

Mashup 工具的分析与研究 *

纪姗姗^{1,2} 李宇¹ 周强¹

¹ (中国科学院国家科学图书馆 北京 100190)

² (中国科学院研究生院 北京 100049)

【摘要】探讨 Mashup 工具的研究现状及功能特点。在介绍 Mashup 工具相关技术、研究成果及已有分类模型的基础上,提出一个新的分类模型,包括数据层融汇工具、表示层融汇工具及企业级融汇工具,然后分类讨论 Mashup 工具的主要功能、关键技术及代表产品。

【关键词】Mashup 工具 分类模型 企业级 Mashup 系统架构

【分类号】G250.7

Research on Mashup Tools

Ji Shanshan^{1,2} Li Yu¹ Zhou Qiang¹

¹ (National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

² (Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

【Abstract】After revealing the current research status and features of Mashup tools, this paper introduces some relevant technologies and classification models of Mashup tools. And based on this, the paper designs a new classification model composed of Data Mashup tool, Presentation Mashup tool and Enterprise Mashup tool. In the same time, it analyzes the main functions and crucial technologies of typical Mashup tools.

【Keywords】Mashup tool Classification Enterprise Mashup Architecture

1 引言

Mashup 是指将不同来源的数据或服务进行组合,从而构建出一种具有新型功能的网络应用^[1,2]。自 2005 年采用 Mashup 技术的 Housingmaps.com^[3] 上线以来, Mashup 引起了广泛的关注并得到迅速的发展,许多网站如 Flickr、Amazon、Yahoo 等将 Mashup 服务建设作为提升自身竞争力的重要策略。Mashup 的主要特点是:支持用户对来自机构内或第三方已有资源和服务的集成重用,缩短服务系统构建周期;支持没有技术背景的科研用户自我构建个性化服务系统,使科研用户真正从单纯的网络信息消费者变成消费、生产以及网络服务构建等多重身份。在应用实践上,如果对 Mashup 用户有较高的技术要求,将严重影响 Mashup 技术和服务的推广,因此开发智能化工具成为当前 Mashup 研究与建设的重要内容。根据 Forrester 分析家 G Oliver Young 的预测,到 2013 年企业级的 Mashup 技术应用市场规模将有望达到 7 亿美元^[4],广泛的用户需求以及潜在的商业价值将不断推动 Mashup 工具的研究与建设。

收稿日期: 2009 - 10 - 16

收修改稿日期: 2009 - 10 - 27

* 本文系中国科学院 2009 年新增能力项目“知识服务集成融汇关键技术研究”的研究成果之一。

2 Mashup工具研究现状

2.1 Mashup技术方法研究

当前,关于 Mashup 的技术研究主要集中在 Mashup 数据获取与呈现技术、数据融汇模型、Mashup 组件规范、Mashup 运行环境规范、信息安全管理与用户认证机制等方面。其中,MIT 图书馆和 MIT 人工智能实验室项目 SMLE^[5]主要研究屏幕抓取技术、语义融汇技术以及融汇信息可视化呈现技术,该项目提出的 Timeline、Exhibit、Semantic Bank 等技术被很多 Mashup 工具所采用。OpenAjax 联盟则联合统领用户、相关开源工程、贡献 Ajax 技术工具的公司,积极推动基于 Ajax 技术的 Mashup 开发环境、Mashup 组件 Widget 定义、运行环境以及跨平台交互机制、安全管理机制等方面的规范化建设,帮助用户提升 Mashup 能力,促进 Ajax Ecosystem 的建设,已经推出了 OpenAjax Hub 2.0 等标准规范^[6]。日本神户大学、中国科学院计算所研究人员分别提出 Mashup Tree^[7]、嵌套表^[8]等融汇数据模型,支持对 Web 服务连接、数据抽取、数据融合等过程控制以及对新生成综合数据对象的描述。由 Blaine Cook、Chris Messina、Larry Halff 及 David Recordon 共同发起的 OAuth^[9]协议为 Mashup 的 API 认证提供了一种解决方案,目前包括 Twitter、Google、MySpace 以及中文豆瓣网等都采用了 OAuth。文献 [10] 对 Mashup 关键技术作了比较详细的分析。

在 Mashup 工具研发方面,包括 BM、Microsoft、Google、Yahoo、Intel 等在内的众多著名国际组织机构、软件公司纷纷加入,开发出了针对不同层次用户的产品。其中 Yahoo Pipes 是一个非常强大的互联网数据源聚合平台,覆盖了目前互联网上常见的全部数据源。Microsoft Popfly 和 Google Mashup Editor 适合于普通用户,可以用来创建自己的小程序,也可以利用现有组件创建 Mashup,并且与其他用户共享。Intel MashMaker 提供 Mashup 网浏览器插件、组件创建工具、应用创建工具、Mashup 引擎等。BM、Xignite、Kapow、JackBe 等推出了自己的企业级 Mashup 系统。欧盟 FP7 也设立了专门项目 Romulus^[11],推出了开源融汇工具 DERI Pipes、My Cocktail 以及支持融汇服务组合的新版本 Liferay。

2.2 Mashup 工具类型

(1) 按获取方式分类

按获取方式不同,当前的 Mashup 工具可以分为以下三类:

免费工具,这些工具一般都是支持在线免费使用,有的工具同时发行提供技术支持的商业版本。比较流行的产品有 Yahoo Pipes、Microsoft Popfly、Intel Mash Maker、iGoogle、Netvibes、Dapper、Afrous 等。

开源工具,它们通过发布开放源码,支持社区用户在遵循相关策略下的自由使用以及根据个性需求进行二次开发。目前比较有影响的开源工具包括 SnapLogic、Aptarx、WSO2 Mashup Server、Jitterbit、Intalio Mashup、Portaneo Posh 等。有些工具文档齐全,并建立了比较完善的开源社区,提供代码维护、版本升级等服务。

商业工具,这些产品需付费使用,主要面向企业用户,它们提供了比较全面的功能,包括安全管理控制功能,支持复杂的企业资源与服务整合需求。代表产品包括:BM Mashup Center、Denodo、JackBe Presto、Kapow、Twinsoft Convertigo、Serena Mashup Composer 等。

(2) 按用户对象分类

不同用户群体对融汇工具的功能需求有所不同,按用户对象可以将 Mashup 工具分为两类:

面向普通用户的 Mashup 工具 (Consumer Mashups),提供简单易用功能,支持用户自我实现多数据源连接、数据获取与融合,支持基于可视化界面的服务组合、新应用系统构建、发布,支持用户个人资源、服务与社会网络之间的交互等。这类工具通常不具备身份认证、信息安全控制管理功能。

面向企业级用户的 Mashup 工具 (Enterprise Mashup),它除了具备第一类工具的基本功能外,还支持机构内开发人员灵活方便增加新的数据源,如机构内部自有数据源、遗留系统资源等;支持对第三方系统、平台中的融汇服务的导入以及将本地融汇服务导出到其他第三方平台上;封装信息安全控制与服务管理功能,支持基于授权的融汇服务调用。企业级融汇工具适合建立面向机构、区域用户的集成融汇服务支撑平台。

(3) 按功能特点分类

根据功能的不同可以将集成融汇工具分为 4 类:

资源适配器型工具 (Adapter),即支持用户从各类资源中抽取信息并与自身的数据格式形成映射,以建立新的服务组件。

数据仓储型工具 (Repository),用于建立 Mashup 组

件仓储系统,支持对融汇对象的注册、存储、发现、调用、管理,存储内容可能包括资源对象、组件 (Widget)对象、自定义对象、特定格式文件等。用户可以浏览、调用这些对象,甚至可以对其进行评价、描述及推荐。

服务组合型工具 (Transformation/Aggregation),支持建立可视化交互界面,支持用户通过鼠标拖 拉等方式选择资源、服务及组件,并按一定流程进行编排,以构建新的服务系统。

应用表现型工具 (Presentation),支持各类融汇组件在一个统一的界面中进行显示、组合,以构建新的服务应用。

资源适配型及数据仓储型工具偏重于各类组件的生成、存储及管理,属于列表器 (Catalogue)类别。服务组合型及应用表现型工具的主要功能采用 Catalogue中的各类资源、服务生成新的应用,属于编辑器 (Editor)类别。Hoyer等人建立了一个 Mashup 工具功能分类矩阵,如图 1 所示:

功能	Editor	Front-End/ Presentation	Google Mashup Editor Microsoft Live Spaces Pageflakes Netvibes iGoogle	Intel Mash Maker	IBM QEDWiki Extensio TIBCO	Nexaweb Enterprise Web 2.0 Suite SAP Research Rooftop	Procession Bungee Labs Bungee Connect Denodo IBM WebSphere sMash WSO2 Mashup Server
	Transformation/ Aggregation		Microsoft Popfly	Yahoo! Pipes		IBM DAMIA SnapLogic	Dabble DB KnowNow ESS
	Repository		Microsoft Popfly Pageflakes Turtilla Netvibes iGoogle	Microsoft Live Spaces Yahoo! Pipes		Procession IBM QEDWiki	
	Catalogue	Adapter	Dapper Dapper Factory		Kapow Technologies Mashup Server	IBM Mashup Hub	
				Intel Mash Maker	RSSBus System One Apatar Data Mashup Integration	Denodo WerkLight RatchetSoft Ratchet-X	Dabble DB KnowNow ESS
			Consumer		Enterprise		
			用户对象				

图 1 Mashup 工具分类矩阵^[12]

3 Mashup 工具分析

图 1 所示的分类模型虽然对 Mashup 工具进行了细致的划分,但不能绝对地把一个工具归为某个类别,如 Yahoo Pipes 既属于数据仓储型工具,又具有服务组合型工具的功能,这就导致了分类结果的不明确。本文从另一个角度按照工具的功能将其分为三个类别:数据层融汇工具 (Data Mashup Tools)、表示层融汇工具 (Presentation Mashup Tools) 及企业级融汇工具 (Enterprise Mashup Tools)。

3.1 数据层融汇工具

数据层融汇工具主要实现数据获取、数据融合及 Mashup 输出等功能^[13]。利用数据层融汇工具,用户可从多种数据源中获取数据对象。有的工具只能从特定格式的数据源中获取数据,如 DERIPipes^[11]只能对 XML 或 RDF 数据进行融汇。有的工具支持对多种数据源的访问,包括数据库、CSV、Excel 表格等本地数据源,以及网络数据源 Web Service (SOAP/REST)、RSS/Atom Feeds,同时支持多种数据描述规范,如 XML、HTML、JSON、E-mail 文本文件等。连接不同的数据源采用不同数据获取技术,包括 Web Feed 技术、API 调用技术及屏幕抓取技术^[14]。获取数据后,Mashup 工具需要对这些格式各异的数据进行相关处理,包括数据转换 (Transform)、映射 (Map)、过滤 (Filter)、连接 (Join)、排序 (Order) 等,最后构建出一个新的综合化的数据对象。为了方便地实现数据融汇,大部分 Mashup 工具提供了简易的编辑环境。有些是拖曳式的图形编辑界面 (如 Popfly),有些则是嵌入到 Eclipse 等开发环境的插件,用户需通过编程来实现数据融汇操作,有些开源工具同时支持以上两种功能。经过前两步操作,Mashup 工具生成一个综合数据对象,可以通过何种形式对它进行输出、保存、发布,不同工具之间存在一定差异。常见的输出形式有以下几种:

(1) 输出融汇组件 Widget,如表格组件 (Data Grid Widget)、图形组件 (Chart Widget)、地图组件 (Map Widget)、时间轴组件 (Timeline Widget)、展览馆组件 (Exhibit Widget) 等。

(2) 输出为分配有 ID 的文件,如 RSS/Atom、RDF、TXT、二进制文件等。

(3) 输出为 Web Service,如基于 SOAP/WSDL 的服务、基于 REST/WADL 的服务等,支持第三方开放调用。

典型的数据层融汇工具包括 Yahoo Pipes、Microsoft Popfly、Denodo、SnapLogic、Intalio | Mashup、Apatar、Jitterbit、Talend Open Studio、WSO2 Mashup Server 等。其中,Apatar^[15]是一个开源的数据融汇工具,它提供了可视化的操作界面以及支持资源连接、数据处理等多种操作组件。Apatar 支持对多种数据源的访问,包括数据库、HTML、XML、RSS、Excel、RDF,还提供 Amazon、Flicker、Salesforce.com 等 API 调用组件。在

数据处理方面,提供的数据操作有 Aggregate、Distinct、Filter、Join及 Split等。与传统的 ETL工具相类似,用户无需编码,只要选择合适的组件并进行配置就能完成数据流程的设计,执行数据流脚本完成数据融汇过程。Apatar的输出类型可保存为本地的数据,如数据库文件、XML等,但它没有提供 Web Service发布功能。Apatar的优点是数据处理功能比较完善,可以嵌入到其他开发平台完成更复杂的应用。缺点是数据输出类型少、没有安全控制等功能。

3.2 表示层融汇工具

表示层融汇工具通过封装融汇组件 Widget,支持用户以可视化方式选择多个融汇组件并进行个性化组合,从而实现用户自我构建新型服务系统的目标。表示层融汇工具主要包括两方面功能:

(1)支持 Widget封装,即创建图形化融汇组件 Widget。Widget是指完成特定功能的 Web应用,用脚本语言描述,可以嵌入到网页中与其他 Widget一起构成新的网页。选用 Widget不仅可以融合来自多个系统的应用,还能以个性化的形式将其显示。在 Web2.0环境下,Widget以其轻量级、个性化、容易制作、功能丰富、界面美观等特点取得了快速发展。

(2)支持 Widget集成组合,包括 Widget组件发现、组件向个人页面嵌入、组件外观属性定制、组件在页面中的布局调整、页面风格配置、页面访问控制以及页面发布共享管理等功能。同时,表示层工具还应具备组件 Widgets的跨平台导入、导出以及用户评论、标签、分级等社会网络管理功能。

目前,比较流行的表示层组件工具有 iGoogle、Netvibes、Afious。其中,Netvibes^[16]是一个基于 Ajax技术创建的表示层融汇工具。它允许信息发布者及应用开发人员创建 Widget,并保存在 Netvibes生态系统中,支持用户持久对 Widget的访问调用。Netvibes的目标是支持用户自我构建个人网络开始页面(Startpage),满足用户日常数字生活需要。Netvibes生态系统已经收集到 17.5万个 Widget,由来自 115个国家的 1.5万个用户贡献,每月提供的 Widget调用服务超过 5亿次^[16]。在 Netvibes网页上,用户可以将 Widgets、Feeds、社交网络、E-mail Video及 Blog信息进行集成显示,形成个性化的个人首页。为了统一平台上 Widget的标准,Netvibes推出了 Widget协议 UWA(Universal Widget

API),该协议适用于多个 Widget平台,包括 Netvibes、iGoogle、Windows Vista及手机平台 iPhone。

3.3 企业级融汇工具

企业级融汇工具的主要特点是功能非常完善,不但包含数据层、表示层工具的相关功能,而且可以根据企业需求增加一些新功能,如信息安全管理、授权使用控制等。对于企业级的融汇来说,一个完整开发过程要从资源的获取开始,经过数据融汇、Widget封装等一系列操作形成 Widget,最后实现 Mashup界面的显示^[17]。关于企业级融汇工具的体系架构, Gartner提出一个 8层模型^[18],但其比较复杂,文献[18]提出的 5层模型则更有代表性,如图 2所示:

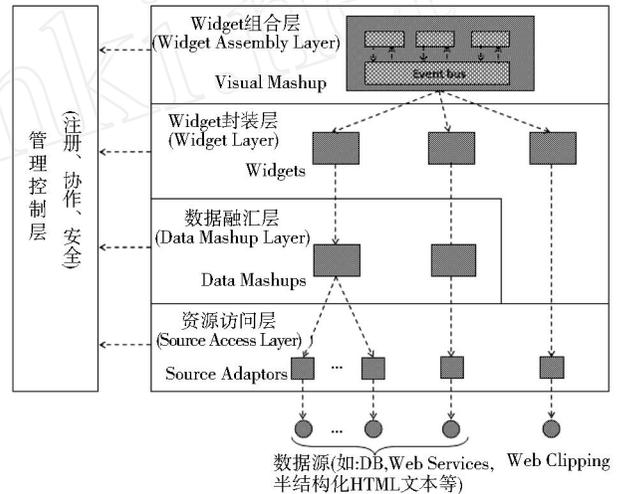


图 2 企业级 Mashup 开发工具参考架构^[19]

(1)资源访问层 (Source Access Layer)实现与关系数据库、Web服务、RSS、网页等数据源的连接,每种数据源需要开发相应的适配器 (Adaptor),经注册后支持上一层调用。

(2)数据融汇层 (Data Mashup Layer)开发可视化操作界面,支持用户将多来源数据融合成新数据对象,数据融汇操作将依照针对特定需求而事先定义的工作流进行。

(3)Widget封装层 (Widget Layer)负责将融汇后的数据与可视化呈现工具融合,生成规范的表示层融汇组件。

(4)Widget组合层 (Widget Assembly Layer)支持对 Widget聚合 (Aggregation)构建个性化页面。

(5)管理控制层 (Common Services)负责以上 4个层次中的数据管理、操作控制,包括组件注册 (Registry)、协

作 (Collaboration) 及安全 (Security) 三大部分。注册是指对各种组件 (数据源连接组件、数据操作组件、Widgets) 的注册、检索及管理, 此外还包含了版本控制及社区反馈的功能。协作指用户之间的 Mashup 共享及推荐。安全包含了对用户的登录管理及角色管理等内容。

目前, 有代表性的企业级融汇工具主要有 BM Mashup Center, Kapow, JackBe Presto, Twinsoft Convertigo, Serena Mashup Composer 等。其中, JackBe Presto^[20] 产品包括以下组件: Mashup 服务器、Mashup 设计器及连接器。Mashup 服务器用于连接资源及发布服务, Mashup 设计器提供创建 Mashup 组件的编辑环境, Presto 连接器则是一个可以与 Portal 服务器集成、实现 Mashup 显示的工具。在管理控制方面, Presto 使用了一个高可用的数据库 (MySQL) 来处理失败, 还具备安全集成并实现了 Active Directory (AD)、LDAP 及 Computer Associate 的 Netegrity 工具上的单点登录 (SSO)。除此之外, JackBe 还通过其 Mashup 开发社区提供了免费的 Mashup 培训和支持, 用户可以共享自己创建的 Mashup。

4 结 语

作为 Web2.0 环境下的新产品, Mashup 工具尚处于探索发展阶段, 随着相关技术的发展及标准规范的建立, 新的 Mashup 工具在功能上日趋完善、兼容性更强、操作上也更简单, 但仍有一些问题需要关注, 包括: 资源访问时的质量控制问题, 既要避免对资源提供方造成知识产权侵犯, 又要应对错误或恶意的数据进行过滤, 以防其影响工具的融汇性能; 用户登录时的安全认证问题, 用户身份的确定是对系统的安全保障, 也是用户享受个性化服务的基础; 资源与服务的共享性问题, 加强各组件的规范化建设, 支持其他 Mashup 制造方的调用、组合, 把计算机、手机等多类型终端作为服务的接入形式, 真正形成一个 Mashup 生态系统。

参考文献:

- [1] Wikipedia [DB/OL]. [2009 - 09 - 10]. [http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)).
- [2] Koschieder A, Torres V, Pelechano V. Elucidating the Mashup Hype: Definition, Challenges, Methodical Guide and Tools for Mashups [R/OL]. [2009 - 09 - 10]. <http://www.integrator.net/mem2009/papers/paper14.pdf>
- [3] Housingnaps [EB/OL]. [2009 - 09 - 10]. <http://www.housingnaps.com/>.
- [4] Forrester. Enterprise Mashups to Hit \$700 Million by 2013 [EB/OL]. [2009 - 09 - 10]. http://www.readwriteweb.com/archives/forrester_enterprise_Mashups.php.
- [5] SMLE Project [EB/OL]. [2009 - 08 - 24]. <http://smile.mit.edu/>.
- [6] Introducing OpenAjax Hub 2.0 and Secure Mashups [EB/OL]. [2009 - 09 - 10]. <http://www.openajax.org>
- [7] Sohei Ikeda, Takakazu Nagamine, Tomio Kamada. Application Framework with Demand - Driven Mashup for Selective Browsing [C]. In: *Proceedings of the 10th International Conference on Information Integration and Web - based Applications & Services*, Linz, Austria. New York, NY, USA: ACM, 2008: 33 - 40.
- [8] Wang G L, Yang S H, Han Y B. Mashroom: End - user Mashup Programming Using Nested Tables [DB/OL]. [2009 - 07 - 28]. <http://www2009.org/proceedings/pdf/p861.pdf>
- [9] Awood M, Balfanz D, Bounds D, et al. OAuth Core 1.0 Revision [EB/OL]. [2009 - 09 - 10]. <http://oauth.net/core/1.0a>
- [10] 李峰, 李春旺. Mashup 关键技术研究 [J]. *现代图书情报技术*, 2009 (1): 44 - 49.
- [11] ICT ROMULUS Report: D4.1 - Annual Report on Mashup Integration [R/OL]. [2009 - 08 - 24]. http://www.ict-romulus.eu/c/document_library/get_file?p_l_id=2266&folderId=29495&name=DLFE-4583.pdf
- [12] Hoyer V, Fischer M. Market Overview of Enterprise Mashup Tools [C]. In: *Proceedings of the 6th International Conference on Service - oriented Computing*, Sydney, Australia. Berlin, Heidelberg: Springer Verlag, 2008: 708 - 721.
- [13] Giusy Di Lorenzo, Hakim Hacid, Hye - Young Paik, et al. Data Integration in Mashups [EB/OL]. [2009 - 08 - 24]. http://www.sigmod.org/sigmod/record/issues/0903/p59_surveys_hacid.pdf
- [14] 李春旺, 肖伟. 集成融汇: 概念、模式与应用 [J]. *现代图书情报技术*, 2008 (12): 22 - 26.
- [15] Apatar [EB/OL]. [2009 - 09 - 10]. <http://www.apatar.com/>.
- [16] Netvibes Overview 2009 [EB/OL]. [2009 - 09 - 10]. <http://www.slideshare.net/culturebuzz/netvibes-overview-09>.
- [17] Hoyer V, Stanoesvka - Slabeva K, Janner T, et al. Enterprise Mashups: Design Principles Towards the Long Tail of User Needs [EB/OL]. [2009 - 08 - 24]. http://iknow.seforge.org/classic_bib/classic_paper/672a52067c7b76846587732e/inproceedingsreference.2009-04-22.7620820818
- [18] Reference Architecture for Enterprise 'Mashups' [EB/OL]. [2009 - 08 - 24]. http://liquidbriefing.com/pub/Harmonia/IndustryAnalysts/reference_architecture_for_e_151491.pdf
- [19] Javier Lopez, Alberto Pan, Fernando Bellas, et al. Towards a Reference Architecture for Enterprise Mashups [EB/OL]. [2009 - 08 - 24]. <http://www.sistedesess/TJISBD/Vol-2/No-2/articles/Lopez.pdf>
- [20] JackBe Presto: The Enterprise Mashup Platform [EB/OL]. [2009 - 08 - 24]. http://www.jackbe.com/downloads/Jackbe_Presto_Overview.pdf

(作者 E-mail: jishanshan@mail.las.ac.cn)