



研究と国際標準

石崎 俊 慶應義塾大学
ishizaki@sfc.keio.ac.jp

このシリーズの第1回はNECの平野さんの「うるう（閏秒）」の話でしたが、よもやま話なので、特に続きというわけではなく、第2回目としては研究と国際標準について、私自身の専門分野を例にして書いてみたいと思います。

■ 研究活動(パターン認識, 自然言語処理)

私は大学に所属していて、その研究・教育活動の主なテーマは自然言語処理やパターン情報処理です。人間の使う言葉と音声、それに画像をコンピュータで扱いますが、その中でも自然言語処理関係では国内の学会活動だけでなく、国際会議のための海外出張を毎年多く経験しています。このような分野の特徴として、人間の認知機能と関係しているので複雑で高度なことと人間の情報処理の仕組みがよく分かっていないために、なかなか技術的な進歩がはかどりません。コンピュータが人間並みの能力を獲得するにはさまざまな学問分野、特に認知科学や脳科学などの基礎研究がまだかなり必要な段階です。ロボットに実装して高度な性能を発揮するにはもう少し時間がかかりそうです。

■ 関連する国際標準化

そのような状況ですので、パターン認識手法や言語理解の手法そのものの国際標準化についてはまだ技術レベルの成熟が高まっていないと考えられます。しかし、研究のための材料となるデータや評価法については標準化の動きが始めています。たとえば、自然言語処理の基礎データとしての言語コーパス（新聞などの大量の電子化言語データ）は日本語ばかりでなく英語や多くの国の言語を対象に集められており、そのデータ構造などの標準化が進められています。また、文書の内容の意味を表現するための概念記述法としてはW3CのセマンティックWebを目指した活動もあります。これは機械翻訳を基礎にしたものですから、さまざまな言語を対象とするので国際標準化になじむ可能性があります。

音声情報処理の分野では、情報処理学会試行標準として、カーナビの音声認識システムの評価法などがいくつか公開されています。音声認識では言語ごとにシステムを作るので言語依存性が高く、国際標準に持っていきの切り口には工夫が必要です。一方、画像や音声の国際標準としてはISO/IEC JTC1 SC29のMPEGやJPEGが有名

ですが、情報圧縮法などの標準化とパターン認識法の標準化とではだいぶ趣を異にしています。

■ JTC1総会(テクノロジーウォッチ)

昨年11月中旬に、カナダのバンフでJTC1総会がありました。会議中にテクノロジーウォッチというワークショップがあり、JTC1における今後標準化の可能性のあるテーマがいくつか選ばれて講演されました。このワークショップでは、グリッドコンピューティング、モバイルIT（ヘルスケア関係）、プライバシー技術、インテリジェントシステムなどのテーマがあり、日本からの講演もありました。この最後のテーマが私の研究に近いのですが、カナダのアルバータ大学の若い先生がインテリジェントシステムの国際標準化の可能性について講演していました。パターン認識はそのような分野の中核となるものですが、講演でもやはり国際標準化は時期尚早という結論でした。実は上記のようなパターン認識のためのコンピュータ環境やデータについての標準化の可能性について話を聞いたかかったのですが、パターン認識の手法そのものの標準化についての話が中心だったようです。

■ 文字の標準化

一方、私自身の標準化活動から考えて見ますと、文字コードの国際標準化を中心に10年以上の活動を続けてきました。それは、JTC1の中ではSC2が担当しています。その中で漢字の標準化については、SC2 WG2 IRGというグループが担当しています。この標準化活動を始めた当初は、自然言語処理などのそれまでの専門分野とはほとんど関係がなかったのですが、しばらく前からは漢字フォント関係で標準化とからんだ研究や調査などを行っています。標準化そのものは研究業績になりにくいのですが、私のところでしばらく前にこの分野で論文博士が1人出て、いまでも博士課程の学生がいて研究を進めています。

■ 標準化教育

今年の2月の初めには、大学などにおける標準化教育についての国際会議が東京であったので私も参加してきました。日本ではまだ専門教育として定着していないかもしれませんが、アメリカ、カナダ、中国、韓国などではすでにいろいろな経験を積んでいるそうです。重要なテーマですので今後の発展が期待されています。

(平成18年3月1日受付)