

自作XML文字列解析ライブラリでリアルタイム・データをGETする

古城 隆

今日の天気 - 10月26日(月)		時間帯(時)	0-6	6-12	12-18	18-24
晴時々曇	降水確率		---	---	0%	0%
最高気温7℃ 前日差(+2.9) 最低気温1℃ 前日差(-)	風		北西の風 やや強く 後 西の風 海上では 北西の風 強く			
	波		3メートル うねりを伴う			
明日の天気 - 10月27日(火)		時間帯(時)	0-6	6-12	12-18	18-24
曇り時々雨	降水確率		0%	10%	10%	50%
最高気温13℃ 前日差(+) 最低気温6℃ 前日差(+3.8)	風		南の風 後 南東の風 やや強く 海上では 後 南東の風 強く			
	波		3メートル うねりを伴う			

図1 Weather Hacksが提供するオープン気象データをGETしてみる

ウェブ・ページからのデータ抽出方法を解説します。前章で述べたとおりウェブ・ページから取得できるデータのフォーマットは大きく二つあり、機械抽出に向くWeb APIと、人間が見る用のウェブ・ページに分かれています。本章では、人間が見る用のウェブ・ページからデータを抽出する事例を紹介します(図1)。

本章で解説するデータ抽出方法は、mbed FRDM-K64Fボードをベースに解説しますが、ほかのマイコンでも同じ考え方で利用できます。

リスト1 自家製のXML解析関数…これを使ってHTML記述のデータを取得することも可能

```

/* 指定されたタグを見つける */
const char *XML_getTag(const char *p,
                      const char *name)
{
    const char *find;
#define TAG_SIZE 20
    char tag[TAG_SIZE] = "<";
    if(p == NULL) return NULL;
    find = strstr(p, strcat(tag, name));
    /* "<"+タグ名 で文字列サーチ */
    if(find) {
        find = strchr(find, '>');
        /* ">" の後ろまで次のサーチのポインタを進める */
        return find + 1;
    } else return NULL;
}

```

(a) 簡易タグ発見用XML_getTAG関数

自作したXML解析関数

● HTML記述もデータ抽出は可能

ウェブ・サーバからHTML言語で提供されているデータにも、マイコン・システムで利用できそうなデータがたくさん含まれています。今回筆者は、XMLデータから目的データを取り出す関数(リスト1)を自作しました。HTMLによるデータから目的のデータを取り出すのも同じような方法で可能です。

ただしHTMLのデータは本来、人間がブラウザで見るためのデータです。ユーザの見やすさの改善やそのほかの理由で、ある日突然、データやハイパーリンクの構造が変更されたりすることもあります。

また、基本的にデータの著作権は提供元にありますので、注意してください。

● HTML解析でめんどくさい点…リンクをいくつもたどっていかないといけない

HTMLによるデータは、ブラウザ表示用のデータなので、ページの情報の中には表示フォーマットに関する指定などマシンが読み取るデータとしては不要なものもたくさん含んでいます。また、多くの場合、目的のデータを含むページに到達するのにいくつかのハ

```

/* 指定されたタグの要素を取り出す */
const char *XML_getElement(const char *p,
                          char *element, int size)
{
    const char *find;
    if(p == NULL) return NULL;
    find = strstr(p, "</");
    /* 終了タグをサーチ */
    if(find) {
        strncpy(element, p, min((find-p), size));
        /* 終了タグの直前までを要素として取り出す */
        element[min((find-p), size)] = '\0';
    } else *element = '\0';
    return find;
}

```

(b) タグの要素取得用XML_getElement関数