

## 情報技術あれこれ (その3)

以前からメール文書を効率的に整理する方法を考えている。その動機はメールソフトの変更である。職場を移る度に、或いはそうでなくとも職場のメールソフトの契約が切れる度に、それに依存したメールデータの整理、保存方法も変更せざるを得ず、結果的に以前のソフトで送受したメールデータを引き出すことが非常にわずらわしくなる。仕事のやり方がメールベースとなり、大量のメールの整理が旧来の書類整理に代わっているの、これは困る。

そこで何年か前に、保存したメールファイル (eml形式) を整理するスクリプトを書こうと思い立った。仕様は簡単なもので、メールを送受信者名、タイトル、添付ファイル名も追記したプレーンテキストに変換するとともに、ファイル名を送受信日時+メールタイトルとするだけである。

これができるとメールの自動整理が格段にやりやすくなる。例えばメール送信者名と組織名、案件名を紐付けたリストをつくれれば自動的に対応するフォルダをつくりそこにメールファイルをふるい分けするなど簡単に言うことができるし、形態素解析やAIでもっと興味深い処理も行えるだろう。

ところがこれが簡単な作業と思っていたのだが、インターネット上を行き交う文書の仕様の奥深さを知るにつけ、それに面食らった。

インターネットでは初期からその通信記号としてASCII文字列しか扱えなかったため、添付ファイル等のバイナリデータもわざわざテキスト変換している。その方式が複数ある上に、文書構造がマルチパート化していて、それぞれにヘッダとボディがあり、適用される変換方式が異なってよいことになっている。また変換対象の文字コードも、JIS、Unicode、派生したものを含む多くの種類があり指定しなければならない。HTMLメール文は、HTML書法として別に定められている。OS固有の「ファイル名として使えない文字」もある。

...途中で挫けそうになった。

一方、この仕様は誰がどのように策定し配布しているのかというと、RFC (Request for Comment「意見などを下さい」と呼ばれる一連の文書の中にあるという。いまや全世界の(殆どの?)人々が利用している仕様にしては頼りない名前のある場所にあるのではないか?

このように複雑怪奇な世界を垣間見たいま、私の感想は是非ICT教育の基礎的な題材として採用すべきというもの。普遍的なものになったインターネットの歴史やメールの仕組み、プログラミングのヒント、参加型プロジェクトなど多くの有用な側面に触れることができるのではないかと思う。

どうでしょう...。そんな網羅的な教科書があればこんな苦勞はしなかったでしょうに。

(耐寒材料チーム 席研究員 安中 新太郎)

\* \* \* \*

表紙左上記号 ISSN 2432-2652の説明

国際的なコード番号であるISSN (International Standard Serial Number : 国際標準逐次刊行物番号)は、ISSN ネットワークが管理する、逐次刊行物を識別するための固有の番号です。この番号は国立国会図書館ISSN日本センターから付与されたものです。