

多言語により検索可能な鉄道技術用語辞典*

デモ 8

土屋隆司 野末道子 松岡彰彦

鉄道総合技術研究所

1 鉄道技術用語辞典

現在、鉄道総合技術研究所では、日本の鉄道技術において使用される語彙を体系化することを目的に、鉄道技術用語辞典の編纂を全所的な事業として実施している。具体的には、鉄道の各技術分野ごとに使用頻度の高い語彙を抽出し、用語間の関連を整理した上で、各用語の重要度に応じた内容説明を与えるという方法を取っている。本用語辞典は、印刷物としての出版以外に、CD-ROM やインターネットを使った提供も予定している。本稿では、インターネット版鉄道技術用語辞典における多言語化の試みについて述べる。

2 5ヶ国対訳表と多言語検索

近年、鉄道技術の世界においても研究・開発における国際協力の重要性に対する認識が高まりつつある。国際的な情報交換や情報共有の円滑化のためには、各国で使用されている語彙の整理と対訳表の作成は不可欠のものである。国際的な鉄道技術用語辞典編纂の試みは、UIC(=Union Internationale des Chemins de Fer, 国際鉄道連合)においても実施されているが、残念ながら、現状では西ヨーロッパの一部言語に限定されたものになっている。我々は抽出した日本語の語彙集合に対して、UIC や中国鉄道部等の協力を得て、日、英、仏、独、中の各国語による見出し語対訳を作成した。今後、これを用語辞典の本体とともに、WWW によって提供する予定である。WWW において多言語混在文書を提供する試みはまだきわめて少ないが、技術的にはいくつかの選択肢が提案されている。我々は、多言語データのエンコーディングについ

ては、ISO2022 準拠のエスケープシーケンスで各言語の文字集合を切り替えて使用する方式を採用した。この方式が現在利用可能な多言語テキスト処理方法として最も問題が少ないと判断したためである。ISO2022 のデータは、W3-mode on Mule、Arena I18N などの多言語対応のブラウザでは表示可能である。しかし、これらのブラウザ自体は、一般にはそれほど普及していない上、多言語のフォントをブラウザ側で用意する必要がある等の問題点もある。より普及度の高いブラウザでの表示を可能にするため、図書館情報大学で開発された多言語表示用ゲートウェイシステムである MHTML [1] を導入し、対訳表の表示手段のひとつとして当サイトにおいて運用することを計画している。用語解説については現在は日本語のみではあるが、外国語による検索機能も試験的に実装した。これは、日本語入力環境のない端末からの利用や日本語が母国語でないユーザによる利用を考慮したものである。検索プログラムは、ユーザが入力した検索文字列と言語情報に基づいて対訳表を検索し、該当する用語とその解説を提示する方法を取っている。本システムは、研究所内での試験運用を経た上で、適当な時期にインターネットでの公開を行なう予定である。

謝辞

本システムの検討、開発にあたり、図書館情報大学の MHTML プロジェクトの皆様に御協力をいただきました。ここに謝意を表します。

参考文献

- [1] 前田、藤田、阪口、杉本、田畑、「WWW 文書のための多言語ブラウザとそのゲートウェイサービス」情報処理学会情報学基礎研究会 44-1, 1996

*Railway technical term dictionary with multilingual support. Ryuji Tsuchiya, Michiko Nozue, Akihiko Matsuoka, Railway Technical Research Institute, 2-8-38 Hikari-cho Kokubunji-shi Tokyo, Japan