

# PerlでXML::XPathを使いXMLからエントリを読み出す

2012/08/22

RSSのパーズにXML::XPathを使う雛形を残す。LWP::Simpleで取得、XML::XPathでパーズ、DateTime::Format::HTTPで日付の整形。

2013/04/06

こちらのドキュメントは管理都合によりwww関連ドキュメントのネームスペースから移動しました。

## ニコニコニュースの記事一覧最新情報切り出し

ニコニコニュースで提供されているRSS( <http://news.nicovideo.jp/media/article/1?rss=2.0> )はこんな感じ。



rssタグの下のchannelタグの下に、itemタグが繰り返し出現する様子がわかります。

itemタグに含まれるtitle,link,pudDateタグを取り出せば、ニュースタイトル、リンク、日付を取得できそうです。

## コード

[newsnicorss.pl](#)

```
use Encode;
use LWP::Simple;
use XML::XPath;
use XML::XPath::XMLParser;
use DateTime;
use DateTime::Format::HTTP;

my $rss = LWP::Simple::get(
    'http://news.nicovideo.jp/media/article/1?rss=2.0' );
my $xml = XML::XPath->new( $rss );
my $nodeset = $xml->find('/rss/channel//item');

foreach my $node ( $nodeset->get_nodelist() )
{
    my $mdt = (DateTime::Format::HTTP->parse_datetime(
        $xml->find('pubDate',$node)->string_value() )->ymd('-'));
    my $title = encode( 'euc-
jp', $xml->find('title',$node)->string_value() );
    my $desc = encode( 'euc-
jp', $xml->find('link',$node)->string_value() );
```

```
    printf("%s\t%s\t%s\n", $mdt, $title, $desc);  
}
```

端末でEUC-JPのコードを使っているのでEncodeを使って変換しています。

## 実行結果はこんな感じ

```
$ perl newsnicorss.pl  
2012-08-21 和む！ 癒される！ ほっこりする！ とにかくかわいい、天然の"天才役者"たち?動物  
動画 http://news.nicovideo.jp/watch/nw345984  
2012-08-21 福島に届かぬ"原発反対"の声 社会学者 開沼博さん<"どうする?原発"インタビュー  
第5回□ http://news.nicovideo.jp/watch/nw345722  
2012-08-16 "技術者として考えれば、原発の安全性に絶対はない" 元原子力プラント設計技術者、  
後藤政志さん<"どうする?原発"インタビュー第4回□  
http://news.nicovideo.jp/watch/nw340431  
2012-08-11 "誰の犠牲もなく、すぐ脱原発できるのか" ジャーナリスト・評論家、武田徹さん<  
"どうする?原発"インタビュー第3回□ http://news.nicovideo.jp/watch/nw336754  
2012-08-08 "生きるとは、他の生命を奪うこと" サバイバル愛好家 カメ五郎が語る"自然ライブ動  
画"の極意くう□主インタビュー第3回□ http://news.nicovideo.jp/watch/nw333494  
2012-08-06 "早く原発潰さないと安心して死ねない" ルポライター・鎌田慧さん<"どうする?原発"  
インタビュー第2回□ http://news.nicovideo.jp/watch/nw330855  
2012-08-01 "世の中が変わる瞬間を見てみたい" ロフト代表 平野悠<"どうする?原発"インタ  
ビュー第1回□ http://news.nicovideo.jp/watch/nw326358  
2012-07-31 "またやってね!" "泣かせないで" 夕刊ニコニコニュース"休刊"を惜しむ声  
http://news.nicovideo.jp/watch/nw325357  
2012-07-30 "神秘的空間"の広告利用は是か否か? "絶対領域"広告化に賛否両論  
http://news.nicovideo.jp/watch/nw324444  
2012-07-27 視聴者から嘆きの大弾幕 ニコニコニュース"お休み"のお知らせ  
http://news.nicovideo.jp/watch/nw322405  
2012-07-26 斬新な"踊る天気予報"にユーザーも驚愕 台湾のお天気アイドルが生出演  
http://news.nicovideo.jp/watch/nw321389  
2012-07-25 "いじめられた経験がある"が76.8% 『夕刊ニコニコニュース』アンケート  
http://news.nicovideo.jp/watch/nw320300  
2012-07-24 "ドッキリかと思ったよ" 視聴者も驚いたイチローの電撃移籍  
http://news.nicovideo.jp/watch/nw319279  
2012-07-23 "買う" "欲しい" "5万円で" "勝手に入るゴミ箱"に購入希望コメント殺到  
http://news.nicovideo.jp/watch/nw318320  
2012-07-23 "アントニオ猪木酒場が好きだった" 戦友が明かした"闘うジャーナリスト" 日隅  
一雄さんの意外な素顔 http://news.nicovideo.jp/watch/nw318111  
2012-07-20 女性のツボは職業系制服!? 働く女性が彼氏に求める"コスプレ"ベスト10  
http://news.nicovideo.jp/watch/nw316591  
2012-07-20 ニコニコ町会議"スタート地点" 鳥取県八頭町に高まる期待  
http://news.nicovideo.jp/watch/nw316507  
2012-07-19 "でも、お高いんでしょう?" "男の勝負服"を女性が選んでくれるサービスに賛否両  
論 http://news.nicovideo.jp/watch/nw315522  
2012-07-18 "幸せいっぱい" "かーちゃん無双" ニコ生ユーザーが芥川賞 直木賞の受賞者を祝福  
http://news.nicovideo.jp/watch/nw314521  
2012-07-17 "生活保護が浮いて見えるのは、世の中が地盤沈下したせい" 派遣村"村長"の湯浅誠氏が  
ナマポ・バッシングを斬る http://news.nicovideo.jp/watch/nw313526  
2012-07-17 失敗を必死でごまかすネコの仕草に、ユーザーが"かわええ"と絶賛
```

<http://news.nicovideo.jp/watch/nw313512>  
2012-07-17 「とにかく何かを作りたい！情熱が止まらないニコニコ動画の職人達」  
<http://news.nicovideo.jp/watch/nw313028>  
2012-07-13 「天使突抜」に「油谷向津具上」 "キラキラネーム"が地名にも進出!?  
<http://news.nicovideo.jp/watch/nw311199>  
2012-07-12 「注目する競技は？」の質問に「フィギュア」の回答 ニコ動ユーザーはロンドン五輪に"興味ない"  
<http://news.nicovideo.jp/watch/nw310008>  
2012-07-11 巨大な"初音ミク"が渋谷に出現!? ニコ生ユーザー「ちょっと渋谷に行ってくる」  
<http://news.nicovideo.jp/watch/nw308915>  
2012-07-10 鳩山元首相の処分軽減に驚く声も <http://news.nicovideo.jp/watch/nw307992>  
2012-07-09 女性キャスターの指導で視聴者も「北斎ヨガ」にチャレンジ  
<http://news.nicovideo.jp/watch/nw306884>  
2012-07-07 「明日から水道水で顔を洗う！」 『ジョジョ』荒木飛呂彦氏の若さの秘訣にユーザーも見習う？  
<http://news.nicovideo.jp/watch/nw305127>  
2012-07-06 動画アワード2011審査員賞受賞 手妻師、藤山晃太郎「ニコ動も伝統芸能も、わかる人同士で盛り上がるハイコンテキストカルチャー」<うp主インタビュー第2回  
<http://news.nicovideo.jp/watch/nw304565>  
2012-07-05 "神の粒子"はツンデレだった？ ヒッグス粒子を萌えキャラ風に解説してみた  
<http://news.nicovideo.jp/watch/nw304025>  
\$

[技術資料](#), [Perl](#), [XPath](#), [Simple](#), [DateTime](#)

From:

<https://wiki.hgotoh.jp/> - 努力したWiki

Permanent link:

<https://wiki.hgotoh.jp/documents/perl/perl-009>

Last update: **2024/11/01 16:30**

