# 医療とコンピュータ

Vol.9 No.12

# 株式会社日本電子出版

http://www.epj.co.jp/medcom/



### 林 賢紀 農林水産研究情報センター

# はじめに

ワープロや表計算ソフトを使って日々さまざまな文書 が作成され,またパソコン通信のログやダウンロードし たWWWのページ,電子メールなども増えて行きます。 そして,増えれば増えるほど整理が難しくなるものです。 同じメーカのワープロソフトなどの文書は一つのワープ ロソフトなどで管理できますが,メーカやファイルの形 式が異なるとそれもできないことが多いです。

そこで,イントラネットなど大規模な環境では文書管 理システムを導入して文書の共有や検索を効率的に行え るようにしていますが,高価なソフトウェアが多く個人 で導入するには不向きです。

ここでは,NamazuとConceptBase Search20という安価な2 つのソフトウェアを使ってパソコン上の文書をうまく管 理する方法を紹介します。Namazuはフリーソフトの日本 語全文検索エンジンで,いくつかのファイルをインスト ールする手間はありますがすべて無償で入手できます。 また,設定も手軽です。もう一方のConceptBase Search20は 有償ですが,さまざまな文書ファイルをまとめて検索す ることができ,同時にその文書の要約も作成します。い ずれのソフトウェアも,一台のパソコン上の文書をまと めて検索することができます。 ることを第一に目指した日本語全文検索エンジン」で、 UNIX・Win32(Windows95/98/NT)・OS/2上で動作します。 コマンドラインや専用クライアントから実行できるほか, CGI(Common Gateway Interface)としても動作するので、 WWWの全文検索システムとしても使えます。

専用のクライアントはMule・Tcl/Tk・JAVA・Win32用があ ります。いずれも、フリーソフトウェアで、GPL2(Gnu General Public License version 2)に従って配布されており, 動作も軽く使いやすいので多くのサイトで活用されてい ます。バージョンアップも煩雑に行われ,多くのユーザ の意見を基により高性能になってきています。UNIX用は それぞれはプラットフォームに合わせてコンパイルして バイナリを作成する必要がありますが,Win32用はすでに 有志の手でコンパイル済みのものが配布されていますの で,より手軽に試すことができます。

ここでは、Windows95上のテキストファイルを検索する 目的で、Namazuを導入してみます。

#### まず、いくつかのファイルをダウンロードします。

いずれのファイルも,バージョンやURLは原稿執筆時 (1998年9月)のものです。最新版は「全文検索システム Namazu for Win32」(http://www.tama.or.jp/%7Ekenzo-/Namazu/) から、ダウンロードできます。

Namazu for Win32

http://www.tama.or.jp/%7Ekenzo-/Namazu/nmz 1206.exe

広瀬健一氏によってWin32用にコンパイルされたNamazu

# Namazuを使ったテキストファイルの検索

Namazuは、高林哲氏によって開発された「手軽に使え

です。

実行すると、自動的にインストールが始まります。 ・kakasi

http://www.tama.or.jp/%7Ekenzo-/Namazu/ka225w11.exe 日本語の分ち書きに使われるソフトウェアです。 実行すると、自動的にインストールが始まります。

cygwinb19.dll

http://www.tama.or.jp/%7Ekenzo-/Namazu/cwdllb19.lzh cygwin32のライブラリです。namazuやkakasiの実行に必 要です。

解凍後、cygwinb19.dllが生成されるので、

c:¥windows¥system にコピーしてください。(WindowsNT の場合は c:¥winnt¥system32)

Perl for Win32

http://www.ActiveState.com/ftp/Perl-Win32/Release/ Pw32i316.exe

テキスト処理言語PerlのWin32用です。

実行すると、自動的にインストールが始まります。 ・nkf

http://www.mirai.or.jp/%7Ecow/soft/nkf32162.lzh 日本語漢字コードの変換ツールnkfです。

解凍後に生成される nkf32.exe は, c:¥windows などにコ ピーしてください。

Search-s for Namazu

http://www.na.rim.or.jp/%7Es\_yam/library/archive/srchs072. exe

山下誠二氏が作成したNamazuのWin32用クライアントです。 実行すると、自動的にインストールが始まります。

全てのファイルのダウンロードとインストールが終了 したら、環境変数を設定します。autoexec.bat の最後に、 以下の記述を加えてください。sysedit.exe を使うと, autoexec.bat などの設定ファイルを編集することができま す。[スタート] - [ファイル名を指定して実行]で sysedit.exe と入力してください。自動的にautoexec.batを開きます。

autoexec.batを書き加えたら,Windowsを再起動して今加 えた設定を有効にすることを忘れないで下さい。 PATH=c:¥peri¥bin;c:¥usr¥local¥bin;c:¥usr¥local¥namazu¥bin;%path% SET NAMAZUCONFPATH=c:¥usr¥local¥namazu¥lib¥namazu.conf SET KANWADICTPATH=c:¥usr¥local¥lib¥kakasi¥kanwadict SET ITAIJIDICTPATH=c:¥usr¥local¥lib¥kakasi¥itaijidict

この環境変数で、Namazuのコマンドのあるディレクト リやkakasiが使う辞書の場所などを設定しています。Perl for Win32やNamazuを c:¥perl や c:¥usr¥local¥namazu 以外のデ ィレクトリやドライブにインストールした場合は、上の記 述をインストールしたディレクトリに変更してください。

これで、Namazuでデータベースを作成する準備が整い ました。

データベースを作るには、MS-DOSプロンプトから mknmz というコマンドを使います。このコマンドを実行 すると、指定したディレクトリにあるテキスト形式 (HTMLファイルもOKです)のファイルを読み込み、イン デックスファイルを生成します。

Namazuでは、日本語の分ち書きツールであるkakasiとその 辞書を元に、それぞれのファイルの文章を単語に分解して います。特定分野のファイルを扱うなどの場合は、辞書を それに特化したものにすると検索精度が向上します。 c:¥docs にあるファイルをデータベース化して検索したい

場合は、MS-DOSプロンプトから

mknmz -O c:¥usr¥local¥namazu¥index c:¥docs

のように入力します。(図1)-O c:¥usr¥local¥namazu¥index をいうオプションで、インデックスファイルを出力するディ レクトリを c:¥usr¥local¥namazu¥index にすることを示して います。出力先のディレクトリを指定しない場合は、カレン トディレクトリにインデックスを出力します。いくつかの

§MS-DOS ブロンブト 自動 💽 []] 国 📴 🖸 🔐 🗗 🗛

:¥usr¥local¥namazu¥index>mknmz -0 c:¥usr¥local¥namazu¥index c:¥docs\_

#### **図**1





叉4

データベースを作成する場合は、インデックスファイルの出力先となるディレクトリを変更してください。図2が、 インデックスを作成している様子です。ファイルが多い 場合は、それなりに時間がかかります。

作成が終了すると、データベースにしたファイルの数やインデックスの大きさ、キーワードの数などの情報が表示されます。同じ情報はc:¥usr¥local¥namazu¥index¥NMZ.logにも納められています。

インデックスされたファイルを検索するには、MS-DOS プロンプトからコマンドを入力する方法もありますが、 「search-s for Namazu」を使う方が便利です。図3が「search-s for Namazu」を起動した画面です。[キーワード:] のところ



に検索するキーワードを入力して[検索]ボタンを押すと、 検索を開始します。検索結果は、図4のように表示され、 テキストファイルならばファイル名、HTMLファイルなら <TITLE>タグの部分とそのファイルの簡単な要約やURLが 表示され、さらにタイトルやURLをクリックすると、目 的の文書全体が表示されます。この例では,過去に作成 したWebページやテキストファイルを検索対象にしてい ます。異なるデータベースを検索する場合は、メニュー バーの[検索]から[検索パスの設定]で、インデックスファ イルのあるディレクトリを指定し直してください。

一度データベースを作成した後に,検索対象としてい るディレクトリにファイルを追加した場合は,インデッ クスを作り直す必要があります。MS-DOSプロンプトから もう一度 mknmz -O c:¥usr¥local¥namazu¥index c:¥docs と入力 すると,今度は追加したり更新されたファイルだけをイ ンデックスに追加します。これを自動的に行いたい場合 は,autoexec.batの末尾にmknmz -O c:¥usr¥local ¥namazu¥index c:¥docs のようにコマンドを追加します。 これでWindowsを起動するときは常にインデックスを作る ようになります。

見い文書デ <sup>、</sup> ファイル(E)	ータベース管理ソール(自動処理 文書データベース( <u>D</u> ) 自動ダ	停止中) 処理( <u>A</u> ) 表示( <u>V</u>	)	)°(Щ)		<u> </u>
文書テ	~ 外~ス   状態		属性	) 77N/\$	自動処理	]
			実行内容	B	時刻	
	文書データベース作成	_			<u>×</u>	
	文書データベース名(№)	tzhaya				
	作成元のフォルダ( <u>S</u> )	N:¥TZHAYA			<u>後期(1))</u> ]]	
		<ul> <li>サフワオルダマ</li> <li>ローナヨイン</li> </ul>	対象にしない(E)			
	言語	<ul> <li>● 日本譜(型)</li> </ul>	() 央諸(E)			
				ОК	キャンセル	
			追加(	) 変更E	)	
			<b>2</b> 5			

💷 CBSea	rch							-	
771N( <u>F</u> )	編集( <u>E</u> )	表示⊙	検索( <u>S</u> )	履歴①	オフ <sup>*</sup> ション( <u>0</u> )	<u>ሳለን⊧ን</u> ሙ	^/l/7°( <u>H</u> )		
奥く新規	検索1>							_ 🗆 🗵	
l Q I	抽出語句	編集へ	7 y	Ø				関連語拡張	
	図書館目創	録の検索							
- 00									
<=									
$  \Rightarrow  $									
結果									
	文書データ	N-2	ローカル旧オ	語: tzha	ya DB				
			_	_	_	_	_		
									1
									111.

図6

# ConceptBase Searchによる各種 ファイルの検索

先の例では、フリーの日本語全文検索エンジンである Namazuを使ってのテキストファイルのデータベース化と 検索を行いましたが、今度は一太郎文書やMicrosoft Word、 PDFなど多様な文書ファイルの検索ができるソフトウェア を紹介します。

ジャストシステムから発売されているConceptBase Search20では、テキストや電子メールをはじめ、ワー プロや表計算、HTMLなどの文書ファイルを対象に検 索できます。

対応している文書フォーマットは、

- ・一太郎
- ・三四郎
- HTML
- ・テキスト
- Microsoft Word/Excel
- Lotus1-2-3
- さらに追加文書フィルタを導入することで
- Microsoft PowerPoint
- PDF
- Microsoft Word97/98

で作成したファイルも検索対象にすることができます。

これらのファイルが検索対象にできれば、業務などで使 用する多くの文書に対応できそうです。

追加文書フィルタは、ジャストシステムのWebサイト からダウンロードできます。

まず、「文書データベース管理ツール」を起動してデータベ ースを作成します。メニューバーの[文書データベース]から [新規作成]を選択します。ここで、文書データベース名とデ ータベースにするフォルダを設定します。(図5)複数のフ ォルダで文書データベースを作成する場合は、データベー スを設定した後に[フォルダ]タブを開いて[追加]ボタンを押 して、文書データベースに追加するフォルダを指定します。

この状態では、まだ対象とするフォルダを設定したのみ で文書データベースそのものは[未作成]になっているので、 メニューバーの[文書データベース]から[作成]を選択しま す。「ただちに文書データベースを作成しますか?」と確認 のメッセージがでますので、[はい]を選択すると文書デー タベースの作成を始めます。定期的に、文書データベース を更新したい場合は、[自動処理]タブを開いて設定します。 文書データベースの作成が終わると、[状態]が[作成済み] にかわります。これで、文書データベースの作成が終わ りました。

作成した文書データベースを検索するときは、 「CBSearch」を起動します。起動したら、まず検索対象と する文書データベースを選択します。メニューバーから [検索] - [文書データベース選択]から選びます。

文書の検索は、単語はもちろん自由に文章を入力して も自動的に解析を行って、目的の文書を選び出します。 図6がCBSearchを起動した画面と、検索語を入力している 様子です。左上のボタンを押すと、検索を開始します。

図7は「図書館目録の検索」をキーワードにした検索結果 です。ウインドウ上部に該当する文書のタイトルなどの情 報が,下部にその本文が表示されます。関連の語句は青色 で表示される他,文章の要約を表示させることができます。

CBSearchでは、検索結果として表示された文書ファイ ルを、その文書を作成したソフトウェアを使って編集す ることもできます。文書を選択してリターンキーを押す か、メニューバーから[ファイル] - [ツール起動]を選択す ると、その文書を作成したソフトウェアが起動します。

## まとめ

今回は、全文検索ソフトウェアを利用してパソコン上 での文書管理を行う手法を2つほど紹介しました。

Namazuは、初心者にはちょっと取っつきにくい面もあ りますが、自分の手元にある文書を全文検索させるには 手軽なソフトウェアです。ホームページの検索エンジン などにも応用できます。さらに、Namazuの開発者や利用 者の間でメーリングリストもあり、バグの報告や利用事

- 🗆 × CBSearch ^//7℃H (新規検索1) 図書館目録の検索() 作成者名 +  $\mathcal{D}$ \$"情報 図書館録 図書情報 9)書籍の検索 result of searching for 図書館 N. ٩. ▼∭₩ ♀!♀↓ 図書館 + トと司は書きまでもなく多様な情報にアクセスすることができますが、書種に関す では初見ません。通常、本にこいての情報を得ようとするときは書店や包囲道路を すが、インターネントでも同様です。 コモいるように、アメリカなどでは美会図書館や大学図書館の内部で運用されて オプンライン目着のPACの中にの。Pkble Access Catalogガルイシターネントに早く ています。また日本国かでも大学図書館を中心にし、Kつかの図書館がOPACを 経由で公開、利用することができます。 「、書店や出版社でも新刊書種などの案内や検索、注文などのサービスをイン 「かっています。 れてし 用オン 条件 1= ヨハの図書館目録 日本国内では多くの ミーネットトーブ E 国内の図書第日録 日本国内では多くの大学図書館が目録を電子化していますが、そのうちのしくつかはイン ターネット上でも公開しており、だれでも自由に利用することができます。 国内のインターネット上の図書館へは ・筑波大学附属図書館「他大学・観閲図書館のサーバ」 http://www.tulips.tsukuba.ac.jp/other/japan.html 日本国内だけでなく海外の國主館の一覧へもリンクしていま。

口-加旧本語:tzhaya DE

例、機能の追加なども活発に議論されています。

ConceptBase Searchでは「文書の類似性」に着目し、検索 したい言葉と検索対象になる文書との間でできるだけ類 似したものを引き出すように設計されています。このた め、高度な日本語の解析機能を持ち、入力したキーワー ドから類義語や関連語を表示するような機能を持つなど かなり高い機能を持っています。図8では「図書館目録の 検索」を分解し、「図書館」というキーワードの類義語を 表示させています。また、扱うことができる文書の種類 が多いのも特徴です。

このようなソフトウェアの活用により、パソコン上での 文書管理などを効率よく行うことができるかと思います。

ご意見やご質問などは、tzhaya@affrc.go.jp までどうぞ。

#### 参考

#### 全文検索システム Namazu

http://saturn.aichi-u.ac.jp/%7Eccsatoru/Namazu/intro.html.ja ジャストシステム ConceptBase総合ページ http://www.justsystem.co.jp/cb/index.html 日本語全文検索エンジンソフトウェアのリスト http://www.kusastro.kyoto-u.ac.jp/%7Ebaba/wais/othersystem.html

CBSearch ファイルE)編集E)表示W)検索S)履歴	田 オブションの ウ	ንለ/ኑኻና//					
文章入力へ       ⑦ ② ②         重み(4/-)       指出語句         1       校未         1       日本         1       日本	← <b>注自加←</b> 新規2週加 育態業 両取2時	類義語がスト	_ □ × 関連語拡張				

文書データベース