

林産試験場の文献閲覧システムについて

八 鍬 明 弘

キーワード：データベース，文献閲覧，刊行物，機関誌

はじめに

林産業界では、多くの企業が依然として続く厳しい経営状況の中、新たなビジネスチャンスの創出や経営改善の必要に迫られています。このような状況において、林産業界に対して有用な情報提供を行なうことが林産試験場の急務の一つです。

現在、林産試験場では、林産業界向けの情報提供の手段として、林産試験場研究成果発表会、林産技術交流プラザ、木製サッシフォーラムなどの各種イベントを実施しているほか、刊行物の配布やWebサイト(以下ホームページ)の運用を行っています。

このなかで、ホームページによる情報提供の効果ですが、トップページへのアクセス件数は平成10年6月から13年6月19日現在までで42,799件あり、月平均1,200件ほどになります。その対象は林産業界に限らず不特定多数となりますが、件数は機関誌(月400件程度発送)の約3倍です。しかも、印刷や発送の必要がないので、低コストで効果の高い方法といえます。また、利用者側からすると、保管するための書庫を必要とせず、必要なときに必要な情報が取り出せるという利点があります。ただし、利用者には自ら情報を得ようとする意思と、インターネットを利用できる環境が必要になります。

このため、現状では利用者が限られてしまっていますが、12年4月に実施した林産業界向けの「林産試験場における今後の試験研究業務に対するアンケート調査」では、全体の約41%の企業が林産試験場の刊行物および監修誌を情報収集源としていることに加え、64%の企業が現在インターネットを業務に活用している、もしくは今後利用予定であるという結果が出ています¹⁾。このことから、林産業界においても利用者は確実に増加していくはずです。

そこで、林産試験場では12年度に、これまでの刊行物の中から定期的に発行してきた機関誌(林産試だより、

林産試験場月報、林産試験場報、林産試験場研究報告)に掲載された文献を林産試験場のホームページで閲覧できるシステムの開発を行ないました(図1)ので、その概要と使用方法を紹介します。

なお、このシステムの開発は、「北海道緊急地域雇用特別基金」を活用した「木材関連情報普及促進事業」によって業務委託して実施しました。

システムの概要

このシステムでは、利用者はインターネットを介してパソコンの画面上で林産試験場の文献を閲覧できます。これを活用しやすくするためには、数多くの文献の中から必要とする情報を短い時間で効率よく得るための検索システムが重要になります。

従来から文献検索の方法としては、書誌情報(タイトル、著者、発行年など)を検索対象とするシステムが広く普及していますが、検索用データの作成や検索速度が速い反面、文献に記載されている内容から検索することができないので、必要とする情報になかなかたどり着けない場合もあります。そこで、書誌情報による検索システムに文献の全文を検索対象とするシステムを加えて、検索効率を上げることを考えました。

このための具体的な作業は、12年3月までに発行した機関誌に掲載された文献約3,200件23,000ページ(表1)を電子化する文献データ化と、書誌情報と全文検索システムおよび文献の全文閲覧システムを開発することでした。開発にあたり、利用者の立場から、検索方法が単純で効率よく行なえること、できる限り掲載誌面のレイアウトイメージと近い形で文献データが表示されること、表示が極端に遅くならないこと、また、サービス提供者の立場から、データの更新方法が複雑にならないことに留意しました。

(1)文献データ

機関誌に掲載された各文献は、PDFファイル形式の

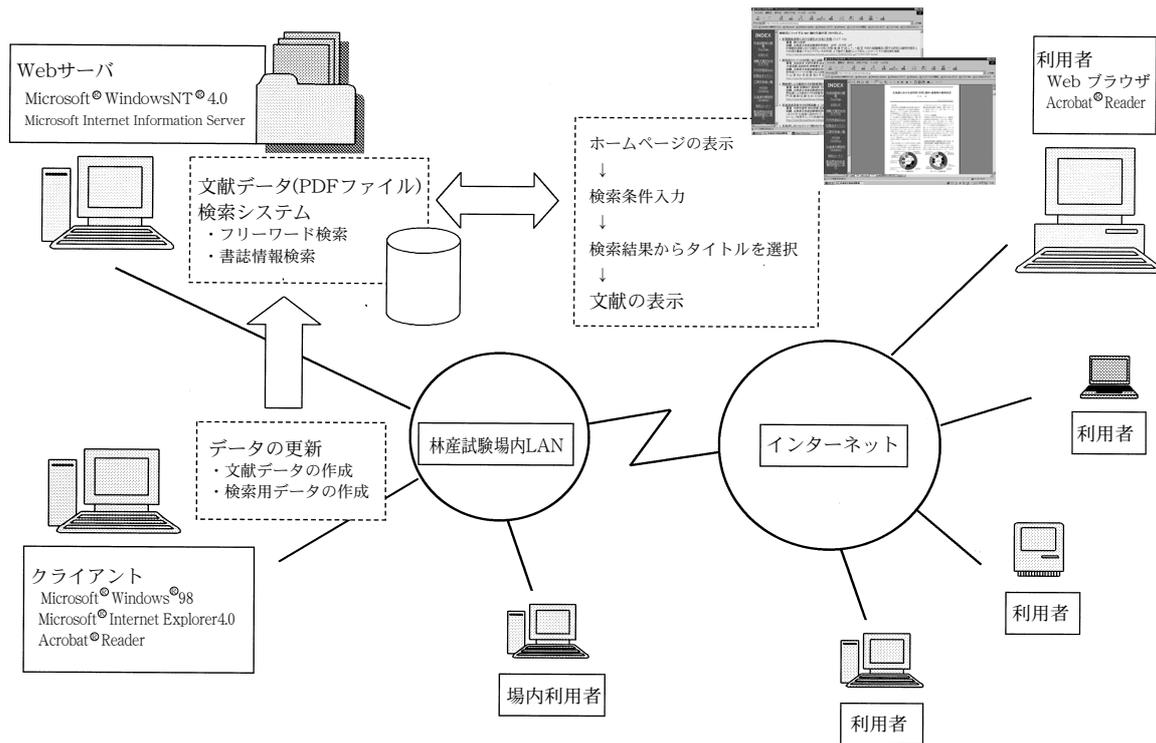


図1 文献閲覧システムの概要

表1 文献データを作成した件数とページ数

機関誌名	発行年	文献件数	ページ数
林業指導所研究報告, 林産試験場研究報告	1952～	195	6,045
林業指導所月報, 林産試験場月報	1965～1986	1,720	9,850
林産試験場報	1987～	266	2,340
林産試だより	1980～	1,015	4,808
計		3,196	23,043

文献データとしました。PDFファイルとは、アメリカのアドビシステムズ社が開発したファイル形式です。無料配布されているソフト（フリーウェア）「Adobe Acrobat® Reader」を使用することによって、文字・画像・レイアウト情報などのデータを表示・出力できます。

文献データの作成作業ですが、まず誌面をスキャナで画像データとしてパソコンに取り込みました。誌面上には文字と図表・写真などの情報がありますが、文字の部分はOCR(optical character recognition)ソフトでテキスト形式(文字データ)に変換しました。誌面どおりにタイピングすることでも同じ文字データを作成できますが、人為的なミスの防止や作業時間の短縮のためにこの方法を用いました。できあがった文字データと画像データ(図表・写真)は、できる限り掲載誌面のイメージに近づけるようにワープロソフトを用いてレイアウトしました。

古い文献の中には紙質が悪くスキャナで読み取れない文字や、読み取れたとしても現在のかな・漢字に変換できない古いものもありました。これらの場合、OCRソフトを用いず誌面をすべて画像データにしました。しかし、画像データのままでは後述する全文検索ができません。そこで、同じファイル内に、この画像データとは別にタイピングした文字データを追加して、全文検索に対応できるようにしました。

この段階で利用者がデータを閲覧するためには、データを作成したワープロソフトが必要になりますが、そのためにソフトを準備するのは現実的ではありません。また、一般的にホームページの作成に用いられているhtml形式ではレイアウトに制限があり、掲載誌面のイメージに近づけることは困難です。したがって、これらすべてのデータをPDFファイルとして保存することで、これらの問題の解決を図りました。

こうして作成したPDFファイル形式の文献データは、サーバのハードディスク内の文献種類別(林産試だより、林産試験場月報、林産試験場報、林産試験場研究報告)のフォルダ内に格納しました。

(2)検索システム

検索システムは2つの方法で構成されています。一つは、書誌情報(タイトル、著者名、刊行年)による検索方法で、もう一つはフリーワードによる検索方法です。

利用者はどちらかの方法を選択し、検索条件を入力すると該当する文献タイトルが表示されます。以下に一連のプログラムの動作とサービス提供者側の作業を説明します。

①書誌情報による検索

ホームページ上で書誌情報に関する検索条件を入力すると、条件が一致した文献タイトルの一覧が表示されます。

サービス提供者は、あらかじめ検索対象となる書誌情報データをCSV(Comma Separated Value)形式で作成します。このデータ形式はレコードとレコードを改行コードで区切り、レコードのフィールドをカンマコード(,)で区切るデータ形式です。機関誌が発行されるたびに、この書誌情報のCSVファイルを更新する必要があります。

②フリーワードによる検索

ホームページ上で全文を対象とした検索条件を入力すると、条件が一致した文献タイトルの一覧が表示されます。このシステムの特徴は、検索エンジンに「Namazu」を使用していることです。「Namazu」は、使いやすさ、検索スピードの速さなどから多くの企業や団体、自治体で使用されているフリーウェアです。詳細は「Namazu」のホームページ (<http://www.namazu.org/>)を参照してください。

サービス提供者は、機関誌が発行されるたびに「Namazu」のインデックス作成プログラム(インデクサ)を実行してインデックスデータを更新する必要があります。インデクサは、自動的にフォルダ内に格納してある文献データから文字データを抽出し、インデックスと文献タイトルのリンク情報を含むファイルを作成します。

(3)文献の閲覧

このシステムの利用にあたっては、ホームページを見るためのWebブラウザが動作可能な環境であればパソコンやソフトの種類は限定されませんが、文献の閲

覧のためにあらかじめAcrobat Reader4.0以上をインストールしておく必要があります。

検索の結果、表示された文献タイトルの一覧から閲覧を希望するタイトルを選択すると、Acrobat Readerが起動してPDFファイルで保存してある文献データが表示されます。

Acrobat Readerには、文書内の文字検索などの機能があります。詳細はAcrobat Readerのヘルプを参照してください。

使用方法

はじめに、パソコンでWebブラウザを立ち上げ、林産試験場ホームページ (<http://www.fpri.asahikawa.hokkaido.jp>)を開きます(図2)。このページの左フレームの中にある「刊行物Data Base」の項目をマウスを操作して選択(クリック)すると刊行物Data Baseのページが表示されます(図3)。ここで「フリーワードによる検索」または「書誌情報による検索」を選びます。

「フリーワードによる検索」は文中の語句を検索対象にしているので、広範囲な検索が可能ですが、検索結果が多く表示される傾向があるので、得たい情報を表示するまでに時間を要することもあります。

「書誌情報による検索」は、閲覧したい文献の見当がついている場合には有効で、短い時間で得たい情報にたどりつけます。

(1)「フリーワードによる検索(全文検索)」の場合

テキストボックスに「木材」などの単語を入力して検索ボタンをクリックします。対象刊行物をチェックすることで検索のしぼり込みが可能です。

検索条件として、テキストボックスに一つの単語を入力するのが最も簡単な方法ですが、たくさんの検索結果が表示される傾向があります。効率的に検索できるように、このシステムには複数の単語の間に「AND」、「OR」、「NOT」などの文字を入力して検索条件をしぼり込む機能があります。さらに名詞や形容詞に限らずフレーズを構成する単語であれば入力できますので、これらも組み合わせることでより効率的な検索が可能となります。

検索結果には、検索条件と一致した文献のタイトルとファイル名が青字で表示されます(図4)。どちらかを選択するとAcrobat Readerが起動して文献を表示します(図5)。



図2 林産試験場ホームページのトップページ

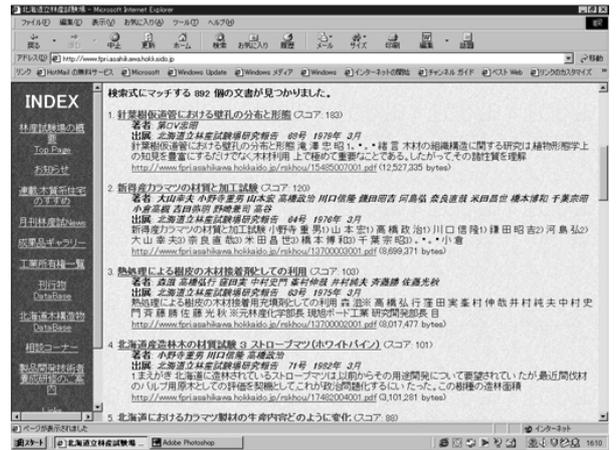


図4 フリーワードによる検索結果

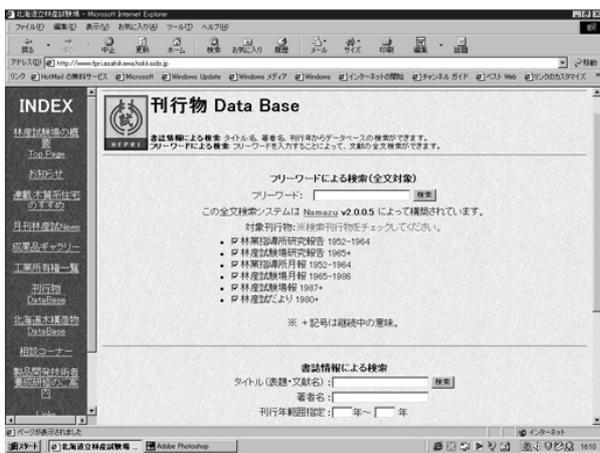


図3 刊行物Data Baseのページ



図5 文献の表示

(2)「書誌情報による検索」の場合

タイトル(表題・文献名), 著者名, 刊行年範囲指定のテキストボックスに検索対象語を入力します。それぞれの条件がAND検索されます。一つのテキストボックスに入力するだけでもかまいません。

検索結果は、検索条件と一致した文献のタイトルとファイル名が青字で表示されます。これは、フリーワードによる検索結果の表示とできるだけ同じデザインとしました。どちらかを選択するとAcrobat Readerが起動して文献を表示します。

おわりに

システムというものは、常に時の流れとともに変化していきます。この文献閲覧システムも陳腐化しないように、データの更新はもちろんですが、システムの新たな利用法なども常に念頭において、必要に応じて形を変えていくような柔軟性を持たせていく必要があ

ると考えています。

なお、以上の動作の確認は、サーバ、クライアントともPC/AT互換機で、サーバのOSはMicrosoft® WindowsNT®4.0, クライアントのOSはMicrosoft® Windows®98, WebブラウザはMicrosoft® Internet Explorer4.0で行いました。

参考資料

1) 竹花邦夫, 遠藤展: 林産試だより, 2000年12月号, 4-7(2000).

(林産試験場 普及課)

AdobeおよびAcrobatはAdobe Systems Incorporated(アドビシステムズ社)の商標です。また、Microsoft, WindowsNTおよびWindowsは、米国Microsoft Corporationの、米国およびその他の国における登録商標または商標です。