

2 これまでの研究成果と今後の計画 ('97、'98)

電子図書館システムに関わる研究テーマは、前章で述べた通りであるが、これらの中から現在研究開発室では、以下のテーマについて重点をおき研究開発を進めている。

- 電子図書館システムのアーキテクチャ
- メタデータ
- 分散型講義ライブラリ
- 概念を用いた学術文献の検索
- プロキシサーバの評価

本章では、これら研究開発室において進行中の研究テーマについて具体的に述べる。

2.1 電子図書館のアーキテクチャ

2.1.1 現在の電子図書館の問題点

電子図書館は研究段階から実用段階へと進み、プロトタイプのシステムから現実に利用されるためのシステムへと進化を続けている。そして、電子図書館のコンテンツが増えるにつれ、従来のプロトタイプではありえなかった新たな問題に直面している。その中でも最大の問題と言えるのが規模の問題である。我々は今後の電子図書館の規模に関する考察を行い、新たな電子図書館のアーキテクチャに関する将来像を探ることを目的として研究を行っている。

本学の電子図書館のコンテンツは年々増加し、現在では60万ページ以上のコンテンツを持ち、今後は電子形態(CD-ROMなど)での出版社側からの提供が増えているためその増加率はますます加速されると考えられる。

さらに、電子図書館の内部に蔵書を持つのではなく、利用の権利を購入し、適時出版社の用意したサーバにあるデータを取得するようないわゆるオンラインジャーナル形式も電子図書館のコンテンツの一部として広く利用されることが望まれている。この方法は出版者がデータのコントロールを容易に行えるという利点があるものの、構築した電子図書館の内部にデータを蓄積しないため、図書館側での自由な操作がむずかしいという問題があげられている。

オンラインジャーナルのように自らが構築した電子図書館サーバに存在しないデータを閲覧することも今後の電子図書館には必要な機能である。特に複数の出版元が多数のオンラインジャーナルを発行している現在ではオンラインジャーナルへのアクセス手法はそれぞれ異なっているのが普通である。ユーザはそれぞれのオンラインジャーナルをアクセスする場合には、それぞれのジャーナルに対して登録の申請を行い、さらにそれぞれのジャー

ナルのインターフェースに従い検索や閲覧を行う必要がある。この操作は誰もが容易に行えるわけではなく、電子図書館の利便性に関する大きな問題点の一つと数えられている。

電子図書館の基幹となるコンピュータハードウェアは年々高速化し、大規模なデータの処理が行えるようになってきているとはいえ、現在のコンテンツの拡大速度に勝ってはいない。そのため、現在一般に行われているような大きな電子アーカイブを構築するような電子図書館ではなく、小さなアーカイブに分散させた電子図書館というものが望まれていると考えられる。

これらの問題を考慮すると次世代型の電子図書館のアーキテクチャは自らが分散を指向することが重要になると考えられる。

2.1.2 分散型電子図書館

前述のように我々は複数のサーバを統合して取り扱い、規模の拡大やメンテナンス、その他の問題を解決する方法が今後の電子図書館において必須であると考え、その研究を進めている。現在一般的に利用されている電子図書館のサーバは単一のサーバマシンを用いて巨大なデータベースを検索するように構築されている。しかしながら、この手法ではデータ数の増加に対応するためには、ハードウェアの拡張が不可欠になり、将来的には現実的ではない。また、外部のサーバを利用したり、オンラインジャーナルを利用する場合には別の方法を利用する必要があるなど、利用者の利便性にも問題がある。そこで、ユーザが利用するインターフェースを提供するためのサーバとは別個に複数のデータベースを構築し、これらを有機的に扱う新しいタイプの電子図書館を提案する。

この電子図書館システム(図 2.1.1)はユーザとのインタラクションを担当する UI 部分(1)とデータを蓄積し、検索や取得の要求に答える DB 部分(3)そして、相互に接続する中間層部分(2)に分割される。中間層を設けることにより、UI 部分を改編することなく、データベースの種類を増加や新しい機能の拡張を行うことが出来る。

- **UserInterface**

ユーザからの要求を受け、データの提供を行うためのモジュールである。現在は Web ブラウザのインターフェースを用いるのが一般的であるが、今後はモバイル端末、バーチャルリアリティなど用途に応じたさまざまな拡張が望まれる。

- **DataBase**

データベースモジュールは検索要求に従い蓄積されているデータを検索し、その結果を返す。データベースの検索プロトコルには Z39.50 や SQL といった形式が考えられるが、すでに稼働しているシステムを組み込む場合には、独自のプロトコルを持っている場合がある。そのため中間層を用意することにより、それぞれのデータベースに適合した検索要求を生成し、検索結果を汎用の形式に変換する必要がある。

- **MiddleLayer**

中間層は UI と DB の 2 つの部分の部分を整合させるために存在し、DB 毎に構築する必要がある DB インターフェース部分とユーザインターフェースに依存する UI インター

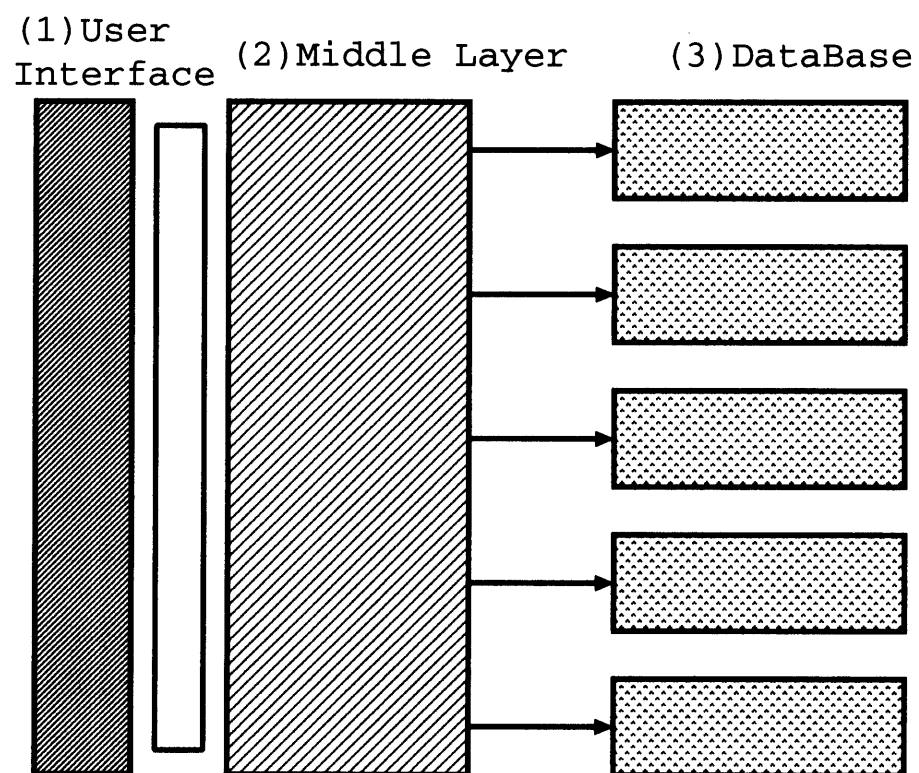


图 2.1.1 分散型電子図書館

フェース部分にわけられる。2つの中間層は密に結合し、要求や返答を伝達する。中間層を設けることにより、ユーザインターフェース、データベースともに拡張が容易になり、将来の拡張が簡単になると言える。

2.1.3 コレクションベースの電子図書館

そこで我々は今まで電子図書館において、フラットに扱って来たデータをコレクションという単位で管理することにより、このコレクションベースで小型の電子図書館を複数構築することを提案する。

ここで述べるコレクションとはある主題に沿って収集された資料の集合であり、大きな規模を持たないものとする。それぞれのコレクションはある主題に基づいたデータとして独自に管理出来る。そのためこれらの一つ一つにおいて、データベースのサーバを割り当てることにより分散管理が可能になる。サーバを分散することにより、それぞれのコレクションの検索速度は飛躍的に高速化され、より有効な情報をより高速に得ることが可能になる。

小規模に設計されたシステムは安価な機材でも十分な性能を持ち、構築が容易になるため、このようなシステムを多数構築することが今後の電子図書館の普及においては重要なポイントになる。

さらに、このコレクションはネットワークを介してアクセスできる場所にあれば、一つの場所で集中管理を行う必要はない。そのため、複数の組織がそれぞれ小さなサーバを持つことにより、多数のサーバを同時に利用することが可能になり、みかけ上巨大な電子図書館システムを構築できることになる。

また、利用頻度の高いコレクションに関してはミラー(コピー)を作成し組織として管理し、利用頻度の低いコレクションに関してはリモートからの利用にとどめるなどの方法により、電子図書館構築に必要な資源を減らすことが可能になる。

しかしながら、ユーザにおいてどのコレクションを選択するべきかという問題は新たな問題となるため、この問題を解決する方法が必要になる。そこで、前述した分散型アーキテクチャに基づきこの問題を解決することを提案している。

2.1.4 横断的な検索手法

このような分散型のサーバを構築した場合には、複数のサーバを横断的に検索し、その結果を統合してユーザに提示する必要がある。本システムにおいてこのような横断的な検索を実現するのは、中間層の仕事である。中間層はUIからの情報をそれぞれのDBに対するモジュールへと伝達する。その時に、モジュールが判断を行い自分のコンテンツのカテゴリと検索要求を照らしあわせることにより、DBへの検索を要求するかどうかを判断する。

検索を行った場合には、結果をまとめUIモジュールへと返答することにより、検索を終了する。データをうけとったUIモジュールはランクづけなどの処理を行い、UIへとデータを伝達し、UIが検索結果の画面を構成する。

このような分散処理を行うことにより、データやデータベースの量が増えた場合にも容易に追加、削除が可能になり、全体を疎な結合を持つ一つのシステムとして構築することが出来る。

2.1.5 遠隔サーバ間の接続

大学やその他の研究機関を中心とし、一般の図書館に至るまで、電子化の波は避けられない。その場合には貴重な資料、参照用資料といったものから順次電子化され、デジタル化された状態で提供されることが考えられる。このような場合に、複数の図書館で同様の資料をデジタル化し、アーカイブとして蓄積することは電子化あるいは保存のためのコストが重複するため効率が悪い。そこでライセンスの許諾を条件に現在行われている相互貸借をさらに進めた形態での相互アクセスは電子図書館の重要な機能となると考えられる。

この場合には資料の検索に利用するためのメタデータの統合や特別なクライアントに依存しないデータ形式での配布など、いくつかの問題があげられる。また、遠隔サーバが停止している場合には検索が出来なくなるため、すべてのサーバが均質になる必要はないが、複数のサーバで同じ情報を持ち、冗長性をあげることによりアクセスを保証する仕組みなどが必要になると考えられる。

2.1.6 Proxy によるオンライン情報の管理

proxy サーバは外部アクセスを代行するサーバであり、全てのクライアントからの接続要求を集約し、一括してアクセスをおこなう。そのため、この proxy サーバ上でさまざまな機能を持つことにより、より拡張されたシステムを構築することが可能になる。この場合には、proxy サーバも電子図書館システムの一部になると考えられる。

オンラインジャーナルの管理には主に2つの方法が行われている。一つはパスワードであり、発行元が指定する ID とパスワードとを利用するほうほうである。また一方では IP アドレスの指定を行い、組織の内部であるか、外部であるかを判別する方法である。IP アドレスの指定を行う場合には、組織外部にいる場合には電子図書館のコンテンツの一部であっても、オンラインジャーナルは利用できないことになる。また、ID やパスワードは組織として1つ与えられることが一般的であるため、現在のシステムではそのままユーザに伝えることは問題が生じることが多い。

- アカウントの管理

組織単位での申込をおこなったオンラインジャーナルでは、個人単位ではなく、組織ごとに ID やパスワードが発行される場合がある。しかしながら、この ID やパスワードは組織を離れると原則上利用してはいけないため、直接ユーザに伝えるのは問題がある。

そこで、proxy サーバを介することにより、ユーザはそれぞれ、自組織の人間であることを proxy サーバに対してパスワードをいれることにより認証を行う。その後に proxy サーバが自動的に組織に割り当てられたアカウントとパスワードを利用してオンラインジャーナルのサーバに対してアクセスを行うという方法が考えられる。

この場合には、組織を離れたユーザのアカウントを proxy サーバ上で削除するだけで、オンラインジャーナルにもアクセスできなくなり、アクセス制限が有効に働いていると言える。

● IP アドレスの変換

オンラインジャーナルをアクセスする場合において、ID ではなく、IP アドレスによって認証を行う場合が存在する。これは組織ごとに割り当てられている IP アドレスのブロックが違うことを利用した方法であり、その組織が存在しているネットワークからのアクセスを認めるようにするものである。

この場合には、組織のネットワーク内部からのアクセスはパスワードによらずに行えるため、問題ないが、外出先など外部のネットワークに接続している場合には IP アドレスが違うため接続することができない。

この場合も、proxy サーバとして自組織のサーバを指定し、パスワードなどで認証を行うことにより、自組織のサーバを経由してアクセスを行うことになる。そのため、通信速度などの問題は残るものの、IP アドレスによって識別されるオンラインジャーナルへのアクセスが可能になる。

● アクセスの保存

シームレスな電子図書館ではそれぞれの図書館に全てのコンテンツを持つのではなく分散管理し、ユーザの要求に従い外部からのアクセスを行う。この場合には従来の図書館と違いコンテンツを独自に持たないためサービスの休止、遅延の増大などが発生することがある。そのため、proxy サーバを利用し、ユーザのアクセスしたデータを一次的に記憶することによりコンテンツの再利用を計ることが出来る。この場合、アクセス履歴の調査により特に利用頻度が高いものは独自にデータを保存することも考えられる。このような機能は今後の分散型図書館の基本機能として必要になると考えている。

このように、proxy サーバは今後のオンラインジャーナルを筆頭とする外部リソースを利用するにあたり、必要不可欠な要素となりうる可能性をもっている。現在もほとんどのインターネット接続組織が proxy サーバを設置し、運用していると考えられるが電子図書館を構築するにあたっては Web トラフィックの削減を目標としたものとは違う図書館のサーバとしての proxy サーバを構築する必要があると考えられる。

2.1.7 まとめ

現在、これらのアイデアに従い分散型の電子図書館の設計を進めている。設計とともに実証システムを構築し、実際に運用することによりシステムの優位性を示したいと考えている。特に分散型のコンセプトおよび、proxy サーバを統合し、外部資源を積極的に利用することを想定した電子図書館の構想は現在考えられている電子図書館像との違いがあるため、実証実験が不可欠であると考えている。よって、これらのシステムを設計し、構築することは電子図書館の将来に対し大きな貢献が出来ると考えている。