Namazu 開発物語
~ フリーソフトウェアを作ろう ~

高林哲　　馬場肇　　樋口千洋　　吉田智子

1998 年 12 月 18 日

1　概略

Namazu[1] は、手軽に使えることを第一に目指した日本語全文検索システムである。GPL に従ったフリーソフトウェアとして公開されており、Linux においても、Plamo Linux でのオンライン・ドキュメントや、Doc-CD で提供されるドキュメントの全文検索に採用されるなど広く利用されている。本発表では Namazu の開発の経緯、および Plamo Linux, Doc-CD での応用例を紹介する。

2　Namazu とは

2.1　特徴

- 手軽に使える全文検索システムソフトウェア
- 多彩な検索手法が利用できる
  - AND 検索, OR 検索, NOT 検索,
  - 括弧によるグループ化,
  - 前方一致, 中間一致, 後方一致, 正規表現,
  - フレイズ検索, フィールド指定の検索
- 検索速度はほぼ 1 秒以内
- 日本語の処理にきめ細やかな配慮を施している
- 文書の形式に応じたフィルタ処理が可能
- 複数のインデックスを同時に検索可能
- 文書の追加/更新/削除を反映したインデックスのアップデートが可能
- UNIX, Win32, OS/2 の環境で動作する

[1] <http://saturn.aichi-u.ac.jp/~ccsatoru/Namazu/>

2.2 生い立ち

「ロック専門の検索エンジンが欲しい」という私 (高林) の個人的欲求がそもそもの始まりです。最初に作ったものは 1,000 ファイル 5M バイトの検索に 30 秒もかかるひどい代物でした。

調べてみると SSE, freeWAIS-sf といった全文検索システムがあることがわかりましたが、どちらも私の趣味に合わない部分があったため、プログラミングの勉強も兼ねて自分で作ることにしました。1997 年の夏の話です。

しばらくして動くものができあがったので linux-users メイリングリストに宣伝したところ、馬場肇さんとやまだあきらさんから反応がありました。以来、両氏の協力のもとに開発が進められ、7 月の末に GPL に従ったフリーソフトウェアとして一般公開する運びとなりました。

リリースすると間もなく多くの方からバグ報告や機能追加などの要望のメイルが届き始めました。なかにはわざわざパッチを送ってくださる方もいらして、ソースコード付きで配布することの意義を実感した次第です。

ユーザの数が増えるにつれてバグの発見は迅速なものとなり、プログラムは徐々に安定していきました。それと同時に、私が考えていなかったようなアイディアが提案されたり、ときには新しい機能を実装したコードそのものが寄贈されることもしばしばありました。

メイリングリストを開設してからは開発はいっそう活発なものとなり、宴会も 2 回ほど催されました。Namazu によって得た人との繋がりは何よりも貴重なものだと思っています。

なお、 Namazu の履歴は<http://saturn.aichi-u.ac.jp/~ccsatoru/Namazu/history.html>に詳しくまとめられています。

2.3 関連ソフトウェア

- namazu.el by やまだあきらさん, まつもとゆきひろさん, 馬場肇さん
- bnamazu by 馬場肇さん
- NamazuJAVA by まつむらのぞみさん
- pnamazu by 古川令さん
- Tknamazu by 広瀬健一さん
- Search-s by 山下誠二さん

3 Doc-CD での応用

Doc-CD は Linux 関連の各種ドキュメントを中心に収録した CD-ROM。Linux Users ML に過去に投稿された 3 万通を越えるメールを中心に、20 種類の ML のメール、RFC など、CD-ROM 二枚分の情報が含まれており、ちょっとした初期設定だけでこれらを Namazu を使って全文検索できるように工夫してあります。

Vol.1 が、'98 年 10 月 15 日にぷらっとホームから、定価 2,000 円で発売開始されています [2] 。そのほか、日本 Linux ユーザ会 (JLUG) からも通信販売されています[3] 。

[2] <http://www.plathome.co.jp/rw/5.html>
[3] <http://Doc-CD.linux.or.jp/>

4 Plamo Linuxでの応用

Plamo Linux[4] は、全文検索システム Namazu を標準で搭載しています。そして、Plamo Linux の/usr/doc の下の数多くの JF のドキュメント群や Plamo Linux 固有のドキュメント群などを対象に予め索引を作成してあります。そのため、ユーザが Plamo Linux を使用中に困った事が起こったら、Namazu を用いてキーワードで検索し、解決のためのヒントを即座に得る事ができます。利用環境としては、X 上で動作する Tknamazu や Mule 上で動作する namazu.el が提供されています。もちろん、WWW を用いた利用もできます。

5 発表者略歴

高林哲 (たかばやし さとる) 1976 年、静岡県浜松市に生まれる。愛知大学経営学部在学。日本イソターネット協会会員。度重なる不運により道を誤り、1997 年 6 月より全文検索システム Namazu の開発に着手。誕生日と血液型は筒井康隆氏と同じだったりする。 Mike Oldfield の熱心なファンでもある。

馬場肇 (ばば はじめ) 1971 年生まれ。現在、京都大学大学院理学研究科 博士後期課程 在学中。たまりゆく電子メールの山に苦労した結果、日本語の使える全文検索システムに興味を持ち、1996 年に KAKASI を用いた freeWAIS-sf の日本語化を行う。Namazu との出会いは、1997 年夏に高林氏が Linux Users メーリングリストに流したアナウンス記事を見て直接連絡したのがきっかけ。その後、特に開発の初期の段階に関わる。Doc-CD の制作メンバー。著書に、1998 年ソフトバンク発行の『日本語全文検索システムの構築と活用』(通称「なまず本」) がある。

樋口千洋 (ひぐち ちひろ) 筑波大学博士課程化学研究科中退。その際、修士号強奪。現在某化学系メーカーに勤務。UNIX との出会いは在学時の BSD 2.9 on PDP-11、しかし以後 UNIX とは縁遠く、1993 年になって Linux と出会い感激する。Linux を如何にして考えるための道具として活用できるかを模索する毎日である。

吉田智子 (よしだともこ) UNIX 関係の記事の執筆に加えて、1998 年 4 月からは、ノートルダム女子大学と京都大学にて、情報リテラシー教育に携わっている。Namazu とは、1998 年春に馬場氏から紹介された、Plamo Linux と Doc-CD で利用されていたものとして出会いファンとなる。1998 年秋には、ルート訪問友の会のページ、および Wnn 関連ドキュメントのページで、Namazu による検索エンジンを使えるようにした。これらのページは、http://www.tomo.gr.jp/ から利用できる。著書に、1997 年ソフトバンク発行『Linux/FreeBSD 日本語環境の構築と活用』(通称「ぼうず本」) 他がある。

[4] <http://www.linet.gr.jp/~kojima/plamo/>