要使 Allow/Disallow 行与一个 URL 相匹配，规则路径就必须是 URL 路径大小写相关的前缀。例如，Disallow: /tmp 就和下面所有的 URL 相匹配：

```
http://www.joes-hardware.com/tmp
http://www.joes-hardware.com/tmp/
http://www.joes-hardware.com/tmp/pliers.html
http://www.joes-hardware.com/tmpspc/stuff.txt
```

3. Disallow/Allow前缀匹配

下面是 Disallow/Allow 前缀匹配的一些细节。

- Disallow 和 Allow 规则要求大小写相关的前缀匹配。（与 User-Agent 行不同）这里的星号没什么特殊的含义，但空字符串可以起到通配符的效果。
- 在进行比较之前，要将规则路径或 URL 路径中所有“被转义”的字符（%XX）都反转为字节（除了正斜杠 %2F 之外，它必须严格匹配）。
- 如果规则路径为空字符串，就与所有内容都匹配。

表 9-3 列出了几个在规则路径和 URL 路径间进行匹配的例子。 233

表9-3 robots.txt路径匹配示例

规则路径	URL路径	匹配吗?	注 释
/tmp	/tmp	✓	规则路径 == URL 路径
/tmp	/tmpfile.html	✓	规则路径是 URL 路径的前缀
/tmp	/tmp/a.html	✓	规则路径是 URL 路径的前缀
/tmp/	/tmp	×	/tmp/ 不是 /tmp 的前缀
	README.TXT	✓	空的规则路径匹配于所有的路径
/~fred/hi.html	/%7Efred/hi.html	✓	将 %7E 与 ~ 同等对待
/%7Efred/hi.html	/~fred/hi.html	✓	将 %7E 与 ~ 同等对待
/%7efred/hi.html	/%7Efred/hi.html	✓	转义符是大小写无关的
/~fred/hi.html	~fred%2Fhi.html	×	%2F 是一个斜杠，但斜杠是种特殊情况，必须完全匹配

前缀匹配通常都能很好地工作，但有几种情况下它的表达力却不够强。如果你希望无论使用什么路径前缀，都不允许爬行一些特别的子目录，那 robots.txt 是无能为力的。比如，你可能希望禁止在用于 RCS 版本控制的子目录中爬行。除了将到达各 RCS 子目录的每条路径都分别枚举出来之外，1.0 版的 robots.txt 方案无法提供此功能。