

図書・各種資料を含む情報共有について

小林 昭夫

総合研究開発機構(NIRA) 政策研究情報センター

Information sharing system which handles various types of documents and records

Akio Kobayashi

National Institute for Research Advancement(NIRA)

Abstract

In most libraries, people can search for books and periodicals using computerized bibliographic databases. However, the information needed may not necessarily be found in the books or periodicals searched. Valuable and useful information can often be found in such documents as conference materials and pamphlets. If searching for these documents through bibliographic databases were possible, then libraries could respond to the ever-increasing demands for more information. In this paper, I propose a Bibliographic Information Sharing System as a new means of information search to be used by libraries. The Application Service Provider (ASP) model was introduced to simplify system-designing. I believe that the ASP model will enable us to share information stored in different locations and that the model will be useful for various types of information sharing.

1. はじめに

情報を検索することを考えた場合、組織内の蔵書管理システムや各種のデータベースあるいはインターネットのディレクトリ型サイトやサーチエンジンのサイトの利用が考えられる。蔵書検索やデータベース利用とは別にポータル・サイトやサーチエンジンも多く利用されている。

本稿は図書検索による情報提供の考え方ををベースに蔵書のみでなく各種資料まで検索対象に含め、より広い範囲の情報を簡易な操作で提供するシステムを考える。資料範囲は一般図書・定期刊行物に限らず、パンフレット、会議資料、カタログ、新聞記事の切り抜き等の各種資料を想定する。

取り扱い資料範囲が広いため登録される情報は多義にわたると考える。このため一つの組織や部署程度の同一業務範囲内が適正ユーザー規模となろう。対象としては資料や蔵書等の点数があまり多くなく中小規模の組織内での利用やセクション内あるいはいくつかの出先を含む範囲を想定している。

情報を必要としている人は、書籍や定期刊行物などの形態にこだわっているわけではなく、どのような形態のものであれ情報を得ることが重要である。情報を提供できれば問題ないが、少なくとも情報へのポインター（所在情報）を示すことでも情報の手がかりを与えられるわけであり情報の所在も重要であると考え。

2. 情報共有計画

情報共有システムを構築するにあたり最初に情報共有計画を立案するものとする。収集する資料範囲を決定し、登録する項目（図書で言えば書誌項目）の選定を行う必

要がある。登録項目において、あまり項目数を増やすと入力作業量が多くなりシステムの維持に多くの労力が必要となる。このためできる限り項目を少なくする方向を考える。

原則的には Dublin Core の基本 15 エlement (表 1) を標準とすべきと考える。15 エlement すべてを登録項目とするのではなく、この中より必要な Element のみを抽出して運用を行う。

表 1. Dublin Core の 15 Element [1]

No.	ラベル	要素名	No.	ラベル	要素名
1	Title	タイトル	9	Format	形 式
2	Creator	著者あるいは作者	10	Identifier	資源識別子
3	Subject	主題および キーワード	11	Source	情報源(出处)
4	Description	内容記述(概要)	12	Language	言 語
5	Publisher	公開者(出版者)	13	Relation	関 係
6	Contributor	寄与者	14	Coverage	対象範囲
7	Date	日 付	15	Rights	権利管理
8	Type	資源タイプ			

図書・各種資料を含む情報共有システムは情報そのものでなく情報へのポインターを示すタイプの登録も視野に入れているため Dublin Core の基本 15 Element 中には定義されていないが「所在情報」を付加するものとする。最小単位の例としてタイトル、内容記述、日付を考え所在情報を付加する。(表 2)

表 2. 最小登録構成例

No.	ラベル	要素名
1	Title	タイトル
4	Description	内容記述(概要)
7	Date	日 付
—	Location	所在情報

最小登録構成例に限らず内容記述(概要、description)の登録内容が検索時に重要となってくると予想されるが本 Element の登録は大きな課題と考える。書籍のようにタイトルのみの登録である程度資料の性質を現すことができるものもあるが、会議資料等のようにタイトルのみでは内容が現しにくい性質の資料群も存在する。資料の概要を入力できれば理想的であるが、資料範囲が広いとすべての資料について要約を作成することは大きな労力を必要とする作業であり、コスト的にも難しいと考える。

概要作成は作成者による個人差が発生する可能性もあり、目次や資料の先頭から規定文字数の入力、規定文字数までの名詞の入力という誰でも可能な手法も考えられる。また、目次ページ等を OCR (光学式文字読みとり) 処理を行い内容記述とする方法も考えられる。文字認識は完全ではないものの現行製品の精度は 95~99% と言われており誤認識を考慮しても、検索利用においては実用になり得ると考える。

3. 情報提供形態

資料検索のユーザーインターフェースはホームページ閲覧で利用されているブラウザを考える。ブラウザであれば各人の操作の習熟度が高く気軽に検索が行える利点がある。利用者はブラウザより登録情報の検索やメンテナンスを行う。(図1、図2)

情報共有サーバーとしてASP (Application Service Provider) サーバーモデルを考える。ASPはモデルとして採用し事業者側にサーバーを設置しても良いし、組織内部に自前のサーバーを設置しても問題ない。ASPモデルであるためサーバーに接続できれば組織内で容易に情報共有が可能であり、サーバーがインターネットに接続したりリモートアクセスの機能を提供できるなら遠隔地よりの利用も可能となろう。

登録情報を維持するために新規入力・更新・削除などのメンテナンス作業が必要となる。一般的にはサーバー側にデータベースシステムを実装しデータを管理するのが望ましいが、図書・各種資料を含む情報共有システムでは厳密なSQLでの検索条件設定の機能が無くても、指定した文字列の部分一致検索のみで必要な情報を提供できる場合が多いと考える。

データメンテナンスにおいても対象の情報の規模が大きくないため Microsoft Excel 等を使用しパソコン側で管理することとする。新規入力や更新・削除を Excel 上でおこなうこととし、この Excel ファイルを原本とする。同時複数人でのメンテナンスはできないが、規模が小さいことを考えると複数人の専従者がメンテナンスを行うことはないと考える。複数人で入力した場合は最後にファイルをマージする。

データの登録は Excel の原本を CSV 形式ファイルで保存し、ブラウザの画面を経由してサーバーに送信するものとする。サーバー側でデータは毎回すべて上書きで置き換えることとする。このことによりデータのメンテナンスを簡易に行うことが可能となり、データベースメンテナンス系のシステム開発が不要になると考える。

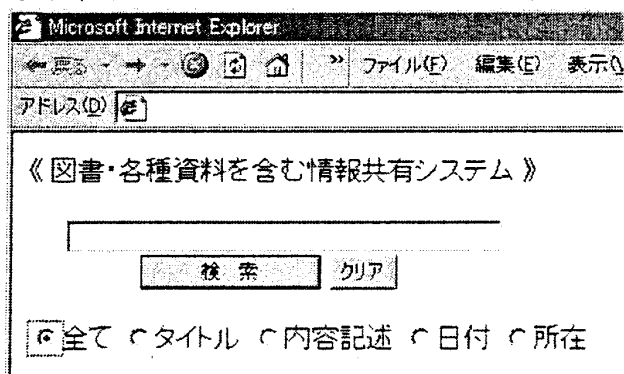


図 1. 検索画面例

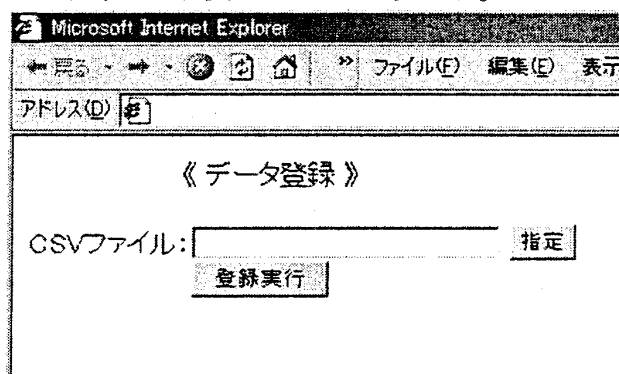


図 2. データ登録画面例

4. 実行環境

サーバー側の動作環境を考えると UNIX 系 (BSD, Linux 等) か Windows かということになる。UNIX では UNIX + apache 構成が一般的であり Windows では IIS (Internet Information Server) となる。apache は web (HTTP) サーバープログラムの名称であり、IIS もホームページなどのサービスを提供する仕組みとなる。この HTTP サーバーより呼び出される CGI (Common Gateway Interface) や ASP (Active Server Pages) で実現する。

サーバー側のデータファイルが 1 データ 1 行になっており UNIX 標準の grep という

文字列検索コマンドを使用して CGI を Perl で記述すると以下の例示となる。(図 3)

```
#!/usr/local/bin/perl
my $filename = "/tmp/data";          # データファイル
my $keyword = $results{'KEYWORD'};  # 実際には form から取得となる
printf "<html>%n";
printf "<body>%n";
open(RESULT, "grep $keyword $filename|") || die;  # 検索実行
while (<RESULT>) {
    printf $_;                        # 結果の書き出し
    printf "<br>%n";
}
close(RESULT);
printf "</body>%n";
printf "</html>%n";                  # HTML ヘッダー等略している
exit(0);
```

図 3. grep コマンドを用いた例

図 3 の通り非常に簡易に記述することができる。項目を指定した検索はできないがこのスクリプトレベルでもサービス提供が可能である。データベースシステムにデータを登録している場合は grep コマンドの代わりに SQL でデータベースに問い合わせをおこなうこととなるが、原理は grep コマンドと同様である。

組織内部の限られた用途であれば内部でサーバーを構築することも考えられる。例として安価な構成を考えるとサーバー: DOS/V パソコン(10 万程度、Windows 不要)、OS: FreeBSD, Linux(無料)、HTTP サーバー: apache(無料)、データベース: MySQL(無料)でも構築が可能である。内部利用のみであればセキュリティにそれほど気を使う必要も無いと考える。

5. 終わりに

図書・各種資料を含む情報共有システムは各種資料等の情報共有を目指すものである。応用分野については各種用途が考えられる。

- ・ 部署での資料管理、資料共有
- ・ 本支店や工場間での資料共有
- ・ 組織内での資料管理、資料共有
- ・ 公的機関の文書検索サービス
- ・ 小中学校・保育園・幼稚園の書誌情報共有

情報共有システムは簡易にシステムが構築可能で運用が簡単であるという特徴を持つが、組織内すべてや外部機関とのデータ共有、インターネットにサーバーを接続するのであれば細かいデータベース周りの実装やセキュリティに配慮したシステムの構築の必要がある。

情報共有システムでは最初に情報共有計画を立案し資料範囲や登録項目、登録内容等を決定することが重要である。運用においてはなるべく作業量の増えない方向を考え、誰もが簡易に情報を管理でき、しかも共有できる環境を目指していく。

参考文献

[1] Dublin Core Metadata Element Set について - 現在の状況と利用例. 杉本重雄. 図書館情報大学.