

分散電子図書館における情報アクセス機構

3 X - 9

安達 淳 松村 敦
 学術情報センター研究開発部

1 背景と目的

広域分散した情報環境で、画像を含む大規模なマルチメディア文書情報のデータベースを構築する際には、対象となる情報資源の識別、検索、同定等が問題になる。この問題の典型は、インターネット上で分散協調して運用される電子図書館のような応用で見ることができる。本稿での電子図書館とは、広義にとらえ、マルチメディア文書を蓄積してネットワーク上に提供するデータベースシステムと考える。

[設計上の検討事項] インターネット上で動作する電子図書館の設計には一般に、(a) サーバの分散化、(b) 情報の一意性と永続性、(c) 情報の粒度、(d) 位置透明性、などが問題になる。

本稿では大規模な分散マルチメディア情報システムにおける情報オブジェクトの識別、同定、検索等の問題を解消するための手法を扱う。この目的のため URS というオブジェクト参照の仕様の記述方法を提案し、これを解決してオブジェクト識別子を生成し、オブジェクトへのアクセスを可能とすることを試みた [2]。また、学術情報センターで運用している電子図書館システム NACSIS-ELS [1][5] へ適応して、応用上の問題を検討したので報告する。

2 オブジェクト識別子と URS の提案

[URI] IETF で検討されている URI [3][4] は、既存の URL 等も包含するものとして提案されており、インターネット上の名前空間を記述する構文(基本的に URL のような形式)、使用する文字セットなどを決めている。Internet のドメイン名と同様に、URI は識別子を与えるオーソリティにより、一意性が管理されることになる。

[実装上の課題] 大規模なデータベースの形成に着目して、上記のページ画像やもっと粒度の細かいオブジェクト間のハイパーリンクの作成には、データ作成時に必ずしも URL が判明しないとい

(利用者からのキーワード)

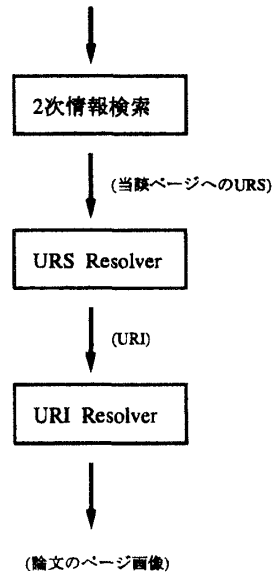


図1 電子図書館における検索のながれ

う本質的な問題がある。またリンク先の URL の変化に対する本来的な脆弱性がある。

提案手法では、当該オブジェクトの公知の属性を記述するというにより、リンク先のオブジェクトを指定するという手法をとる。

このようにリンク先に影響されないデータ作成仕様は、大量の文献をデジタル化する場合に処理手順の観点から必須の要件であると考えている。

[URS の提案とその実装] 情報アクセスの過程で、オブジェクトへの参照に URS (Uniform Resource Specification, 情報オブジェクト記述) と URI (Uniform Resource Identifier) を採用する。URS は、参照するオブジェクトの持つ標準的な属性を記述する方法であり、オブジェクト固有の情報の組として構成するものである。また、URI はオブジェクト識別子であるが、まだ標準的な方式が定まらない段階なので、システム内で一意に決定するための暫定的な付与方式を採用することにした。

以上のものを用いて、二次情報からページ画像を得る過程は図1のようになる。二次情報データベースの各書誌レコードには当該論文の URS が格納されている。NACSIS-ELS での実装では、URS は他の書誌的な情報から自動的に作成することが可能である。

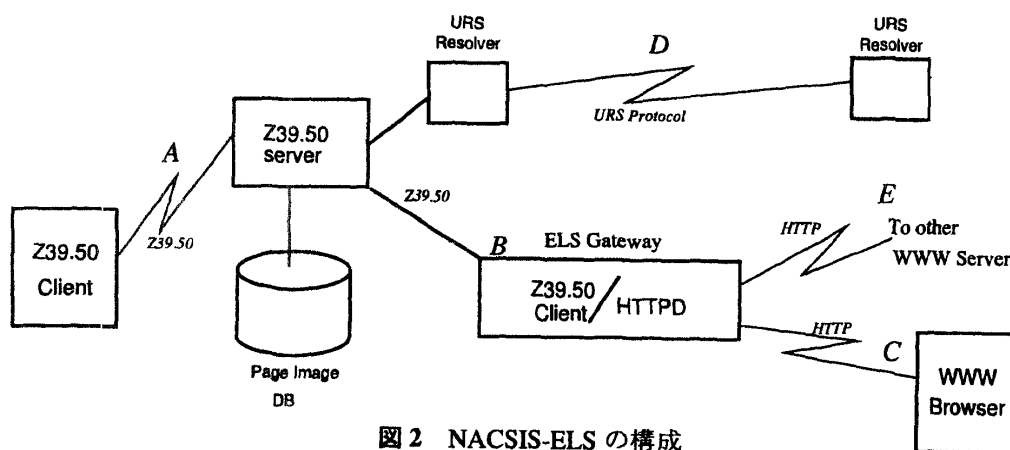


図2 NACSIS-ELSの構成

次にこの URS が URS resolver に与えられてシステム内での識別子 URI が確定する。さらにこれが URI resolver によって解決され、物理的な記憶場所が判明し、URL が得られる。以上のメカニズムによって、分散した情報ベース間で名前を解決し、情報獲得を実現しようとしている。

[実装方法] 図2に示したものは、現在の学術情報センターの NACSIS-ELS の構成である。サーバでは、情報検索プロトコルの Z39.50 を使用している。このプロトコルを直接使用したクライアント(図中(A))が用意されている。残念ながら、Z39.50 はインターネット上でまだ多用されていないので、もう一つのアクセス方法として、WWW クライアントからのアクセス経路も提供されている(図中(B), (C))。

後者のアクセス機能の実現のために、ソフトウェアの構成としては、Z39.50 クライアントと HTTPD を統合したような中間サーバ機能を設け、これを「ELS ゲートウェイ」と呼んでいる。(C)の標準的な WWW ブラウザからこのゲートウェイにアクセスし、CGI 機能を用いて Z39.50 サーバからのデータベース検索等の結果を HTML に変換し、WWW ブラウザに表示している。

現在、この構成の中で、URS Resolver は Z39.50 サーバの配下に置かれて動作している。これを(D)に示すように、他サイトの URS Resolver との間の交信により、URI/URL を解決し、(B)のゲートウェイに返してやり、この結果、(C)の WWW ブラウザからは(E)のようにして最終的なオブジェクトにアクセスを可能にする予定である。

3 考察

現在動作している URS Resolver は、URS 第1版の機能面の実装と検証が主たる目的であった。今後、これを分散サーバ間で協調動作させるための

機能設計とプロトタイプ作成を行う計画である。

その他、分散 URS Resolver には、(1) URS Resolver 間で相手の所在を確認して、互いの情報をリフレッシュする転送機能、(2) URS Resolver 個々の配下にある URI (URL) 情報のデータベースを圧縮するための方法、それを相互に配送/キャッシングする機能、(3) 情報の一意性、複製可能性等を管理するための仕様とデータベースの設計方法、などの機能が要求され、順次実装を図る。

また、別種の課題として、名前解決のアルゴリズムをルールに沿って高速に行う方法を精緻化する必要がある。現在は単純な辞書を参照して解決している。分散 Resolver では、ローカルに蓄えた情報と他サイトへの問合せの転送を上手に切り分けなければならない、このアルゴリズムの実装とルールベースの処理を一体化することが次の課題になっている。

謝辞 本研究は一部文部省科研費重点領域研究「高度データベース」の助成による。

参考文献

- [1] Jun Adachi, H. Hashizume: NACSIS Electronic Library System: Its Design and Implementation, Proceedings of The International Symposium on Digital Libraries 1995, pp.36-41, Tsukuba, Japan (August, 1995).
- [2] 安達淳: 情報ネットワーク環境における分散オブジェクト管理の一手法, 「高度データベース」平成8年度研究成果報告会講演論文集2, pp.32-38 (January 1997).
- [3] RFC1630, "Universal Resource Identifiers in WWW", (June 1994).
- [4] RFC1737, "Functional Requirements for Uniform Resource Names", (December 1994).
- [5] URL: <http://www.nacsis.ac.jp>.