

図書情報の電子化と WWW サービスシステムの作成と評価

RDBMS と WWW の連携による情報サービスシステム

武山 慎一
a625@ccedu.sendai-ct.ac.jp

海野 啓明
kaino@cc.sendai-ct.ac.jp
仙台電波工業高等専門学校

矢島 邦昭
yajima@cc.sendai-ct.ac.jp

近年電子図書館あるいはデジタル図書館に関する研究は盛んに行われているが、これらのシステムには図書館と言う概念にとらわれたものが多いと思われる。

本研究では RDBMS による電子化された図書情報と、インターネットで提供されている膨大な量の情報を有機的に結合することで、より高度な情報サービスを提供するシステムを開発する。

巨大データベースから知識獲得を高速に行う手段として盛んに研究されているデータマイニングの概念と、ネットワークエージェントの技術を結びつけ、新たにインフォメーションマイニングと言う概念を提案すると同時に、この概念の実装モデルであるライブラリアンエージェントについて述べる。

Electrification of Library Information, and make and evaluate the WWW Service System Information Service System in cooperation with RDBMS and WWW

Shinichi TAKEYAMA
a625@ccedu.sendai-ct.ac.jp

Keimei KAINO
kaino@cc.sendai-ct.ac.jp
Sendai National College of Technology

Kuniaki YAJIMA
yajima@cc.sendai-ct.ac.jp

In recent years, study of an electronic library or a digital library is very expanded. It is considered that these systems were caught by the concept called library.

In this study, the system which provides more advanced information service is developed by combining organically the processed books information by RDBMS, and the information on a huge quantity currently provided in the Internet.

We are closely connected with the concept of the Data mining currently briskly studied as a means to perform knowledge acquisition at high speed and a network agent's technology are connected from huge data base. So We proposal that the concept of information mining is proposed newly, the librarian agent who is the mounting model of this concept is described.

1. はじめに

電子図書館あるいはデジタル図書館に関する研究として、日本とフランスが共同幹事国となって進めているG7電子図書館プロジェクト¹⁾や、国立国会図書館を中心に行われているパイロット電子図書館プロジェクト²⁾等が挙げられるが、いずれも図書館の機能の電子化を中心としている。これらの研究ではインターネット上でWWW(World Wide Web)を用いたのサービス実験などが行われているが、WWWと密接に連携しているとは言い難い。また、学術情報センター電子図書館サービス(NACSIS-ELS)では実際にサービスが行われているが³⁾、学協会の発行する学術雑誌のページをそのまま画像データとして蓄積し、書誌情報とともに検索できるようにした情報サービスに限定されている。

視点を変えればWWWはそれ自身が巨大な図書館であると考えることができる。図書館にある図書情報とWWWで提供されている情報を有機的に結合させることにより、より多くの有益な情報が得られる可能性がある。

仙台電波高専の図書館には、5万冊あまりの蔵書があるが、これらの図書情報は図書目録を主とした紙のカードに書き移して管理している。図書目録は電子化された目録情報から、これまでの紙のカードにして管理している。その為、図書館司書の業務にかかる負担も大きく、利用者は非効率的な図書検索を強いられている。これらのことから、電子情報の有効利用と業務の簡略化が求められている。

そこで本研究では

1. RDBMS (Relational DataBase Management System) を用いた図書情報の電子化
2. WWW を利用することによるユーザインターフェイスの向上
3. WWW を利用することによる情報発見率の向上

を目的とする。

これらに対し、近年盛んに研究されているネッ

トワークエージェントの技術を用いることで、図書館司書の業務支援、図書館利用者のインフォメーションマイニング(Information Mining)支援とするものである。

2. インフォメーションマイニング

近年、巨大データベースから知識獲得を高速に行う手段として、データマイニング(Data Mining)がデータベースと人工知能の境界分野で注目されつつある⁴⁾。

データマイニングは、巨大データから必要な知識データを発見する技術のことをす。われわれの生活の中では、本やテレビなどのマスメディア、WWW, Netnews, E-mail, また、人との会話を通じて日々の生活の中で文字、画像、音声情報が情報として入力される。これらの情報から自分が必要な情報だけを見つけ出すことは、情報化社会といわれる現代においては非常に困難になってきている。たとえば何か欲しい情報をインターネット上で探す場合、大抵はInfoseek⁵⁾やYahoo⁶⁾などのWebページ検索サービスを利用するだろう。

多くの検索エンジンはWebページ中に含まれる文章からキーワードで指定された語を探し出し、検索者に対して報告する。しかしながら、これらのシステムは単語が一致するページの検索であり、情報の一致ではない。

また、現在多くのシステムでは画像検索は実現されていない。これらの問題を解決するためには文書解析や画像解析、インデックス付け等の技術が必須である。

そこで我々は、現在主流となっている指定キーワード一致型の検索方法とは異なるアプローチとして、サーチエンジンや辞書サービスを利用して得た情報から、情報検索者に最適化された情報をもたらすための技術を提案する。この技術をインフォメーションマイニングと名づけた。

3. WWW と RDBMS

ユーザインターフェイスがシステムと密接な関わりを持つことは言うまでもないが、新たなユー

ザインターフェイスを考察するために、現在使われているシステムの問題点について考える必要がある。

まず、現在主流となっている C/S(Client/Server)システムにおいて DB(DataBase)を利用する場合におきる問題点について検討する。また、WWWから外部のプログラムへアクセスする手法として主流となっている CGI(Common Gateway Interface)が持つ問題もあわせて検討する。

3.1 Client/Server システムの問題

今日では多くの企業で DB(DataBase)を中核とした C/S システムが構築、利用されているが、これらのシステムには共通した特徴がある。それは使用するハードウェア、アプリケーションプログラム、クライアントとサーバ間の処理(データ)分散、サーバの負荷分散等が全て設計者に任されているという点である。このため以下のような問題が発生している。

- 設計者によりサーバインターフェイスやクライアントのシステム要件にばらつきがある。
- データの構造情報やサーバのインターフェイスが公開されていない場合や特殊な設計の場合、ユーザは自由にアクセスできない。このため、純粋な EUC(End User Computing)を行うことができない。
- パフォーマンスの確保のためにクライアントの負荷を高くしたり、クライアント数を制限したりして、システムの拡張性を犠牲にしている場合がある。
- C/S システムの開発生産性を向上させるために採用したミドルウェアやアプリケーション開発ツールが、システムのボトルネックを作り出す場合がある。

以上のような問題から、現在利用されている C/S システムは、分散システムとしてうまく機能していない場合がある。また、システムの透過性が低く、拡張も難しくなっている。このため、業務の生産性の向上が図れない場合が多くなっている。

3.2 WWW と RDBMS の連携

上記の問題を解決するため、本研究では RDBMS と WWW を連携させ、サーバの処理には Web サーバ + RDBMS、クライアントには Web ブラウザを利用するシステムを作成・評価する。

WWW の基本的な理念は、サービスを提供する側もユーザ側も同じ技術と利便性を得られる点にある。ユーザは WWW の技術に対応したブラウザを用意すれば、どの Web サーバにもアクセスできる。これに対しサービスを提供する側はユーザが情報を取りに来る為、Web サーバ上に最新の情報を用意するだけでよい。

以上の事から、C/S 間の分離が徹底でき、以下のようなメリットが享受できる。

- サーバの構築や拡張が容易に着手できる
- クライアントに左右されない為、クライアントを管理する必要がない

3.3 CGI を使用する場合の問題

CGI は Web サーバからプログラムを呼び出すインターフェイスである。機能的には単体のコマンド発行に等しく、プログラム間の連携や状態管理などの様にシステム管理を意識したものではない。この為、CGI だけで既存システムにアクセスすることは、アプリケーションを効率的に動作させたり、セキュリティに依存したサービスの構築などの面で機能が不十分である。また、CGI は単なるインターフェイスであり、効率のよいプログラミングを支援する技法や開発ツールは見当たらない。従ってアプリケーション開発の生産性や保守性の面で困難が予想される。

また、以下の点からも CGI プログラム部分と既存システム部分を分離する必要がある。

- CGI プログラムでは DB セッションを保持できないため、連続して DB にアクセスする対話型プロシージャを処理できない。この為、CGI プログラムを起動するたびに新しく DB との冗長な SQL のやり取りが必要となる。
- DB アクセスに対する認証は基本的に DB マネージャ側で行うため、認証完了までの Web サーバと DB 間の負荷が高くなる。

Web サーバと DB は CGI で直接リンクさせることはできるが、セキュリティ上の理由から静的に 1 対 1 にロックされるため両サーバの分散が難しい。また、ネットワークポロジの変更や DB の移動の影響をアプリケーションが直接受けることになる。

このため、システムの拡張性を確保しようとすると、システムの設計が難しくなる。

4. ライブラリアンエージェント

現在ネットワークエージェントに関する研究は、ネットワーク社会を支援する新しい知能メディアとして盛んに行われている⁹⁾。

一般に言うネットワークエージェントは、人間の代理としてネットワーク上の情報収集や流通を支援するソフトウェアエージェントである。ここでは主にインターネット上の情報収集を行うインターネットエージェントとしての機能、更にデジタル (あるいはバーチャル) 図書館司書としてのライブラリアンエージェント (Librarian Agents) 機能について考察する。

4.1 システム概要

WWW と DB を連携させる際の問題を解決するために現在多くの技術が開発されているが、未だ標準化されていない。SQL-92 に対応したものや、独自拡張したもの、Meta-HTML を利用したものなどがある。

本研究ではプロトタイプとして Windows NT Server に付属の IIS(Internet Information Server)と Microsoft SQL Server を連携させるシステムを作成・評価し、今後のシステム構築の基

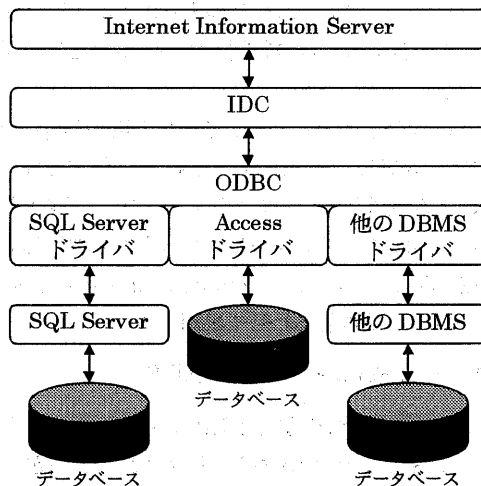


図 1 WWW+RDB システムの概念図

礎データとする。

図 1 にシステムの概念図を示す。

IIS では IDC(Internet Database Connector)と呼ばれる機能が提供されており、IIS から DB へアクセスする際に使用されるインターフェイスである。これは、ODBC(Open DataBase Connectivity)を利用して接続されるようになっている。

この方式を選択した主な理由として、ODBC が SQL Server 以外にも多くの RDBMS でドライバが用意されていることが挙げられる。今後正規のシステムを作成する際に技術を応用でき、今後予定している Java ベースのシステムへの移行も容易であると考えられるからである。

これにより、JDBC を利用したプラットフォームに依存しないシステムとすることができる。また、SQL-92 の機能拡張として、ODBC をもとにした SQL/CLI(Call Level Interface)が追加されていることから、今後 SQL3 使用に準拠した RDBMS が出てきた場合にも対応が容易であると思われる。

クライアントは Web ブラウザならば、どのブラウザでも使用可能である。

4.2 ライブラリアンエージェント

ライブラリアンエージェントは、図 2 に示すよ

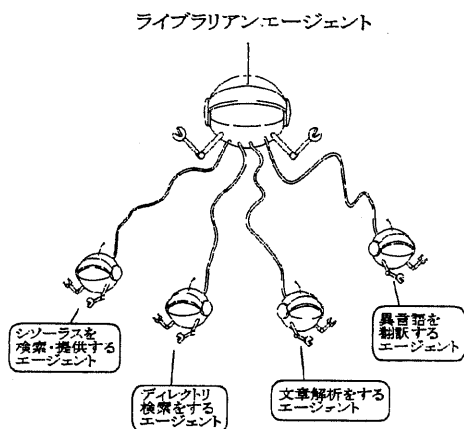


図 2 ライブラリアンエージェント

うな複数のサブエージェントからなるマルチエージェントシステムである。

サブエージェントにはシソーラスを検索・提供するエージェント、ディレクトリ検索をするエージェント、文書解析をするエージェント、異言語を翻訳するエージェントなどが存在する。

これらの内、ディレクトリ検索をするエージェントは、インターネット上ですでに稼動しているサーチエンジンを対応づけることができる。また、シソーラスを検索・提供するエージェントや異言語を翻訳するエージェントは CD-ROM などで出版されている辞典類や、インターネット上で提供されているサービスが応用できると考えられる。これらのエージェントが相互に連絡を取り合うことにより、ユーザが探そうとしている情報を分析・予測し、検索する。

文書解析をするエージェントはライブラリアンエージェントの核となるエージェントである。サーチエンジンから得られたアドレスをもとに Web 文書を解析し、ユーザが必要としている文書なのかどうかを判定する。

4.3 ユーザインターフェイス

ユーザインターフェイスとしては様々なものが考えられる。図書カードの検索をイメージしたカ

ード型検索⁹⁾や、CG 司書¹⁰⁾などが既に研究されている。

インターフェイスとして、いずれも利用者が感じる心理的抵抗は少ないと思われるが、エージェントとしての特性を考えた場合、CG による擬人化したインターフェイスがより望ましいと思われる。

5. 進捗状況

プロトタイプシステムとして PC サーバ 1 台に IIS と SQL Server を構築し、クライアントとして PC, Macintosh, UNIX マシンを使用した。

先ずデータに関しては、図書館の電子化に際して、既存のデータをどのように取り込むかが大きな課題となった。

既に電子化されているデータは独自の形式のテキストファイルであるため、データを変換し DB へ格納するためのプログラムの開発から始めた。開発言語には Visual Basic 4.0 を使用し、ODBC を使用してデータ操作をするものとした。

ユーザ及び WWW・RDBMS 間のインターフェイスに関しては、WWW サーバから RDBMS へアクセスするページを作成した。

現在はテスト段階のため、非常にオーソドックスなキーワード検索型のインターフェイスになっている。また、CG による擬人化したインターフェイスの開発も同時に進めている。

エージェントに関しては、ライブラリアンエージェントの 1 機能である文書解析をするエージェントの開発に取り組んでいる。

6. 終わりに

本研究は文部省によって平成 9 年度から始められた MUP(Multimedia University Pilot)事業の「マルチメディア教育研究用データベース利用システム」構築の一環として行われている。

参考文献

- 1) <http://www.ndl.go.jp/G7/index.html>
- 2) <http://www.cii.ipa.go.jp/el/index.html>

- 3) <http://www.nacsis.ac.jp/els/els-j.html>
- 4) 福田,森本,森下,徳山:データマイニングの最新動向, 情報処理, Vol.37, NO.7, pp597-603 (1996)
- 5) <http://www.infoseek.co.jp/>
- 6) <http://www.yahoo.co.jp/>
- 7) 磯辺裕一:Web サーバーと既存システムの連携, UNIXUSER Vol.5 No.12(1996)
- 8) 西田豊明:ネットワーク社会とエージェントー擬人化された人工システムー, 情報処理, Vol.38, No.1, pp10-16 (1997)
- 9) 服部文夫:ネットワークエージェントによる情報収集と流通, 情報処理, Vol.38, No.1, pp30-35 (1997)
- 10) 野田尚志,旭敏之,芝勝徳:カードメタファ図書検索ユーザインターフェースの実験的評価, 情報処理学会第 53 回全国大会, 2T-6 (1996)