

Duckling:支持 e-Science 的协作平台与应用框架

南 凯 董科军 谢建军

中科院 计算机网络信息中心

【摘要】支持 e-Science 的协同工作环境套件 (Duckling) 是面向 e-Science 及中科院信息化发展的需求研究开发的一套基于互联网的工具软件包。Duckling 是专为科研团队提供的综合性资源共享和协作平台,目标是集成数字化的硬件、软件、数据、信息等各类资源,组织跨地域、跨单位的科研人员,构成一个高效易用的网络化环境,支持和促进信息化时代的新型科研活动方式。Duckling 从 2008 年 11 月发布第一个正式版本 1.0 到 2010 年 3 月的最新版本 2.0,逐步发展为不仅是一个网络协同工具集,还是一个科研应用开发框架。Duckling 使用自定义的标记语言 DML 来存储文本信息,通过文档库 CLB 来实现文档资料的统一存储和管理,通过虚拟组织管理工具 UMT 来进行用户管理,支持单点登录和应用定制化的访问控制,通过 Duckling 应用集成框架 (DAIF) 支持 Portlet 方式的门户集成。Duckling 2.0 目前采用 Apache License Version 2 许可证开放源码,可以在 SourceForge 下载 (<http://duckling.sourceforge.net>)。

【关键词】协作平台、应用框架、e-Science、Duckling

1. 引言

信息技术的发展不断催生新的科研方法和手段,e-Science 是未来科研活动的发展方向[1]。在互联网的支持下,跨地域、跨组织、跨学科的大规模科研协作成为可能,当科学家进入这片新的天地后,他们对相关的软件工具提出了越来越高的要求。一方面,使用者对工具的期望是有助于更容易或更高效地完成工作,而不是本末倒置地在工具自身上花费大量的时间精力。另一方面,与目标越符合的专用工具越受欢迎,因

为通常专用的工具比通用的更有效。对科研人员来说，第二个问题更为突出，不同学科领域的差异使得科研应用具有很强的定制性，通常 e-Science 应用软件的开发都需要结合相应的领域知识。泛泛的通用工具可能一开始引起了科研人员的兴趣，但由于缺乏真正解决日常科研活动中实际问题的效用，往往难以形成使用粘性。

基于互联网的协作平台已经相当普遍。无论是在学术界还是公众的社交网络，各种各样的虚拟组织如雨后春笋，协作平台或其他形式的网上交流站点成为必需的工具。然而，对于开展科研工作的严肃业务而言，一般的协作平台或网上社区等软件却仍远远不能满足要求。前面提到的两方面的问题在这里形成某种程度上的冲突：成熟的、精雕细刻的软件需要在大量用户的参与下较长时间去磨合，通常这需要协作平台具有一定的通用性来适应足够规模的用户群，而这与科研业务的专业性难以在较短时间内同时满足。在中科院「十一·五」信息化建设专项中我们承担的支持 e-Science 的协同工作环境套件[2]的研发工作，就面临这样「鱼与熊掌」的困难。

本文介绍了我们提出的一种基于 Web 的协作平台和应用框架，命名为「Duckling」，希冀能以此为基础不断扩展形成能服务于 e-Science 的解决方案。下面第 2 节定义我们的目标并解释了一些关键的思路和设计原则的选择，第 3 节给出了 Duckling 的体系结构，第 4 节介绍了已有的若干应用实例，一些相关工作在第 5 节中作了简要讨论，最后是总结并展望了 Duckling 下一步的发展。

2. 目标与设计原则

本文的工作是在中科院的背景和支持下开展的，主要目标用户是中科院的科研人员。中科院有覆盖几乎全部自然科学各个领域的 100 多个研究所，分布在全国各地的二十多个城市。当然 e-Science 绝对是一个全球性的理念，我们目前把此项工作的范围定义在中科院内部，只是为了更有针对性和更容易入手与检验。应该容易看出，我们接下来讨论的很多问题都是可以延伸到更大的应用范围，或许我们将来会在另一篇独立的文章中再行讨论。

因此，我们的目标就是研究和开发一个基于网络的协作平台，科研人员将把越来越多的每日的科研活动放在这个网上的环境中进行，它将逐步改变科研人员使用计算机和软件的方式，并带来很多超越现在的新功能。这个平台首先要在中科院的各研究所中得到应用。下面列出我们定义的这个协作平台的一些核心指标：

(1) 平台的用户是以团队为基本单位，可以是现实中的组织或机构，也可以是由有着共同兴趣的陌生人组成的虚拟社区。

(2) 平台可以部署为供大量团队使用的公共服务，也可以部署在一个团队私有的环境中单独使用。

(3) 用户可以随时随地、尽可能方便地使用平台的服务。

(4) 用户可以为平台增加功能，可以灵活定制自己的使用方式。

(5) 平台可以抵御用户的恶意攻击。

(6) 用户和系统的行为是可记录的，并可提供给授权的应用。

针对上述目标，我们结合近年来互联网、移动通信、云计算[3]等技术的快速发展，提出了以下一些设计原则：

(1) 完全基于 Web，在浏览器中可以使用全部功能；

(2) 对不同的终端有相适应的用户界面，目前典型的是计算机和智能手机；

(3) 有良好的可伸缩性，支持软件即服务 (SaaS) 方式的大规模服务；

(4) 可以通过分布式部署来提升性能和可靠性；

(5) 平台提供基本的交互界面、信息存储和应用间协作的服务，支持应用的热部署和安全隔离。

3. Duckling 的体系结构

基于上述考虑，我们设计了一套支持 e-Science 的协同工作环境套件，命名为

Duckling[4]。Duckling 从 2008 年 11 月发布第一个正式版本 1.0 到 2010 年 3 月的最新版本 2.0，逐步发展为不仅是一个网络协同工具集，还是一个科研应用开发框架。Duckling 使用自定义的标记语言 DML 来存储文本信息，通过文档库 CLB 来实现文档资料的统一存储和管理，通过虚拟组织管理工具 UMT 来进行用户管理，支持单点登录和应用定制化的访问控制，通过 Duckling 应用集成框架（DAIF）支持 Portlet 方式的门户集成。Duckling 2.0 目前采用 Apache License Version 2 许可证开放源码，可以在 SourceForge 下载（<http://duckling.sourceforge.net>）。

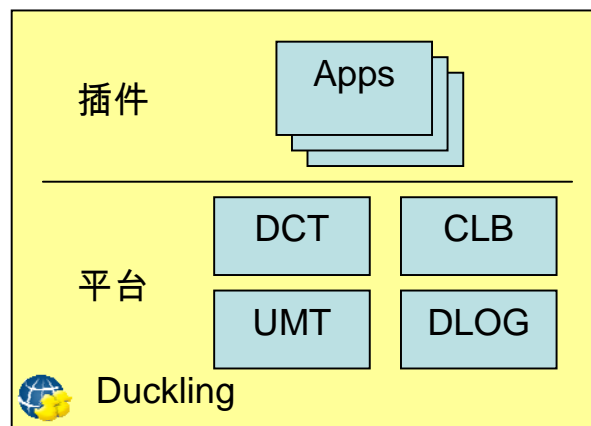


图 1 简化的 Duckling 体系结构

简化的 Duckling 的体系结构如图 1 所示：整体上分为两层，下层是平台层，提供协作平台的基本功能和对应用插件的管理；上层是应用层，通过插件的方式实现各种各样的应用功能。平台层包括 4 个主要模块：DCT 是文档协同工具，提供了网页形式的基本交互界面，支持通过所见即所得的编辑器以 wiki 方式在线编辑页面、输入信息、发布内容；CLB 是文档库，提供统一的文件存储和管理服务；UMT 是虚拟组织管理工具，提供面向团队的用户管理功能，并支持单点登录；DLOG 是日志服务。插件的接口目前采用的是 Portlet。

4. 应用实例

我们已经基于 Duckling 开发了一些具体的科研应用，会议服务平台和文献管理工

具就是针对科研活动中组织学术会议和管理科技文献这两方面的需求分别实现的应用插件。

会议服务平台（CSP）是一套基于 Duckling 核心服务开发，用于会务组织，信息发布与沟通，信息获取、资源组织管理及共享的平台软件，是会议组织者、会议代表、管理部门三者间的桥梁与纽带。主要提供多会议创建、用户注册与管理、在线支付、摘要提交与管理、酒店预订与管理、信息发布与浏览、功能与界面的定制，站内消息与反馈沟通等功能。

文献管理工具（LMT）是用于科研团队内部进行文献管理与共享的工具，依据用户特定学科领域内的文献检索与管理需求，进行自动的文献数据获取，针对相应的结果进行集中管理与共享，同时为用户提供文献检索及详情查看、检索历史保存、个人收藏、文献全文上传/下载、在线预览、文献导入、共享与推荐、文献源及检索周期设定等功能。

下面以「能源微生物功能基因组学虚拟实验室」为例介绍 Duckling 在一个科研团队中应用的实例。

生物质能源是新型清洁能源的典型代表，在生物质能源的研发过程中，科研工作者往往需要在一个集成的协同科研平台上分析大量不同能源微生物的海量基因组和功能基因组等数据，并支持不同地域的科研工作者在研究分析过程中以多种方式进行交流和讨论，同时还可以使用协同科研平台集成的高性能计算资源进行计算模拟以及结果的可视化。

基于 Duckling 提供的核心服务，中科院计算机网络信息中心和中科院青岛生物能源与过程研究所一起共同研发了能源微生物功能基因组学虚拟实验室，形成了面向能源微藻的协同科研环境，建立了围绕 454、Solexa 等新一代测序仪的，从数据产生、数据规整、数据拼接到数据分析的全科研流程，功能包括：微藻基因数据质量控制、基因拼接、基因预测、基因注释、专业文献推送、实验室管理和团队协作等。利用能源微生物功能基因组学虚拟实验室所提供的各类服务功能，可以在一个集成的信息化

科研环境实现生物质能源技术研究的全过程化，从而加快了生物质能源的研发速度。



图 2 Duckling 应用实例：能源微生物功能基因组学虚拟实验室

5. 相关工作

面向科研需求开展的相关软件平台和工具的研发近年来受到普遍重视，在网格（Grid Computing）和 e-Science 等方向上各国都有一些典型的项目，例如美国 TeraGrid 的科学网关（Science Gateway[5]）、英国的虚拟研究环境（Virtual Research Environment, VRE[6]）、欧盟的虚拟实验室（Virtual Laboratory for e-Science[7]）等等。Duckling 与它们相比，共同的地方是都与一些特定学科领域进行密切结合来开发应用；Duckling 的特点是在应用框架和开发平台方面做了很大努力进行扩展。

SAKAI[8]是一个社区开源项目，聚焦于构建核心协作功能的资源，开发一个协同工具箱，用于教学以及即时协作。该项目提供一个框架，一方面「开箱即用」；另一方面，通过增加一些特别的工具组件，使之也能应用于特定领域。Liferay[9]是一个基于 Portlet 的门户框架，Duckling 目前也是采用了 Portlet 作为插件开发接口。

6. 总结与展望

Duckling 目前具备了一些网络协作平台的核心功能，并在中科院的若干研究所中开发了特定学科领域的应用。作为一个应用框架，Duckling 才刚刚起步，对第三方应

用开发人员的支持还远远不够，开源之后，相信能够有助于推动在这方面的发展。下一步 Duckling 平台层的 4 个模块将进一步完善和精练 形成规范化的应用插件开发接口（Duckling API），预计将于 2010 年底正式发布。同时，我们将循着本文第 2 节描述的目标和设计原则继续推进 Duckling 在性能和安全性方面的发展。

参考文献

- [1] Tony Hey and Anne E. Trefethen. The UK e-Science Core Programme and the Grid. Future Generation Computer Systems. 2002, 18(8):1017-1031.
- [2] 南凯，李华飏，董科军等. 支持 e-Science 的协同工作环境. 科研信息化技术与应用，2008，1(1):35-40.
- [3] Brian Hayes. Cloud computing. Communications of the ACM, 2008, 51(7):9-11.
- [4] <http://duckling.escience.cn/>, 2010
- [5] <http://www.sciencegateway.org/>, 2010
- [6] Roadmap for a UK Virtual Research Environment: Report of the JCSR VRE Working Group (2004),
http://www.jisc.ac.uk/uploaded_documents/VRE%20roadmap%20v4.pdf
- [7] <http://www.vl-e.nl/>, 2010
- [8] <http://sakaiproject.org/>, 2010
- [9] <http://www.liferay.com/>, 2010

【作者简介】

南凯 男

职 称：中科院 计算机网络信息中心 研究员

职 务：中科院 计算机网络信息中心「协同工作环境研究中心」主任

研究领域：分布式系统、网络协同工作环境

个人简介：1996 年北京大学本科毕业；分别于 1999 年和 2006 年中科院计算所硕士和博士毕业。主要的研究成果有协同工作环境套件（Duckling）和科学数据网格中间件（SDG）。

联络电话：86-10-58812301

联络邮箱：nankai@cnic.ac.cn

谢建军 男

职 称：中科院 计算机网络信息中心「协同工作环境研究中心」助理研究员

职 务：中科院 计算机网络信息中心「协同工作环境研究中心」技术主管

研究领域：协同软件

个人简介：1998 年 9 月～2002 年 7 月 就读于北京师范大学 获学士学位

2002 年 9 月～2005 年 7 月 就读于北京师范大学 获硕士学位

2005 年 7 月～至今 中科院计算机网络信息中心工作

联络电话：86-010-58812322

联络邮箱：xiejj@cnic.cn