

鉄道総研におけるトータルドキュメンテーションシステムの開発

5B-3

野末道子 松岡彰彦 池田宏

鉄道総合技術研究所

1. はじめに

鉄道総研には、コンサルティングや研究テーマなど各種の報告書がある。これらの中には、同様な書式でたびたび類似の文書が作成されるものも多く、これらは検索、文章の再利用の頻度が高いと言える。しかし、研究者は印刷文書のみ提出し、電子テキストの提出は義務づけられていない。そのため、①他の研究者との文書の共有ができない、②電子テキストの増加により個人レベルでの報告書管理さえ困難、③異動、出向、退職などにより資料の紛失、等の問題が挙げられる。

現在、電子テキストの共有化は、直接的なフロッピーディスクの交換などを除き不可能な状況にある。そこで、蓄積、検索、ダウンロードの機能を持つ報告書の全文データベースシステムにより、文書の共有化、個人レベルでの文書管理支援を図っていく必要性が生じている。

全文データベースの構築には、①全文が電子的に蓄積されているために欲しい情報がすぐに見られる（即時性）、②全文テキストのダウンロードにより文献の編集が簡単になる（文献の再利用）、③抄録作成の過程で生じる、情報の選択もれがない（概念の網羅性）、④検索のためのキーワードが多く含まれている（キーワードの網羅性）といった利点がある。

本研究所では、この報告書の作成支援、検索、および再利用の流れを統合したトータルドキュメンテーションシステムを試作した。ここで蓄積している報告書の分析と、システムの概略について以下で述べる。

2. 報告書の項目整理

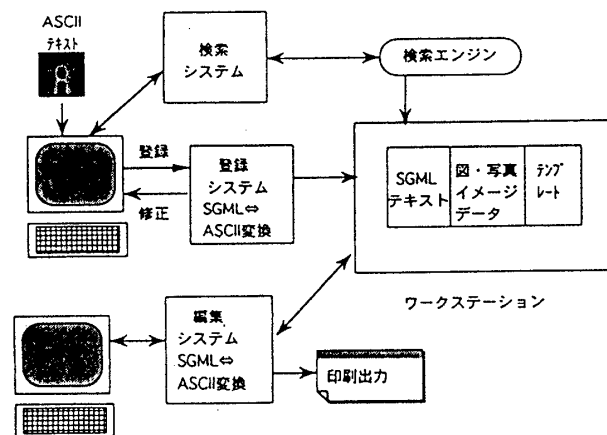
原著論文について神門¹⁾は、構成要素カテゴリを設定し分析している。鉄道総研における報告書についても、この構成要素カテゴリのように共通の項目がある。

例としては、「目的」、「方法」、「実験」、「結果」、「考察」等の報告書の一般的な枠組みと、「実験日時」、「場所」、「供試品名」等の項目というようになる。

報告書の要件としては、読みやすい文章であることは当然として、報告書として必ず含まれるべき項目が記述されなくてはならない。また報告書によっては、供試品、場所、日時、等の項目内容は出現していても、これらの項目について別の用語が利用されている場合、もしくは項目名が明記されない場合がある。そこで、報告書間で、文章が構成される際の枠組み、項目の共有化を行うことは、報告書の規格化、質の向上に有効である。以下では、この枠組みと項目をテンプレートと呼ぶものとする。さらに、テンプレートを用いて検索することにより全文データベースの検索効率を向上させることができる。

3. システム構成

システムの構成を第1図に示す。文書の蓄積・検索はワークステーション上で行い、研究者の個々の端末の対話画面から、登録、検索、編集、印刷を行うことができる。



第1図 システム構成

3.1 テンプレートの利用

現在登録の対象として検討している報告書は、鉄道材料関係のコンサルティング報告書である。この報告書の目次（約70件）の枠組み及び項目の分析を行いテンプレート

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

ト項目を決定した。テンプレートの一部分を第2図に示す。このテンプレートのうち%部分の項目については、登録の際、記述を義務づける必須項目とした。一方●部分の項目は、報告書の内容に応じて著者が選択、組み合わせ、及び登録が可能とした。

% 受付番号	% 受理日	% 発表コード
% 取扱種別	% 特許扱い	% テーマ番号
% テーマ名	% 発表者人数	% 発表者・分担比率
% 要題和文	% 要題英文	
% 抄録		
% 目次		
% コンサルティング対象会社		
% コンサルティング対応日時・機関		
% 依頼内容		
● 事故概況		
● 場所		
● 事故路線		
● 事故車両		
● 事故発生日時		
● 事故概要		
● 環境・気象条件		
% 供試品		
● 供試品の名称		
● 供試品の形状・仕様		
● 供試品の材質		
● 供試品の形式		
● 供試品のメーカー		
% 供試品の使用条件(環境)		
● 損傷・事故状況		
% 試験調査		
● 仮説		
% 調査試験手法		
● 調査項目		
● 測定・検査器材		
● 調査場所		
% 試験調査結果		
● 推察・推定根拠		
● 過去の事例		
● 事故原因(推定原因)		
● 再現試験		
● 対策・対応措置		
% 考察		
● 予測・勧告		
% 参考文献		
● 参照引用文献		

第2図 テンプレート(鉄道材料関係コンサルティング)

個々の文献は、それぞれが使用したテンプレートを持っており、検索の際に文献の目次情報として、ブラッキングすることができる。

3.2 文章の登録

文書の登録は、報告書の著者が作成したASCIIテキストの文章原稿部分及び表を登録する。このテキストは、章・節などの文章構造を自動分析する。文章構造及び表の記述には、日本化学会欧文誌等と同様^{2) 3) 4)}、文書記述言語であるSGMLを利用した。なお、文章構造の自動分析結果を著者に提示しこの修正を行うと共に、テンプレートの項目が記述されているかどうかのチェックを画面上で行う。

また、図、写真についてはイメージデータとして蓄積し文章中のどの部分からの参照であるかという情報をSGMLにより登録文章に記述する。

3.3 検索システム

検索システムでは、登録された報告書のうち、現在は

テキスト部分について検索可能となっている。

SGMLで蓄積された文書はさらに全文データベースの検索系へ登録される。ここで開発している検索系では、高速の全文サーチ検索及び、論理構造を利用した高度な検索テクニックを用いることができる。また、検索結果の適合度順出力ができるようになっている。このほかSGMLで認識している文章構造、及びテンプレートの項目を組み合わせた検索についても可能である。

一方、全文データベース検索上の問題点としては、当該文献中において重要な概念を表現していないキーワードも多く含まれることから、ノイズが多く含まれることになる。知的検索手法については、今後の検討の課題としていきたい。

3.4 編集・印刷

検索された結果の出力、及び報告書の印刷編集を行う際は、テキストのダウンロードを行う。

SGML形式文書は、印刷・出力の必要性に応じて、TEXもしくはDTPソフトを用いて、利用者自身(文書作成者、及び文書の利用者)が利用し、加工できるようになっている。SGMLで記述されている文書のTEXへの変換、及びMacintosh上で編集可能としている。

4. まとめ

今回対象とした鉄道材料分野については、比較的テンプレートを決定し導入が容易であった。このほかの報告書や研究論文についてもテンプレートの分析を行い、これに従って全文の管理を行うことは有効であると考えられる。現在はまだ、文章の登録、全文検索、目次情報検索、電子編集という基本的な機能について試作を行った段階である。今後は実際の利用に供することで、各種の問題点を検討していきたい。

[参考文献]

1. 神門典子. 構成要素カテゴリを用いた情報メディアの内部構造分析の試み: 原著論文を例として. 東京, 慶応義塾大学, 1991, 237p. 修士論文.
2. 石塚他. 日本化学会欧文誌のSGML形式全文データベースの構築・印刷そして検索. 情報処理学会研究報告, 情報学基礎. Vol. 93, No. 39, p. 1-8, May, 1993.
3. 安達淳. 特集: 電子図書館と学術雑誌. 電子図書館と学術雑誌. 情報の科学と技術. Vol. 44, No. 5, p. 247-253(1994)
4. 高須淳宏. SGMLと全文データベース. 情報の科学と技術. Vol. 43, No. 12, p. 1093-1103(1993)