

Web データベース仮想統合システムの構築

池田 隼人

現在、Web 上には数多くのデータベースが公開されている。しかしながら、それらはデータ自体には価値はあるものの、十分利用されているとは言えない状況にある。その理由は2つあり、1つはどのようなデータベースがあるのかわからないこと、もう1つはデータベースにアクセスするための技術が整備されていないことである。そこで、本研究では近年急速に整備が進んでいる Web 技術を活用し、データベースへの透過的なアクセスを確保することで、データベースへの視認性を向上させることを目的とする。

本研究で対象とするデータベースは、筑波研究学園都市で公開されている物理、化学、農学等、多種多様な約 2000 のデータベースである。これらは各研究機関の研究成果やデータを研究者間で共有する目的で作成されたものである。しかしながら、これらのデータベースは作成された時期が異なるため、CGI プログラムによって作成された古いタイプのものであれば、Web API を定義して作成された新しいタイプのものもある。さらに、古いタイプと新しいタイプそれぞれに様々なパターンが存在するため、単純に考えればデータベースの数だけプログラムを作成する必要があり、非常に煩雑である。そこで、本研究では、検索語の送信や検索結果の返戻方式などの様々な仕様を1つに集約することでシンプルに扱えるようなシステムを構築した。

まず、新しいタイプのデータベースに対しては、OpenSearch に対応しているデータベースのみを対象とすることで、1つのプログラムに限定した。一方、古いタイプのデータベースに対しては、多様なパターンの中から必要な情報だけを取り出す Web スクレイピングという手法を使うことで、返戻形式を2つのパターンに集約した。それに加え、本研究ではデータベース登録機能を実装し、個々のデータベースの仕様を記述することで、登録したデータベースを単一の操作で扱えるようにした。

以上から、本研究によって仕様を記述するだけで Web 上の多種多様なデータベースをあたかも1つのデータベース群であるかのように仮想的に統合することが可能になった。その結果、透過的なアクセスを確保することができ、データベースへの視認性を向上させることができた。

(指導教員 宇陀則彦)