん)
  - CodeFest 2005
- リアルタイム検索 (竹迫)
  - ファイルシステムへフックをかける (WebDAV / Samba ...)
    - ファイルの追加・更新・削除
  - 差分更新リストの管理
    - インデックス対象になっていないものは grep 検索
    - cronなどの定時処理で差分リストをインデックス化
    - 検索時に grep と namazu 検索の結果マージ

ご清聴ありがとうございました



これからも Namazu をよろしくお願いします