

先端科学技術（原子力分野）の理解増進に向けた 情報発信システムの構築

(財)高度情報科学技術研究機構 佐藤 治

報告概要

先端科学技術（原子力分野）に関する情報を分かり易い形式でまとめ、インターネットを通じて社会に幅広く発信することを目的として、過去 20 年にわたり原子力百科事典 ATOMICA が整備されてきた。ATOMICA は現在では原子力分野の信頼性及び専門性の高いデータベースとして、研究者、大学関係者、マスメディア関係者、国会議員など幅広い層に利用されている。

こうした状況に鑑み、財団法人高度情報科学技術研究機構（以下、RIST）は平成 21 年度から、文部科学省より ATOMICA を継承し、自主財源に基づいて整備、運営を行っていくこととした。継承した時点で ATOMICA には 2500 件余りのデータ（事典の個別コンテンツ）のストックがあったが、更新時期が古く内容が現状に即していないか又は正確とは言えないデータが多数含まれており、また、管理・運営にも古い設計に基づくシステムが使用されていて情報の更新に時間がかかるという課題を抱えていた。

そこで、これまでの管理・運営システムを根本から見直し、データベースを迅速に更新することができる効率的なシステムの構築と管理ツールの開発を行うとともに、既存の ATOMICA データの個別的精査・分類を行い、更新対象データをメンテナンスの可能な規模にすることによって、継続的に信頼性の高い情報を社会に発信していくための技術的基盤を整備することとした。この目的のため以下の作業を行った。

(a) リアルタイムのデータ提供システムの構築

これまでのシステムでは、データの更新・追加に際してシステム内のすべての公開用ファイル（HTML 形式）を再生成し、公開用サーバに転送・保存する方法を採用しており、データ更新・追加の結果を迅速に反映した情報を提供することが困難であった。

そこで、ATOMICA システムの設計を根本から見直し、①データ更新・追加後に直ちに登録作業を行ってデータベースサーバを常に最新の状態に維持し、②利用者がアクセスした際にその都度データベースサーバからデータを抽出して HTML データを生成し利用者のウェブ・ブラウザに出力するシステムを構築した。これによって、更新・追加の結果をリアルタイムに反映したデータを利用者に提供することが可能となった。

(b) データ管理ツール等の開発

データの頻繁な更新を可能とし、かつ効率的にメンテナンスを可能とする利便性の高い管理ツールを開発した。具体的には、平易なユーザインターフェイスを整備するとともに、概略以下に示す機能を組み込んだ。

- ・更新作業に必要な更新前データのテキスト及び図表のダウンロード
- ・更新済み原稿の書式等の自動チェック及びアップロード
- ・更新データのアップロード後の検索用インデックス作成処理
- ・既存のデータ・用語解説の削除または非表示処理
- ・データベースとリンク情報のバックアップ及び復元処理

また、ATOMICA ではデータ本文中の専門用語を用語辞書にリンクする仕組みとなっているが、これまでの管理ツールではこの作業を完全自動化し、データ更新・追加の度に ATOMICA のすべてのデータに対して本文の全文検索を実施してリンクを作成していた。この全文検索による用語辞書へのリンク作業には多大な作業時間が必要とされ、こまめな更新を妨げる主たる要因となっていた。そこで、データ作成時に執筆者が本文中に含まれる専門用語の中から適切なキーワードを選定し、このキーワードのみをリンクの対象とする方法を採用してリンク作業に要する時間の短縮を図った。

さらに、ATOMICA データと用語辞書の登録用データの書式（標準値、許容範囲、違反した場合の処置）を定義するとともに、これに基づく自動チェック機能を整備し、データベースに登録する際の誤入力の防止を図った。これらの措置を通じて登録作業の効率を向上させ、メンテナンスに係る労力の小さい管理ツールを整備することができた。

(c) 既存データの個別的精査・分類

更新対象データをメンテナンスの可能な規模にするため、既存の ATOMICA データの内容を個別的に精査し、更新の要否、緊急性、データの重要性等の判断に基づいて、以下のカテゴリに分類した。①更新対象とし更新間隔が小（概ね5年以下）、②更新対象とし更新間隔が大（概ね6年以上）、③更新の対象外（原子力に関連が薄いもの、過去の出来事の解説など）、④削除するもの（有意性が低いもの、他のデータと内容が重複するものなど）。このうち、③に該当するデータについては、本文表示画面に更新対象としない旨を明示するようにした。④に該当するデータに関しては、当面は非表示扱いとし、データそのものは保管しておくこととした。

また、今後年間 150 件程度の更新ペースでデータの最新性を維持することを想定して、更新対象とする①、②のカテゴリのデータを約 1100 件の規模に絞り込んだ。この作業の結果、全体の半数を超えるデータを更新の対象外とせざ

るを得なかったが、これは現状におけるデータの取り扱いを示すものであり、原子力を巡る今後の情勢変化、情報ニーズの変化等に対応して柔軟に運用を行うことを予定している。なお、非表示のデータに関しては、データを直ちにデータベースから消去するのではなく、当分の間は保存しておき、有益な情報が含まれているものに関してはできるだけ他の登録データに内容に移転してから、消去する予定である。

また、用語辞書についても約 3000 件の用語解説の全数点検を行い、記述の全体的な改訂が必要なもの、記述の部分的な改訂・追記が必要なもの、用語の選定・用語間の相互関係に問題があるもの、記述内容が辞書風にまとめられていないもの（解説風の記述方法で最後まで読まない用語の定義が分からないもの）、軽微な修正を要するもの（用語・字句の変更など）を摘出した。

この調査研究で実施した新たな ATOMICA システムの構築、管理ツールの整備、及び既存の ATOMICA データの精査・分類、用語辞書の点検を通じて、RIST が先端科学技術（原子力分野）に関して継続的に信頼性の高い情報を社会に発信していくための技術的基盤を整備することができた。今後は、この成果を活用して ATOMICA システムの整備、運営を行い、原子力分野の先端科学技術に対する国民各層の理解の増進に役立てる予定である。