

★ RSS Parsing in PHP & Interaction with CSS

This document available at
<http://brionews.com/cs4350tutor/>

RSS Formats

RSS is a family of web feed formats used to publish frequently updated digital content, such as blogs, news feeds or podcasts. [wikipedia]

- Really Simple Syndication (RSS 2.0)
- Rich Site Summary (RSS 0.91, RSS 1.0)
- RDF Site Summary (RSS 0.9 and 1.0) (on **RDF** specs.)
- RSS formats are specified in XML
- Users of RSS content use programs called feed "readers" or "aggregators"

Q. #1 - 2 html - css



Metadata, Semantic Web, RDF, OWL, DAML


Read my article at: http://brionews.com/cont/a_web_spider.php

Metadata

- Without the Metadata the Web simply doesn't work
- The programmed choice of the Web is to question not the texts but the Metadata.

The Semantic Web is the representation of data on the World Wide Web.

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in co-operation." -- Tim Berners-Lee, James Hendler, Ora Lassila - The Semantic Web, *Scientific American*, May 2001.

- DTD + SGML (ISO standard)
- Resource Description Framework (RDF) 
- DARPA Agent Markup Language (DAML)
 - *"The goal of the DAML effort is to develop a language and tools to facilitate the concept of the Semantic Web"* [daml.org]
- Ontology Web Language (OWL) - Relations, Cardinality, Equality, Symmetry
- N-Triples (subject, predicate, object)
- Ontology, definition: study of the existence; a branch of the metaphysics related to the nature of the human being; from the ancient Greek root: ont- "*being*".

Metadata, Semantic Web, RDF, OWL, DAML

```
<rdf:RDF xmlns:a="http://www.daml.org/2001/01/gedcom/gedcom#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:daml="http://www.daml.org/2000/12/daml+oil#">
  <a:Individual rdf:ID="@I1@"
    a:givenName="Michael Anthony"
    a:surname="Dean"
    a:sex="M">
    <a:spouseIn rdf:resource="#@F1@"/>
    <a:childIn rdf:resource="#@F2@"/>
    <a:birth rdf:resource="#event1820"/>
    <daml:equivalentTo rdf:resource="mailto:mdean@bbn.com"/>
  </a:Individual>
  <a:Birth rdf:ID="event1820"
    a:date="7 Aug 1961"
    a:place="Wyandotte, Wayne Co., Michigan"/>
  ...
</rdf:RDF>
```

RSS 1.0 - Example

```
<?xml version="1.0"?>
```

```
<rdf:RDF
```

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

```
  <channel rdf:about="http://www.xml.com/xml/news.rss">
```

```
    <title>XML.com</title>
```

```
    <link>http://xml.com/pub</link>
```

```
    <description> XML.com features a rich mix of information and
      services for the XML community.
```

```
    </description>
```

```
    <image rdf:resource="http://xml.com/universal/images/xml_tiny.gif" />
```

```
➔ <items>
```

```
  <rdf:Seq>
```

```
    <rdf:li resource="http://xml.com/pub/2000/08/09/xslt/xslt.html" />
```

```
    <rdf:li resource="http://xml.com/pub/2000/08/09/rdfdb/index.html" />
```

```
  </rdf:Seq>
```

```
</items>
```

```
</rdf:RDF>
```

RSS 2.0 - Example

```
<?xml version="1.0"?>
```

```
<rss version="2.0">
```

```
→ <channel>
```

```
  <title>Liftoff News</title>
```

```
  <link>http://liftoff.msfc.nasa.gov/</link>
```

```
  <description>Liftoff to Space Exploration.</description>
```

```
  <language>en-us</language>
```

```
  <pubDate>Tue, 10 Jun 2003 04:00:00 GMT</pubDate>
```

```
  <lastBuildDate>Tue, 10 Jun 2003 09:41:01 GMT</lastBuildDate>
```

```
  <docs>http://blogs.law.harvard.edu/tech/rss</docs>
```

```
  <generator>Weblog Editor 2.0</generator>
```

```
  <managingEditor>editor@example.com</managingEditor>
```

```
  <webMaster>webmaster@example.com</webMaster>
```

```
→ <item>
```

```
  <title>Star City</title>
```

```
  <link>http://liftoff.msfc.nasa.gov/news/2003/news-starcity.asp</link>
```

```
  <description>How do Americans get ready to work with Russians  
  </description>
```

```
  <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
```

```
  <guid>http://liftoff.msfc.nasa.gov/2003/06/03.html#item573</guid>
```

```
</item>
```

RSS 2.0 - Example

```
➔ <item>
  <title>Space Exploration</title>
  <link>http://liftoff.msfc.nasa.gov/</link>
  <description>Sky watchers in Europe, Asia, and parts of Alaska and
    Canada will experience a partial eclipse of the Sun</description>
  <pubDate>Fri, 30 May 2003 11:06:42 GMT</pubDate>
  <guid>http://liftoff.msfc.nasa.gov/2003/05/30.html#item572</guid>
</item>
➔ <item>
  <title>The Engine That Does More</title>
  <link>http://liftoff.msfc.nasa.gov/news/2003/news-VASIMR.asp
    </link>
  <description>Before man travels to Mars, NASA hopes to design
    new engines that will let us fly through the Solar System more
    quickly. The proposed VASIMR engine would do that.
    </description>
  <pubDate>Tue, 27 May 2003 08:37:32 GMT</pubDate>
  <guid>http://liftoff.msfc.nasa.gov/2003/05/27.html#item571</guid>
</item>
</channel>
</rss>
```

Atom 1.0 - Example

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">

  <title>Example Feed</title>
  <subtitle>A subtitle.</subtitle>
  <link href="http://example.org/" />
  <updated>2003-12-13T18:30:02Z</updated>
  <author>
    <name>John Doe</name>
    <email>johndoe@example.com</email>
  </author>
  <id>urn:uuid:60a76c80-d399-11d9-b91C-0003939e0af6</id>

  → <entry>
    <title>Atom-Powered Robots Run Amok</title>
    <link href="http://example.org/2003/12/13/atom03" />
    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
    <updated>2003-12-13T18:30:02Z</updated>
    <summary>Some text.</summary>
  </entry>

</feed>
```

Introduction End - Keywords You Must Know

Keywords

- Semantic Web
- Resource Description Framework (RDF)
- N-Triples (subject, predicate, object)
- Ontology Web Language (OWL)
- DARPA Agent Markup Language (DAML)
- DTD - SGML
- RSS 0.90 - 0.91 - 0.92 - 0.93
- RFC 4287 - 4685
- RSS 1.0
- RSS 2.0 (Harvard copyright)
- RSS 3.0
- Atom 1.0

People

- Ramanathan V. Guha
- Dan Libby
- Tim Berners-Lee
- James Hendler
- Ora Lassila
- Sean Palmer - Christopher Schmidt - Cody Woodard

Keywords

- Syndication
- Podcasting
 - Autocasting
 - Blogcasting
 - Mobilecast
 - Peercasting
 - Video Podcasts
 - Godcasting

Apps & Websites

- Mozilla Thunderbird
- NewsGator Online
- BitTorrent, Broadcatching
- Bloglines

RSS Aggregators (no effort required)

Websites one can read

- rssfeedreader.com
- my.yahoo.com
- bloglines.com
- Google Reader - www.google.com/reader/
- newsisfree.com
- Rawdog - rawfish7.tripod.com/rawdog.html
- + dozens more, probably hundreds

Websites generating scripts to use on your website

- newsisfree.com
- feedroll.com/rssviewer/
- www.rss-info.com/en_rssinclude.html
- gregarius.net/
- + dozens more

RSS Parsers (some effort needed)

Languages

- PHP + Pear
- Python
- Perl
- Javascript
- and, yes, Asp, Java too

Probably others

opensource Scripts

- feed2js
- magpie
- CaRP
- etc. etc.

Da Code - Yo-Hoo!

feed2js - Opensource, based on Magpie

maricopa.edu

previously at :

jade.mcli.dist.maricopa.edu/feed/

fork or something :

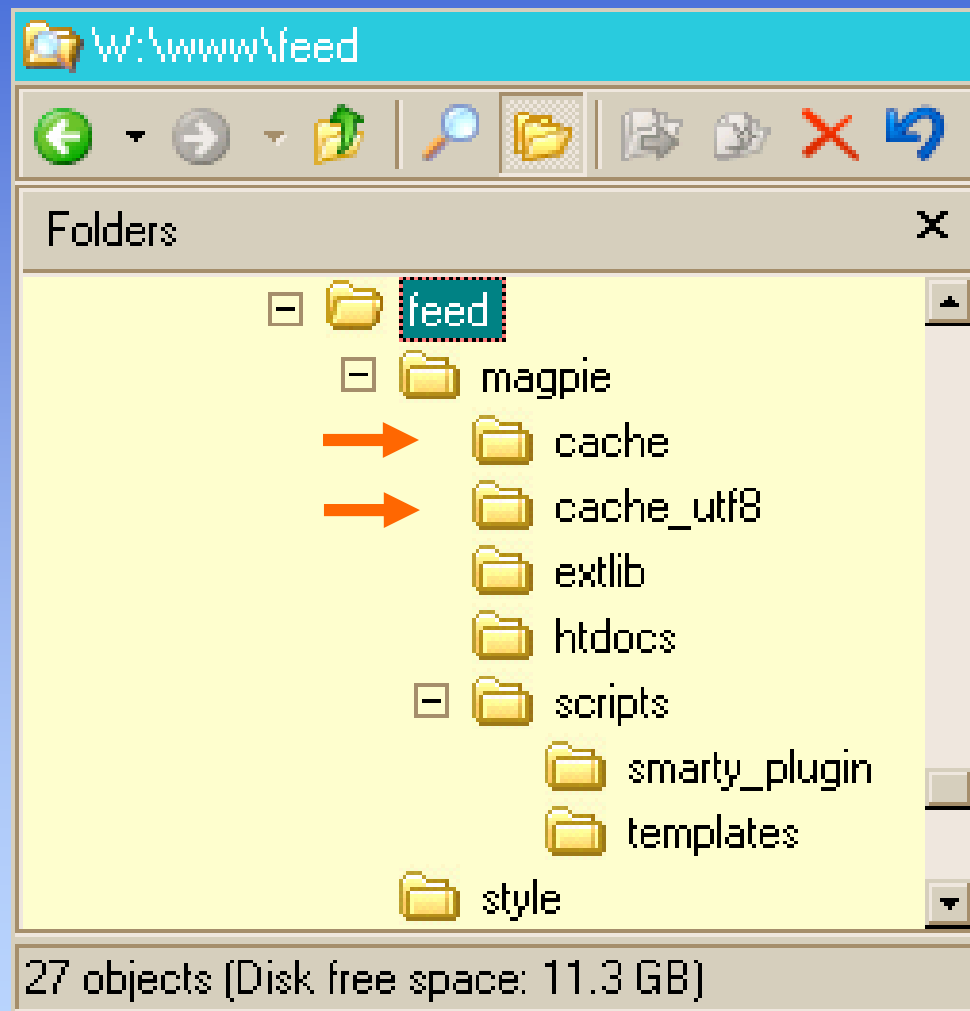
<http://feed2js.org/>

It offers both :

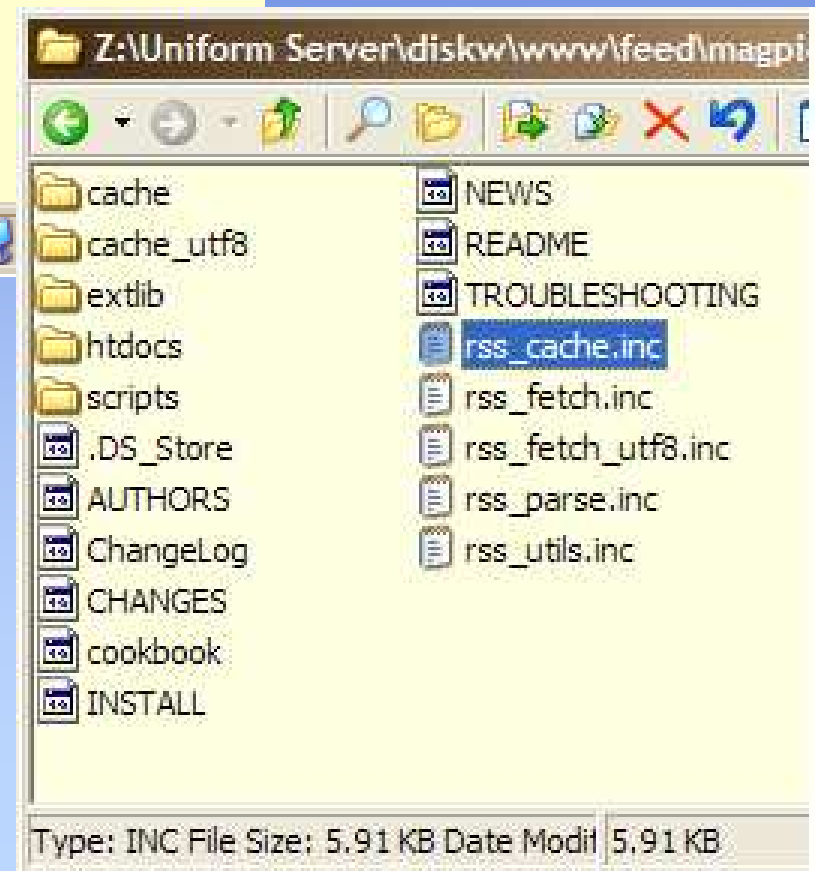
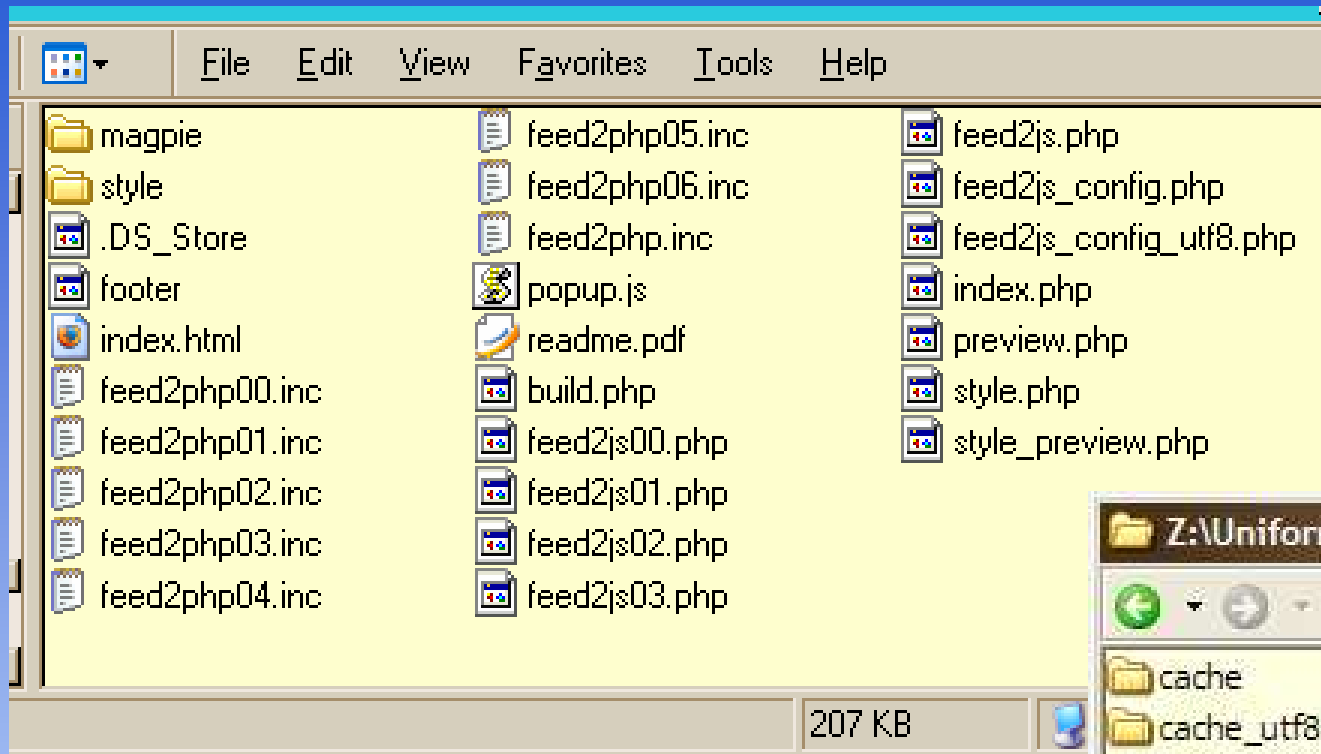
- free service, on site parsing, sends back CSS formatted links
- run your own version

Download it !

Da Code - Yo-Hoo!



Da Code - Yo-Hoo!



RSS Formats

- a. `feed2js.php` - the main workhorse script; see below for configuration details.
- b. `feed2js_config.php` - local configuration, set the file paths for Magpie, etc. Accessing this as a separate include will make future updates less tedious (a simple replacement of the `feed2js.php` file)
- c. `nosource.php` - a simple error file if the script is not provided an URL for the RSS feed.
- d. `build.php` - a version of our build page which will allow your web site users to easily generate their proper JavaScript that works from your installation of Feed2JS.
- e. `style.php` - a version of our style tool which will allow your web site users to select and modify CSS styles for their feeds.
- f. `style_preview.php` - generates the styled previews
- g. `styles` - sample style sheets used for the preview feature of the build script.
- h. `magpie` - a working version of version 0.71 of Magpie RSS.

feed2js_config.php

```
<?php
```

```
➔ define('MAGPIE_DIR', 'w://www/feed/magpie/');
```

```
require_once(MAGPIE_DIR.'rss_fetch.inc');
```

```
require_once(MAGPIE_DIR.'rss_utils.inc');
```

```
// value of 2 optionally show lots of debugging info but breaks  
JavaScript
```

```
// This should be set to 0 unless debugging
```

```
define('MAGPIE_DEBUG', 0);
```

```
➔ define('MAGPIE_CACHE_AGE', 3600); // $vince$ : 7200 is advised
```

feed2js.php

```
require_once('feed2js_config.php');

// trap for missing src param for the feed, use a dummy one so it gets
// displayed.
if (!isset($src)) $src = "http://jade.mcli.dist.maricopa.edu/feed/
nosource.rss"; // $vince$ : actually this URL appears obsolete

// check for utf encoding type
if (!isset($utf)) $utf = 'n';

if ($_GET['utf'] == 'y') {
    define('MAGPIE_CACHE_DIR', MAGPIE_DIR . 'cache_utf8/');
    // character encoding
    define('MAGPIE_OUTPUT_ENCODING', 'UTF-8');
} else {
    define('MAGPIE_CACHE_DIR', MAGPIE_DIR . 'cache/');
}
```

feed2js.php

```
// GET VARIABLES -----  
// retrieve values from posted variables  
  
// flag to show channel info  
if (!isset($chan)) $chan = 'n';  
  
// variable to limit number of displayed items; default = 0 (show all, 20 is a safe  
bet to list a big list of feeds)  
if (!isset($num)) $num = 0;  
if ($num==0) $num = 20;  
  
// indicator to show item description, 0 = no; 1=all; n>1 = characters to display  
// values of -1 indicate to displa item without the title as a link  
// (default=0)  
if (!isset($desc)) $desc = 0;  
  
// flag to show date of posts, values: no/yes (default=no)  
if (!isset($date)) $date = 'n';
```

feed2js.php

```
/* flag to open target window in new window; n = same window, y = new win-  
dow, other = targeted window, 'popup' = call JavaScript function popupfeed to  
display in new window (default is n) */
```

```
if (!isset($targ)) $targ = 'n';
```

```
if ($targ == 'n') {
```

```
    $target_window = ' target="_self";
```

```
} elseif ($targ == 'y') {
```

```
    $target_window = ' target="_blank";
```

```
} elseif ($targ == 'popup') {
```

```
    $target_window = ' target="popup" onClick="popupfeed(this.href);return  
false";
```

```
} else {
```

```
    $target_window = ' target="" . $targ . "";
```

```
}
```

feed2js.php

```
// flag to show feed as full html output rather than JavaScript, used for alternative
// views for JavaScript-less users.
//   y = display html only for non js browsers
//   n = default (JavaScript view)
//   a = display javascript output but allow HTML
//   p = display text only items but convert linefeeds to BR tags

// default setting for no conversion of linebreaks
if (!isset($html)) $html = 'n';

$br = '';
if ($html == 'a') {
    $desc = 1;
} elseif ($html == 'p') {
    $br = '<br />';
}
}
```

feed2js.php

```
// optional parameter to use different class for the CSS container
if (isset($css)) {
    $rss_box_id = '_' . $css;
} else {
    $rss_box_id = "";
}
```

End of output config parms

feed2js.php

```
// PARSE FEED and GENERATE OUTPUT -----  
// This is where it all happens!  
  
// Fetch the data, thanks Magpie  
$rss = @fetch_rss( $src );  
  
// begin javascript output string for channel info  
$str = "<div class=\"rss_box\" . $rss_box_id . \">\n";
```

feed2js.php

```
if (!$rss->channel['link']) {
    // error, nothing grabbed
    $str.= "<p class=\"rss_item\"><strong>Error! Error!</strong> No data was
found for RSS feed $src or no items available for this feed</p>/div>\n";
} else {
    if ($chan == 'y') {
        // output channel title and description
        $str.= "<p class=\"rss_title\"><a class=\"rss_title\" href=\"\" . trim($rss-
>channel['link']) . \"\" target=\"\" . $target_window . \"\">\" . strip_returns($rss-
>channel['title']) . "</a><br /><span class=\"rss_item\">\" . strip_returns($rss-
>channel['description']) . "</span></p>\n";
    } elseif ($chan == 'title') {
        // output title only
        $str.= "<p class=\"rss_title\"><a class=\"rss_title\" href=\"\" . trim($rss-
>channel['link']) . \"\" target=\"\" . $target_window . \"\">\" . strip_returns($rss-
>channel['title']) . "</a></p>\n";
    }
}
```

```
// Walk the items and process each one
$all_items = array_slice($rss->items, 0, $num);
foreach ( $all_items as $item ) {
    if ($item['link']) {
        // link url
        $my_url = $item['link'];
    } elseif ($item['guid']) {
        // feeds lacking item -> link
        $my_url = ($item['guid']);
    }
    if ($desc < 0) {
        $str.= "<li class=\"rss_item\">\n";
    } elseif ($item['title']) {
        // format item title
// $my_title = strip_returns($item['title']);
        $my_title = addslashes(strip_returns(substr($item['title'],0,82))); // $vince$
        // create a title attribute. thanks Seb!
        $title_str = substr(strip_returns(strip_tags($item['summary'])), 0, 255) . '...';
        // write the item
        $str.= "<li class=\"rss_item\"><a class=\"rss_item\" href=\"\" . trim($my_url) .
        \"\" title=\"\$title_str\" . $target_window . ">\" . $my_title . "</a><br />\n";
    }
}
```

```
} else {
    // if no title, build a link to tag on the description
    $str.= "<li class=\"rss_item\">\n";
    $more_link = " <a class=\"rss_item\" href=\"\" .trim($my_url) . \"\" . $star-
get_window . ">&laquo;details&raquo;</a>";
}
// print out date if option indicated and feed returns a value.
// Use the new date_timestamp function in Magpie 0.71
if ($date == 'y') {
// NOT working now on RSS 2.0 feeds, help Magpie!!
//     $pretty_date = date($date_format, $item['date_timestamp']);

// use the old approach until Magpie is fixed
    $in_date = "";
    $rss_2_date = $item['pubdate'];
    $rss_1_date = $item['dc']['date'];
    $atom_date = $item['issued'];
    if ($atom_date != "") $in_date = parse_w3cdtf($atom_date);
    if ($rss_1_date != "") $in_date = parse_w3cdtf($rss_1_date);
    if ($rss_2_date != "") $in_date = strtotime($rss_2_date);
    if ($in_date == "") $in_date = time();
```

```
// format date
```

```
if (strlen($rss_1_date) == 10) {  
    // hack to catch dc:date in format 200x-XX-XX  
    $pretty_date = date("F d, Y", strtotime($rss_1_date));  
} else {  
    $pretty_date = date($date_format, $in_date);  
}
```

```
$str.= "<span class=\"rss_date\">posted on $pretty_date</span><br />\n";  
}
```

```
// output description of item if desired
```

```
if ($desc) {  
    // Atom/encoded content support (thanks David Carter-Tod)  
    if ($item['content']['encoded']) {  
        $my_blurb = $item['content']['encoded'];  
    } else {  
        $my_blurb = $item['summary'];  
    }  
}
```

```
// strip html
if ($html != 'a') $my_blurb = strip_tags($my_blurb);

// trim descriptions
if ($desc > 1) {

    // display specified substring numbers of chars;
    // html is stripped to prevent cut off tags
    $my_blurb = substr($my_blurb, 0, $desc) . '...';
}
$str.= "document.write("" . addslashes(strip_returns($my_blurb, $br)) . "");\n";
}
$str.= "document.write('$more_link');\n";
}
$str .= "document.write('</div>');\n";
}
```

```
if ($html != 'y') {
    // Render as JavaScript
    // START OUTPUT
    // headers to tell browser this is a JS file
    if ($rss) {
        header("Content-type: application/x-javascript");
    }

    // Spit out the results as the series of JS statements
    echo $feedcheck_str . $str;

} else {
    // for HTML output, strip out JavaScript, strip escaped quotes
    $str = preg_replace("/document.write\(\\'/", "", $str);
    $str = preg_replace("/^\\)\;/", "", $str);
    $str = stripslashes($str);
}
```

```
// Now write a basic page with the feed as content
echo "<html><head><title>RSS Feed: " . $rss->channel['title'] . '</title>
    </head><style type="text/css">';
echo "#content { margin: 10px 20px;}
p { font-family: verdana, arial, sans-serif; font-size: 12px; margin-top:0;
    margin-bottom:1em;
    }
h1 { font-family: verdana, arial, sans-serif; font-size: 24px; margin-bottom: 2px;
text-align:center;}
</style>
</head>";

    echo '<body bgcolor="#FFFFFF"><div id="content">';
    echo '<h1>RSS Feed for ' . $rss->channel['title'] . '</h1><p>Note: Content for
this RSS feed is provided as a text alternative to inline RSS feeds that may not dis-
play on all browsers.</p>';
    echo "$str</div></body></html>";
}
?>
```

```
<?php
```

```
require_once('feed2js_config.php');
```

```
// check for utf encoding type
```

```
if (!isset($utf)) $utf = 'n';
```

```
if ($_GET['utf'] == 'y') {
```

```
    define('MAGPIE_CACHE_DIR', MAGPIE_DIR . 'cache_utf8/');
```

```
    define('MAGPIE_OUTPUT_ENCODING', 'UTF-8');
```

```
}
```

```
else {
```

```
    define('MAGPIE_CACHE_DIR', MAGPIE_DIR . 'cache/');
```

```
}
```

```
// GET VARIABLES -----
```

```
$chan = 'title';
```

```
$desc = 0;
```

```
$date = 'n';
```

```
if (!isset($num)) $num = 3;
```

```
// n = same window, y = new window, other = targeted window, 'popup' = call
JavaScript function
// popupfeed to display in new window - (default is n)

if (!isset($targ)) $targ = 'y';

if ($targ == 'n') {
    $target_window = 'target="_self"';
} elseif ($targ == 'y') {
    $target_window = 'target="_blank"';
} elseif ($targ == 'popup') {
    $target_window = 'target="popup" onClick="popupfeed(this.href);return false"';
} else {
    $target_window = 'target="' . $targ . "'";
}
```

```
// flag to show feed as full html output rather than JavaScript, used for alternative
// views for JavaScript-less users.
//   y = display html only for non js browsers
//   n = default (JavaScript view)
//   a = display javascript output but allow HTML
//   p = display text only items but convert linefeeds to BR tags

// default setting for no conversion of linebreaks
//if (!isset($html)) $html = 'n';
if (!isset($html)) $html = 'y';

$br = ' ';
if ($html == 'a') {
    $desc = 1;
} elseif ($html == 'p') {
    $br = '<br />';
}

// PARSE FEED and GENERATE OUTPUT ----- This is where it all happens!
// Fetch the data, thanks Magpie
$rss = @fetch_rss( $src );
```

```
// begin javascript output string for channel info
$str = "\t<div class=\"rss_box\">\n";

if (!$rss->channel['link']) {
    // error, nothing grabbed
    $str.= "\t<ul class=\"rss_item\">\n\t\t<li>This site could be down</li>\n\t\t<li>$src</li>\n\t\t<li>No data was found or no items available for this feed</li>\n"; // $vince$
}
else
{
    $str.= "\t<div class=\"rss_title\"><a href=\"\" . trim($rss->channel['link']) . \"\" \" . $target_window . ">\" . substr(strip_returns($rss->channel['title']),0,74) . "</a></div>\n"; // $vince$

    // begin item listing
    $str.= "\t<ul class=\"rss_item\">\n";

    // Walk the items and process each one
    $all_items = array_slice($rss->items, 0, $num);
```

```
foreach ( $all_items as $item ) {
    if ($item['link']) {
        // link url
        $my_url = $item['link'];
    }
    elseif ($item['guid']) {
        // feeds lacking item -> link
        $my_url = ($item['guid']);
    }
    if ($item['title']) {
        // format item title
        $my_title = strip_returns($item['title']);
        // create a title attribute. thanks Seb!
        $title_str = substr(strip_returns(strip_tags($item['summary'])), 0, 255) . '...!';
        $title_str = str_replace( "\"", "", $title_str); // $vince$
        // write the item
        $str .= "\t\t\t<li><a href=\"\" . trim($my_url) . \"\" title=\"\"$title_str\" \" .
            $target_window . ">\" . $my_title . "</a>\n";
        $str = str_replace( "&", "&";, $str); // $vince$
    }
}
}
```

feed2php.php

```
$str .= "\t\t</ul></div>\n";  
echo $str;  
?>
```

To Read

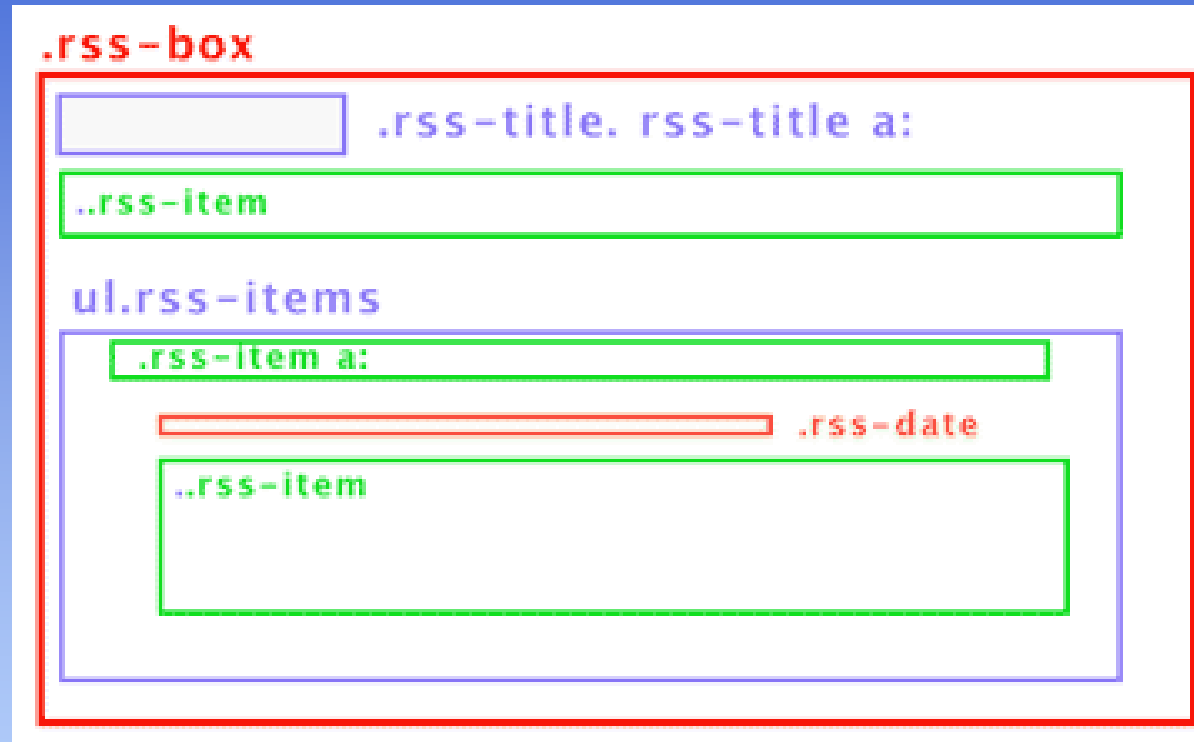
[/www/feed/](http://www/feed/)

[/www/feed/magpie/](http://www/feed/magpie/)

[/www/feed/magpie/extdocs](http://www/feed/magpie/extdocs)

[/www/feed/magpie/extlib/Snoopy.class.inc](http://www/feed/magpie/extlib/Snoopy.class.inc)

Default Style

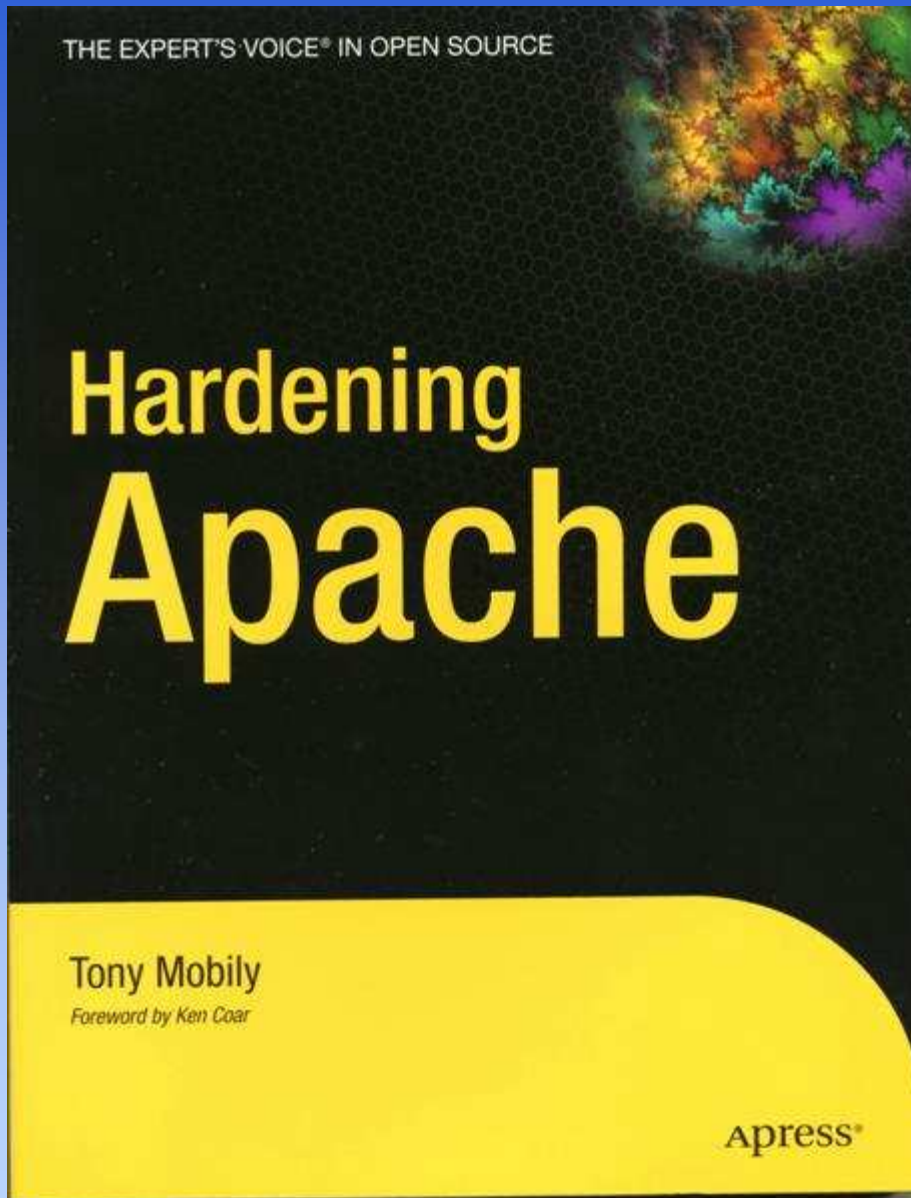


What I Write

```
<?php
    $src = 'http://www.securityfocus.com/rss/vulnerabilities.xml';
    include 'feed/feed2php.inc';
?>
```

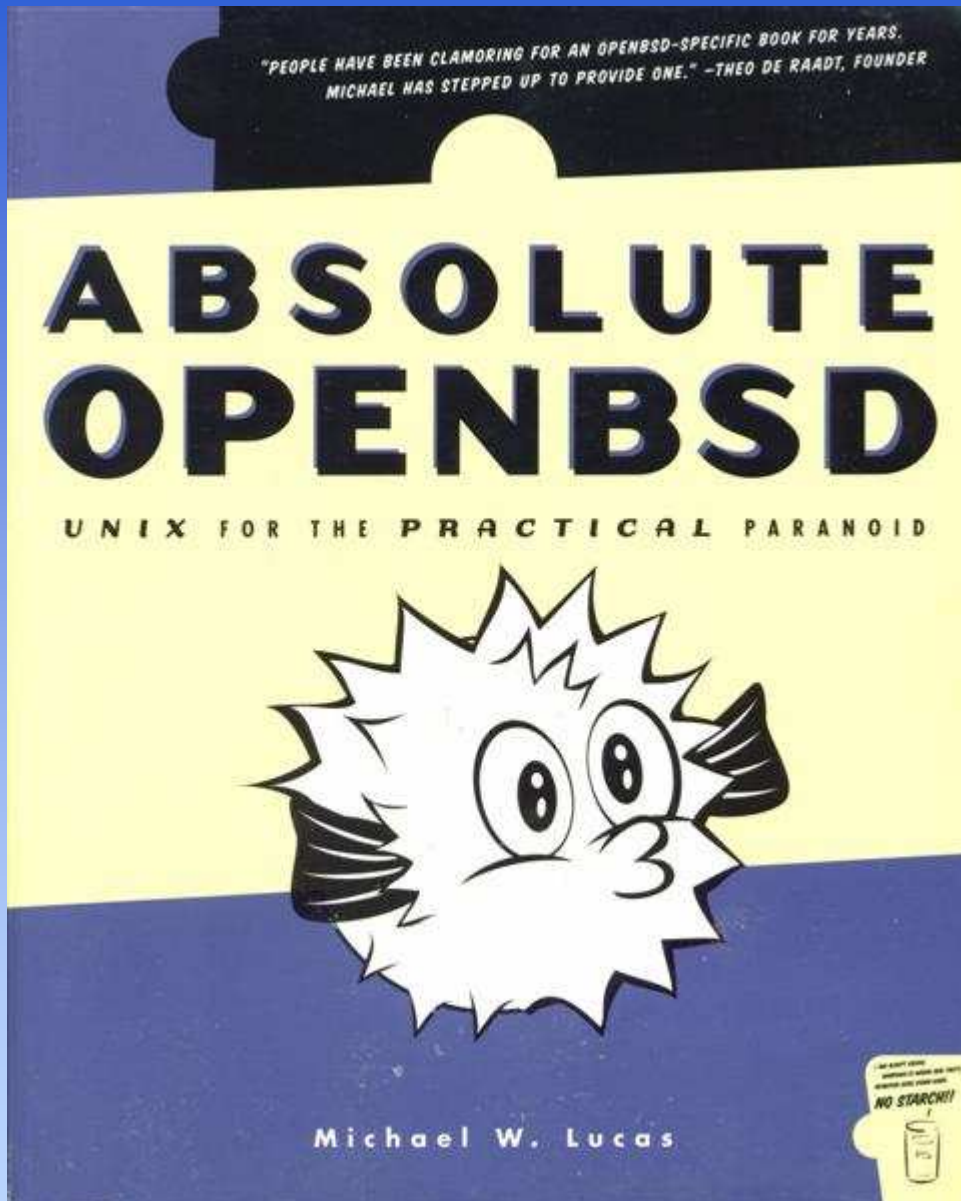
What I Get

```
<div class="rss_box">
    <div class="rss_title"><a href="http://www.securityfocus.com"
        target="_blank">SecurityFocus Vulnerabilities</a></div>
    <ul class="rss_item">
        <li><a href="http://www.securityfocus.com/bid/22757" title=" GnuPG Signed
Message Arbitrary Content Injection Weakness ..." target="_blank">Vuln: GnuPG Signed
Message Arbitrary Content Injection Weakness</a>
        <li><a href="http://www.securityfocus.com/bid/22933" title=" Xine DirectShow
Loader Remote Buffer Overflow Vulnerability ..." target="_blank">Vuln: Xine DirectShow
Loader Remote Buffer Overflow Vulnerability</a>
        <li><a href="http://www.securityfocus.com/bid/22534" title=" Wordpress Tem-
plates.PHP Cross-Site Scripting Vulnerability &gt;&gt; Advertisement
&lt;&lt;&lt; ALERT: How A Hacker Launches A Blind SQL Injection Attack Step-
by-Step! - White Paper Blind SQL Injection can deliver total control of your server to a
hacker..." target="_blank">Vuln: Wordpress Templates.PHP Cross-Site Scripting Vulner-
ability</a>
    </ul></div>
```



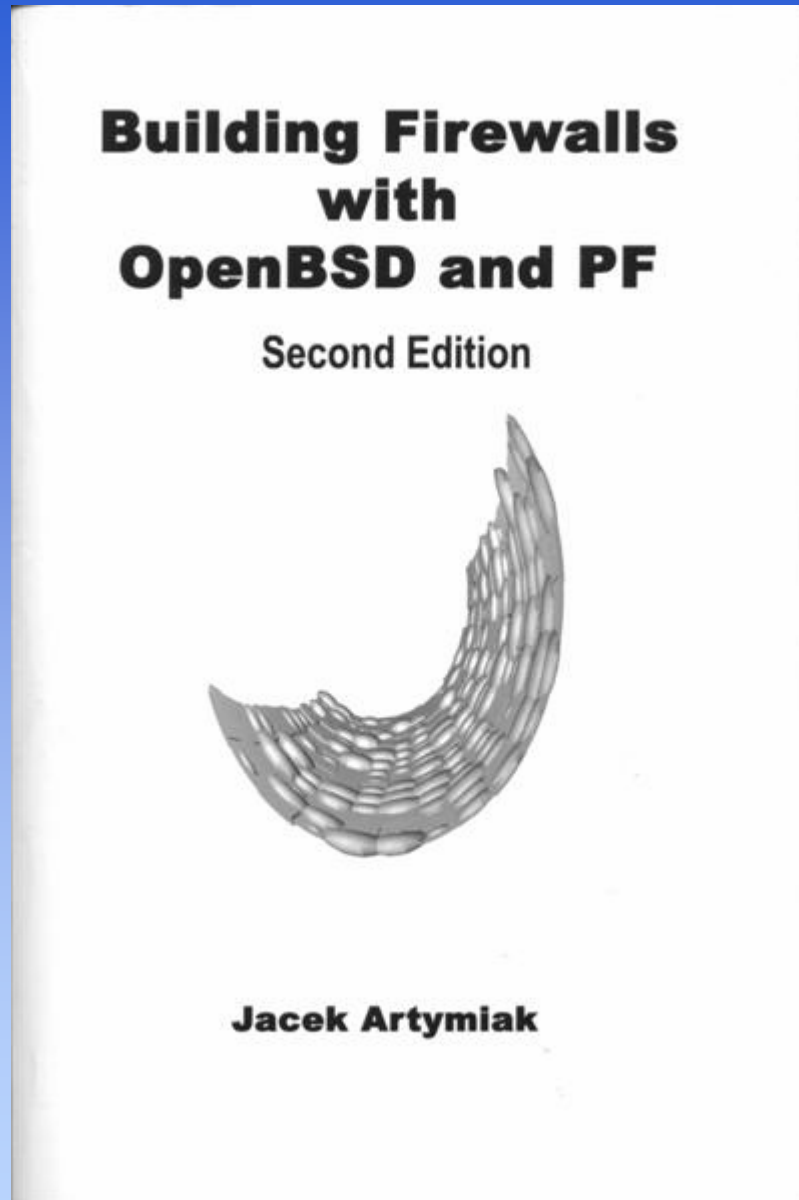
amazon.com

book	6.59
S. & H.	3.49
<hr/>	
Total	10.08



amazon.com

book	14.41
S. & H.	3.49
<hr/>	
Total	17.90



amazon.com

book	35.00
S. & H.	3.49
<hr/>	
Total	38.49