

鉄道総研における文書管理システムの開発

4D-1

松岡 彰彦 野末 道子 池田 宏
 (財) 鉄道総合技術研究所

1. はじめに

鉄道総研では、国鉄時代から国内外の鉄道技術資料をはじめとする数多くの文書情報の保管・管理を行ってきた。しかし、これらの多くは印刷物資料であり、その利用頻度に比べて保管には大きなスペースを必要とする。また、管理が個々の担当者に任されているため、担当者が不在の場合に検索ができない等の不都合が生じている。

我々はこれらの問題を解決し、既存資料の有効活用を計るため、全文データベースを利用した文書の蓄積から検索・再編集までを統合的に管理するための試作システムの開発を行った。

本報告では、主にシステムの構成および機能について概説する。

2. システム構成

本システムは図1に示すように各データベースを管理するためのサーバ（WS）とそれにアクセスするための各アプリケーションからなるクライアント（PC）によって構成される。

(1) ドキュメント管理サーバ

ドキュメント管理サーバには文書管理、図表管理、全文の3つのデータベースがある。基本的には文書管理と図表管理の各データベースは、文書データと図表データのファイル管理を行うためのものであり、リレーショナルデータベースを用いて索引情報などを管理する。また、文書管理データベースでは、オリジナル文書（各種ワープロ文書）と構造化文書（SGMLに準じたタグ付き文書）による2種類の管理形式があり、後者には図表データファイルへのリンク情報なども組み込まれている。

一方、全文データベースには文書内の全文に対して、フリーキーワードによる検索が行えるように構造化文書の全テキスト情報（タグを除く）を独自の形式で変換（学習）し、登録管理する。

(2) ドキュメント管理クライアント

PC上で稼働するドキュメント管理クライアントは、サーバ上の各データベースへのアクセス（登録、検索）と検索結果を利用して文書を作成する（編集）

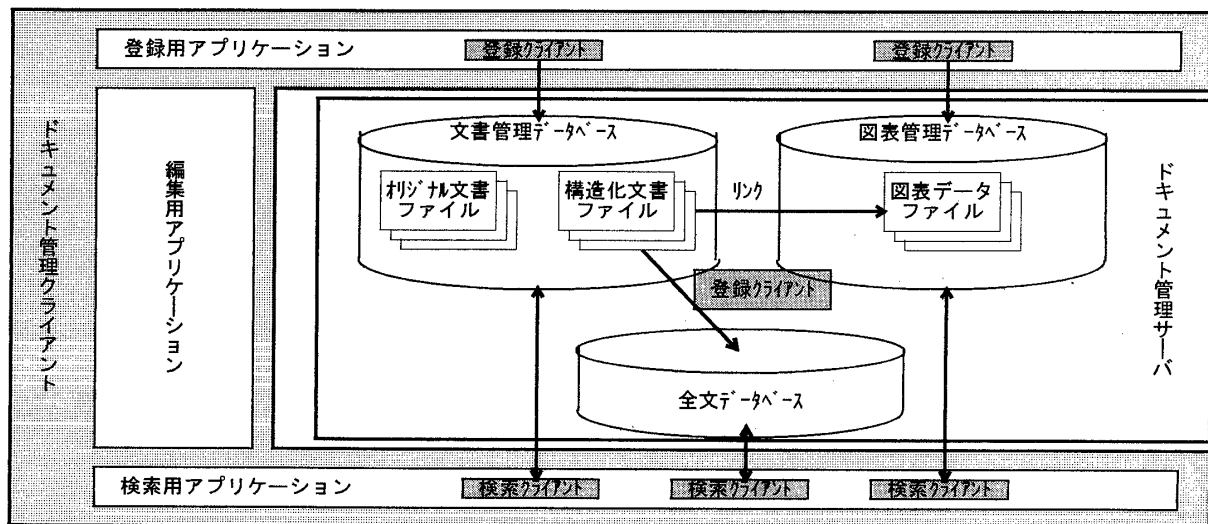


図1. システム構成

Development of Document Management System in RTRI.
 Akihiko Matsuoka; Michiko Nozue; Hiroshi Ikeda.
 Railway Technical Research Institute.
 2-3-38 Hikaricho, Kokubunji-shi, Tokyo 185 JAPAN

3つのアプリケーションから構成される。また、登録用/検索用の各アプリケーションにはそれぞれのデータベースに対応するクライアントが用意されており、個々のデータベース毎にアプリケーションを構築する事が可能である。

3. 機能

3.1 登録

(1) オリジナル文書

本システムでは、対応するアプリケーション（ワープロ）を使用すれば再利用（編集、印刷）ができる文書ファイルをオリジナル文書と呼ぶ。また、図表データファイルを併せ持つワープロ文書については、それらを1つのファイルにまとめてオリジナル文書とすることもできる。文書管理データベースにはシステムが自動付与する管理番号とともに表題、著者名などの索引情報、文書ファイルの格納情報が登録される。但し、文書ファイルの実体は各利用者毎に定められたWS上に分散して格納される。

(2) 図表データファイル

本システムで取り扱う図表データは、上記のアーカイブファイルとは異なり、図、写真などのイメージデータファイル、あるいはロータス、エクセルなどの表データファイルを文書ファイルとは独立した個々のデータとしてデータベースに登録したものを指す。このため、図表データの登録には文書データの登録と同じように索引情報を付与する必要がある。

(3) 構造化文書

構造化文書はオリジナル文書、あるいはOCRで読み取った印刷文書をベースにしてテキスト部分を抜き出した後、文章構造解析プログラムを介してタグ情報を付与する等の編集処理によって作成される。作成した構造化文書は索引情報を付与した後に単独、あるいはオリジナル文書を伴って文書管理データベースに登録することができる。また、登録された構造化文書のテキストデータは自動的に、全文データベースに登録される。

3.2 検索

(1) 文書管理データベース

登録文書に付与する管理番号には形式（オリジナ

ル文書、構造化文書、全文DB文書）が異なる同一文書を管理する共通文書番号と各文書に固有な個別文書番号の2つがあり、文書管理データベースでは各索引情報をキーとして共通文書番号が検索される。

また、共通文書番号をキーとしてオリジナル文書、構造化文書をダウンロードすることができる。

(2) 図表管理データベース

基本的な図表データの検索方法としては、上記の場合と同様に索引情報をキーとして図表データ番号を検索し、これを元に図表データファイルをダウンロードする方法が取られる。また、検索した構造化文書の図表データへのリンク情報を使って、同様な処理を行うことができる。

(3) 全文データベース

全文データベースの基本的な検索は、利用者が入力した文字列を登録文書の中から探し出し、結果として、共通文書番号、あるいはその文脈（段落）テキストをクライアントに送信する。

3.3 編集

編集処理には、構造化文書を作成するテキスト編集モードと印刷イメージ（レイアウト情報を含む）で文書を作成するDTP編集モードの2つがあり、前者で作成したものは構造化文書、後者で作成したものはオリジナル文書として新規に登録することができる。（但し、既登録文書の更新は不可）

また、編集処理においては文書作成の補助機能としてテンプレートを利用することができ、テンプレート項目をキーとした全文データベースの構造化検索を可能とした。

4. おわりに

当研究所で管理するコンサルティングや研究報告書の中には、同様な書式で類似の文書が作成されることが少なくない。また、一定書式での報告書作成が求められることも多いため、本システムが有効に活用できると考えている。

参考文献

- [1] 野末 道子、松岡 彰彦、池田 宏。「鉄道総研におけるトータルコミュニケーションシステムの開発」．情報処理学会第49回全国大会講演論文集(1) p. 1-383 (1994)．