

■宮澤彰 情報社会相関研究系 教授

【タイトル】

学術情報の総合目録データベースを作成

【本文】

図書館の目録と言えば、かつてはカードを作成していました。コンピューターが導入され、目録情報をネットワークで検索できるようになり、利便性が向上しましたが、その裏にはいろいろな仕掛けも必要です。例えば、ネットワーク上の図書館や所蔵先が、目録情報を1つのデータベースに書き込むようにすれば重複して入力する手間も省けて、より大規模な総合目録を作成することができます。こうして、20年ほど前に、国内の大学図書館を中心とする図書・学術雑誌について総合目録データベースを作成するためのシステム（NACSIS-CAT）を開発しました。以来、こうした総合目録作成に関連する課題についての研究を続けてきました。

インターネット上のデータの目録づくり

目録というのは、いわばデータに関してのデータですから、図書のメタデータとも言えます。さて、インターネットの時代になると、ネットワーク上にも、さまざまな論文や研究資料、文書などの情報資源があふれかえるようになりました。どこに何があるかという情報を目録（メタデータ）として記録して、誰もが使えるようにしておけば、有用な情報を検索したり、発見したりするのが容易になります。そのメタデータをどうつくったら良いのかという研究をしています。

こうした目的をもった人たちが世界中から集まり、1995年に設立されたDCMI(Dublin Core Metadata Initiative)という団体があり、そこに参加しています。DCMIでは、資料の属性を表すものとして、タイトル(Title)、著者(Creator)、主題(Subject)、日付(Date)、権利(Rights)…など、コアとなる15の基本語彙(要素)を定めています。私は、「著者とは何か」というように、それぞれの定義を決める委員会のメンバーです。これらをコンピューターが理解できる言語で資料に埋め込んでおけば、コンピューターがそれを読んでさまざまな処理ができるようになることを期待されています。世界中の文献や人、概念などに背番号を付けられれば、識別には楽ですが、それは不可能なので、できる限りあいまいさを排除できるような実用的指針を探っているところです。

情報を標準化して流通しやすくする

このほか、図書館の目録を作成するにあたって、派生的に出てきた文字コードの標準化などについても研究しています。漢字には「異体字問題」と言われる問題がありますし、各国の言語にはそれぞれに特殊な問題もあります。まず、世界中の文字をすべてコンピューター上で扱えるように、ユニコードとして知られる国際的な字体の統合化ルールを定めるプロセス作成にもかかりました。さらに、情報分野のJIS規格の整備にも努めてきました。

私の研究テーマの多くは、いずれも情報の表現にかかわるもので、情報を流通させやすく、利用しやすくすることを目指したものと言ってよいでしょう。(取材・構成 塚崎朝子)