

## 《专题：数字资源长期保存》

# 新西兰国家图书馆和大英图书馆的 合作网络资源保存工具项目

新西兰国家图书馆 著 冯浩然 编译

## 1 引言

随着互联网的不断发展，世界上越来越多的文献资源只有通过网络才能获取。尽管这些文献很有价值并且使用方便，但是由于网络资源的版权不明确、稳定性欠佳和经常变化等原因，因此给专业的收集人员和想要获取它的研究人员造成了获取和使用上存在着很大挑战。

为了解决上述问题，新西兰国家图书馆和大英图书馆启动了一个设计和开发网络资源保存工具的项目，在图书馆环境下，通过用户协作，网络资源保存工具支持在线资源的筛选、收割和质量评估。网络资源保存工具支持选择性网络收割，选择者只要确定需要收割整个网站还是网站上某个特定学科、某一重要事件或主题的内容就可以了。这是一个优质、精细的网络内容收割工具。

网络资源保存工具是网络收割（使用一种叫做爬行器的软件在万维网中采集指定区域内网站的快照，包括该网站上的图像和文档）实践的最新进展。该工具的出现是一个进步，能够确保世界上的数字资源为后代子孙所使用，而不会因网页内容的临时性甚至是网站消亡而消失。

## 2 新西兰国家图书馆同大英图书馆之间的合作

新西兰国家图书馆同大英图书馆之间的合作是由国际互联网保存联盟（International Internet Preservation Consortium, IIPC）资助的。面对收集网络资源的巨大挑战，国际互联网保存联盟设想了一个桌面解决方案，这个方案能让各个组织不需要理解高深的技术就可以进行网络资源收割。

新西兰国家图书馆和大英图书馆愿意制定一个能够管理网络资源收割整个过程的解决方案，并且所有联盟成员和其他组织都能使用的。这个计划全部由新西兰国家图书馆和大英图书馆资助，其它 IIPC 联盟成员会在项目初期提供协作。

## 3 国际背景

数个国家图书馆、文化机构和财团已经意识到了收割网络资源的价值，而且他们已经开始实施了一些保存项目。这些组织包括 PANDORA 协会（澳大利亚国家图书馆及其合作伙伴）；英国网络存储协会（UK Web Archiving Consortium, UKWAC，包括大英图书馆，英国国家档案馆，威尔士国家图书馆，苏格兰国家图书馆，英国联合信息系统委员会，英国惠康基金会）；美国国会图书馆；法国国家图书馆；北欧网络存储协会（Nordic Web Archive，包括丹麦国家图书馆，冰岛国家图书馆，挪威国家图书馆，瑞典国家图书馆）；加拿大国家图书馆和档案馆；互联网档案馆（Internet Archive）。

上述大多数组织已经加入了国际互联网保存联盟。国际互联网保存联盟致力于网络资源保存方面的创新、工具开发、结果共享以及网络资源保存技术的进步。目前绝大多数网络资源收割活动是由国际互联网保存联盟及其成员组织完成的。

## 4 网络资源保存工具项目

本项目有两个正式目标：

- （1） 为国际互联网保存联盟制订出一个联合规范

- (2) 设计和开发出网络资源保存工具
  - a. 满足新西兰国家图书馆的需要
  - b. 满足大英图书馆的需要
  - c. 可扩展,从而满足澳大利亚国家图书馆和其他国际互联网保存联盟成员的需要

网络资源保存工具作为一个企业级规模的解决方案,它的设计是以用户为中心,并且实现了同其它组织系统的互操作。用户在它的帮助下不用具备深入的网络资源收割技术知识就能完成选择、描述和收割网络资源的工作。这一工具是可评估的,有工作流程,在确定存档内容后将对这些内容进行管理,包括许可、选择、描述、收割和质量评价。

网络资源保存工具能够为资源采集的各个流程提供支持,包括:选择网络资源;寻找可以收割的资源并让公众可以在收割过后能顺利地获取;对资源加以描述;判断资源的类别和范围;定时进行网络资源收割;对收割的资源进行质量评价并选择采集或放弃该资源;把采集的资源保存到数字知识库或档案库中。

随着网络资源保存工具的测试完成,它于2006年9月以一个开源项目进入了网络资源保存领域。用户可以从网上下载该工具,网址是<http://webcurator.sf.net/>。

## 5 网络资源保存工具在资源采集组织中的作用

随着越来越多的人认识到了网络资源的价值,人们对网络资源收集的活动也更为关注。而图书馆和档案馆在保存数字资源满足科研人员的短期、中期和长期需求中发挥着重要作用。

越来越多的科研人员在研究中使用网络资源,他们需要的是对这些资源的永久引用。然而永久引用问题会给所有要求可靠的、具有学术性的学科和领域产生影响。其他一些科研人员在学习过程中实际利用了网络资源,他们需要互联网上资源的引用状态的永久记录。

网络资源保存工具能够帮助那些负责资源采集的组织(例如国家图书馆、大学图书馆、专业图书馆、档案馆和研究图书馆)执行采集活动。

- (1) 它是协作式的(在组织内部或者各个组织之间都可使用)。
- (2) 它为组织提供一种具体方法用来收集网络资源。
- (3) 它考虑了国家利益,具有学术性,关注特定主题或者热门研究领域,能够引起组织内研究人员的兴趣。

下列是选择性网络资源收割及其潜在的科研价值一些例子:

- (1) 政党网站(在一次大选期间): 政治研究。
- (2) 约会、交友、博客或者社区网站: 性别研究或社会网络。
- (3) 娱乐或专业的邮件目录或网站: 休闲研究或产业研究。
- (4) 贸易或内容分享网站: 电子商务或市场营销。

## 6 结论

新西兰国家图书馆馆长 Penny Carnaby 说:“能与大英图书馆和国际互联网保存联盟一同参与这个旗舰项目是非常激动人心的。它能为在数字资源保存领域带来好处,对此我们深感荣幸”。

在2006年西班牙举行的欧洲数字图书馆大会上的第六次国际数字资源保存研讨会,新西兰国家图书馆和大英图书馆的代表发布和演示了该工具。项目小组当前的工作是让更多的人有机会了解和使用这一开源工具。

编译自: National Library of New Zealand. Web Curator Tool: a joint open-source project by the National Library of New Zealand and British Library. <http://www.ndl.go.jp/en/publication/cdnla0/058/584.html>. [2007-3-27]

(李忠霞 校)