

理论完备，深入阐释了标记系统的价值和原理

实践性强，全面讲解了标记系统的架构设计方法和实现细节

华章程序员书库

Tagging: People-Powered Metadata for the Social Web

标签 标记系统设计实践

(美) Gene Smith 著

张军 陈军亮 译



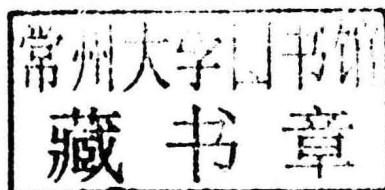
机械工业出版社
China Machine Press

Tagging: People-Powered Metadata for the Social Web

标签 标记系统设计实践

(美) Gene Smith 著

张军 陈军亮 译



机械工业出版社
China Machine Press

本书对标记系统这一概念的内涵和外延进行了系统化的、深入浅出的阐述。从什么是标记系统、标记系统有什么价值，到标记系统的架构和与其他分类系统的对比，再到标签的呈现方式和标记系统的实现细节，作者都用通俗易懂的语言进行了阐述，并附有详细的示例和具体的案例研究。本书的每一章都涵盖了标记系统的一个方面，主要内容包括：标记系统的模型、价值、架构，标签的分类、可视化、管理方法，最后介绍标记系统设计方法。本书带领读者对标签和大众分类法进行深入思考，最终的目标是帮助理解标签和设计标记系统。

Authorized translation from the English language edition, entitled TAGGING: PEOPLE-POWERED METADATA FOR THE SOCIAL WEB, 1E, 9780321529176 by SMITH, GENE, published by Pearson Education, Inc, publishing as New Riders, Copyright © 2008.

All rights reserved.

CHINESE SIMPLIFIED language edition published by CHINA MACHINE PRESS/HUAZHANG CO., Copyright © 2012.

本书中文简体字版由美国Pearson Education培生教育出版集团授权机械工业出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

封底无防伪标均为盗版

版权所有，侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2011-6870

图书在版编目（CIP）数据

标签：标记系统设计实践 / (美) 史密斯 (Smith G.) 著，张军，陈军亮译. —北京：机械工业出版社，2012.6

(华章专业开发者丛书)

书名原文：Tagging: People-Powered Metadata for the Social Web

ISBN 978-7-111-38598-1

I. 标… II. ① 史… ② 张… ③ 陈… III. 网页制作工具—程序设计 IV. TP393.092

中国版本图书馆CIP数据核字 (2012) 第111530号

机械工业出版社 (北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑：秦 健

三河市杨庄长鸣印刷装订厂印刷

2012年6月第1版第1次印刷

186mm×240mm · 13.75印张

标准书号：ISBN 978-7-111-38598-1

定价：59.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88378991，88361066

购书热线：(010) 68326294，88379649，68995259

投稿热线：(010) 88379604

读者信箱：hzjsj@hzbook.com

译者序

什么是标签 (tag)？标签有什么用处？这两个问题乍一听起来似乎很好回答：“标签不就是那些给内容加的关键词吗？加上关键词以后，查找起来会方便一些。”这样的答案大概反映了大多数普通用户对标签的印象。

诚然，上述的确是标签的作用之一。然而，这就是全部吗？标签真的就像表面上看起来这么简单吗？并非如此。自从有人类历史以来，人们就一直致力于对信息进行记录、整理和分类，从远古时代的结绳记事，到如今我们耳熟能详的图书馆联机公共检索目录 (OPAC)、Yahoo分类目录、Google搜索引擎，无不是人们面对信息的爆炸式增长为信息的管理而做出的努力。标签和以其为基础的大众分类法 (folksonomy，也译作分众分类或众分类法) 的出现，为人们在迅速增长的信息流中快速准确地检索到所需信息提供了一种新的可能。相比于传统的层级式分类，大众分类法产生的分类是一种扁平结构的由用户生成的分类结构，比传统分类有更好的实时性和灵活性，但同时也存在数据可能会混乱或只有少数活跃用户发表意见等问题。依据不同的系统目标、用户需求和使用情景，标记系统 (tagging system) 可以独立，也可与传统分类法结合，帮助用户完成标记、整理和分享信息的目标。从2008年本书的英文原版出版至今已有4年，越来越多的系统将标签和大众分类作为重要功能引入其中，这足以说明其价值。

本书的作者Gene Smith对标签这一概念的内涵和外延进行了系统化的、深入浅出的阐述。从什么是标签、标签有什么价值，到标记系统的架构和与其他分类系统的对比，再到标签的呈现方式和标记系统的实现细节，作者都用通俗易懂的语言进行了阐述，并附有详细的示例，书的末尾还附有具体的案例研究。本书具有很好的启发性，能够引发读者对于标签和大众分类法更为深入的思考；并且体

量适中、文字通俗，不会让人望而却步。因此，我们建议读者可以在忙碌的工作和学习之余，泡上一杯清茶，以轻松愉快的心情阅读本书。

本书的致谢、引言、第1~5章、第7章和附录由张军翻译，第6章由陈军亮翻译。由于时间和水平所限，书中可能会存在翻译不太妥当的地方，欢迎读者提出宝贵意见。

译 者

2012年3月

于大连海事大学

william@dlmu.edu.cn

前言

“当新鲜感逐渐消失……我认为标记（tagging）[⊖]会从根本上改变我们的信息图景。”

——Jon Udell

大约三年前，我在信息架构协会（Information Architecture Institute）[⊕]的邮件列表上问了一个简单的问题：“你们有些人可能没有注意到Furl、Flickr和Del.icio.us所提供的使用用户自定义标签来组织和分享信息的服务……对这种非正式的社会化分类应该起个什么名字呢？”这个问题唤起了大家对于标签、大众分类法（folksonomy）和其他种类的源于用户的元数据（metadata）的持久兴趣。

标记正在改变我们寻找、使用和分享信息的方式。正如Jon Udell所说，它会对我们的信息图景产生长久的影响。

本书记录了这些改变，是一本关于标签的指南：什么是标签、如何标记，甚至包括为什么标记。我尽可能多地使用现实世界的示例和研究。较之于抽象的概念，你会看到更多实际的应用。

本书的终极目标是帮助你理解标签和设计标记系统，以供你和你的用户使用。无论你是Web设计师、开发人员、信息架构师、用户体验开发人员，还是产品经理，都能够在本书中找到有用的概念和示例。当然，本书的重点是标记，但也包

⊖ 为了准确达意，译者对于tag和tagging的翻译作了如下处理：1) tag在作为名词出现时，翻译成“标签”，纯粹地指代标签这个事物；2) tag作为动词出现时，翻译成“标记”，表示添加标签的这种行为，意为“打标签”或“为……加标签”；3) tagging与动词性tag类似，含有动作意味，也翻译成“标记”。——译者注

⊕ 信息架构协会（<http://iainstitute.org/>）是一个信息架构方面的专业组织。——译者注

含一些信息架构、社会化软件和界面设计方面的讨论，希望这些内容对你也有所启发。

本书每一章涵盖了标记的某个方面的主题，从一般到专业，如下所示：

- 第1章介绍了标记，列出了由三部分组成的标记系统模型（这个模型在全书中都会用到），并解释了为什么标记很重要。
- 第2章探究了标记的价值：用户从标记中获得的体验回报和构建标记系统时可预期的投资回报。
- 在第3章中，你将了解到标记系统的架构。你为系统定义的规则和关系会影响到系统中的所有活动。
- 第4章的主题涉及标签充当元数据。你会了解到大众分类法，还有将标签与传统分类法以及其他分类系统相结合的方案。
- 第5章讨论导航与标签的可视化，包括深入探究当今无处不在的标签云（tag cloud）。
- 第6章将会展现标记界面的真相。你将会了解到加标签的模式、不同类型的标签推荐，以及标签管理的特性。
- 在第7章中，你将了解到标记系统的技术设计，包括数据模型、脚本和用以实现书中其余部分所述创意的其他工具。
- 本书的3个附录是案例研究，介绍特定种类的标记系统：
- 附录A介绍了社会化书签标注应用中的标记。你可以在其中看到简短但很有趣的Del.icio.us（第一个协同标记应用）的历史。
- 附录B是关于媒体分享的。介绍Flickr、YouTube、SlideShare与其他媒体分享网站的相似点和本质区别。
- 附录C的主题是个人信息管理中的标签，包括了诸如Microsoft的Photo Gallery和BlueOrganizer之类的示例。

我力图将最好的想法和示例尽可能详细地收录在本书有限的篇幅中，但是网站设计的变化是迅速而频繁的，在你读到这本书的时候，我相信书中用到的一些示例可能已经更新了。如果遇到这种情况，你可以将其当做一个机会来对比网站过去和现在的运行方式。你也可以顺便访问我的网站 (<http://genesmith.ca>) 来告诉我，以便我能够掌握这些改变的线索。

致 谢

本书很大程度上来源于我在信息架构和用户体验的会议上与朋友们的交谈或分享的演讲。我能够立即想到的朋友包括Thomas Vander Wal、Rashmi Sinha、Peter Morville和Peter Merholz，但其实还有很多其他的朋友。

感谢那些在本书撰写过程中接受我采访的人：James Melzer、Alex Wright、David Weinberger、Tim Spalding、David Millen、Mike Migurski、Peter Van Dijck和Timo Hannay。也非常荣幸能有Tom Coates、Moritz Stefaner和Jane Murison为本书所做的慷慨贡献。

很多人给予我鼓励来开始撰写这本书：Christina Wodtke、Lou Rosenfeld，特别是Liz Danzico阅读了最初的提纲，并帮我联系了Peachpit出版社。当开始写作时，我在nForm的同事们给了我所需的时间和空间。Jess McMullin和Yvonne Shek在我写作过程中让我们的业务保持蒸蒸日上，并充分信任我。我非常感谢他们。

ChrisFarnum 是本书的技术编辑。如果你在书中发现了错误或疏漏，那Chris一定是曾经试图让我修正这些问题。第7章的改进要感谢Jonathan Snook和我的同事Toby Spendiff发挥了自己的技术专长。Joe Lamantia和Johanna Dietrich阅读了本书的初稿并提出了很多极好的建议。

Peachpit(New Riders)出版社的团队（成员包括Michael Nolan、Becky Winter、Cliff Colby和其他许多我没有见过的人）在整个项目中的表现都非常出色。我非常高兴Jeff Riley凭借自己的丰富经验使我始终沿正轨前进，保持写作切题。感谢Jeff。

在1982年一个寒冷的日子，我的父母（Ellen和Bob）带回家一台Apple II计算机。那是非常重要的一天。感谢他们——因这件事，也因他们为我所做的一切。

我也非常感谢我的孩子们（Noah和Avery），他们具有与生俱来的灵感和富有感染力的激情。

最后，如果没有Jen Beverly，我将无所适从。她阅读了每一章，在我想要退缩的时候鞭策我，在我整夜撰写书稿之后允许我睡到很晚才起。这些都是她在怀孕和后来我们的小儿子Evan快要出生期间所做的（Evan是在此书写到大约第4章或第5章之间的时候出生的）。她是一个无与伦比的好伴侣。

目 录

译者序	
前言	
致谢	
第1章 什么是标记	1
1.1 标记是如何起作用的	3
1.1.1 标记系统的基础模型	4
1.1.2 当今的标记系统	6
1.1.3 标记的3个视角	12
1.2 标记为何重要	16
1.2.1 标记很流行	18
1.2.2 标记是多面的	18
1.2.3 标记是社会化的	18
1.2.4 标记是灵活的	19
1.2.5 标记适用于流	19
1.3 小结	20
第2章 标记的价值	21
2.1 标签能为你做什么	21
2.2 体验回报：使用标记的5个动机	23
2.2.1 易用性	23
2.2.2 管理个人信息	24
2.2.3 协作与分享	27
2.2.4 娱乐	28
2.2.5 自我表达	29
2.3 投资回报：商业上的7个益处	31
2.3.1 辅助协同工作	31
2.3.2 获取描述性元数据	31
2.3.3 增强可寻性	32
2.3.4 增加用户参与	32
2.3.5 识别模式	34
2.3.6 强化现有分类成果	34
2.3.7 激发创新	34
2.3.8 另有一事：让你的工作符合系统目标	35
2.4 小结	37
第3章 标记系统架构	39
3.1 用户、资源和标签：探索我们的三部分标记模型	41
3.1.1 第1部分：用户	42
3.1.2 第2部分：资源	46
3.1.3 第3部分：标签	50
3.2 标记实践：现实世界的例子	54
3.2.1 4个标记系统及其架构选择	54
3.2.2 5个常见的标记陷阱（及其解决办法）	55
3.3 小结	61
第4章 标签、元数据和分类系统	63
4.1 面向大众的元数据	64
4.1.1 三种元数据：描述型、管理型和结构型	65
4.1.2 作为元数据的标签	66
4.2 传统分类法和受控词汇表	67
4.2.1 受控词汇表	68
4.2.2 传统分类法	72
4.3 分面	76
4.3.1 理解分面	76
4.3.2 分面标记的两个具体实现：	

Buzzillions.com和Mefedia	77	6.2.3 动作分隔的系统	125
4.4 大众分类法	82	6.3 标签推荐	127
大众分类法的4个特征	84	6.3.1 三种推荐方式	127
4.5 元数据生态系统中的标签	86	6.3.2 推荐的价值如何	131
4.5.1 元数据垃圾问题	87	6.4 标签管理	132
4.5.2 中等规模的问题	90	6.4.1 编辑和删除标签	132
4.5.3 速度分层问题	91	6.4.2 批量编辑和拆分	132
4.5.4 一个生态学解决方案	92	6.4.3 管理还是忽略	133
4.6 小结	93	6.5 小结	135
第5章 导航与可视化	95	第7章 技术设计	137
5.1 标签云	95	7.1 数据模型	138
5.1.1 标签云：基础知识	96	7.1.1 简单标记模型	139
5.1.2 制作标签云	97	7.1.2 协同标记模型	140
5.1.3 扩展标签云	101	7.2 标签云	143
5.2 导航标签	104	7.2.1 按比例缩放	143
5.2.1 中心点浏览	105	7.2.2 线性缩放	146
5.2.2 热门度	107	7.2.3 分类缩放	149
5.2.3 过滤	109	7.3 FreeTag	149
5.3 地理标记	111	7.3.1 FreeTag基础知识	149
5.4 小结	116	7.3.2 FreeTag云	152
第6章 界面	117	7.3.3 使用FreeTag和Ajax产生标签 推荐	153
6.1 标记界面中的模式	118	7.4 小结	160
6.1.1 添加并标记资源	119	附录A 案例研究：社会化书签 标注	161
6.1.2 批量标记	120	附录B 案例研究：媒体分享	181
6.1.3 只标记已存在资源	121	附录C 案例研究：个人信息管理	193
6.2 标签输入	123	索引	203
6.2.1 快速与简洁性	123		
6.2.2 字符分隔的系统	124		

第1章 什么是标记

在本章中，你将了解到：

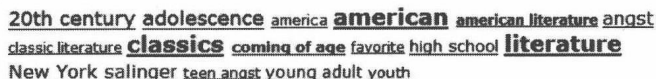
- 一个标记系统的基础模型。
- 5种不同的标记系统。
- 标记与个人信息管理、信息架构和社会化软件的关系。

步入美国康涅狄格州（Connecticut）丹伯里市（Danbury）的公共图书馆，你会发现堆满图书的普通书架排列齐整。小说作品按照作者姓氏的字母顺序排列。对非小说类的标题，则与北美成千上万的其他图书馆一样，按照“Dewey十进制”分类进行排列。

但是访问丹伯里图书馆的在线分类目录（<http://danburylibrary.org>），你就会发现它与典型的图书馆有所不同。

如果搜索《The Catcher in the Rye》，结果中不仅会出现一个图书编号，同时也会出现相关图书和标签（如“adolescence”、“angst”、“coming of age”和“New York”之类的关键词）的列表来描述J. D. Salinger的经典小说（见图1.1）。点击标签“angst”（焦虑），就会出现一个与焦虑有关的标题列表，例如，《The Bell Jar》、《The Stranger》和《The Virgin Suicides》。如果你查阅《The Stranger》，会看到诸如“existentialism”、“philosophy”和“French literature”的标签。

图1.1
《The Catcher in
the Rye》的标签



20th century adolescence america american american literature angst
classic literature classics coming of age favorite high school literature
New York salinger teen angst young adult youth

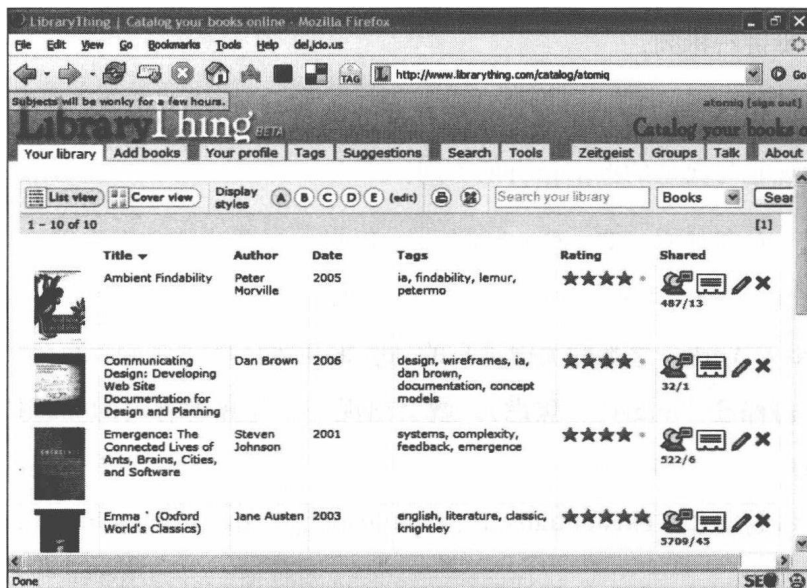
每个标签都是链接到另外一些书籍和思想的“跳板”。点击“existentialism”，会看到Kafka和Dostoyesvky的作品，而点击“French literature”则会看到Proust、Flaubert和Dumas的作品。每个标签都会根据不同的关系将书籍组合成一个集合。

如果看到Dumas的《The Count of Monte Cristo》，你可以尝试查看标签“revenge”。这样你就会看到Herman Melville的经典作品《Moby Dick》、Neil Gaiman的漫画小说《The Sandman》和Olivia Goldsmith的《The First Wives Club》，这是一个涉猎“鸡仔文学”^①的非常好的起点。你可以将这种在标签和图书中间来回跳跃的过程无限地延续下去。

这些标签并不是丹伯里图书馆馆员的功劳，而是LibraryThing（见图1.2）的功劳。LibraryThing是一个供超过20万来自世界各地的读书爱好者检索和标记（添加描述性关键词）自己个人书库的网站。LibraryThing的成员已经为1500万册书籍添加了近2000万个标签，使其成为北美的第二大“图书馆”。

图1.2

人们可以在上面分享个人图书收藏的网站：LibraryThing (<http://www.librarything.com>)



① 鸡仔文学 (chick lit) 指由女性撰写并且主要面向二三十岁的单身职场女性的文学作品。主要通过轻松幽默的方式讨论现代女性关注的话题。“chick lit”最初是美国大学里对“女性文学传统”这一课程的俚语表达。——译者注

LibraryThing是众多使用标签的网站和应用之一，标签是利用普通用户提供的关键词进行信息组织的一种新兴方法。像丹伯里图书馆那样的社区图书馆可以访问LibraryThing的数据来帮助读者发现和了解新书。

1.1 标记是如何起作用的

如果你是一个爱读书的人，可能会在LibraryThing上注册。你会注意到：在添加了一两本书之后，你还可以添加一些词语来描述每一本书。

假如你为《Pride and Prejudice》添加了一些标签，例如“Jane Austen”、“19th century”、“English”、“fiction”和“romance”，又为《Middlemarch》、《Catch-22》、《Mastering Regular Expressions》、《Don't Make Me Think》和其他一些喜欢的图书添加了标签。

那么，你现在完成了5件事情：

第一，你创建了一种方法来根据自己的语言和兴趣浏览自己的图书收藏。你可以给《Pride and Prejudice》添加诸如“19th century”和“romance”的描述性标签，但也可以使用“brooding male protagonist”、“happy ending”、“gift from Mom”，或者对你有用的其他任何描述。

第二，通过为图书分配多个标签，你建立了多个不同的途径来找到它。这实质上相当于将同一本书放在了两个书架上。当你下一次寻找妈妈的礼物（gift from Mom），或者有美好结局（happy ending）的故事时，你会找到《Pride and Prejudice》。

第三，你的标签成了为图书建立的群体标签池（pool of tags）中的一部分。标签将个人和群体的知识联系在一起，大家的标签合在一起构成了对每本图书的一种群体共同认识（“19th century”和“romance”最终成为《Pride and Prejudice》的两个最热门标签）。

第四，你的标签会成为一种将《Pride and Prejudice》与其他图书联系起来的潜在的新方法，或者，如果你使用了已经存在的通用标签，那么你也加强了已经存在的这种联系。

第五，你的标签会成为一些小“钩子”，可以将其他使用标签的网站上的信息“拉”到一起。可以用“romance”标签将来自于Technorati、Flickr、Del.icio.us和其他Web2.0网站的内容聚合在一起。

上述5方面特性的结合使标记成为一种独特而强大的信息组织工具。在本书中，你会看到可以如何使用标签来管理你自己的信息，也会看到标记的社会属性以及它是如何应用于在线社区中的。

在本书结尾部分，你将会了解设计一个标记系统的来龙去脉。你可以通过标记充分发挥用户的专业知识并加以利用。你也会看到一些有关设计社交应用的内容。

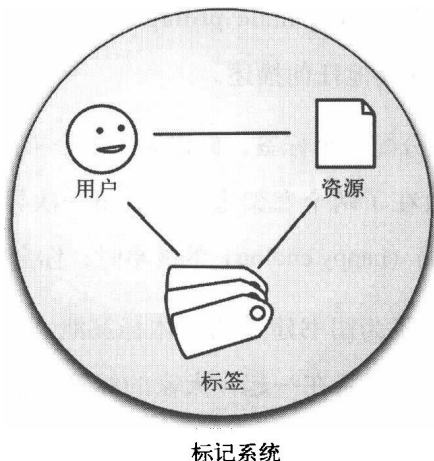
下面我们开始正式介绍内容。

1.1.1 标记系统的基础模型

可以将标记行为分解，从而形成一个相当简单的模型：用户在系统中可以对诸如照片和网页之类的资源加标签（见图1.3）。

图1.3

用户在标记系统
中给资源加标签



1. 用户

将使用标记系统的人称为“用户”（有时也会称其为“标记者”），他们创建标签，有时也添加资源。

虽然用户有各种各样的兴趣、需求、目标和动机，但有一点是确定的：他们不是因为标记特别棒或者很有趣才进行标记的，而是试图完成某个更大的目标——例如分享一张照片或者标出一篇文档以方便以后再找到它。加标签就可以完成这样的目标。

2. 资源

“资源”是用户贴标签的对象。资源可以是任何东西——一本书、一个网页、一段视频，甚至一个地点。只要我们有办法惟一地识别某事物，它就可以被标记。

在每个标记系统内部，资源通常有一些共同的属性。在LibraryThing中，资源就是图书。在照片分享网站Flickr (<http://flickr.com>) 中，资源就是照片。

3. 标签

用户添加的关键词就是标签。因为标记是不受限制的，所以标签可以是有关任何方面的：可以是对资源主题、所在位置、预期用途的描述、一种提醒，或者完全不同的其他东西。标签可以是单独的词（如“有趣”），也可以是短语（如“妈妈的礼物”）。不同的人有不同的标记特征：某些人做的标记富于表现力，而有些人的标记则只是描述性的。标记系统容许（甚至是鼓励）这种差异性。

标签本质上是关于资源的元数据（metadata）。根据美国国家信息标准组织（National Information Standards Organization, NISO）的定义，元数据是“用来描述、解释、定位某个信息源，或者使其易于获取、使用或管理的结构化信息”。换言之，标签是“关于信息的信息”。我们会用第4章一整章的篇幅来讨论标签这种元数据。