

の応用例の紹介にあてられている。読者は、RNA配列の2次構造が確率的文脈自由文法でうまくモデル化でき、その構造の予測に形式言語の学習アルゴリズムが役に立つという事実を知って驚くであろう。このように、最近、計算論的学習理論の研究者がデータマイニングなどの領域に続々と進出しており、強力な手法を提供している。また逆に、こうした応用研究からの要請が基礎理論の動機づけになり、新しい学習モデルや学習の手法が提案されるという流れも起きている。基礎研究と応用研究の1つの理想的な形がここにできつつある。本書は、この1冊で基礎から最新の研究の潮流をも堪能することができる良書である。

(瀧本英二／東北大学)



馬場 肇 (著)

Namazuシステムの構築と活用 －日本語全文検索徹底ガイド－

ソフトバンクパブリッシング, 2001年, 267p.
2,800円 (税別), ISBN 4-7973-1641-1

本書は、2000年11月号の「情報処理」にも掲載されたオープンソースソフトウェアであり、最近では、Vine Linux等にもバンドル搭載されるようになった全文検索システムNamazuの解説書である。本書の特徴は、導入方法を解説してあるだけでなく、基礎的な全文検索の要素技術や、検索エンジンの技術的な仕組み、検索にまつわりつく自然言語処理等も初心者にも理解しやすいということである。このため、Namazuを利用しようとする人だけでなく、これから情報検索を学ぼうとする人達にも格好の入門書である。情報検索理論や自然言語処理理論の類の書籍には難しい専門用語や数式が立ち並び、とかく最後まで読むのが難しいものである。しかし、CD-ROMも付属しており導入に必要な文書とソフトがあるので、本書に書いてある手順に従って導入をしていけば、全文検索システムというものが大筋身につくのではないかと思う。本書は、以下の構成になっている。

- 第1章 全文検索システムとは何か
- 第2章 全文検索システムの基礎知識
- 第3章 日本語全文検索システムNamazuとは
- 第4章 Namazuのインストール
- 第5章 Namazuの基本的な使い方
- 第6章 NamazuをWebサーバに組み込む

- 第7章 検索システム運用の実際
- 第8章 Namazuを個人ベースで活用する
- 第9章 Namazuの高度な使い方
- 第10章 全文検索システムの詳しい構造
- 第11章 その他の全文検索システムの紹介
- 添付CD-ROM 各種検索システム構築ツール

第1章では、全文検索システムとは何かをWeb検索を例にとり解説している。

第2章では、全文検索システムを利用するにあたって、実際問題となる、古典的検索手法の基礎で完全一致型の成熟した技術であるブーリアン検索を利用したサーチャーのための検索式の作成手法を図解で解説している。また、高度な手法として部分一致検索、正規表現検索、ユニークな近傍検索 (NEAR検索)、重み指定検索等の例も含めて、いかにして検索要求に応じた結果を取得するかを解説している。

第3章では、Namazuのソフト自体の入手方法、導入手順、仕組み、関連情報の流れているメーリングリスト、また、Namazuの導入事例の紹介をしている。フリーウェアとはいえ、官公庁、一般企業、マスコミ・出版関係、大学・教育機関等、多々の導入事例がある。実際に導入を検討している人は、いかに普及しているかが分かり、自分で運用をしてみる際の参考になる。

第4章では、UNIX系のプラットフォーム、Windows系プラットフォームで導入する際に必要なインストール前に準備する必要のあるツールと実際のインストール手順を事細かに掲載している。

第5章では、さまざまな用途のインデックスファイル (転置ファイル) を作成するmknmzコマンドの使用法、またその設定ファイルであるmknmzrcの利用方法等利用できるまでの作業が述べられている。

第6章では、NamazuをWebサーバに組み込むためのCGIの設定手順について述べられている。Webで利用できるフリーウェアのCGIを利用する際には、手間取ることが一般的に多いであろうが、NamazuのCGIに固有のトラブルシューティングも記述されている。

第7章では、大量の文書が流通して蓄積された情報の宝庫であるメーリングリストのアーカイブをHTMLに変換して公開しNamazuを利用して検索する方法について述べられている。さらにcronを用いたインデックス作成の自動化方法などメンテナンス手法も書いてありメーリングリストの管理者の人にお勧めである。

第8章は、Namazuの個人ベースの利用方法に関する解説の章である。Mule/XEmacsからもNamazuは利用できるが、クライアントの設定、さらに、メールソフト、Webプロキシとの連携方法についても手順を示してある。これらによって、個人で所蔵した文書を検索

し、ローカル環境で使うことができるようになる。文書検索システムを何に利用するのは、当然問題になる。近頃、自分でWebサイトを開設して日記などを公開している人も多いであろうが、日記システムとの連携の例の紹介があるのでぜひご覧いただきたい。

第9章は、Namazuの高度な使用方法の技術的紹介である。本システムには、インデックス作成、検索語ログの収集、文書のクラスタリング等の機能がある。いずれも、膨大なテキスト情報から必要な情報を検索するには大事な技術である。それぞれインストール後には、Unixコマンドのように使用できるので説明が分かりやすい。最近ブームになっているWeb検索で使用されているPageRank手法によるページの人気度を利用したランキング手法を測定するためのコマンドも用意されている。このprnmzコマンドは、自分で検索エンジンを開発している人がベンチマーキングに利用することもできる。

第10章では、Namazuシステムの構造だけでなく、

全文検索システムの構造とその要素技術、またシステムの評価方法等を初心者でも理解できるように解説がなされている。

第11章では、Namazu以外にも多々あるフリーの全文検索システムと商用の全文検索システムの導入サイトの紹介とそれぞれの特徴について紹介している。フリーソフトの方では、実装の方法などを比較したりしているサイトやシステム選択のコツ等も紹介している。

本書は、Namazuシステムと検索エンジンの主要な技術について余すことなく解説している。本システムは現在も進化しているが、メーリングリストで最新の情報を収集して、知識を深めることができる。さらに自らがWeb検索サイトを構築、運営したりメーリングリストで議論したりすれば、全文検索システムに関する開発や研究に充実した経験が得られるであろう。

(小金沢雄一 / (株) 東芝)



おふいすらん



最近、携帯電話を新しいのに変えた。もちろんメール機能付き。周りの友人はすでに使っていたから「今ごろ？」と冷たい反応だった。最近の携帯電話はメール以外にもさまざまな便利な機能がついている。マニュアルを読んであれこれと用途を考えた。地域情報の配信サービス、カメラもついている。Web機能もあるので飲み会のとき終電確認ができるな、とか。しかし実際には使えなかったりするだ。この間の飲み会のときのこと。すっかり話して夢中になっていた私は「あれが最終だから」と、まわりに促され、友人と2人で階段をかけあがった。しかし無常にも電車の扉は閉まって発車してしまった。友人とタク

シーで帰ろうか、と話していたところに心配したもう1人の友人から「〇〇からまだ電車があるんじゃないか」と電話。沈着冷静な友人が携帯を使って検索した結果、どうやらまだ電車があるらしいことが判明して無事に目的駅まで電車で帰ることができた。私は友人から電話がかかってきたときも携帯で検索などと思いつかなかった。私には携帯などの便利なツールよりも「頼りになる友人」の方が貴重な存在であるようだ。余談だが私はよく姉に「何度も電話したのに。」と怒られる。私はよく携帯を携帯するのも忘れる。

(五味奈津子 / 調査研究部門)