XML in the real world

XML in the *real* world
真实世界中的XML

☺ *Agent Zhang* ☺

(章亦春)

2006.10

♡ RSS is *cool*!

RSS 很酷哦！

RSS

➥ Really Simple Syndication

Tired of checking your favorite news and blogs sites everyday?

厌倦了每天去查看你最喜爱的那些新闻和博客网站了吧？

Let me tell you
how RSS can *save* you.

让我来告诉你
如何让 RSS 拯救你。

Open Google Reader:
the *first* thing that I do everyday.

打开 Google Reader：
我每天做的第一件事情。


Google Reader (33) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/reader/view/
Go
Google Reader LABS
agentzh@gmail.com | Settings | My Account | Help | Sign Out
Home
All items (33)
Starred items
Shared items
Add subscription
Browse »
All subscriptions: [only list updated]
journal (6)
Allison's Journal (1)
chromatic's Journal
Pugs (5)
沉思之骨
Cherry的共享空间
fglock's Journal (8)
miyagawa's Journal (5)
Perl.com
perl.perl6.compiler G... (14)
一个人生活
与咩咩分享生活点滴
Manage subscriptions »
Home
Recently starred - FEAR-less Site Scraping Perl.com Many web programmers talk
A look at what's new
Allison's Journal (1)
My other journal This is Allison's other journal. Not a replacement, just an addition. Oh yeah, I've got quite a bit of Parrot stuff to catch ...
See more from journal (6) »
miyagawa's Journal (5)
Screen scraping w/ XPath and CSS Selector in Action
Here is more example code that demonstrates how to use XPath and CSS Selector to do screen scraping without using nasty ...
HTML::Selector::XPath Per discussions in CSS Selector in Perl, I made a quick perl module HTML::Selector::XPath, which is available at ...
Anyone visiting Tokyo in October? We're planning our 7th technical meeting of Shibuya.pm in the middle of October. If anyone reading this has a plan to visit ...
Pugs (5)
The Benefits of Smart Linking Today's learning
Done
Open Notebook

Let's read chromatic's *latest* journals
in Google Reader...

让我们在 Google Reader 中阅读 chromatic
最近的日记.....


Googl...
Goog...
learn...
learn...
Journ...
use P...
It's ...
Journ...
2006.10
Google Reader LABS
agentzh@gmail.com | Settings | My Account | Help | Sign Out
Home
All items (28)
Starred items
Shared items
Add subscription
Browse »
All subscriptions: [only list updated]
journal (6)
Allison's Journal (1)
chromatic's Journal
Pugs (5)
沉思之骨
Cherry的共享空间 (3)
fglock's Journal
miyagawa's Journal (5)
Perl.com
perl.perl6.compiler G... (14
一个人生活
Manage subscriptions »
chromatic's Journal
Feed settings...
Expanded view
List view
All items [show only new] - Mark all as read - Refresh
It's Full of Sharps!
Oct 4, 2006 (18 hours ago)
by chromatic
Ahh, so radical simplicity is why dynamic languages on static VMs are important. Some days, I even miss Tcl/Tk.
Add star
Share
Email
Mark as read
Edit tags: journal
C is not Perl
Oct 3, 2006 (2 days ago)
by chromatic
Stevan asked me to write some XS code for Class::MOP, and it sounded like more fun than painting my deck or writing the non-fiction book, so I dodged segfaults for a couple of hours this afternoon.
The Perl XS API has its own kind of consistency. It's never really stuck in my brain, but it's just familiar enough that I sort of know what I want.
Previous item
Next item
more than 20 items
Done
Open Notebook

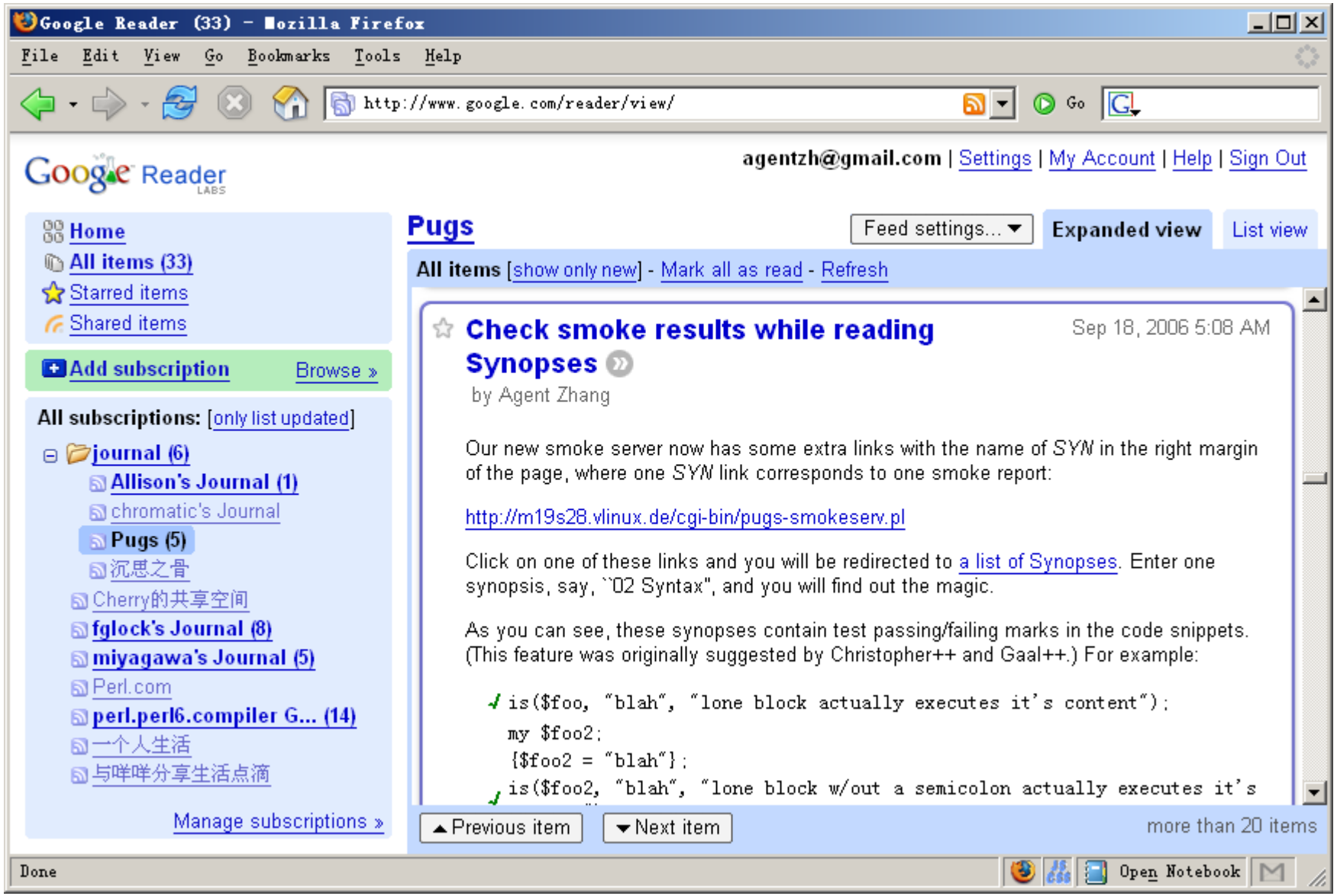
Let's take a look at the *original* journal item on the use.perl.org site...

让我们来看看
use.perl.org 站点上的原始日记.....


Journal of chromatic (983) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://use.perl.org/~chromatic/journal/31190?from=rss
Go
Search use Perl
Go
use Perl;
All the Perl that's Practical to Extract and Report
FAQ | About | Submit Story | Older Stories | Topics | Polls | Journals | Authors | New Modules | Preferences | Logout
Logout Preferences Password ~agent (5836) ~chromatic (983)
Journal of chromatic (983)
Info | Relation | Journal | Friends | Fans | Foes | Freaks | Tags | Bookmarks
agent's Journal | Write in Journal | Delete/Edit Entries | Top 10 | Edit Preferences | Friend's Journals
Monday October 02, 2006
03:07 PM C is not Perl [#31190]
Stevan asked me to write some XS code for Class::MOP, and it sounded like more fun than painting my deck or writing the non-fiction book, so I dodged segfaults for a couple of hours this afternoon.
The Perl XS API has its own kind of consistency. It's never really stuck in my brain, but it's just familiar enough that I sort of know what I want.
chromatic (983)
chromatic
(email not shown publicly)
http://wgz.org/chromatic/
Done
Open Notebook

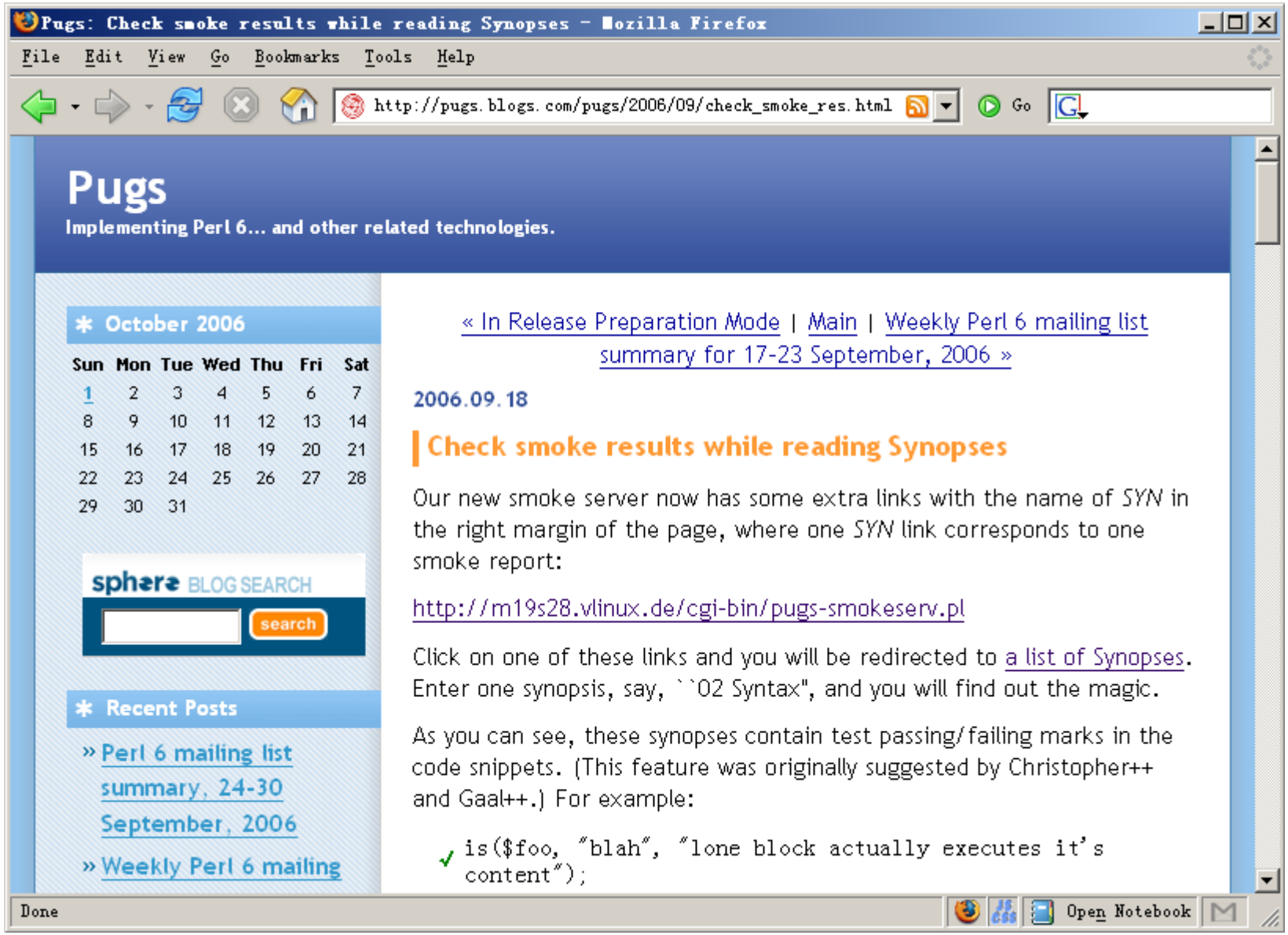
Now let's turn to the *latest* Pugs blog posts in Google Reader...

现在让我们在 Google Reader 中转向新的 Pugs 博客文章.....


Google Reader (33) - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/reader/view/
Go
Google Reader LABS
agentzh@gmail.com | Settings | My Account | Help | Sign Out
Home
All items (33)
Starred items
Shared items
Add subscription
Browse »
All subscriptions: [only list updated]
journal (6)
Allison's Journal (1)
chromatic's Journal
Pugs (5)
沉思之骨
Cherry的共享空间
fglock's Journal (8)
miyagawa's Journal (5)
Perl.com
perl.perl6.compiler G... (14)
一个人生活
与咩咩分享生活点滴
Manage subscriptions »
Pugs
Feed settings...
Expanded view
List view
All items [show only new] - Mark all as read - Refresh
Check smoke results while reading Synopses
Sep 18, 2006 5:08 AM
by Agent Zhang
Our new smoke server now has some extra links with the name of SYN in the right margin of the page, where one SYN link corresponds to one smoke report:
http://m19s28.vlinux.de/cgi-bin/pugs-smokeserv.pl
Click on one of these links and you will be redirected to a list of Synopses. Enter one synopsis, say, ``02 Syntax", and you will find out the magic.
As you can see, these synopses contain test passing/failing marks in the code snippets. (This feature was originally suggested by Christopher++ and Gaal++.) For example:
is($foo, "blah", "lone block actually executes it's content");
my $foo2;
{$foo2 = "blah"};
is($foo2, "blah", "lone block w/out a semicolon actually executes it's
Previous item
Next item
more than 20 items
Done
Open Notebook

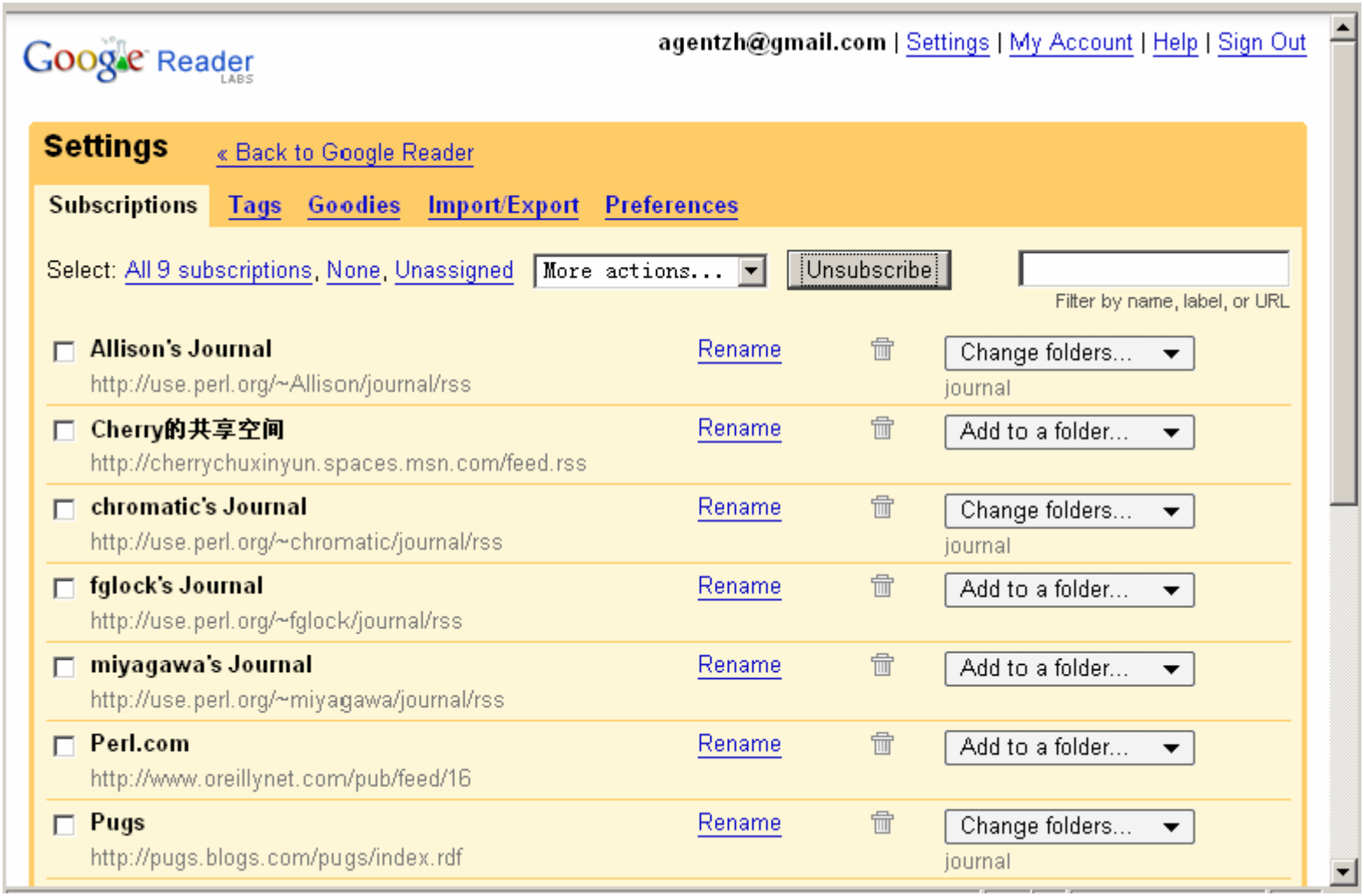
The *original* post I published onto the pugs.blogs.com site...

我最初发布在
use.perl.org 站点上的帖子.....


Pugs: Check smoke results while reading Synopses - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://pugs.blogs.com/pugs/2006/09/check_smoke_res.html
Go
Pugs
Implementing Perl 6... and other related technologies.
October 2006
Sun Mon Tue Wed Thu Fri Sat
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31
sphere BLOG SEARCH
search
Recent Posts
» Perl 6 mailing list summary, 24-30 September, 2006
» Weekly Perl 6 mailing
« In Release Preparation Mode | Main | Weekly Perl 6 mailing list summary for 17-23 September, 2006 »
2006.09.18
Check smoke results while reading Synopses
Our new smoke server now has some extra links with the name of SYN in the right margin of the page, where one SYN link corresponds to one smoke report:
http://m19s28.vlinux.de/cgi-bin/pugs-smokeserv.pl
Click on one of these links and you will be redirected to a list of Synopses. Enter one synopsis, say, ``02 Syntax", and you will find out the magic.
As you can see, these synopses contain test passing/failing marks in the code snippets. (This feature was originally suggested by Christopher++ and Gaal++.) For example:
✓ is($foo, "blah", "lone block actually executes it's content");
Done
Open Notebook

The XML magic *behind* the curtain...

幕后的 XML 魔法.....


Google Reader LABS
agentzh@gmail.com | Settings | My Account | Help | Sign Out
Settings
« Back to Google Reader
Subscriptions
Tags
Goodies
Import/Export
Preferences
Select: All 9 subscriptions, None, Unassigned
More actions...
Unsubscribe
Filter by name, label, or URL
Allison's Journal
http://use.perl.org/~Allison/journal/rss
Rename
Change folders...
journal
Cherry的共享空间
http://cherrychuxinyun.spaces.msn.com/feed.rss
Rename
Add to a folder...
chromatic's Journal
http://use.perl.org/~chromatic/journal/rss
Rename
Change folders...
journal
fglock's Journal
http://use.perl.org/~fglock/journal/rss
Rename
Add to a folder...
miyagawa's Journal
http://use.perl.org/~miyagawa/journal/rss
Rename
Add to a folder...
Perl.com
http://www.oreillynet.com/pub/feed/16
Rename
Add to a folder...
Pugs
http://pugs.blogs.com/pugs/index.rdf
Rename
Change folders...
journal

RSS feed for *chromatic*'s journals...

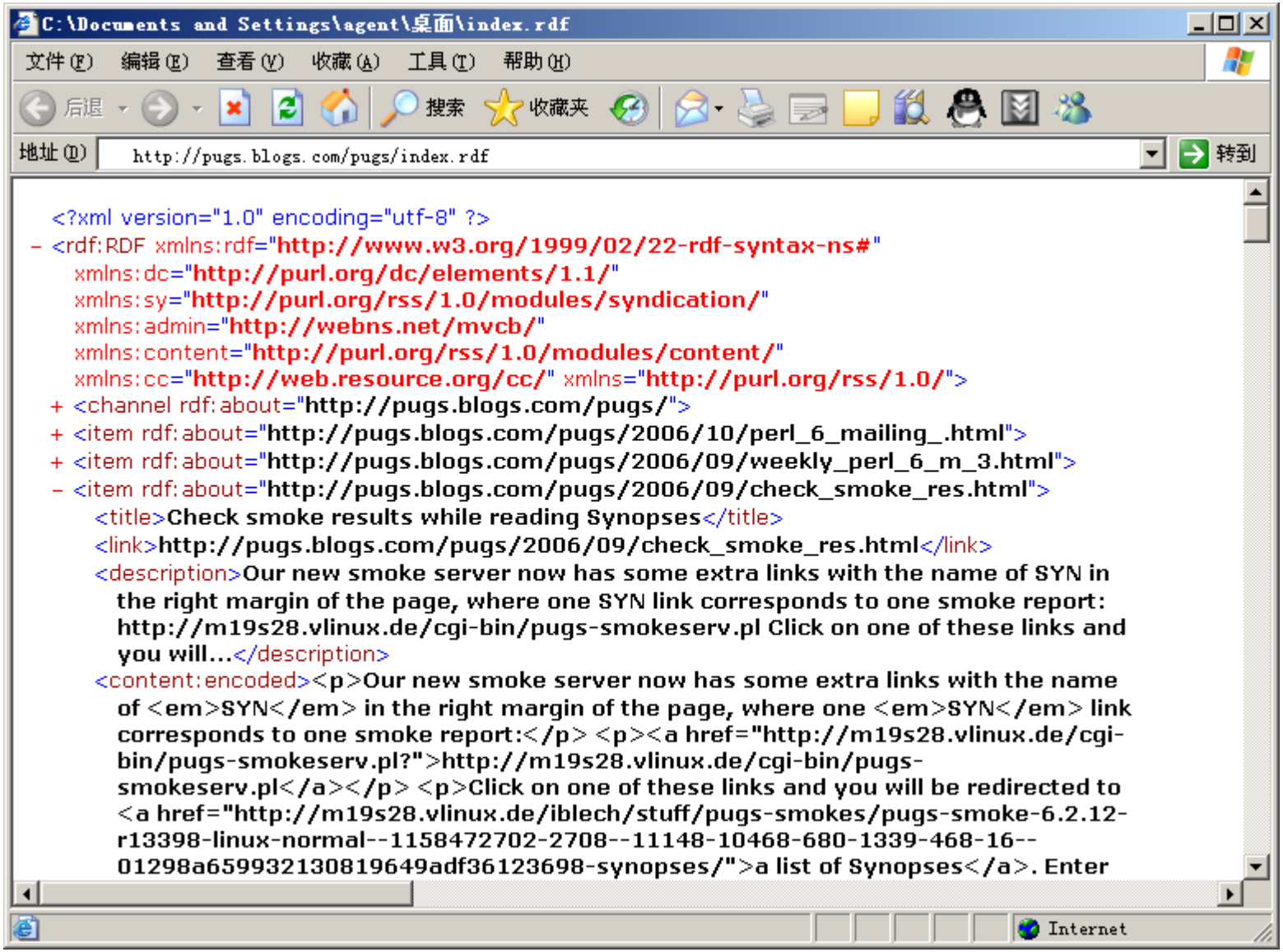
chromatic 的日记的
RSS 反馈.....

地址(D) http://use.perl.org/~chromatic/journal/rss 转到

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns="http://purl.org/rss/1.0/" xmlns:slash="http://purl.org/rss/1.0/modules/slash/"
    xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:syn="http://purl.org/rss/1.0/modules/syndication/"
    xmlns:admin="http://webns.net/mvcb/">
  + <channel rdf:about="http://use.perl.org/~chromatic/journal/">
  + <image rdf:about="http://use.perl.org/images/topics/useperl.gif">
  - <item rdf:about="http://use.perl.org/~chromatic/journal/31213?from=rss">
      <title>It's Full of Sharps!</title>
      <link>http://use.perl.org/~chromatic/journal/31213?from=rss</link>
      <description><p>Ahh, so <a
        href="http://www.learningpython.com/2006/10/02/ironpython-hello-world-
        tutorial/">radical simplicity is why dynamic languages on static VMs are
        important</a>. Some days, I even miss Tcl/Tk.</p></description>
      <dc:creator>chromatic</dc:creator>
      <dc:date>2006-10-04T07:45:54+00:00</dc:date>
      <dc:subject>journal</dc:subject>
    </item>
  - <item rdf:about="http://use.perl.org/~chromatic/journal/31190?from=rss">
      <title>C is not Perl</title>
      <link>http://use.perl.org/~chromatic/journal/31190?from=rss</link>
      <description><p> <a href="http://use.perl.org/~stevan">Stevan</a> asked me to
        write some XS code for <a href="http://search.cpan.org/perldoc?
        Class::MOP">Class::MOP</a>, and it sounded like more fun than painting my deck
        or writing the non-fiction book, so I dodged segfaults for a couple of hours this
        afternoon.</p> <p>The Perl XS API has its own kind of consistency. It's never
```

RSS feed for
our *Pugs* blog site...

Pugs 博客站点的
RSS 反馈.....


C:\Documents and Settings\agent\桌面\index.rdf
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://pugs.blogs.com/pugs/index.rdf
转到
<?xml version="1.0" encoding="utf-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:sy="http://purl.org/rss/1.0/modules/syndication/"
xmlns:admin="http://webns.net/mvcb/"
xmlns:content="http://purl.org/rss/1.0/modules/content/"
xmlns:cc="http://web.resource.org/cc/" xmlns="http://purl.org/rss/1.0/">
+ <channel rdf:about="http://pugs.blogs.com/pugs/">
+ <item rdf:about="http://pugs.blogs.com/pugs/2006/10/perl_6_mailing_.html">
+ <item rdf:about="http://pugs.blogs.com/pugs/2006/09/weekly_perl_6_m_3.html">
- <item rdf:about="http://pugs.blogs.com/pugs/2006/09/check_smoke_res.html">
<title>Check smoke results while reading Synopses</title>
<link>http://pugs.blogs.com/pugs/2006/09/check_smoke_res.html</link>
<description>Our new smoke server now has some extra links with the name of SYN in the right margin of the page, where one SYN link corresponds to one smoke report: http://m19s28.vlinux.de/cgi-bin/pugs-smokeserv.pl Click on one of these links and you will...</description>
<content:encoded> Our new smoke server now has some extra links with the name of SYN in the right margin of the page, where one SYN link corresponds to one smoke report: http://m19s28.vlinux.de/cgi-bin/pugs-smokeserv.pl Click on one of these links and you will be redirected to a list of Synopses. Enter
Internet

♡ AJAX, our *good* friends!

AJAX，我们的好朋友！

AJAX

➥ Asynchronous JAvaScript and XML

☺ Let's open *Cherry*'s Qzone blogs...

让我们打开 Cherry 的 Qzone 博客.....


瓶子里的雏菊~ [11854905.qzone.qq.com] - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://u13.qzone.qq.com/cgi-bin/cgi_client_entry.cgi?uin=11854905
转到
瓶子里的雏菊~ http://11854905.qzone.qq.com
装扮我的空间
添加关注
主 页
日 志
音乐盒
留言板
相 册
迷你屋
个人档
互动
音乐收藏 MORE
1 许愿池的一
2 好人有好一SHE
3 马德里不一
4 love is一
5 because一
6 be care一
7 恋爱达人一
互 动 MORE
最近访客
好友收藏
solon
邀请开通
章亦春
邀请开通
网络日志 MORE
[个人日记] 我不要再叫猪猪了~!~~ 评论(6)
[个人日记] 好恶心~ 评论(2)
[个人日记] 心情好的时候多写点~~~~ 评论(1)
[个人日记] 开学的第二个周末~ 评论(2)
[个人日记] 郁闷 评论(2)
[个人日记] 堕落啊~~~~ 评论(6)
[个人日记] 开学了哦~ 评论(6)
[个人日记] 很奇怪~ 评论(4)
留言板 MORE
低调的走了过...
我也来踩哈子...
Internet

☺ *Click* one of the articles and enter it...

点击其中的一篇文章进入……

瓶子里的雏菊~ [11854905.qzone.qq.com] - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏夹

地址(D) http://u13.qzone.qq.com/cgi-bin/cgi_client_entry.cgi?uin=11854905 转到

瓶子里的雏菊~ http://11854905.qzone.qq.com 装扮我的空间 添加关注

主 页
日 志
音乐盒
留言板
相 册
迷你屋
个人档
互动

[个人日记]

转载到我空间 复制本文网址 隐藏签名档

我不要再叫猪猪了~!~~

发表时间：2006年9月22日 12时58分 评论(6)

突然发现，越来有这么多人都曾经或者现在或者将来有可能被称呼为"猪猪"!!!月饼的空间已经用猪猪和猪头做标志了,Sunday也被叫过猪???看来,这个跟了我N年的外号现在已经泛滥的不行了，比传染病还要流传的快，大街上喊一声猪猪,最起码有N+1个回头的，我不要这个外号了，我要改外号!!!!!以后不要再喊我猪猪了~~~~郁闷了~~~~~估计我老是减肥减不下来就是因为大家老喊我猪猪,就喊肥了!!!!!

以后谁喊我和谁急!~~~~~~

都说大笑容易长皱纹，可我就是喜欢＂哈哈哈＂~~~~

完毕 Internet

What happened *behind the curtain* when we're performing these actions?

在我们执行这些动作的时候，
幕后都发生了哪些事情？

Here is the underlying HTTP traffic between the *Qzone site* and my *IE browser* recorded by HTTP::Proxy...

这里有 HTTP::Proxy 模块记录下的 Qzone 站点与我的 IE 浏览器之间的底层 HTTP 通信.....

```
[16:04:56]  GET http://u13.qzone.qq.com/cgi-bin/cgi_client_entry.cgi?uin=11854905
[16:05:40]  GET http://u13.qzone.qq.com/proxy.html
...
[16:09:37]  GET  http://b1.qzone.qq.com/cgi-bin/blog/blog_signature.cgi?uin=11854905
[16:10:00]  GET  http://b1.qzone.qq.com/cgi-bin/blog/blog_get_category.cgi?
                 uin=11854905
[16:10:00]  GET http://imgcache.qq.com/qzone/proxy.vbs
[16:10:02]  GET  http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?
                 uin=11854905&blogid=39&flag=0
[16:10:04]  GET  http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?
                 uin=11854905&blogid=39&archive=-2
...
```

Most of the HTTP requests were initiated by the JavaScript code running in your *web browser*.

这些 HTTP 请求中的大部分是由运行在你的网络浏览器中的 JavaScript 代码发起的。

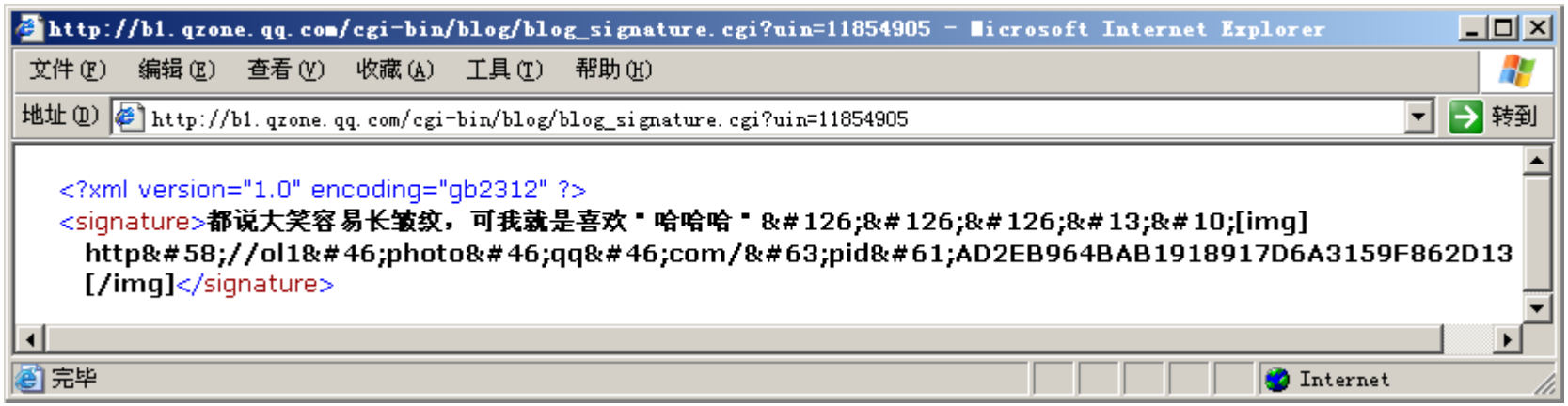
Let's *check* some of
the HTTP requests by hand...

让我们来手工查看一下
其中的几个 HTTP 请求.....

XML data for Cherry's *signature*

Cherry 的个性签名所对应的 XML 数据

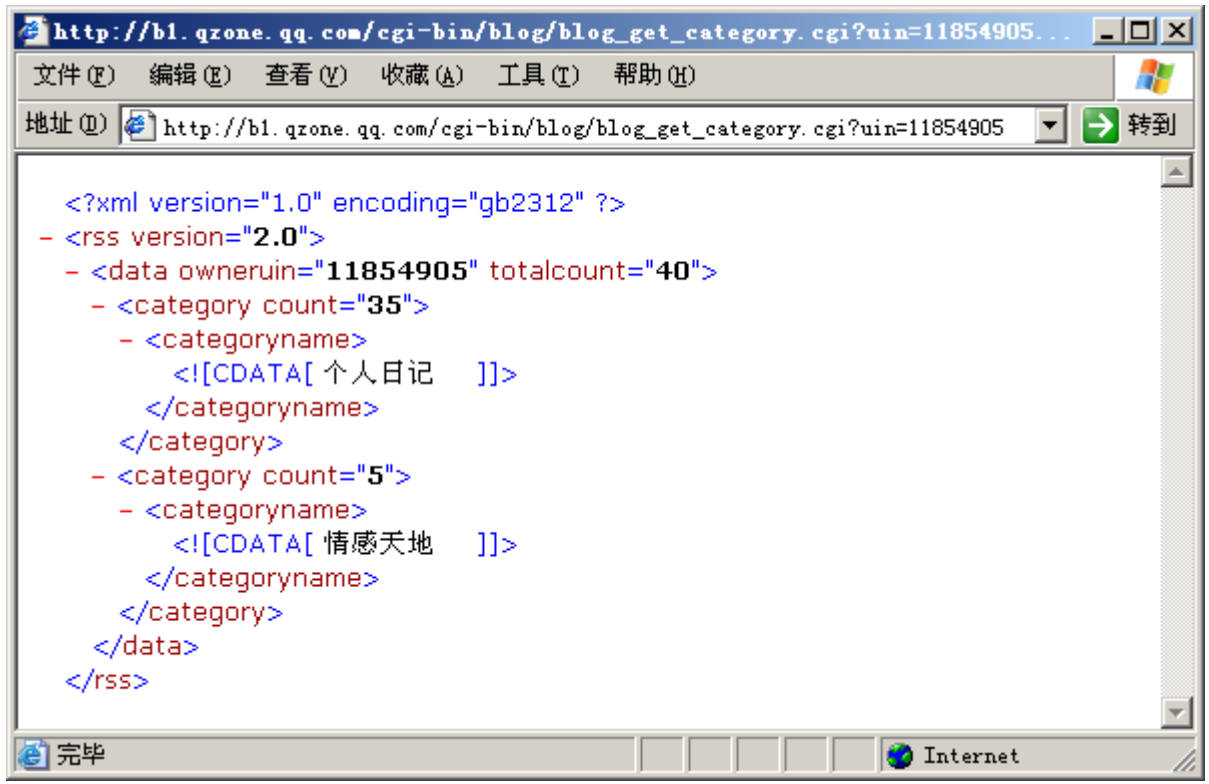
➥ http://b1.qzone.qq.com/cgi-bin/blog/blog_signature.cgi?uin=11854905


http://b1.qzone.qq.com/cgi-bin/blog/blog_signature.cgi?uin=11854905 - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://b1.qzone.qq.com/cgi-bin/blog/blog_signature.cgi?uin=11854905
转到
<?xml version="1.0" encoding="gb2312" ?>
<signature>都说大笑容易长皱纹，可我就是喜欢＂哈哈哈＂~~~
[img]
http://ol1.photo.qq.com/?pid=AD2EB964BAB1918917D6A3159F862D13
[/img]</signature>
完毕
Internet

XML data for Cherry's article *category* list

Cherry 的文章类别列表所对应的 XML 数据

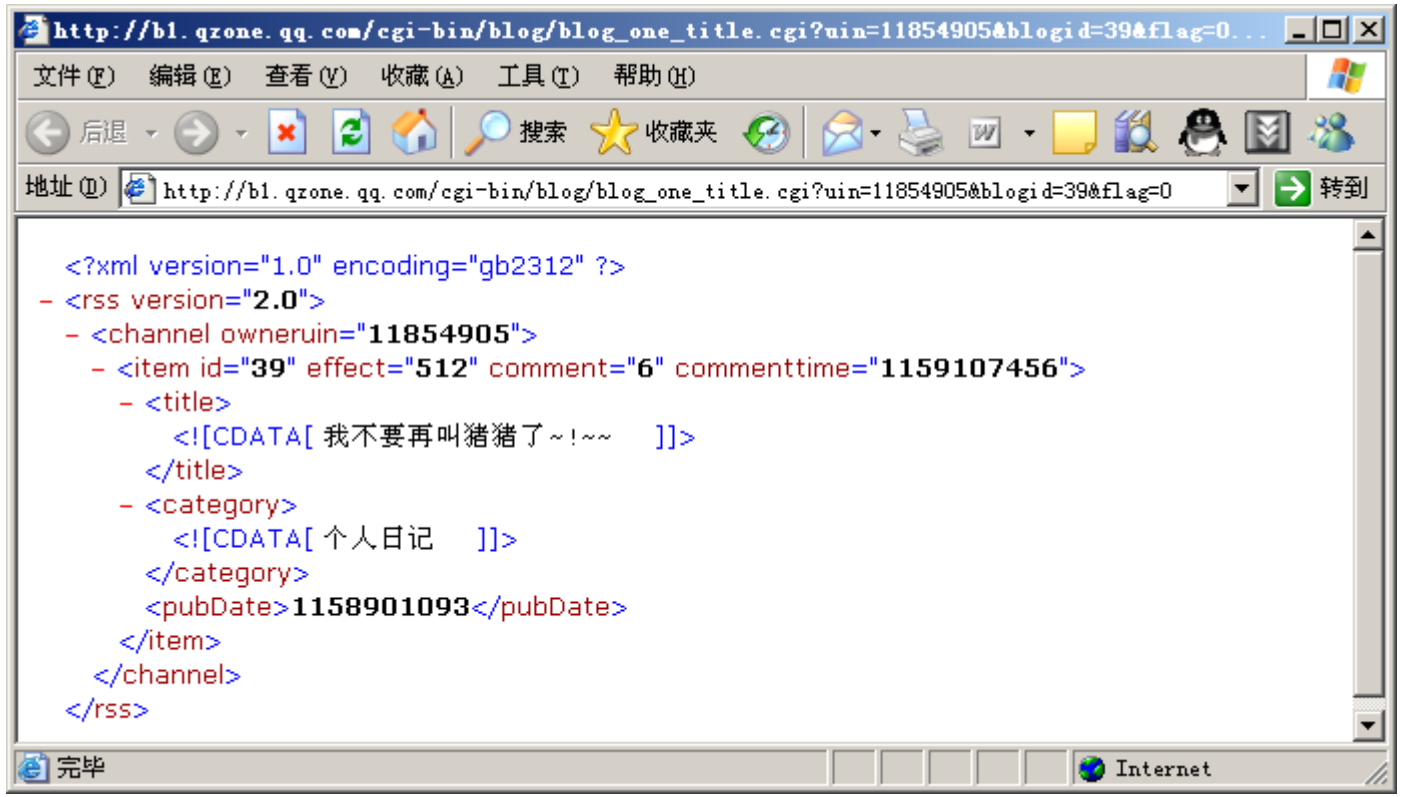
➥ http://b1.qzone.qq.com/cgi-bin/blog/blog_get_category.cgi?uin=11854905


http://b1.qzone.qq.com/cgi-bin/blog/blog_get_category.cgi?uin=11854905...
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://b1.qzone.qq.com/cgi-bin/blog/blog_get_category.cgi?uin=11854905
转到
<?xml version="1.0" encoding="gb2312" ?>
<rss version="2.0">
<data owneruin="11854905" totalcount="40">
<category count="35">
<categoryname>
<![CDATA[个人日记]]>
</categoryname>
</category>
<category count="5">
<categoryname>
<![CDATA[情感天地]]>
</categoryname>
</category>
</data>
</rss>
完毕
Internet

XML data for the *title* of
Cherry's 40th post (with ID 39)

Cherry 的第 40 篇帖子(标识为 39) 的标题
所对应的 XML 数据

➥ http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905&blogid=39&flag=0


http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905&blogid=39&flag=0...
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905&blogid=39&flag=0
转到
<?xml version="1.0" encoding="gb2312" ?>
<rss version="2.0">
<channel owneruin="11854905">
<item id="39" effect="512" comment="6" commenttime="1159107456">
<title>
<![CDATA[我不要再叫猪猪了~!~~]]>
</title>
<category>
<![CDATA[个人日记]]>
</category>
<pubDate>1158901093</pubDate>
</item>
</channel>
</rss>
完毕
Internet

XML data for the *body and comments* of
Cherry's 40th post (with ID 39)

Cherry 的第 40 篇帖子(标识为 39) 的正文及评论
所对应的 XML 数据

➥ http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905&blogid=39&archive=-2


http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905&blogid=39&property=D...
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
地址(D) http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905&blogid=39&archive=-2
转到
<?xml version="1.0" encoding="gb2312" ?>
- <rss version="2.0">
- <channel archive="-2" uin="11854905" id="39">
- <blog orguin="11854905" voteid="-1">
- <blogcontent>
<![CDATA[突然发现，越来有这么多人都曾经或者现在或者将来有可能
被称呼为"猪猪"!!!月饼的空间已经用猪猪和猪头做标志了，Sunday也被
</blogcontent>
+ <blogautograph>
</blog>
- <item archive="0">
- <comment commentid="0" commenteffect="0" archive="0" commentuin="3502284">
- <commentnick>
<![CDATA[水的伤痕]]>
</commentnick>
- <commentcontent>
<![CDATA[猪头=。=

这个不错……
 谢谢我吧 给你换了个名字]]>
</commentcontent>
<pubDate>2006-09-23 11:11:22</pubDate>
- <commentautograph>
<![CDATA[彪汗的人生不需要任何解释!
 ]]>
</commentautograph>
</comment>
Internet

Our web browser renders these XML data files using HTML templates sent by the Qzone server, and generates the final HTML source.

我们的网络浏览器根据 Qzone 服务器
传过来的 HTML 模板对这些 XML 数据进行渲染,
生成最终的 HTML 源码。

XML data + HTML templates = final HTML source

XML 数据 + HTML 模板 = 最终的 HTML 源码

The *whole* process happens
in our web browser.

整个过程都发生在我们的浏览器内部。

But *where* are
the HTML templates?

但是 HTML 模板究竟在哪里呢？

Let's check the *raw* HTML source sent from the Qzone server

让我们来看看 Qzone 服务器传过来的原始 HTML 源码


瓶子里的雏菊~ [11854905.qzone.qq.com] - Microsoft Internet Explorer
文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)
后退 搜索 收藏夹
地址(D) http://u13.qzone.qq.com/cgi-bin/cgi_client_entry.cgi?uin=11854905
瓶子里的雏菊~ http://11854905.qzone.qq.com 装扮我的空间
主 页
日 志
音乐盒
留言板
相 册
迷你屋
个人档
互动
个人资料 MORE
形象 | 档案 | 统计
相册
互 动 MORE
最近访客 好友收藏
网络日志
[个人日记] 我不要再叫猪猪了~!~~
后退(B)
前进(O)
背景另存为(S)...
设置为背景(G)
复制背景(C)
设为桌面项(D)...
全选(A)
粘贴(P)
创建快捷方式(T)
添加到收藏夹(F)...
查看源文件(V)
编码(E)
打印(T)
刷新(R)
使用网际快车下载全部链接
导出到 Microsoft Office Excel(X)
添加到QQ自定义面板
转换为 Adobe PDF
主页 我的空间
显示该页的 HTML 源文件。

```
...
<!-- 日志//-->
<div id="tpl_blog_b" class="mode_table" style="display:none">...
 <table cellSpacing="0" cellpadding="0" width="100%" class= ...
    [%repeat_0 match="/rss/channel/item" repeat_num="10"%]
    <tr><td class="index_blog_btd">
    [<a href="#"  onclick="openCategory('[%=@type%]');return false"
    ...title="点击进入分类">[%=@category%]</a>]
    <a href="#" title="[%=@title%] -- 发表于 [%=@pubTimeString%]"
    onClick="openBlog('[%=@archive%]','[%=@id%]');return false">
    ... </a></td>
    <td class="info">评论(<span class="hit">[%=@comment%]</span>)</td>
    [%_repeat_0%]
 </table>
</div>
...
```

You see,
it's a *client*-side HTML template!

你看，
这是一个客户端的 HTML 模板！

It's the JavaScrip code that *grabs* the XML data from the web and fills it into the HTML templates automatically, resulting in the final appearance we see in the browser.

所以是 JavaScript 代码自动从网上获取 XML 数据并将之填入到 HTML 模板中，最终得到我们在浏览器中看到的效果。

Then why can't we do XML data grabbing *ourselves*?

那么为什么我们就不可以自己去攫取 XML 数据呢？

For example, we can obtain the data
for all of Cherry's articles
by simply changing the URL!

比如，我们可以通过简单地修改网址
得到的有文章的数据！

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=38 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=38 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=37 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=38 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=37 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=36 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=38 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=37 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=36 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=35 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=39 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=38 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=37 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=36 &flag=0
http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=35 &flag=0

...

http://b1.qzone.qq.com/cgi-bin/blog/blog_one_title.cgi?uin=11854905& blogid=0 &flag=0

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=38 &archive=-2

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=38 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=37 &archive=-2

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=38 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=37 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=36 &archive=-2

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=38 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=37 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=36 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=35 &archive=-2

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=39 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=38 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=37 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=36 &archive=-2
http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=35 &archive=-2

...

http://b1.qzone.qq.com/cgi-bin/blog/blog_commentlist.cgi?uin=11854905& blogid=0 &archive=-2

It means that...

now we can *directly* access Qzone's database,

completely *bypassing* its cumbersome HTML interface!

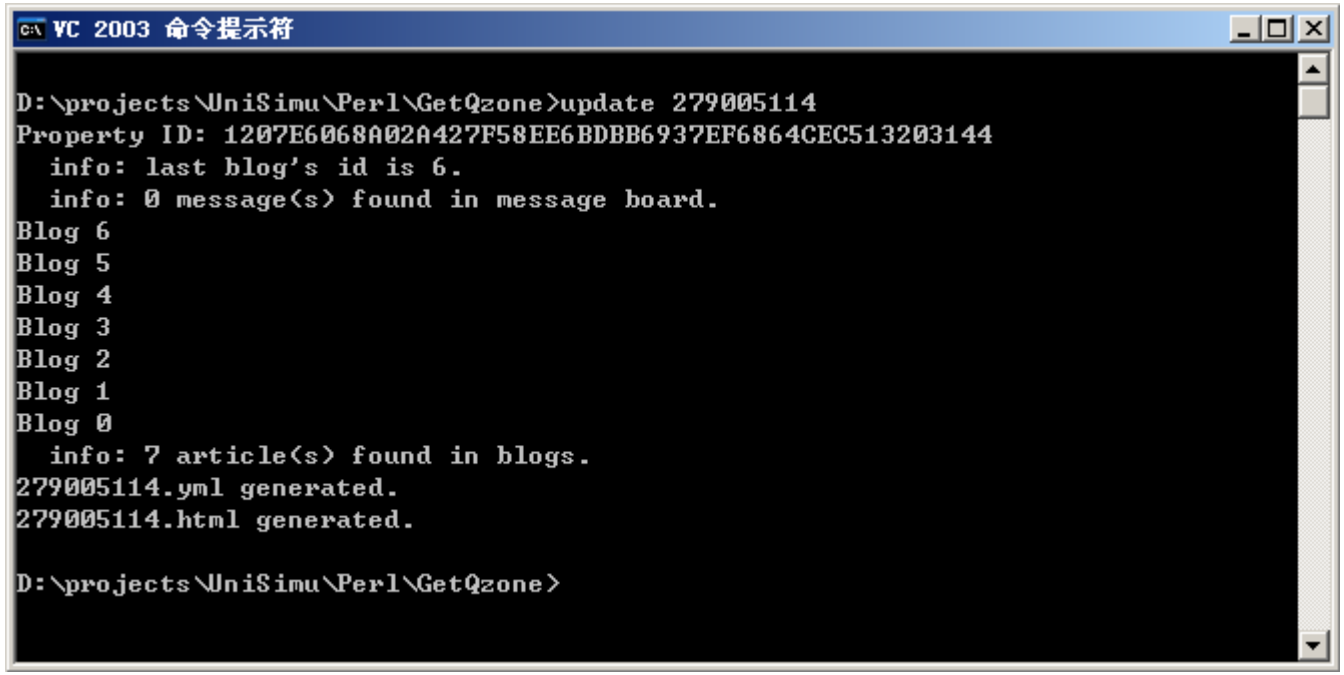
这意味着.....

现在我们可以直接访问 Qzone 的数据库

完全绕过它那笨重的 HTML 界面！

Let's write a tiny Perl *script* to do *all* these tricks for us!

让我们来编写一个小小的 Perl 脚本来为我们实现所有这些把戏。


VC 2003 命令提示符
D:\projects\UniSimu\Perl\GetQzone>update 279005114
Property ID: 1207E6068A02A427F58EE6BDBB6937EF6864CEC513203144
  info: last blog's id is 6.
  info: 0 message(s) found in message board.
Blog 6
Blog 5
Blog 4
Blog 3
Blog 2
Blog 1
Blog 0
  info: 7 article(s) found in blogs.
279005114.yml generated.
279005114.html generated.
D:\projects\UniSimu\Perl\GetQzone>

One sample output of the program
for *Cherry*'s Blogs

➥ http://perlcabal.org/agent/cherry.html

该程序针对 Cherry 的博客的一次典型输出


褚鑫昀 - QZone 空间文章汇总 - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://perlcabal.org/agent/cherry.html
Go
褚鑫昀 - QZone 空间文章汇总
文档生成日期
个人信息
留言板
所有文章
个人日记
我不要再叫猪猪了~!~~
好恶心~
心情好的时候多写点~~~~
开学的第二个周末~
郁闷
堕落啊~~~~
开学了哦~
很奇怪~
随想~
我又回来了~
我会做土豆泥了~~~~
小怀念下~~~
终于可以写点东西了~
我~~~
很开心哦~~~~
刮痧.....
被一个小姐姐戳了下~~~~
我喜欢这样的感觉~~~~
莫名其妙 ~~~~
Done
Open Notebook


褚鑫昀 - QZone 空间文章汇总 - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://perlcabal.org/agent/cherry.html
Go
所有文章
个人日记
我不要再叫猪猪了~!~~
Fri Sep 22 12:58:13 2006
突然发现，越来有这么多人都曾经或者现在或者将来有可能被称呼为"猪猪"!!!月饼的空间已经用猪猪和猪头做标志了，Sunday也被叫过猪???看来,这个跟了我N年的外号现在已经泛滥的不行了，比传染病还要流传的快，大街上喊一声猪猪,最起码有N+1个回头的，我不要这个外号了，我要改外号!!!!!以后不要再喊我猪猪了~~~~郁闷了~~~~~估计我老是减肥减不下来就是因为大家老喊我猪猪,就喊肥了!!!!!
以后谁喊我和谁急!~~~~~~
Comments
水的伤痕 (3502284) - 2006-09-23 11:11:22
猪头=。=
这个不错..... 谢谢我吧 给你换了个名字
自做多情 (167860281) - 2006-09-23 20:29:42
Done
Open Notebook

This program uses LWP::UserAgent to *get* the XML data directly from the web and uses XML::Simple to *parse* it.

该程序使用 LWP::UserAgent 模块直接从网上获取 XML 数据，并利用 XML::Simple 解析之。

No need for Audrey's Template::Extract
to *extract* data from the HTML source.
That is the *power* of AJAX and XML!

不再需要唐凤的 Template::Extract 模块来
从 HTML 源码中提取数据。
这就是 AJAX 和 XML 的威力！

Get the slides today!

http://perlcabal.org/agent/slides/xmlapp/xmlapp.xul
http://perlcabal.org/agent/slides/xmlapp/xmlapp.ppt
http://perlcabal.org/agent/slides/xmlapp/xmlapp.pdf

You need Firefox to access the .xul link above.

Thank you!