9.2.5 User-Agent导向

Web管理者应该记住，会有很多的机器人来访问它们的站点，因此要做好接收机器人请求的准备。很多站点会为不同的用户代理进行内容优化，并尝试着对浏览器类型进行检测，以确保能够支持各种站点特性。这样的话，当实际的HTTP客户端根本不是浏览器，而是机器人的时候，站点为机器人提供的就会是出错页面而不是页面内容了。在某些搜索引擎上执行文本搜索，搜索短语“your browser does not support frames”（你的浏览器不支持框架），会生成一个包含那条短语的出错页面列表。

站点管理者应该设计一个处理机器人请求的策略。比如，它们可以为所有其他特性不太丰富的浏览器和机器人开发一些页面，而不是将其内容限定在特定浏览器所支持的范围。至少，管理者应该知道机器人是会访问其站点的，不应该在机器人访问时感到猝不及防。[16]

9.3 行为不当的机器人

不守规矩的机器人会造成很多严重问题。这里列出了一些机器人可能会犯的错误，及其恶劣行为所带来的后果。

- 失控机器人

  机器人发起HTTP请求的速度要比在Web上冲浪的人类快得多，它们通常都运行在具有快速网络链路的高速计算机上。如果机器人存在编程逻辑错误，或者陷入了环路之中，就可能会向Web服务器发出大量的负载——很可能会使服务器过载，并拒绝为任何其他人提供服务。所有的机器人编写者都必须特别小心地设计一些保护措施，以避免失控机器人带来的危害。

- 失效的URL

  有些机器人会去访问URL列表。这些列表可能很老了。如果一个Web站点对其内容进行了大量的修改，机器人可能会对大量不存在的URL发起请求。这会激怒某些Web站点的管理员，他们不喜欢他们的错误日志中充满了对不存在文档的访问请求，也不希望提供出错页面的开销降低其Web服务器的处理能力。

- 很长的错误URL

  由于环路和编程错误的存在，机器人可能会向Web站点请求一些很大的、无意义的URL。如果URL足够长的话，就会降低Web服务器的性能，使Web服务

注16：如果某站点上有一些不应该让机器人访问的内容，站点管理员该如何控制机器人在其站点上的行为呢？9.4节给出了相关的信息。