

クライアントサーバで実現する統合文書情報システム

—Bibliotheca—

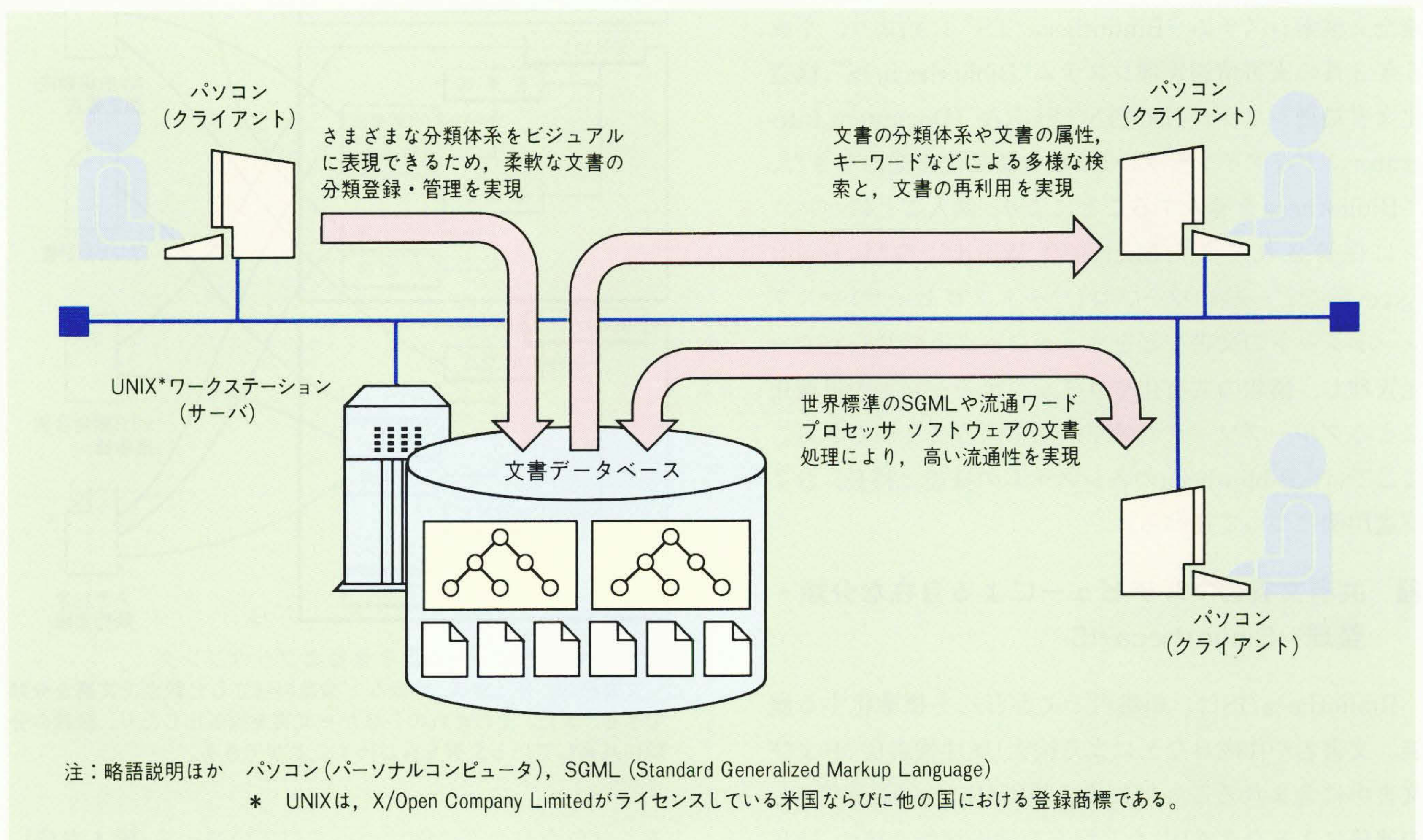
Integrated Document Information System on Client Server System

和歌山 哲* Satoshi Wakayama

浅川 悟志* Satoshi Asakawa

星 幸雄* Yukio Hoshi

山本 貴禎** Kiyoshi Yamamoto



統合文書情報システム“Bibliotheca”によって効率化する文書処理

柔軟な分類体系、多種・多様な検索方法、文書の標準化によって企業の情報の共有化と流通を支援する。

BPR (Business Process Re-engineering) など企業活動の効率改善や活性化のためのコミュニケーションツールとして、コンピュータとネットワークを利用する情報システムの構築が注目されている。現在、企業情報の大部分は、文書という形態をとって企業内を流通し、いわゆるホワイトカラー作業者の企業活動の大部分が文書にかかわる作業に費やされる。しかし、大量に作成された文書は貴重な企業情報の宝庫でありながら有効に再利用されず、逆に「情報の洪水」となっている。これに対し、企業内の数値データの処理の大部分を対象としていたコンピュータを企業内文書処理へ適用することがBPR推進

のかなめとして注目されている。

統合文書情報システム“Bibliotheca”は、作成、保管、流通、再利用といった文書のライフサイクル全体を支援する。文書という非定型な情報を組織内で共用、再利用するためのビジュアルな整理・体系化を実現する“Bibliotheca/IS”，文書に含まれることばを対象に自由なキーワードによる検索を可能にする“Bibliotheca/TS”，および世界標準規格のSGML (Standard Generalized Markup Language) 文書を扱う“DOCINTEGRA”などにより、ユーザーのニーズに沿った文書情報システムが構築できる。

* 日立製作所 ソフトウェア開発本部 ** 日立西部ソフトウェア株式会社

1 はじめに

企業にとって重要な財産である文書を有効に活用する文書情報システムが注目されている。

日立製作所は、クライアントサーバ構成の統合文書情報システム“Bibliotheca”を開発し、平成4年12月の高速全文検索システム“Bibliotheca/TS”に始まり、平成6年3月の文書情報管理システム“Bibliotheca/IS”，構造化文書処理システム“DOCINTEGRA”（Document Integrator：ドックインテグラ）と順次製品化を進めてきた。

Bibliothecaを導入することにより、個人ごとにパソコンに保管されてきたMicrosoft Word^{※1)}やMicrosoft Excel^{※2)}など一般のワープロ（ワードプロセッサ）やスプレッドシートの文書などをネットワーク上のサーバで一元管理し、情報の共有化やコミュニケーションの迅速化によるグループワークの効率向上を図ることができる。ここでは、Bibliothecaの各システムの機能と特長、および適用例について述べる。

2 共有文書のマルチビューによる自在な分類・整理—Bibliotheca/IS—

Bibliotheca/ISは、組織内で文書形式を標準化する機能、文書名や作成日などによる検索（属性検索）、および文書中に含まれることばによる検索（Bibliotheca/TSとの連係による全文検索）など豊富な検索機能を持つ。特長は、マルチビューと呼ぶ文書の分類・整理機能と、マルチビューの階層をたどって文書を探し出すブラウジング機能である。

従来、文書管理では基本的に一つの階層構造の下に文書を分類してきた。しかし、事業を積極的に展開するような活性化された組織ほど、扱う情報が流動的で、観点も多様な場合が多く、共通の分類体系を持つことが難しい。マルチビューによる分類では、ビューごとに異なる観点の分類体系を作成し、一つの文書を複数のビュー下に多重に分類・整理できる。この多重分類を使った検索マルチビューブラウジングでは、一つまたは複数のビューを指定し、その下に分類されている文書を一覧表示す

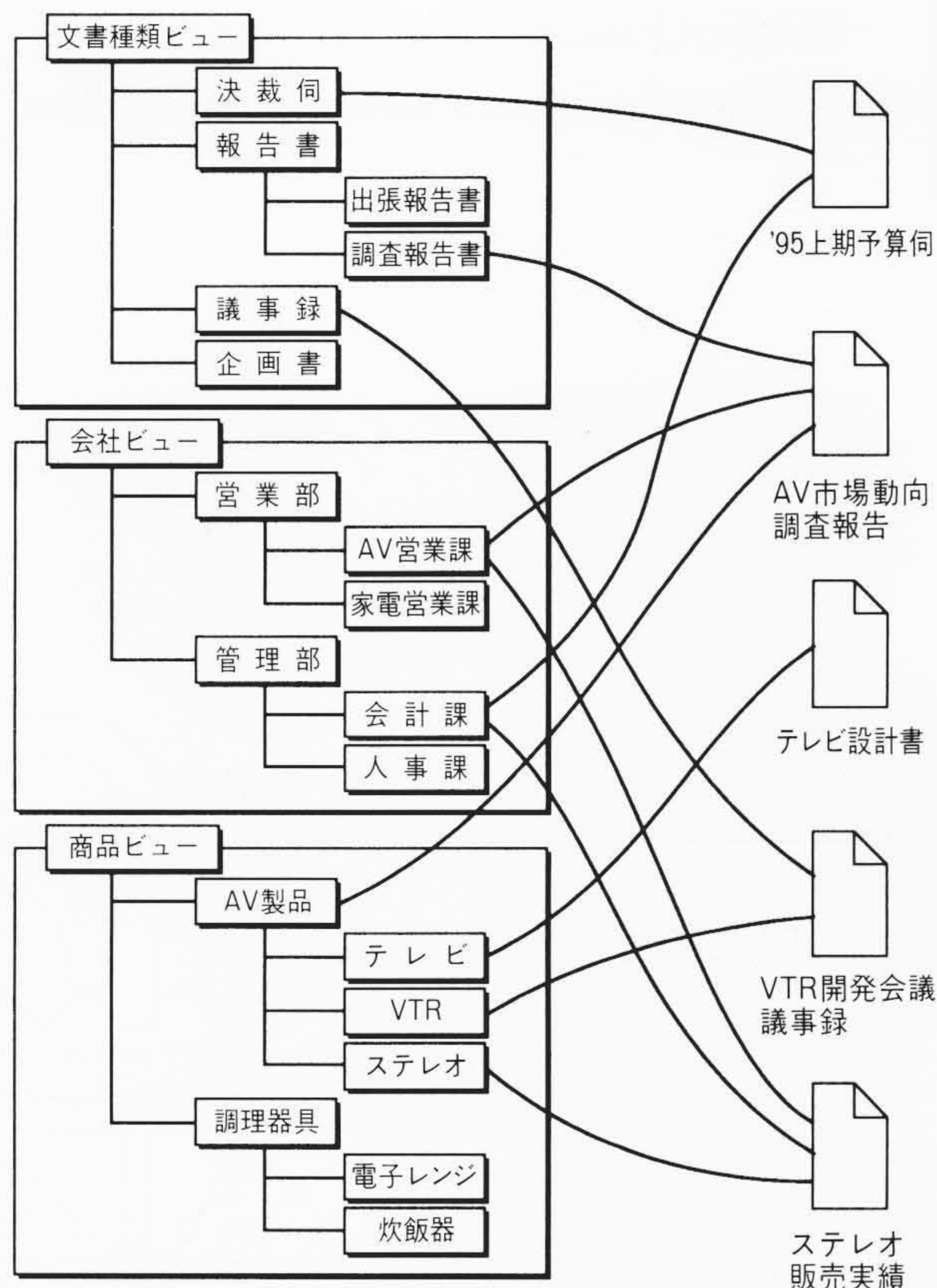


図1 マルチビューによる分類のブラウジング

文書種類、会社組織、商品など複数の独立した観点で文書を分類できる。また、それぞれの分類から文書を探し出したり、複数の分類に共通している文書を探し出すことができる。

る。ブラウジングの例について以下に述べる（図1参照）。

(1) 異なるビューによるブラウジング

一つの文書を複数の異なる分類体系下に整理し、任意の分類体系から文書を探し出す。図1の例では、文書「'95上期予算伺」は、文書種類ビューの「決算伺」、会社ビューの「会計課」のいずれからでも探し出せる。

(2) 上位の分類項目からのブラウジング

分類体系の上位を指定して、下位に分類されたすべての文書群を引き出す。どこに分類されているかあいまいな場合や、対象範囲を広げて下位に含まれる文書を一括して扱いたい場合に有効である。図1の例では、商品ビューの「AV製品」から、文書の「テレビ設計書」、「ステレオ販売実績」などを探し出せる。

(3) 複数のビューを組み合わせたブラウジング

異なる観点の分類体系を組み合わせて文書群を絞り込む。図1の例では、文書種類ビューの「議事録」と商品ビューの「AV製品」を組み合わせて、「AV製品」に関する「議事録」群の中から文書「VTR開発会議議事録」を探し出せる。

※1) Microsoftは、米国Microsoft Corp.の登録商標である。Microsoft Wordは、米国Microsoft Corp.の商品名称である。

※2) Microsoft Excelは、米国Microsoft Corp.の商品名称である。

さらに、マルチビューの指定に検索条件指定を追加すれば、マルチビューによって絞り込まれた文書群に対し、属性検索や全文検索を実行し、欲しい文書を探し出せる。

3 階層型プリサーチ方式による高速全文検索 —Bibliotheca/TS—

3.1 高い検索性能

Bibliotheca/TSの特長は、第一にその検索性能にある。日立製作所独自の階層型プリサーチ方式により、検索タームの存在する可能性がない文書を読み飛ばすことで、フルテキストサーチの検索速度を加速している。

最新バージョンでは文字成分表の構造を変更し、プリサーチの段階で性能を低下させることなく、より正確な検索結果を得られるようにした。またプリサーチの段階で、より対象が絞り込まれるため、テキストサーチで余分にサーチする必要がなく、性能を向上することができる。この結果、大規模の文書データベースでも高い性能を確保することができる。例えば、おおよそのヒット件数を得る高速検索モードでは、100万件の新聞記事データベースで、高い精度の検索結果をわずか1秒で得ることができる。

3.2 異なる表現への展開やあいまいな表現での検索

Bibliotheca/TSは指定されたキーワードに完全に合致する文書だけではなく、記述や表記の異なる表現や同義語に展開しての検索や、ワイルドカードを用いたあいまいな表現での検索にも対応している。異表記については、「コンピュータ」というキーワードを指定すると、「電子計算機」という同義語や「コンピューター」などの異表記に展開して検索することができる。しかも、同義語や異表記に展開した場合でも、多重文字列照合方式で実現しているため、検索性能の劣化はない(表1参照)。

4 世界標準規格SGMLによる文書処理 —DOCINTEGRA—

4.1 グループワークにおける文書作成編集処理の課題

ワープロソフトによって個々の文書作成編集の効率化が実現されたが、文書形式が各社のワープロソフトによって異なっているため文書交換が困難である。また、文書形式が印刷レイアウトを強く意識したものであることから、複数文書を一つにまとめる場合に再編集が必要になるなど、効率の悪いものとなっている。この問題を解決するために、文書交換を目的とした文書表現形式である国際規格(ISO8879) SGMLに基づく文書処理が注目さ

表 1 異表記展開の例

10種類以上の異表記に展開されるキーワードも存在する。しかも、展開した場合でも検索性能の劣化はない。

[illegible]

れている。

4.2 SGMLに基づく文書処理

DOCINTEGRAはSGMLに基づく文書処理であり、**図2**に示すように三つのプロセスを支援する。

(1) 文書設計プロセス

文書の論理構造および文書の体裁を設計するプロセスである。このプロセスで作成された文書論理構造定義(DTD: Document Type Definition), および文書の体裁情報(レイアウトルール)をDOCINTEGRAで利用する。

(2) 文書作成・編集プロセス

文書作成者が文書の内容を論理構造に沿って入力・編集するプロセスである。このプロセスではDOCINTEGRATE/SEにより、文書の構造に対応したタグ付きのテキスト (SGML文書) を負担なく編集することができる。

(3) プレゼンテーションプロセス

文書設計プロセスで作成したDTDとレイアウトルールを利用し、DOCINTEGRA/ALによってSGML文書を自動的にレイアウトし、表示あるいは印刷を行うことができる。

このように文書の論理構造・内容と文書の体裁情報を分離することにより、従来のワープロソフトと比較して次に示す利点が得られる。

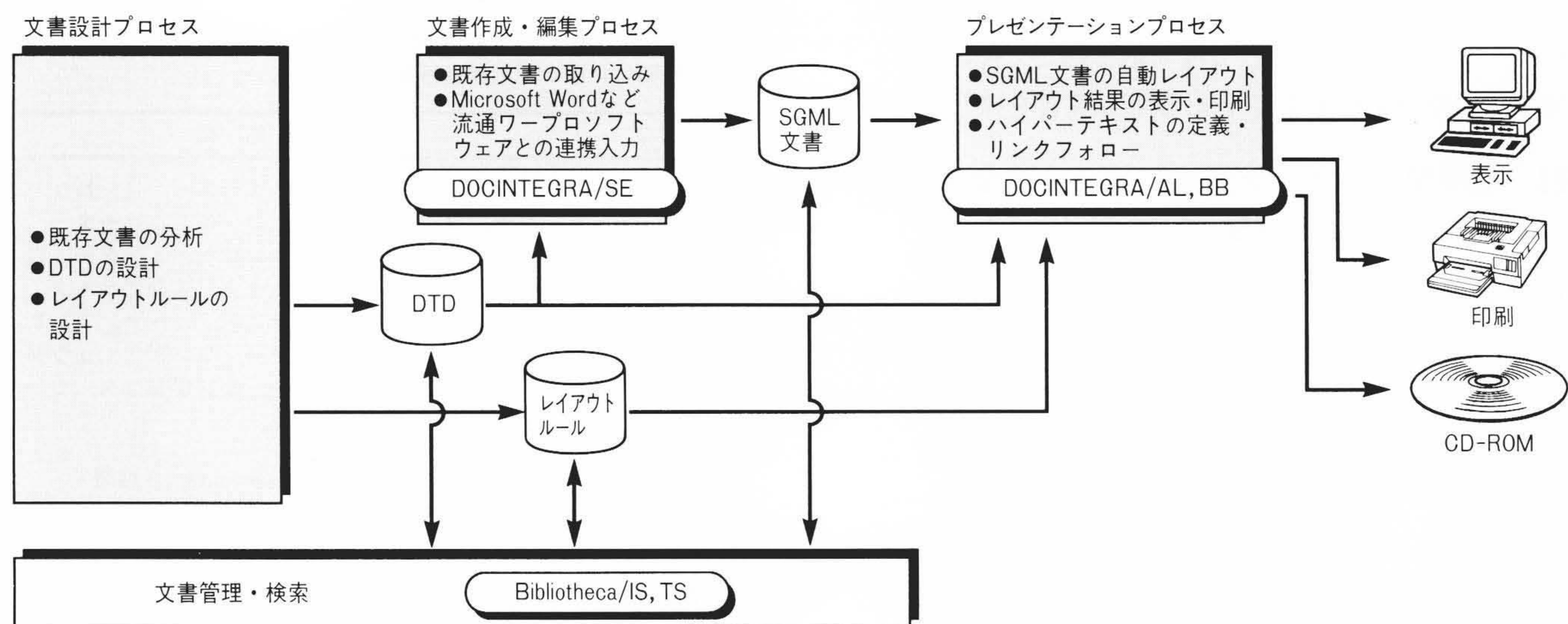
(a) 文書作成効率向上

文書作成・編集で、文書作成者が論理的な内容だけに専念できる。

(b) 共同執筆で文書の論理構造の統一が容易

(c) 検索、情報抽出などほかのシステムとの連携が容易

さらにSGMLによる利点を発展させ、文書の部品化、ハイパーテキスト化などの機能を実現している。またDOCINTEGRAでは、(1)部分化された図・表を画面上ではり付けるだけの簡単な操作で文書構成登録ができ



注：略語説明など (Bibliotheca；ソフトウェア群)，DOCINTEGRA/SE (構造化文書エディタ)，DOCINTEGRA/AL (自動レイアウト) DOCINTEGRA/BB (ブックブラウザ)，SE (Structured Document Editor)，AL (Auto Layout)，BB (Book Browser)，IS (Infoshare) TS (Text Search)

図2 SGMLに基づく文書処理モデル

標準規格SGMLによって文書の流通を促進する。Microsoft Wordなど流通ソフトウェアで作成した文書もSGML文書として取り込むことができる。

る、(2) 文書のブラウジング時に目次・索引の情報から検索できるなど高い操作性を実現している。

5 適用例

地方自治体での法規事務支援システムでは、条例や規則をBibliotheca/ISのマルチビューを使用して分類・管理し、改正作業ではBibliotheca/TSを併用して漏れなく関連する条例や規則を探し出す。さらに、DOCINTEGRAを利用して、SGML文書化した条例・規則の改正案作成、およびレイアウト作業の自動化による作業支援ができる⁴⁾。

製薬・化粧品会社では、厚生省が実施予定しているSGML文書での医薬品、医薬部外品、化粧品の許認可申請に対応する申請業務システムに、DOCINTEGRAを適用している。

日立製作所社内では、ソフトウェア開発本部の設計部署で、仕様書、議事録、レビュー報告書、その他設計ド

キュメントの管理に適用している。品質保証部門では、事故や不良の事例管理と検索サービスにBibliothecaを適用している。また総務部門では、技術情報、運用マニュアル、規格・規則、帳票類の集中管理や複写・印刷サービスへの適用を推進している。

6 おわりに

ここでは、企業の文書処理全般について支援するBibliothecaについて述べた。

今後は、画像、音声、動画などのマルチメディア文書の管理、インターネットと接続した検索システムや電子図書館システム、WAN(Wide Area Network)や移動体通信を介したりモートコンピューティングへの対応などを順次展開していく。また、日立グループウェア“Groupmax”へ技術を展開し、さらに効率の良い文書情報システムを提供していく考えである。

参考文献

- 1) 星 暁雄：文書情報システム，日経オープンシステム (1994.9)
- 2) 手島 透：イメージデータの活用，日経オープンシステム (1994.12)
- 3) 松田，外：文書情報管理システムBibliotheca/IS(1)(2)(3)(4)，第49回情報処理全国大会(1994)
- 4) 栗原，外：行政事務の高度化を目指す文書情報システム，日立評論，76，3，259～264(平6-3)