

将已解析的 robots.txt 文件保存在 WWW::RobotRules 对象中，这个对象提供了一些方法，可以用于查看是否禁止对某指定 URL 进行访问。同一个 WWW::RobotRules 可以用于解析多个 robots.txt 文件。

下面是 WWW::RobotRules API 的一些主要方法。

- 创建 RobotRules 对象

```
$rules = WWW::RobotRules->new($robot_name);
```

- 装载 robots.txt 文件

```
$rules->parse($url, $content, $fresh_until);
```

- 查看站点 URL 是否可获取

```
$can_fetch = $rules->allowed($url);
```

下面这个短小的 Perl 程序说明了 WWW::RobotRules 的用法：

```
require WWW::RobotRules;

# Create the RobotRules object, naming the robot "SuperRobot"
my $robotsrules = new WWW::RobotRules 'SuperRobot/1.0';
use LWP::Simple qw(get);

# Get and parse the robots.txt file for Joe's Hardware, accumulating
# the rules
$url = "http://www.joes-hardware.com/robots.txt";
my $robots_txt = get $url;
$robotsrules->parse($url, $robots_txt);

# Get and parse the robots.txt file for Mary's Antiques, accumulating
# the rules
$url = "http://www.mary's antiques.com/robots.txt";
my $robots_txt = get $url;
$robotsrules->parse($url, $robots_txt);

# Now RobotRules contains the set of robot exclusion rules for several
# different sites. It keeps them all separate. Now we can use RobotRules
# to test if a robot is allowed to access various URLs.
if ($robotsrules->allowed($some_target_url))
{
    $c = get $url;
    ...
}
```

235

下面是 www.marys-antiques.com 的假想 robots.txt 文件：

```
#####
# This is the robots.txt file for Mary's Antiques web site
#####
```