

ORDBによるマルチメディア・ドキュメントの管理

大野邦夫

ohno@inse.co.jp

INS エンジニアリング (株)

〒141 東京都品川区西五反田 4-31-18

佐藤和也

kazuya@approach.co.jp

アプローチ (株)

〒106 東京都港区麻布台3-2-7

マルチメディア情報を管理可能なSGMLデータベースシステムを開発した。データベースのコアには、Informix社のILLUSTRAを用いている。ILLUSTRAはORDB (Object Relational DataBase) であり、RDB用のSQLを照会言語としているが、オブジェクト指向技術によりデータ型の拡張を可能としている。そのため、従来のRDBが、データ型として主に数字と文字しか扱えなかったのに対し、図形、画像、音声、映像などのマルチメディアをデータ型として扱うことに特徴がある。この枠組みを有効に使うことにより、マルチメディアを効果的に管理することが可能なSGMLデータベースが開発された。SGMLデータベースの機能は、ILLUSTRAのオブジェクト・ライブラリであるデータブレードとして実現されている。

Managing Multimedia Document Through ORDB

Kunio Ohno

INS Engineering Co.

4-31-18 Nishi-Gotanda, Shinagawa-ku, Tokyo, 141, Japan

Kazuya Sato

Approach Inc.

3-2-7 Azabudai, Minato-ku, Tokyo, 106, Japan

A SGML Database System that manages Multimedia Information has been developed. Informix ILLUSTRAs which is known as an Object Relational Database Management System (ORDB) is implemented to the core of the database. Though existing RDB manages restricted data types such as figures and characters, ORDB can flexibly extend data types and manages them through their object-oriented architectures. Therefore, ORDB can manipulate multimedia as figures, images, sounds, videos, etc. Using those characteristics, the SGML Database System has been developed. The system is implemented as a DataBlade that is the object based library of ILLUSTRAs.

1. まえがき

マルチメディア情報を一元的に管理可能なSGMLデータベースシステムを開発したので報告する。データベースのコアには、Informix社のILLUSTRA[1]を用いている。ILLUSTRAはORDB (Object Relational DataBase) であり、RDB用のSQLを照会言語としているが、オブジェクト指向技術によりデータ型の拡張を可能としている。そのため、従来のRDBが、データ型として主に数字と文字しか扱えなかったのに対し、図形、画像、音声、映像などのマルチメディアをデータ型として扱えることに特徴がある。

この枠組みを有効に使うことにより、マルチメディアを効果的に管理することが可能なSGMLデータベースを開発した。SGMLデータベースの機能は、ILLUSTRAのオブジェクト・ライブラリであるデータブレードとして実現している。本報告では、その概要を紹介する。

2. 企業情報システム

2.1. イン트라ネットの課題

大企業、中小企業の如何を問わず企業情報システムは必須のものである。古くはMIS (Management Information System) と呼ばれ、その後OA (Office Automation), DSS (Decision Support System), SIS (Strategic Information System) 等と様々な呼び方がなされてきたが、企業情報システムは基本的にはデータと文書の管理システムである。近年多くの場合、データはRDBに、文書はワープロやDTPのファイルで管理されている。

最近のインターネットの普及により、企業情報システムにインターネット技術を導入する試みが顕在化しつつある。これらはイントラネットと呼ばれ、爆発的に普及しつつあるように見えるがこの本質的な性格を見抜く必要がある。イントラネットとは、ウェブ・ホームページを通じて単に情報の配布・検索や参照手段を効率化させたに過ぎない。CALSやISO9000等で問題になる、企業内における情報の作成・承認さらに保管やセキュリティといった面に関して、現状のイントラネットは必ずしも適合しない。

2.2. ワークフロー管理ツールの事例

企業内における情報の作成・承認さらに保管やセキュリティといった機能は、ワークフロー管理ツールによって実現可能である。我々はInterleaf RDM[2]を用いてワークフロー管理を検討し、ISO9000シリーズに適合する文書管理システムの検討を試みたことがある[3]。

Interleaf RDMは、Oracleを検索照会エンジンとして、文書の作成から廃棄に至るライフサイクル全般を管理する文書データベース兼ワークフロー管理ツールである。この製品コンセプトは、Lotus Notesと類似であるがNotesの出荷以前に既に使われていた[4]。さらに当初から三層クライアント/サーバ方式を採用した極めて先進的なシステムであった。

Interleaf RDMは、米国ではそれなりのマーケットを獲得したが、日本での導入は困難な側面を持っていた。大きな問題の一つは管理の厳格さにあった。すなわち、作成された文書の作成日時、承認日時、配布日時などが、その責任者とともに厳格に管理されてしまうシステムは日本の文書処理になじまないという指摘である。

日本の文書処理は、必要な場合には、決済日時の変更などが許されないと困るのである。ところがInterleaf RDMは、逆にそれが売り物なのである。以上の企業文化的な問題以外にも初期のInterleaf RDMに対しては、以下のような技術的な要望が出された。

(1) クライアントがInterleaf製品に限られるのは問題である

(2) クライアント/サーバ間のAPI仕様を公開してほしい

(3) エンジンをOracle以外のRDBに拡張してほしい

その後Interleaf社は以上の要望に対し、Intellect/Business Webと呼ばれる新製品を提供した[5]。これはRDMを内部に含む文書管理ツールで、従来のRDMをウェブ・ブラウザから利用可能とする意欲的なシステムである。上記の問題点に対して、以下のような解決策を提供していた。

(1) 管理対象は、ファイルであれば何でも構わない。但しクライアント側にそのファイルを処理するアプリケーションを必要とする。

(2) 文書管理のAPIとしては、文書管理のコンソーシアムであるDMA (Document Management Alliance) の規格を用いる[6]。

(3) 検索照会エンジンとしてはOracle以外にもマイクロソフトのSQLサーバ等を利用可能とする。

以上から、イントラネットにおけるワークフロー管理のツールとしてはハイエンドの十分な機能を実現していると言える。しかしカスタマイズのためには、CGIからDMAインタフェースが使えるだけであり、現状ではそれ以外のAPIの提供はなされていない。

2.3. 専用システムからオープンシステムへ

上記のInterleaf社の場合を含め、従来のワークフロー管理ツールは、各ベンダーの製品を中核とした専用システムになりがちである。そのため、それを導入すると各種アプリケーションツールの統一といった方向に向かい易く、企業情報システムとしては柔軟性に欠ける。このようなツールを使って成功した事例なども紹介されているが、そのようなやり方が長期的に見ても成功事例であるとは限らない。

本質的な解は、オープンなワークフロー管理ツールの構築にある。そのための最も有望な解は、文書、データ、さらにそれらをコントロールするアプリケーションをオブジェクトとして統合し一元的に管理・運用する仕組みに求められるであろう。具体的には、OMG[7]のCORBA[8]を用いたシステムに求められると考えられる。以下にその理由を説明する。

3. クライアント／サーバ方式の進化

3.1. RPC

CORBAの説明のためには、RPC (Remote Procedure Call) の概念が必要である。RPCとは、文字通り異なるコンピュータ上の手続き(プロセス)の呼び出しである。同一のコンピュータ上の呼び出しであれば、各々の手続きの開始メモリ番地は明白なので実行中のプロセスは直接にサブルーチン、関数などの手続きを呼び出して実行し、結果を得ることが可能である。

RPCの場合は、相手のコンピュータの番地の情報を知ることができないので、手続きの名称や引数の仕様(シグニチャ)を相互に参照するためのスタブと呼ばれる切り口が双方に設けら

れる。クライアントから遠隔の手続きが呼ばれるとスタブを通じて手続き名称と引数が特定のデータ形式で相手のコンピュータ(サーバ)に送られる。サーバでは、得られた手続き名と引数によりその手続きが実行され、結果をスタブ経由でクライアントに送り返す。

3.2. CORBA

CORBAは、以上のRPCにおける呼び出し構文を一元化し、異なるコンピュータ間の呼び出しを可能とし、相互運用を実現するための規格である。呼び出しの構文の規格はIDL (Interface Definition Language) と呼ばれる。サーバ上の各種サービスの呼び出し(リクエスト)と返り値の仕様はIDLにより記述される。

サーバ上のサービスは、以前はオブジェクトサービス、コモンファシリティ、アプリケーションオブジェクトの3種類のカテゴリに分類されていた[9]。

(1) オブジェクトサービス：全てのCORBAオブジェクトの基本となるサービス

(2) コモンファシリティ：アプリケーションオブジェクト構築のための有用なサービスを提供するCORBAオブジェクトの集合

(3) アプリケーションオブジェクト：個別アプリケーションのサービスを提供するCORBAオブジェクトの集合

なお、昨年のOMGの組織改革で、コモン・ファシリティとアプリケーションオブジェクトの中間領域としてドメイン・インタフェースという領域が設定され、業界毎のオブジェクトやビジネスオブジェクトの規格が審議されるようになった[10]。

先に述べたワークフロー管理の機能もIDL化の検討がなされているが、これらはビジネスオブジェクトに関連するドメイン・インタフェースの規格として定められる見通しである[11]。

3.3. 三層クライアント／サーバ方式

CORBAの規格はクライアント・プログラムを作るためのインタフェースの仕様であり、サーバ側の実装を規定するものではない。ところが新規システムの構築の場合を除くと、サーバ側は既存のシステムが稼働しているものであり、CORBAとの融合が課題となるであろう。

既存の企業情報システムは、その多くのデータをRDB上に構築しているのでRDBへの照会機能が要求される。照会機能はオブジェクト・

サービス（最近はCORBAサービスと呼ばれる）で規定されているが、それはOODBへの照会（OQL）をベースにした中途半端な規格となっているので既存のRDBとは整合が取れない。従ってCORBAオブジェクトであるミドルウェアを介させる三層クライアント／サーバ方式とならざるを得ないであろう。Oracle社は、このようなミドルウェアをデータ・カートリッジと呼んでいる[12]。

この場合、クライアントとミドルウェアの間はCORBA準拠のやりとりとなり、ミドルウェアとRDBの間は、従来のSQLでインタフェースされることになる。

4. ORDB

4.1. ILLUSTRATION

さらにSQL3[13]のようにミドルウェアとRDBを一体化するアプローチもあり得る。これは、RDBへの照会言語SQLをオブジェクト指向化するものである。オブジェクト指向への拡張は、クラス定義に相当する新たな型の定義機能とメソッド定義に相当する関数定義機能を付与することになる。これらの機能はデータカートリッジのような三層のミドルウェアでは実現不可能な機能である。このようなデータベースはORDB (Object Relational Database) と呼ばれ、具体的な製品としてはUniSQL[14]とILLUSTRATIONが挙げられる。

通常のRDBに対するORDBの特徴は、マルチメディアをBLOBへのポインタとしてではなく、直接にデータとして管理可能なことが挙げられる。そのため後に詳細に述べるがORDBはマルチメディア管理システムとして非常に優れた性質を持っている。

4.2. データブレード

ILLUSTRATIONはUniSQLと類似のORDBであるが、UniSQLが裸のエンジンであるのに対し、ILLUSTRATIONはデータブレードと呼ばれるライブラリ群を前提にシステムを構築するツールである点が大きく異なっている。

データブレードは、各種分野のアプリケーションのためのデータ型（通常の言語であればクラス）に対応した関数群（通常の言語であればメソッド）のライブラリである。具体的には下記のようなブレードが提供されている。

(1) Time Series DataBlade

このデータブレードは、時系列データや時間データを処理するために作成された。Timeseries（時系列）とCalendar（カレンダー）という2つのデータ型が定義され、それらに関連する40あまりの関数が定義されている。

(2) Text DataBlade

このデータブレードは、文字および文字列に関する情報とテキスト文書とを関連付ける役割を持つ。通常のRDBも文字列型は持っているものでそれらの機能に対する拡張と考えてもよい。最大の特徴は、文書のような膨大なテキストデータをBLOB (Binary Large Object Block) データとして直接扱えることが挙げられる。

(3) Text Conversion DataBlade

このデータブレードは上記のText DataBladeへの追加機能である。従って新たな型は定義されないで関数のみが追加されている。これは、広く普及している、ワープロ、DTPなどのデータ形式をデータベースに取り込むと同時に、相互に変換する機能を持つ。

(4) Two Dimensional (2D) DataBlade

このデータブレードは、位置、距離、図形、イメージなどの情報に対し、2次元空間における演算、操作を加えて何らかの結果を得ようというものである。本データブレードを用いて、地図情報をベースにした検索に役立てることが可能となる。

(5) Three Dimensional (3D) DataBlade

このデータブレードは、上記2Dデータブレードの3次元版である。従って、3次元CADのような立体図形を対象とした演算、操作が可能である。

(6) Image DataBlade

このデータブレードは、image型という新たな型を定義し、SQLを用いて画像データの入力、表示、演算、変換などの操作を扱う関数を保有する。

(7) Basic Image Query DataBlade

このデータブレードは上記のImage DataBladeを機能強化したものである。イメージデータに対するエンハンス、輪郭検出、輪郭の平滑化等の機能を有する。

(8) Web DataBlade

このデータブレードは、ウェブブラウザとデータベースとのインタフェースを容易に構築するためのものである。HTMLで記述されたページに、アイコン、グラフィック図形(2D,3D)、映像、音声、他ページへのリンク等を記述することが可能である。またSQLによる照会をサポートしているので、ウェブブラウザのユーザがILLUSTRAデータベースを検索し、結果として上記の各種データをブラウザ上に得ることが可能である。

(9) VIR DataBlade

これは、サードパーティ(Virage社)により提供されたデータブレードで、類似の画像を参照・検索するための機能を有する。色彩、輪郭形状、テキストなどの属性をウェイト付けして、検索対象と類似度の高いものを抽出させる機能を有する。

4.3. マルチメディア管理

以上のようにILLUSTRAはマルチメディアを格納し、それを検索・参照するには極めて好都合

な環境を提供してくれる。これらの機能は、従来のデータベースでは実現できなかった機能であり、マルチメディアを要素として持つドキュメントの管理システムの構築には理想的なツールであると言える。

5. SGMLデータブレード

5.1. 構成

以上から、我々はマルチメディアを要素として扱える文書管理システムの構築をILLUSTRA上に試みた。そのシステムはILLUSTRAのデータブレードとして実装される。データベース上の文書の枠組みとしては、ISOの標準であるSGMLを用いている。SGMLは文書の論理構造を記述するための言語であるが、見方を変えると文書をオブジェクトの集合体として記述するための言語でもある。従ってデータをオブジェクトとして扱うILLUSTRAとは整合性が良好である。システム構成を図1に示す。

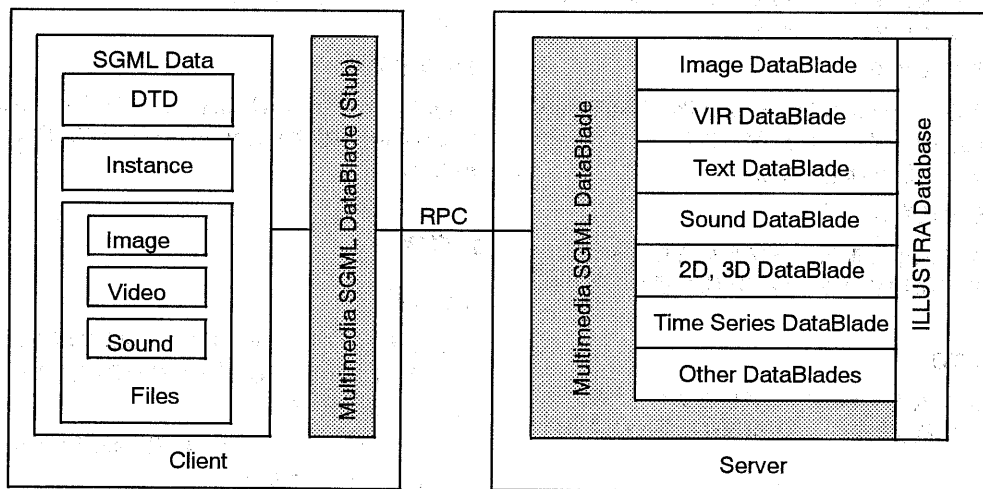


図1 システム構成

5.2. 機能

機能を要約すると下記の通りである。

(1) SGMLデータの登録

(2) SGMLデータの出力(流用/変更)

(3) エレメント単位での出力/登録(出力対象以外にはダミーを埋め込む)

(4) SGMLデータの削除

- (5) DTDからインスタンスの検索／インスタンスからDTDの検索
- (6) インスタンスで利用されているエンティティの検索
- (7) エLEMENTで利用されているエンティティの検索
- (8) 複数のインスタンスによるELEMENTクラス共有
- (9) 上位ELEMENTの検索
- (10) 下位ELEMENTの検索
- (11) 同位ELEMENTの検索
- (12) 日付によるDTD, インスタンス, エンティティ, ELEMENTの検索
- (13) オーナーによるDTD, インスタンス, エンティティ, ELEMENTの検索

5.3. 主な仕様

(1) ディレクトリ管理

SGML文書をツリー構造で管理するUNIXライクなディレクトリである。

(2) SGML文書

文書名, バージョン, 作成日時, SGML宣言へのID, ファイルへのID等を持ち, バージョンの分岐が可能な構造とする。

(3) SGML宣言

システム中に専用テーブルを持ち, テキストのままに格納される。各SGML文書から参照される。

(4) DTD

システム中に専用テーブルを持ち, テキストのままに格納される。各SGML文書から参照される。

(5) エンティティ

文字列, 特殊文字のエンティティ値は文字列として格納する。外部実体の場合は, ファイル種別毎にデータブレードの入出力関数, 格納テーブル (イメージテーブル, 音声テーブルなど) を参照可能とする。現状では, 外部実体としてはイメージだけを扱う。

(6) 外部実体エンティティ

外部実体のオブジェクトIDを用いてデータブレード関数を呼び出し, 各種操作を行うことは可能であるが, この関数を用いて外部実体の取り出し, 格納は不可能である。なお, オブジェクトIDはSGML文書取り出しやエンティティ取り出しのAPIを用いて取得可能である。

(7) ELEMENT部分木

指定ELEMENTから指定階層まで, あるいは最下層までの部分木を取り出すことが可能である。ELEMENTに属するアトリビュートも同時に取り出される。

(8) アトリビュート

ELEMENTのアトリビュートがテーブルに格納される。SGML文書またはELEMENT部分木取り出しのAPIにより取り出される。アトリビュート名と値を指定して適合するアトリビュートのリストおよびそのSGML文書IDのリストを得る。

5.4. データベース構造

前項の仕様を満たすためのデータベース構造を図2に示す。図における四角のブロックはテーブルを示す。図の矢印はリンクしているデータを示す。

5.5. 動作環境

現状では, サーバはSolaris, クライアントはNetScapeが搭載されたWindows95/NTの環境で動作する。

6. 考察

6.1. 特徴

このデータブレードを用いたILLUSTRAシステムは, 従来のSGMLデータベースに比べ以下のような特徴を持つ。

- (1) オーサリングよりは, 検索・参照・配布・管理を指向する。
- (2) マルチメディアを含むSGMLの要素をデータベースで直接扱える。
- (3) ILLUSTRATIONの他のデータブレードを利用して要素データを検索できる。
- (4) オーサリングツール, パーサ, ビューワ等を選ばない。

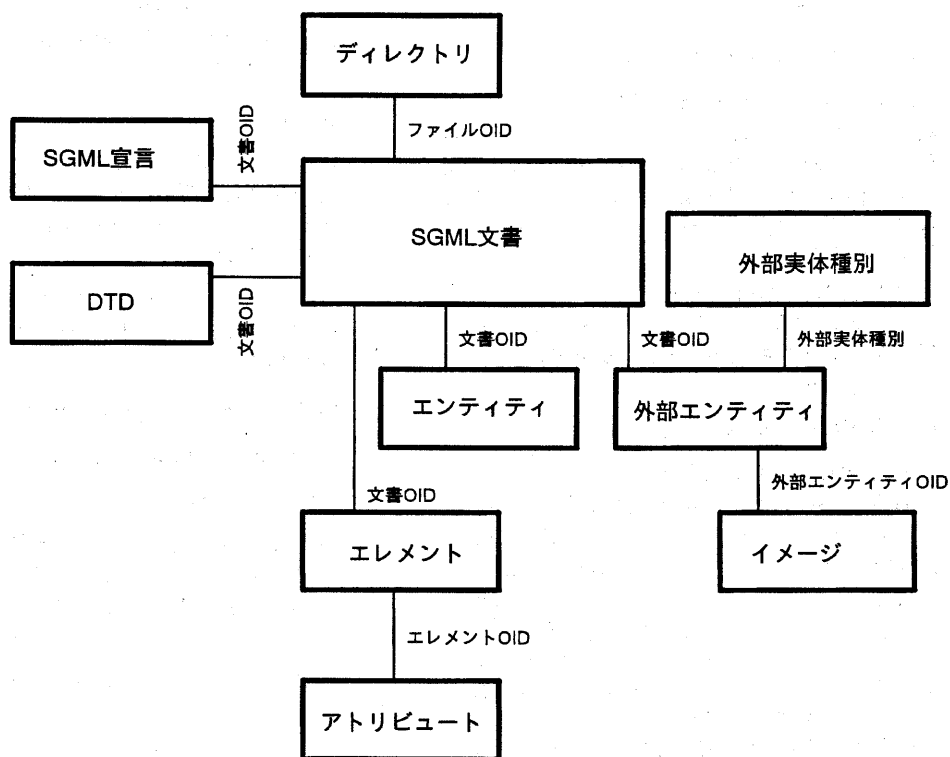


図2 データベース構成

以上の特徴は、従来の製品が、SGMLデータの作成者（マニュアル等の作成者）を対象として開発されていたため、企業内における文書の参照、業務のナビゲーションなどの機能が貧弱であった点を考慮した結果である。

一方、文書の参照、業務のナビゲーションなどに関しては、インターネットの枠組みによる企業内情報システム（簡易なイントラネット）が優れたコスト・パフォーマンスで構築可能なことから爆発的に進展しているが、アクセス制御やデータの保護の面で問題をかかえている。日本でも今後は情報システムに対するハッカーの進入やデータ破壊に備える必要があり、情報をオブジェクトとして一元的に管理するメカニズムが必要であり、本システムはそのような用途に有効である。

6.2. データカートリッジとの比較

データカートリッジとデータブレードの比較はよく議論されるが、この議論の本質は各々の技術の狙いが相違している点に存在する。データカートリッジは、SQL92レベルの従来のRDBにアクセス可能なOMG準拠のミドルウェアである。従って、既にRDBに蓄積された文字や数字の大量の情報を扱うアプリケーションをCORBAベースで構築する場合は、データカートリッジ的な枠組みが必要になる。

それに対し、データブレードは、マルチメディアのような拡張されたデータ型を操作するためのインタフェースを提供してくれる。データカートリッジがマルチメディアを扱う場合は、RDB上ではポインタ経由で扱わざるを得ないであろう。なぜならば、従来のRDBは直接マルチメディアを扱えないからである。

7. 今後の計画

7.1. Universal Serverへの対応

Informix社は、ILLUSTRAのデータブレード機能を、従来のInformix系列のデータベース製品に採り入れ、Universal Serverと呼ばれる新系列の製品への展開を開始している。米国では既に最初の版が出荷され、一部のデータブレードが提供されている。

日本語対応のUniversal Serverの出荷は、本年の秋以降になされると思われるが、我々のデータブレードもUniversal Serverに対応してゆく方針である。

7.2. CORBAへの対応

最近のOMGにおけるビジネスオブジェクト等の進展にともない、APIのIDL化が要求として挙がっている。データベースの世界では、OracleのNCA (Network Computing Architecture) におけるデータカートリッジが引き合いに出され、データブレードとの比較[12]が云々されることが多いが、先に述べた通り、データブレードとデータカートリッジは基本的に異なる技術であり相容れないものではない。本データブレードのようにAPIが定義されているシステムをCORBAと融合させる手法としてはラッピング技術[15]が挙げられる。当面、Orbix[16]を対象に検討したいと考えている。

8. あとがき

以上、ORDB、ILLUSTRAを用いたSGMLデータベースについて紹介した。本システムは、はじめに述べた通り、企業情報システムにおけるマルチメディア情報管理のために用いられることを狙っている。情報管理とは、マルチメディアデータの管理のみならず、CALSやISO9000等で問題になる、業務に関する指示書、説明書、さらには作成・承認等のワークフローの管理、インターネットへの情報の配布など幅広い用途を指向している。今回完成したのは、その基本部分のデータブレードであるが、今後はさらに上記機能を拡張してゆきたいと考えている。以上の拡張機能については、単にSGMLのDTDとして実現されるものと独立したデータブレードとなるものに大別されるであろう。DTD自体も新たなデータブレードとして提供される場合もあるであろう。このような拡張性は、ORDBが持つ大きなメリットである。

文献

- [1] 伊藤他； “DataBlade構築技法” ,BNN,(1996)
- [2] <http://www.interleaf.com/rdms.html>
- [3] 中島、大野； “ISO9000シリーズ用の電子化文書管理システム” , 情報処理学会テクニカルコミュニケーション研究グループ報告 (1994-9-14)
- [4] 大石； “オーバービューオブ Interleaf” , Super ASCII, Vol.3, No.10, p.89, (1992-10)
- [5] <http://www.interleaf.com/ibw-pr.html>
- [6] <http://www.aiim.org/dma/>
- [7] <http://www.omg.org/>
- [8] OMG (共訳)； “共通オブジェクト・リクエスト・ブローカ - 構造と仕様 - CORBA1.1” , 創研プランニング, (1992)
- [9] OMG (共訳)； “オブジェクト・マネジメント・アーキテクチャ・ガイド - OMAG2.0” , 創研プランニング, (1994)
- [10] 大野； “転機を迎えた分散オブジェクト” , Computer Today, No.77, pp.4-14, (1997-1)
- [11] http://www.omg.org/library/schedule/CF_RFP4.htm
- [12] http://www.oracle.com/nca/html/nca_wp.html
- [13] 日経インテリジェントシステム； “ISOなどが標準化進めるSQL3” , オブジェクト指向データベース (日経インテリジェントシステム別冊) pp.72-92, (1994)
- [14] 日経インテリジェントシステム； “RDBとOODBの特徴を兼ね備えた「UniSQL」” , オブジェクト指向データベース (日経インテリジェントシステム別冊) pp.230-245, (1994)
- [15] T.J.Maubray et.al.； “The Essential CORBA” , John Wiley & Sons, pp.231-267, (1995)
- [16] <http://www.iona.com/Orbix/index.html>