

鉄道総研における文書管理システムの開発

4D-2

利用者による評価

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

電子テキストを共有化するために、米国においてはCALIS戦略が推進され、SGMLが標準文書形式として浸透しつつあるが、日本国内においてはまだSGMLに対応したワードプロセッサやエディタなどの環境があまり普及していない。そのため、一次文書を統合環境において管理を行うことは、必要性、要望はあっても実用段階にはまだ発展していない。しかし最近国内でも、マニュアルなどの管理が必要な電気メーカーやソフトウェアメーカーなどではこのようなシステムの導入、開発を行っている例はいくつかある。^{1) 2) 3)}

鉄道総研では、土木、車両、電気、心理などの研究活動の他に現地試験、材料試験などのJRからの依頼によるコンサルティング活動および受託活動を行っている。これらの報告書は、図書室に保管されている他、光ファイルに文書イメージが登録されている。しかし、ワープロなどにより電子化文書として作成されているにもかかわらず、これらの電子化文書が管理されていないために、文書の検索、再利用、共有化などが難しい。そこで管理する文書の形態、管理者、データの登録、利用者といったシステムをとりまく環境を検討し、プロトタイプシステムの開発を行った。今回はこのプロトタイプシステムによる文書作成に対する効果、および利用者による評価を報告する。

2. SGMLを導入した文書管理システムのメリット

今回開発を行った文書管理システムを導入することで、鉄道総研における以下の効果が期待できる。

・文書保管効率の向上

図書室の他に各研究室のロッカーに同一の報告書が保管されている。各研究室から報告書などの文書が簡単に検索できることにより、各研究室で重複保管する必要がなくなり省スペース化できる。

Development of Document Management System in RTRI; The estimation by user.

Michiko Nozue; Akihiko Matsuoka; Hiroshi Ikeda.
Railway Technical Research Institute.
2-8-38 Hikari-cho, Kokubunji-shi, Tokyo 185 Japan

・探索効率の向上

現在、文献検索システムにおいて報告書などをタイトルによって検索でき、イメージによる原文書の取得はできるが、原文書の検索及び電子化文書の取得はできない。そこで今回開発したシステムにより即座に原文書を入手することが可能になる。また、今回開発しているシステムではフルテキストサーチエンジンを利用しているので、インデックスの付与を行わなくても、全文中の言葉を対象として検索を行うことが可能である。

・文書作成者の文書管理を支援することができる。

電子化文書を作成することが当然となっている現在において、文書の管理は個人で行うことも容易ではなくなっており、既存の自分の文章を利用して再編集を行ったり、参考とすることが難しい状況にある研究者が多い。

また、蓄積文書へのSGMLを導入したメリットを以下に挙げる。⁴⁾

・オープンな情報共有化

既存文書の電子化を国際標準や業界標準のデータ形式で行うことにより、他の文書作成者との情報の共有、標準テキスト形式を採用した学会への論文投稿、WWWへの情報提供、プラットフォームに依存しない標準に基づいた情報交換、及び将来的な文書システムへの対応が可能となる。

・文章構造の認識

文書の構造や内容を識別するためのタグを利用して、蓄積効率の向上、全文検索などの検索効率の向上が図れる。

3. 開発システムの特徴

本章では文書の共有化を目的としたシステムは他にもいくつか存在しているが、今回のシステムで特徴的なものについて述べる。なお、システム全体の機能概要については松岡他による本大会の前発表を参照されたい。

本研究所においては、各研究課題毎に出される報告書として、部内向けを対象とした研究業務資料、JR各社向けの研究開発テーマ報告、部外一般に向け

た学会発表や雑誌投稿などがある。これらの資料はその対象によって内容、分量などの差はあるものの同一研究テーマにおける成果をまとめた資料であり、同じ文章が別の報告書に重複して現われることが多い。また、コンサルティング業務において発生することが多い、既存の報告書とほぼ同一形態、類似の文章を元にして作成される報告書があり、これらは定型の書式(一定の章項目のもとに作成される形式。以下、テンプレートと呼ぶ)を持っていると言える。今回のシステムはこのような文書の特徴を踏まえ開発を進めた。このほか、文書管理システムの設計に当たって検討を行った項目について表1に示す。

表1 システム設計にあたっての検討事項

検討事項	本システムでの方針
管理文書種類	・正式に提出された文書(報告書) ・会議資料、議事録等のプロジェクト内で発生した資料
形態	・決まった章項目見出し(定型書式)を持つ資料 ・特に定型書式を持たない資料
管理者	一次文書管理ワークステーションの管理は各部門ごとに任せ、ユーザ管理、プロジェクト管理、セキュリティ管理を行うものとする。
データ登録	データは基本的に文書作成者が登録を行う。
利用者	参加するプロジェクトや、所属する部署に関する文書について、検索を行うことができる。

4. 効果および検索評価

現在構造化文書の検索として一般的に考えられている機能としては1.段落、もしくは文章中の章、節を対象に部分検索する2.章、節タイトル中にキーワードが出現するものを検索する3.階層構造を利用した論理演算が可能、といったものがある。³⁾⁴⁾

この機能に加え、本システムでは特に文書の見出し項目テンプレートをシステムの重要なポイントとしている。この見出し項目一覧は文章を検索する際ブラウジングすることができる上、ある見出し項目を選択し、その下位構造となっているテキストを対象とした文章を検索することができる。(図1)例えば「××車両における○○事故」に関する情報を求めている検索者の質問は、テンプレート中の「事故車両名」という項目を対象として検索を行うことで、より精度の高い検索結果が得られるようになる。

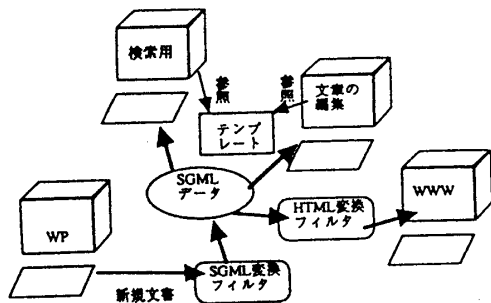


図1 文章の検索/編集

またテンプレートを利用することによる、文章作成、及び文章管理者のメリットも以下のように仮定し、これに対する評価を行った。

[文章作成者]

同一プロジェクトで作業を行う研究者が、どのような項目をトピックとして文章を作成しているのかわかる。また、既存の文章を参考にして、アウトラインプロセッサ的に文章を作成することができる。あらかじめ提示された文章形式で文章を穴埋め式に作成することで、文章の作成が楽になる。

[文書管理者]

必要項目を満たしていない文章についてはシステム側でチェックを行う機能を持たせている。文章の形態(構造)を一定形式(一定レベル)に保つことができることで、文書の質を一定にすることができる。(文章項目における漏れ、ノイズがない)

5. おわりに

現在のシステムでは、登録、検索などの部分でまだ改良を行っていく必要があり、今後積極的に利用者による運用評価を行っていきたいと考えている

また別の課題として、今後は、文章/構文解析等の手法を取り入れた意味的な文書の構造化や⁶⁾⁷⁾⁸⁾、検索用の同義語、シソーラスの構築、管理を行い、より語義的なあいまい検索を行うことができるような機能を取り入れることを検討している。

謝辞 本システムの作成に携わった(株)TESSの渡辺氏、鶴原氏に謝意を表します。

参考文献

1. 平賀正樹, 六条範俊, 熊谷典久. マルチメディア分散情報共有システム: FIND2. FUJITSU, Vol.45, No.5, p.427-435 (1994)
2. CALSの現状と NEC の取組み. Trend & Technology. C&C SYSTEM REVIEW. No.37, p.2-13 (1995)
3. 住田一男, 酒井哲也, 三池誠司, 小野頭司. 電子図書館のための効率的な文書検索; 検索/提示のための文書構造化と抄録作成. デジタル図書館. No.3, p.35-41 (1995)
4. 根岸正光, 石塚英弘編. SGML の活用. 総合マルチメディア選書. 東京. オーム社. 1994. 168p.
5. Macleod, I. A. Storage and retrieval of structured documents. Information Processing & Management. Vol.26, No.2, p.197-208 (1990)
6. 神門典子. 文章特性が異なる複数領域の原著論文の構造. 情報処理学会情報学基礎研究会. FI-33-4, p.17-22 (1994)
7. 三池誠司, 住田一男. 文の意味役割解析に基づく全文検索. 情報処理学会研究報告情報学基礎研究会. FI-34-3, p.17-24 (1994)
8. 青木孝弘, 大門一輝, 他. 文書構造に基づく技術論文検索システムに関する検討. 工学院大学研究報告第66号, p.216-220 (1990)