

J A X A 技術文書管理支援システムについて

JAXA Digital Arcives

田中陽子†
Yoko Tanaka

祖父江真一†
Shinichi Sobue

小田部昌行†
Masayuki Kotabe

生川 剛†
Tsuyoshi Narukawa

1. まえがき

旧宇宙開発事業団 (NASDA) において創立以来蓄積されてきた約 53 万件におよぶ研究開発および開発成果に関する技術文書の蓄積、共有、活用を効率的に行うため、宇宙航空研究開発機構 (JAXA) 情報化推進部では「技術文書管理支援システム」の整備を行ってきた。本システムは、JAXA 職員が PC 上から PDF ファイルなどにより電子化され蓄積された技術文書を WEB ベースで検索、閲覧、貸出依頼を実施し、開発プロジェクトにおける過去のプロジェクトの様々な技術情報の参照等に利用される技術情報蓄積データベースとして整備されている。

本論文では、本技術文書管理支援システムの概要、今後の課題について報告する。

2. 技術文書管理支援システムの概要

2.1 目的

旧宇宙開発事業団 (NASDA) においては、創設以来、約 53 万件におよぶ技術資料、成果報告書等の技術文書を蓄積、保管、管理し、職員がこれらの技術文書の検索、閲覧をネットワーク上で行えるように技術情報の管理・検索・閲覧システムを運用して来た。

しかし、平成 10 年に整備されたそのシステムは、老朽化が進み、増加の一途をたどる技術文書の保管には対応できなくなるとともに、また検索レスポンスが遅いなど利便性にも問題があった。さらに、平成 15 年 10 月の宇宙航空研究開発機構 (JAXA) において制定された情報セキュリティ規程への対応が行われていないため、部外あるいは社外に開示できない文書の電子ファイルが登録できない状況となった。

上記をうけて、利便性及び情報のメンテナンス性向上を図るとともに、開示制限情報の電子ファイルの閲覧サービスにも対応可能なシステムとして、「技術文書管理支援システム (通称：デジタルアーカイブ)」の整備を平成 14 年度から行った。

2.2 システム構成

本システムは、JAXA 職員が電子化され蓄積されてきた技術文書について、WEB ベース (NETSCAPE 4.78) で検索、閲覧、貸出し依頼を実施し、開発プロジェクトにおける過去のプロジェクトの様々な技術情報の参照等に利用されるための情報蓄積データベースとして整備を行った。

本システムは、インターネットプロトコルをベースとした WEB コンピューティング方式を採用しており、ユーザはブラウザソフト (NETSCAPE 4.78) をインストールすればよく、技術文書管理支援システムのアプリケーションの維持・管理を容易に実施することができる。また、技術文書管理支援システムのアプリケーションは、Java Servlet、

Java Server Pages (JSP)、Struts 等により開発を行い、開発の効率化を図った。

また、約 2500 人の職員等が使用するシステムであり、膨大な呼量によるサーバへの負荷が予想され、サーバの負荷を分散するために、SUN マシンの 2 台を用いて、データを蓄積する DB サーバと、アプリケーション及びユーザアクセス用の WEB サーバから構成している。なお、本技術文書管理支援システムは JAXA において外部に開示できない技術情報を保存、管理しているため、これらのサーバは JAXA の内部ネットワーク上に設置され、社内からのみ利用可能としている。ハードウェア構成概要については、図 1 のとおりである。

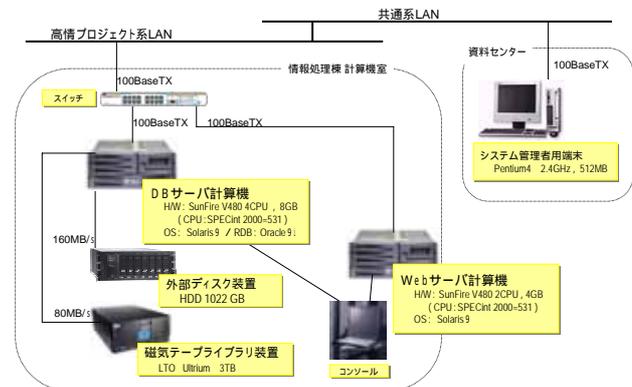


図 1. ハードウェア構成概要

2.3 利用者別の機能

本システムにおいては、技術情報の検索などを行う一般利用者機能のほかに、部外あるいは社外に開示できない文書についてのアクセス権の設定などを行う情報セキュリティ機能、ならびに登録されている技術文書の紙ベースの本文を管理する資料センターとの仲介を行う情報管理担当者機能およびシステム管理者機能を実装している。以下にそれぞれの利用者別の機能を示す。

(1) 一般利用者機能

- ・ 検索：キーワードを入力して、技術文書を文書単位とバインダ単位で検索。
- ・ 電子ファイルの表示、ダウンロード
- ・ 貸出依頼：本システムで管理している文書・バインダの資料センターで借用・閲覧するための機能。
- ・ 登録依頼：文書・バインダを資料センターへ登録依頼するための機能。

(2) 情報セキュリティ機能

- ・ アクセス権管理機能：利用者の電子ファイルへのアクセス権限の設定。

(3) 情報管理担当者機能

- ・ 登録管理：文書・バインダの書誌情報、所在情報、

電子ファイル等を登録。

- ・ 貸出 / 返却機能：文書・バインダの貸出 / 返却管理。
- ・ 廃棄機能：文書・バインダの廃棄処理手続きを行う。

(4) システム管理者機能

- ・ テーブル管理機能：管理項目の内容を一元管理
- ・ 利用者管理機能：ユーザ登録や抹消など
- ・ ユーティリティ機能：バックアップ、ログ取得などシステム維持、管理を目的とした機能

2.4 データ登録内容

本システムにおいては、ID、標題、サブタイトル、キーワード、作成者、作成日、セキュリティ区分、情報ステータス、保有年限、廃棄予定年月などの項目からなる「書誌」情報と、文書を電子化した「電子ファイル」の2種類のデータを本システムに登録している。これらは、一般の図書館情報システムと同様であるが、技術文書の管理において文書間の連携を強めるために、本システムでは「バインダ」という種別を作成している。このバインダには、バインダに綴じられた文書を複数紐づけることができる。したがって、書誌情報には、「文書」と1つ以上の文書を物理的にまとめた「バインダ」の2つの区分がある。

また、文書書誌には文書を電子化した1つ以上の「電子ファイル」を紐づけることができる。電子ファイルには、長期保存を目的とした「PDF ファイル」と、一時的な利用を目的とした Word や Excel などで作成された「アプリケーションファイル」の2つの区分を設け、それぞれ複数の電子ファイルを文書書誌に紐づけることができる。文書、バインダ、電子ファイルの関連は図2の通りである。

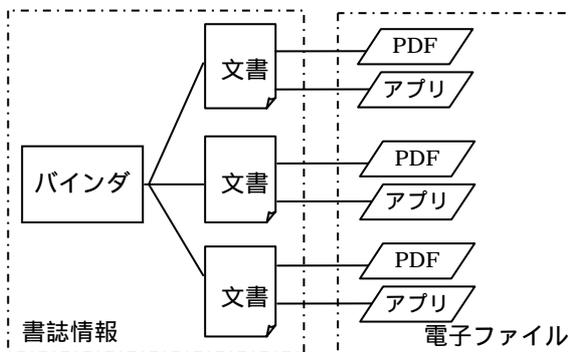


図2. 文書、バインダ、電子ファイルの関連

2.5 検索・閲覧

利用者は、文書とバインダの両方あるいはいずれかを対象として検索を行い、検索結果より文書 / バインダ ID をキーとしてそれらの書誌の詳細画面を表示することができる。文書は電子ファイルが登録されている場合は、検索結果画面および、書誌の詳細画面より電子ファイルを表示・ダウンロードすることができる。図3に検索画面と、検索結果画面を示す。



図3. 検索画面（左）と検索結果画面（右）

2.6 検索エンジン

検索エンジンは将来シソーラス検索を可能とすることを前提としたため Oracle Text を採用し、文書とバインダの「書誌」情報を対象としてテキスト検索を行っている。

実現方法としては、書誌テーブルとは別に全文検索用のテーブルを用い、これに Oracle Text の機能を適用することで、高速な全文検索を実現している。書誌情報は全文検索用テーブルに、キーワードとして集められ、10分おきに Oracle Text 索引が更新される。テーブルを別出しにすることで無駄なバッファを使用することなく、最大限に Oracle Text の機能を使用でき、さらに将来、検索専用サーバへと拡張が可能となる。

3. 今後の課題

今年度から本システムは運用を開始しているが、さらに利便性の向上などを目指し、次の課題に対応した機能改善を検討している。

(1) 全文検索

現在、PDF ファイルの作成は、Word や Excel などのアプリケーションファイルから PDF へ変換を行うのではなく、紙をスキャナで読み込み、イメージとして PDF ファイルを作成しているため、PDF ファイル内の情報を検索することができていない。したがって、検索はキーワードに依存しており、初心者にとっては少し使い勝手が悪い。そこで、OCR を行いテキストにして、書誌に加えて PDF の中身についても検索を可能とし、全文検索機能の使い勝手の向上を図る。

(2) キーワード辞書の作成

文書登録にあたってキーワードで使用する用語をできる限り整理・統一し、キーワード検索の有効性を向上する。

(3) 文書間の紐付け、関連性

本システムに登録されている文書間の紐づけができるようにする。例えば、プロジェクトでは文書体系が決まっており、それらを一括して整然と検索できるようにする。

4. 参考文献

- [1] 技術文書管理支援システムソフトウェアの整備（その3）成果報告書、三菱スペース・ソフトウェア（株）、2004