

大澤 文孝
OSAWA, Fumitaka

RSSで 更新情報を配信する

RSSリーダーの作成とコンテンツ更新時のRSS自動更新

Technology Tools

- Visual Basic .NET
- Visual C# .NET
- SQL Server 2000
- Oracle 9i
- Access 2002
- ASP.NET
- Internet Information Services
- Other:

Level



Samples

・この記事で取り上げたソースコードおよびサンプルプログラムは、付録CD-ROMの¥DOTNET¥FO1_02ディレクトリに収録しています。

¥RSS

本稿で取り上げたRSSリーダーのサンプルプロジェクト

- ・_README.TXT
- サンプルに関する注意書き

はじめに

Webサイトの更新情報を素早く知りたい。そんなときに利用されるのが、RSS (RDF Site Summary) です。

一言で言えば、RSSとは、Webサイトが提供している、

- ・主なコンテンツのURL
- ・更新日時
- ・更新頻度

などが記述されたXML形式のファイルです。

XML形式のファイルですから、RSSは手作業で書いてもかまいません。しかしコンテンツの更新とは別にRSSを用意するのは面倒です。

そこでブログのような日々更新されるWebサイトでは、コンテンツが更新されると同時にRSSも更新されるという自動更新の仕組みが提供されているのが一般的です。

RSSはXML形式のファイルなので、.NET Frameworkから容易に扱えます。

知らなければならないのは、その構造と意味だけです。

そこで本稿では、RSSの構造を説明したのち、具体的に.NET Frameworkで、どのようにRSSを扱えばよいのかを説明していきます。

RSSの構造

まずは、RSSの構造から説明します。

RSSは、図1に示すXMLの構造をとります。RSSファイルの拡張子は何でもかまいませんが、一般的には、「.rdf」か「.xml」です^[注1]。

RSSのルート要素は「<rdf:RDF>」です。

```
<rdf:RDF
xmlns="http://purl.org/rss/1.0/"
xmlns:rdf=
"http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```

RSSで必須となるのは上記の2つの名前空間だけです。

注1) RSSファイルは、XML形式なので、コンテンツタイプは、「application/xml」として返すようにします。

RSSには3種類のバージョンがある

RSSには「RSS 0.9x」「RSS 1.0」「RSS 2.0」の3つのバージョンがあり、書式が異なります(表A)。本稿で説明するのは、「RSS 1.0」です。RSS 1.0は、リソースに対して意味を定義するというRDF仕様(Resource Description Framework)に則られたものです。RSS 1.0のルート要素は、<rdf:RDF>で始まります。

それに対して、「RSS 0.9x」や「RSS 2.0」はRDFではなく、ルート要素が<rss>になるなどの違いがあります。

本稿では、簡単なRSSリーダーを作ってゆきますが、「RSS 1.0」としながら、名前空間に誤りがあるRSSとか、書式が正しくないRSSなども巷には出回っている状況です。そのため、中には読めないRSSも存在する可能性があります(とくにRSS 0.9xやRSS 2.0のRSSは、まったく読めません)。

本稿のサンプルでは、XmlDocumentオブジェクトのXPath機能を使ってパースしているため、少しでも書式が違っていると、うまくXPath式の条件にマッチせずに処理できません。

多少おかしなRSSでもきちんと読めるようにするというのは、思いのほか大変です。

表A：RSSのバージョンと仕様の入手先

バージョン	概要	仕様の入手先
0.9x	Netscape社がMy Netscapeのチャンネル機能を実現するために作ったものの改良版(RDFではない)	http://backend.userland.com/rss092
1.0	RDFに基づいて再設計されたRSS	http://web.resource.org/rss/1.0/spec
2.0	0.9xを拡張したもの(RDFではない)	http://blogs.law.harvard.edu/tech/rss

図1では、もうひとつxmlns:dcを定義していますが、これは「モジュール」と呼ばれるオプションです。詳細は後述します。

channel要素

RSSでは、提供されるコンテンツの内容が「channel要素」で示されます。channel要素では、このサイトのタイトルや説明などが記述されます。

channel要素の「rdf:about属性」が指すのは、このRSSファイルが「何について記したもののなのか」を示すURIです。一般にこの部分は、RSSファイル自身のURLとします。

channel要素の配下には、表1に示す要素を含むことができます。

items要素内に一覧を書き、その実態がitem要素で示される

実際にサイトで提供されているコンテンツの一覧は、channel要素の配下の「items要素」の中にあります。itemsの配下には「rdf:Seq要素」があり、その中にコンテンツ一覧を示す「rdf:li要素」が並びます。

rdf:li要素では、実際のコンテンツの内容は示されません。代わりに、rdf:RDF要素の直下にあるitem要素との結び付けが示されます。

rdf:li要素とitem要素とを結び付けるのが、「rdf:resource属性」と「rdf:about属性」です。

図1に示したように、rdf:li要素のrdf:resource属性で指定したURIと同じ値の「rdf:about属性」をもつitem要素とが繋げられ、それがコンテンツの説明部分となります^[注2]。

item要素では、表2に示す情報を定義できます。表2を見るとわかるように、item要素が、それぞれのコンテンツのタイトルやリンク先を示すものとなります。

機能を追加するモジュール

RSSは、いくつかの要素を追加することができる仕様になっています。追加される仕様のことを「モジュール」と言います。

モジュールを使うには、rdf:RDF要素において、適当な名前空間を指定します。

注2) rdf:resource属性とrdf:about属性で指定するURIは、多くの場合、実際のコンテンツのURLにしますが、そうでなくてもかまいません。ここで指定するのは、rdf:li要素とitem要素を結び付けるための識別子でしかなく、コンテンツのURLとは関係ありません。