

Date : 28/06/2006



虚拟国际规范文档

连接德意志图书馆和美国国会图书馆的规范文档

里克·贝内特 (Rick Bennett)

OCLC 联机计算机图书馆中心
美国. 俄亥俄州. 都柏林市

**克里斯蒂娜·亨格尔-迪特里希
(Christina Hengel-Dittrich)**

德意志图书馆
德国. 美茵河畔法兰克福市

爱德华·T·奥尼尔 (Edward T. O'Neill)

OCLC 联机计算机图书馆中心
美国. 俄亥俄州. 都柏林市

巴巴拉·B·蒂利特 (Barbara B. Tillett)

国会图书馆
美国. 哥伦比亚特区. 华盛顿市

Meeting:	123 Cataloguing
Simultaneous Interpretation:	Yes
<i>WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL</i> 20-24 August 2006, Seoul, Korea http://www.ifla.org/IV/ifla72/index.htm	

摘要

德意志图书馆、美国国会图书馆和 OCLC 联机计算机图书馆中心联合开发了一个个人名称的虚拟国际规范文档(VIAF)，它连接世界各国国家书目机构的规范记录，并将免费在网上提供检索。本项目的目的是要证明自动连接不同国家规范文档

的规范记录之可行性，并演示其优点。美国国会图书馆和德意志图书馆的规范文档和书目文档被用来创建初始的 VIAF，包括 600 多万个名称和 50 多万个连接。本项目的关键方面是自动名称匹配算法的开发，它利用了规范记录和其对应书目记录的信息。我们显示了用算法建立各国家规范文档的个人名称之间连接的实用性，两个规范文档中共有的 70% 的个人名称被自动连接，错误率小于 1%。VIAF 项目的长远目标是要将许多国家图书馆和其他重要来源的规范名称结合起来，形成一个共享的全球规范服务。

引言

国际图联(IFLA)编目组内的一些小组认识到虚拟国际规范文档(VIAF)的可能性 [1]，要使得全世界各个国家书目机构所创建的代表同一个实体的规范记录连接起来，并在因特网上提供使用。这样的 VIAF 实际上会扩展世界书目控制的概念，并将其建立在各个国家书目机构的工作的基础上。它将允许规范形式的国家差异或地区差异并存，从而支持全世界用户对于自己所偏好的语言、文字和拼写方式的不同需要。

目前对万维网之未来的一些设想考虑使用本体来使得万维网更为智能化，可以用机器来自动处理。如果与其他来自文摘索引服务、档案馆、博物馆、出版社等来源的可控词汇和规范文档结合，VIAF 可以成为“语义网”的一个基本组成部分 [2]。图书馆现在面临一个很好的机遇，可以在这个方面作出较大的贡献，并促使这个前景成为现实。对于这个达成共识的前景来说十分重要的是，VIAF 应该是在网上免费提供服务的。

其他一些项目已经考察了规范文档中个人名称的连接问题。LEAF 项目（连接和探索规范文档） [3] 提出要将不同来源（包括图书馆、档案馆、文献和研究中心等）的规范记录连接起来。这些规范记录具有不同的格式，而且其类型的细节和内容的数量也有很大的差异。LEAF 项目设想要在记录载入系统时自动进行连接。由于名称规范记录的来源千差万别，他们发现可以供连接使用的仅有公共信息是名称以及相关参照和与名称关联的日期。由于现在的项目参与机构的名称规范记录经常不含日期，所以他们的名称规范记录的不匹配的错误率非常高，难以接受。

InterParty 项目[4]是一个欧盟资助的演示项目，它要在不同的机构之间创建连接规范文档，主要目的是支持数字版权管理。所设想的 InterParty 系统要提供该系统中各个数据库的统一的检索点，所以它一开始就提供一个集中的检索服务。由于各个数据库中的名称之间的连接是手工确定的，负责制作关联的个人可以输入连接信息。然后，系统可以自动使用这些连接。根据建立连接的机构的情况，系统会判断这些连接是否足够可信。一个参与机构确定的一个连接，未必会被系统中的其他机构接受。该项目允许算法匹配的可能性，但是没有明确表示支持连接能力所必须的技术和数据需求。

VIAF 项目

在 2003 年于柏林召开的国际图联世界图书馆和信息大会上，德意志图书馆(DDB - Die Deutsche Bibliothek)、美国国会图书馆(LC - Library of Congress)和 OCLC 联机计算机图书馆中心(OCLC Online Computer Library Center)一致同意开发一个个人名称的虚拟国际规范文档(VIAF)[5]。VIAF 项目之目标是证明自动连接来自各个不同的国家规范文档记录的可行性，并演示 VIAF 的优点。VIAF 项目将连接 LC 和 DDB 的名称规范记录，形成一个统一的虚拟名称规范文档。OCLC 开发软件，以匹配两个规范文档之间的个人名称规范记录。VIAF 项目的长远目标是连接来自许多国家图书馆和其他权威数据源的规范名称，形成一个共享的个人、团体、会议和地名的全球规范文档服务。

VIAF 项目分五个阶段：

1. 从 DDB 的个人名称数据库(PND - Personennormdatei)和 LC 规范记录建立“增强规范”记录。这项工作还包括确定合适的规范记录进入增强规范记录，并确定载入文件所需要的特殊处理需求。
2. 开发匹配算法，并匹配 PND 和 LC 增强规范记录，以创建 VIAF 的初始版本。这个过程与第一阶段是交替进行的，因为中间的匹配结果可以指出另外一些可以被抽取和用于增强规范记录的信息，并进一步改进匹配结果。
3. 构建一个 OAI 服务器[6]，提供 VIAF 的检索。

4. 维护 VIAF 数据库，要求参与机构添加和修改规范记录和书目记录。这一更新和维护系统将围绕 OAI 请求更新所使用的协议来进行设计。
5. 要检索 VIAF 记录，就需要一个万维网上可以公开访问的用户界面。最后，数据库和界面将支持 Unicode 和多种语言、多种字符功能。数据库的直接请求（例如提供一个 LC 版的名称并请求匹配的 PND 名称作为一个简单的 HTML 连接），可以用于支持语义网功能。

本项目最初侧重于演示 VIAF 的可行性，连接 PND 和 LC 名称规范文档 (LCNAF) 的名称规范记录。到 2005 年 12 月 31 日，LCNAF 包含 420 万条个人名称规范记录。截止到同一日期，LC 创建并对外提供了 930 万条书目记录。

到 2005 年秋天，PND 文档包含 260 万个人名称规范记录。PND 规范文档既用于 DDB 书目记录，也用于巴伐利亚图书馆联盟(BVB - Bibliotheksverbund Bayern) 的书目记录。在两个书目文档之间，有大约 1500 条书目记录与 PND 规范记录关联。

名称匹配问题

开始，VIAF 会作为一个德-英和英-德人名词典。例如，美国的用户检索 J. P. De Valk（LC 确立的名称形式），这个名称会自动“翻译成”Johannes P. De Valk（DDB 确立的名称形式）。正如这个例子所表示的那样，不同的国际编目机构确立名称的形式经常是不同的；反过来，他们也可能使用同样的名称表达不同的著者。很可能，DDB 所确立的名称 J. P. De Valk 完全是另一个著者。

个人名称可以对同一个人有不同的形式，或对不同的人有同一个形式，使得不同规范文档之间的难以可靠地匹配。这两个规范文档的覆盖面有很大的不同，只有一小部分的个人名称同时出现在两个文档中。所以，必须用除了名称本身以外的信息来保证可靠的匹配。在个人名称的规范记录中，经常有个人的生/卒日期。出生日期和去世日期的组合一般就足以区分同名的个人。

为了确认在没有补充信息的情况下匹配规范记录的难度，我们抽取了 LC 和 DDB 规范文档中共有的名称的样本。然后，我们人工判断这些规范记录对是否代表同样的个人。通过这一工作，我们发现，大概有 10% 的个人名称对是不同个人的名称。这样，仅用名称的确立形式的匹配错误率太高，不能接受。因为这两个国家规范文档之间的名称形式不总是相同的，对相似的（不是相同的）名称进行配对，会导致更高的错误率。同样，这个简单的方法也不能匹配确立形式不同的众多名称。

名称匹配的解决方案

很明显，我们需要额外的匹配信息来确认或拒绝可能的个人名称匹配。例如，我们考虑如下 Diane Glynn 的 LC 规范信息：

```
100 10 $a Glynn, Diane, $d 1946-
400 10 $a O'Connor, Diane, $d 1946- $w nna
670    $a Country western dancing, 1994: $b CIP t.p. (Diane
        Glynn) pub. info. (an avid country w. dancer & co-author
        of How to make your man more sensitive)
```

直接可用的数据只有名称和出生日期。有两个题名著录在 670 字段（发现数据的信息源），可以被抽取用于机器处理。实际上，只有某些题名可以可靠地从这些字段里抽取出来。

很明显，书目记录可以作为个人的附加信息。这些书目记录可以被用来作为个人作品的附加属性，以区别其他名称相似的个人。有一个书目记录如下：

```
100 1  $a Glynn, Diane, $d 1946- -
245 10 $a How to make your man more sensitive / $c by Diane and
        Dick O'Connor.
700 1  $a O'Connor, Dick, $d 1938- $e joint author -
```

书目记录包含两种类型的附加信息。书目记录通常有题名等作品层次的信息和 ISBN 等载体表现层次的信息。题名的匹配几乎可以确定支持名称的匹配。

书目记录还有附加的信息，可以应用于该个人的多部作品。这些信息可以帮助进行没有特定题名匹配情况下的著者匹配。关于此类信息，合著者 Dick O'Connor 是一个很好的例子。Dick O'Connor 与 Diane Glynn 合著了不止一部作品，这就有力地支持了规范文档之间的名称匹配。即使同一部作品同时出现在两个国家数据库

中，但是作品在一个数据库中是翻译本，题名匹配也很难自动完成。在这个情况下，合著者的名称在两个数据库中可能很相似，从而确认匹配的结果。

名称包含在主要款目、附加款目、或主题中的所有适合的书目记录经过转换，我们就创建了被称为“衍生规范”的中间记录。这些衍生规范记录与原始的规范记录一起组合，形成增强型的规范记录。因为增强规范记录包含与来自书目记录的名称关联的附加信息，它们可以支持比规范记录本身更严格的匹配过程。

名称匹配确认

简单比较两个国家规范文档的名称，是查找同一个人的合理方法。我们肯定会发现一些名称的变化形式，从而导致个人是同一个人的几率降低。为了自动确认这些个人的匹配，我们在这里采用的方法是：(1) 名称必须兼容，(2) 必须有足够的补充确认信息来确认匹配。

兼容性要求，不存在排除名称代表同一个人的区别。名称可以在完整性上不同，例如 John A. Smith 和 John Allen Smith。这两个名称是兼容的，因为“A”可以代表“Allen”。然而，John A. Smith 和 John B. Smith 肯定不是兼容的，因为其中间的缩写名冲突。在测试兼容性的时候，我们同时考虑名称的规范形式和变异形式。

一旦名称被确定是兼容的，有关这些名称所采集的补充信息就被用来确认匹配。书目文档可能包含许多不同的但相似的题名和许多不同但相似的名称。如果一个名称/题名对在两个文档中都是相似的话，那么很可能这个名称就代表同一个人。这个基本的策略可扩展到从书目记录采集到的其他类型的信息。

单独考虑日期，我们认为是正面的关系。如果日期差别超过一年，名称就被认为是不兼容的，匹配就被拒绝。我们仅允许一年内的日期差别。在 VIAF 的开发中，我们经常会发现某些日期的一些小差别，而附加匹配信息足以确认日期有微小变化情况下的匹配。

在比较两个增强规范记录时，匹配的每一个元素被认为是一个匹配点。匹配点被分为三类：强匹配点、中匹配点、弱匹配点。对于兼容的名称，一个强匹配点就足以确认这些个人是同一个人。强匹配点是题名、ISBN、生卒日期、合著者。仅

有出生日期不足以区分名称，它比较适合作为中匹配点。中匹配点还有个人的作品环境指标，例如所用的出版社、主题范围、个人的角色（例如，插图者或作曲者）。大出版社会出版许多著者的作品，至少有一些著者有相似的名称。多个中匹配点的匹配就足以确认一个匹配。弱匹配点只可以用来区分本来是模糊的匹配。弱匹配点的例子有：语言、主题范围、出版国等。

为了组合匹配点，每一个匹配点被赋予数值记分。对于一个号码（例如 ISBN），匹配要么是精确的，要么就是不匹配，从而记分就是 1（匹配）和 0（不匹配）。对于文字信息（例如题名），记分根据文字信息的相似程度，从 0 到 1。我们用一种三连记分技术。单个的记分被赋以基于长度的权重（强、中、弱），然后相加。如果总记分超出测试过程中确定的阈值，匹配就被确认。在实际的匹配算法中，我们要通过许多记录的测试，允许在这些类别中进行记分的调整；随着更多的规范文档进入系统和我们经验的增加，还会有更多的调整。

建立增强规范记录

上面所描述的技术用来从 PND 和 LC 名称规范记录来创建增强规范记录。LC 书目文档经过处理用于衍生规范文档并增强 LC 规范文档，DDB 和 BVB 书目文档经过处理来增强 PND 规范记录。图 1 简要说明产生增强规范记录的信息流。

对于增强的 LC 规范文档，420 万条规范记录中有 380 万条(90%)可以被增强。只有 260 条记录(60%)用书目记录中的信息得到增强，书目记录总计有 740 万条。其他增强方式还包括从规范记录 670 字段（发现数据的信息源）抽取的 410 万个题名。题名是产生匹配的最重要的增强元素，我们将会从结果部分看到这一点。

对于 PND 规范文档，260 万条规范记录中有 240 万条(90%)得到了增强，但是只有 200 万条(80%)从书目记录增强。其余的 40 万条记录是通过从 PND 规范记录本身抽取的题名得到增强的。

测试匹配技术

VIAF 参与机构支持匹配过程的开发，进行精确性评估和结果鉴定。例如，开始使用了丛编题名，但是后来发现经常错误匹配名称。每一次评估导致一些修正，

增加了匹配的数量，或者减少了错误的匹配。在这段时间内，合理的精确阈值分和记分算法也形成了。我们只是将最后确认的测试介绍一下。

为了确认匹配过程的精确性和有效性，DDB 和 LC 的资深规范编目员对名称匹配样本进行了评估。第一个样本有两个目的，就是确定两个规范文档之间的名称的重合，并发现这些名称对的哪些部分可以通过匹配过程识别。第二个样本用于寻找可以被修复的系统性错误和缺陷并估计总体错误率。

第一个样本有 391 条随机选择的 PND 规范记录。我们自动和手工检索 LC 规范文档，寻找它们的匹配记录。在自动检索的过程中，样本中的 PND 规范记录与所有 LC 规范记录中有共同姓氏的记录配对，产生出 74,000 对供核查的记录。将匹配算法应用于所有 74,000 个名称对，自动匹配了 79 个 PND/LC 规范记录对。

对 391 条 PND 规范数据进行手工评估后，发现还有 35 个名称有对应的 LC 规范记录，但是自动配对过程中，通过姓氏的配对没有发现它们，或者匹配算法没有确认这些匹配。通过这个手工评估，79 个匹配被确认是精确的。使用 PND 样本，我们估计大约有 30% 的 PND 名称也出现在 LC 规范记录中，并且算法可以匹配这些共同名称中大约 70% 的名称。这样，我们估计在两个规范文档中有 80 万个共同的名称，通过自动匹配过程可以识别 55 万个名称。

我们还评估这些结果，以改进名称配对过程。如果只用姓氏，我们要完成每一个匹配，就需要大约有 1000 个名称对需要经历完整的匹配过程。我们用手工匹配测试结果确定，基于姓氏、名字、加上限制日期信息的策略，基本上可以用来粗略估计名称的兼容性。我们发现，这个简单的指标可以确定 95% 的匹配，每完成一个匹配，只需检查 4 对名称。这个指标十分有效，进行一些小调整以后，就可能进一步改进结果。

第二个样本的目的是估计匹配错误率。作为全过程的一部分，该样本测试初步阈值分的合适性，并作必要的调整。在使用阈值分的时候，在接近记分阈值时匹配的错误率高于记分远远大于阈值时的匹配。大多数匹配名称规范记录的记分远远在阈值之上。为了实现最少的人工评估获得最佳的错误率估计，我们根据记分将样本分为四个子样本。手工的评估确认匹配错误，并确定每个子样本的错误率和置信

度。这些部分结果加权并求和，从而找出匹配技术的总体错误率。错误匹配的数量少于一个百分点。

有一个子样本看上去刚好在阈值之下。如果阈值再降低，每三个正确的匹配就会附带一个错误的匹配。我们显然不能再降低阈值。在刚好位于阈值之上的记分范围内，25个匹配中只有一个坏匹配。因为在这个范围内匹配数量相对较少，所以它对错误率的总体影响很小，许多正确的匹配被保留了下来。所以，初步阈值被接受。

构建初始 VIAF

两个来源的增强规范文档要经过匹配算法，产生的记录（无论是否匹配）都要转换成 VIAF 记录。该过程如图 2 所示。在产生的 VIAF 文档中有 630 万条记录，也包括 50 万连接了的记录、370 万条来自 LC 规范文档的不匹配记录以及来自 PND 规范文档的 210 万条不匹配记录。这与根据手工测试而作出的估计十分接近。我们估计，还有 25 万个代表同一个人的规范记录对由于缺少可用信息而不能自动匹配。最终的系统将允许对此类匹配进行手工连接，并允许其他脑力识别匹配。规范记录将包括一个 VIAF 序列号。

图 3 是一个 MARC21 格式 VIAF 记录的例子。因为 VIAF 的主要目的是提供文档之间的连接，VIAF 记录在 700 字段（标目连接款目）中包含每个名称的一个款目，并说明其来源。因为没有唯一的规范名称，我们不用 100 字段（个人名称标目）。当算法确定一个匹配以后，两个连接款目被加入记录中。如果名称不匹配，只有一个 700 字段出现。

补充信息包含在增强规范记录中，作为本地字段(9xx)。图 4 简单描述了增强规范记录中所用的本地字段。为了简化匹配，所有文字信息用 NACO（合作编目计划中的名称规范合作计划）的正规化规则的修订版本进行正规化处理。[7] 特定词的出现次数被记录在 \$9 子字段中。由于这个信息最初是为机器处理而准备的，它不必出现在终端用户的记录显示中。随着以后更多国家规范文档的加入，它们将首先与现有的增强 VIAF 记录进行比对，并在进入 VIAF 记录时纳入更多的匹配。当匹配完成以后，来自匹配记录的增强信息也被合并。

在许多例子中，一个文档中的规范名称匹配另一个文档中的多个规范名称。因为 VIAF 的目的是一对一连接服务，如果有多个匹配出现，匹配是不能被确认的；因为多重匹配的原因，有 7 万个算法匹配被排除了。根据我们的分析，多重匹配至少有两个原因。

第一，在 PND 中有许多未区分的名称，每个这样的名称都与 LC 名称规范文档中两个或更多的区分的名称匹配。根据德国编目规则 RAK-WB，过去德国的编目实践是不区分个人名称的。当 DDB 开始用规范文档进行编目时，这个做法被放弃，DDB 不再创建没有区分的个人名称规范记录。然而，PND 仍然包含许多过去没有区分的名称。DDB 将根据增强规范记录中著录的 LC 和 DDB 匹配的题名，尽可能自动区分有多重匹配的名称，其余部分用人工进行处理。纠正的部分将作为定期更新进入 VIAF，并在匹配的记录之间建立明确的连接。

第二，许多 LC 规范记录反映了 AACR2（《英美编目条例》第二版）的做法，用不同的规范记录来表示同一个人的每一个书目身份（例如笔名），这与 PND 的未区分记录是完全相反的。在这种情况下，LC 对同一个人建立了多个规范记录。根据 RAK-WB 规则，PND 对一个人的所有身份只建立一条规范记录。与没有区分的名称一样，这些“过度区分”的规范记录也产生了问题，但是没有完全另人满意的解决办法。

连接的名称可以直接用于 LC 名称规范记录和 PND 名称规范记录之间的自动翻译。这可以支持语义网或联邦搜索的需要。维护“参见自”根查，可以向用户提供更多的信息。

参与机构的规范文档的规范号或 VIAF 本身的规范号也可以作为 URI（网址）的基础。这提供了今后规范数据网址统一服务的可能性。只要从文献、记录或网站中出现的任何一个 URI 引用出发，用户就可以被指引到 URI 所代表的规范相关的所有资料、记录、资源等等，以及规范记录本身。

进行中的系统

国家名称规范文档和书目数据库是不断变化的。对于一个建立于两个或更多变化中的文档基础上的数据库，连接必须经常重新评估和更新。初始 VIAF 系统的设

计思想和软件正在修改，以便允许记录的连续更新。随着新的书目记录或规范记录不断接受，现有的增强规范记录将被修改，而且跨库的匹配也将重新评估。新的匹配会不断建立，由于数据源记录的变化而不再支持的匹配将被断开。匹配断开以后，先前的匹配历史还将被保留在每一个匹配过的记录中以供参考。

如果可行，今后的 VIAF 系统将利用源数据库拥有者的 OAI 馈源(feeds)。同时，FTP 等较为传统的文档存取手段将会用于项目的测试。

在同一个地点如果有大量的数据，我们可以设想许多不同的方法来存取和使用数据。作为语义网的一部分，连接可以用来把个人名称翻译成终端用户所希望的格式。我们还可以建立一些工具，以支持可选数据库的检索，提供该数据库适合的名称形式。同样，我们还可以构造一些编目和规范控制的工具，标识记录中所含名称的适当形式。当然，VIAF 数据库也是可以直接检索的。

结论

PND 文档已经从这个项目获得了很多好处。在两个文档中的自动匹配测试，使得 PND 能有大量的更新，DDB 期望能通过增强记录对中的匹配题名使得个人名称的区分得到更多的支持。为本项目开发的匹配过程和算法也适用于许多其他应用。我们正在考察有关服务，利用个人名称匹配数据以改进书目信息的存取，并支持参与机构的编目活动。

本项目演示结果表明，自动连接两个国家规范文档的个人名称是可操作的。两个文档中共有的人的规范记录中有 70% 得以连接，错误率小于 1%。用书目记录中的信息补充初试规范记录的策略大大改善了匹配率，减少了错误匹配的数量。规范记录的微小变化也会大大改进匹配。许多失败的匹配是由 670 字段（发现数据的来源）的语法分析失败而产生的。附加结构、避免使用简略的名称和题名、或者直接向源书目记录的连接是十分有用的。直接标识共同的角色或专业（作曲家、插图画家、数学家等）将进一步增强匹配，不管是自动匹配还是手动匹配；正如名称的完整形式（至少作为参照）能有助于匹配一样。

本研究提出了一个令人信服案例，涉及规范控制和规范记录的使用，涉及联网和连接，也涉及为图书馆建立语义网。对于在德国获得或拥有 LC 名称规范文档检

索点的图书馆和图书馆网络，VIAF 可以作为作为一个跨越规范文档的平台，或者将书目数据中的 LC 名称规范文档检索点转录成 PND 检索点，或者实现通过 VIAF 用 PND 标目进行检索。如果将其实施到诸如“欧洲图书馆”(The European Library)门户那样的多国或多语言的门户，VIAF 可以在 LC 名称规范文档和 PND 中自动组合检索语句，将用户指引向两个数据源的相关书目记录。

用现在的匹配技术，我们计划做一个可更新的系统，使用 OAI 馈源来从参与机构的现有个人名称规范数据和书目数据中采集信息。本系统在设计时就考虑了可扩展性，新的参与机构可以共享他们的规范记录和书目记录。由于还没有更多的机构参与 VIAF 项目，我们尚不清楚它在可扩展性方面有什么的局限。

VIAF 项目一直侧重于规范记录的匹配问题。要维护、扩展和实施 VIAF 项目，我们需要长期的服务和管理策略。我们需要做出一些决策，涉及项目的扩展，涵盖团体名称，增加新的参与机构。我们已经计划采纳 Unicode 字符集，以扩展系统的功能。Unicode 可以处理非罗马字符，但是扩展匹配算法也是一个挑战，特别是对于中、日、韩等表意的文字尤其困难。

参考文献

1. IFLA Core Activity: IFLA-CDNL Alliance for Bibliographic Standards (ICABS) <http://www.ifla.org.sg/VI/7/icabs.htm> [May 2006]
2. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities." *Scientific American*, May 17, 2001. <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> [May 2006]
3. LEAF Project, <http://www.leaf-eu.org> [May 2006]
4. Project InterParty: From Library Authority Files to E-Commerce, Andrew MacEwan, http://www.haworthpress.com/store/E-Text/View_EText.asp?a=3&fn=J104v39n01_11&i=1%2F2&s=J104&v=39 [May 2006]
5. VIAF: The Virtual International Authority File, <http://www.oclc.org/research/projects/viaf> [May 2006]
6. Open Archives Initiative - Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html> [May 2006]

7. Hickey, Thomas B., Jenny Toves, and Edward T. O'Neill. "NACO Normalization: A detailed Examination of the Authority File Comparison Rules", *Library Resources & Technical Services*, Vol. 50, No. 3, p. 18-24. [forthcoming]

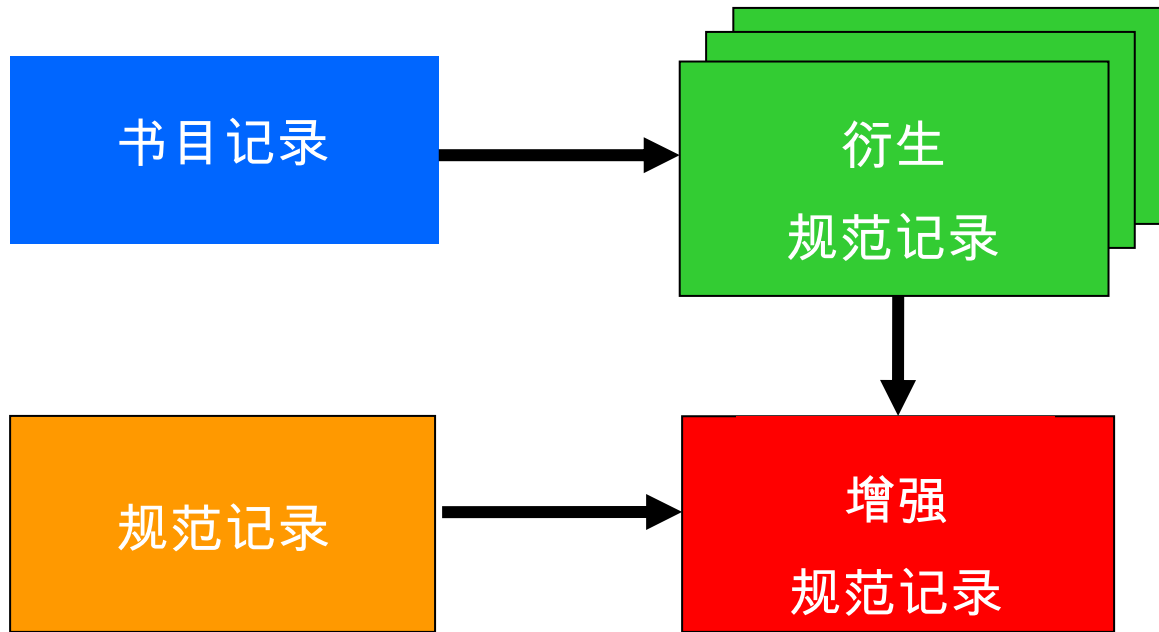


图 1：创建增强规范记录

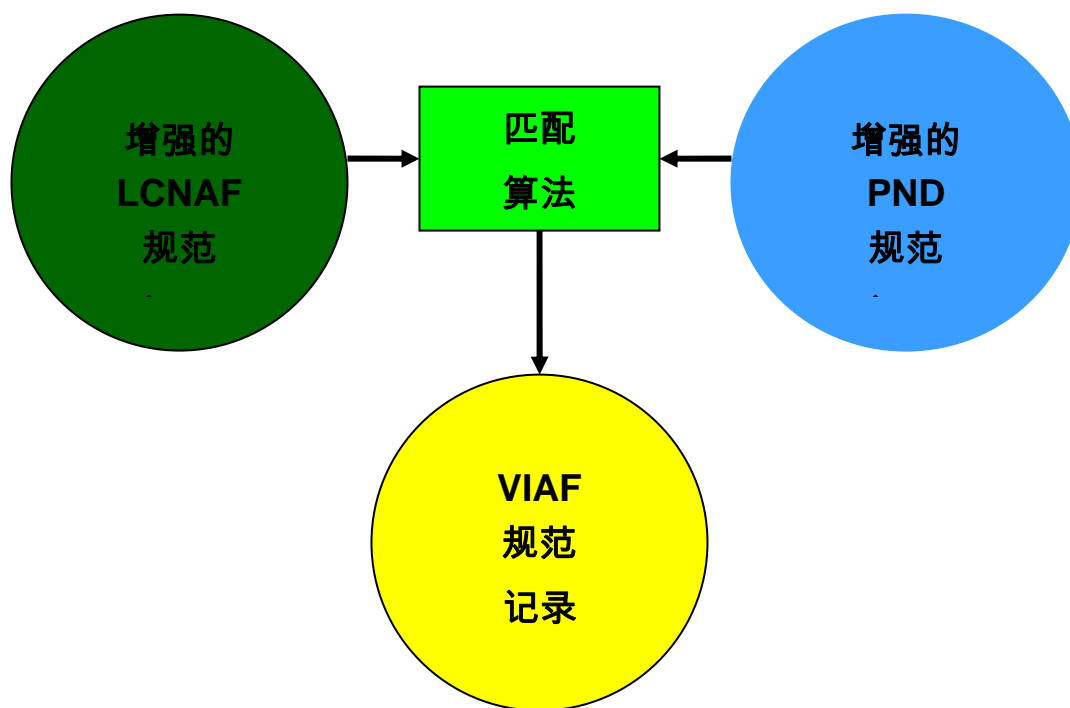


图 2 : 创建 VIAF 规范记录

```

000    nz n
001    viaf 30543
005    20050826163535.0
008    050826n||anannabbn |a aaa
040    VIAF    $c VIAF
400 10    $w nnaO'Connor, Diane,    $d 1946-
700 17 Glynn, Diane,    $d 1946-    $2 DLC    $0 n 94057411
700 17 O'Connor, Diane    $2 DDB    $0 108982424
901    052512920    $9 1
901    349917275    $9 1
901    350215532    $9 1
903    75014386    $9 1
910 11 how to make your man more sensitive    $9 3
910 11 macht eure manner zartlicher    $b liebevolle ratschlage
fur e neues rollenverhalten    $9 1
910 11 macht eure manner zartlicher    $b wie e frau ihrem mann
helfen kann e verstandnisvoll    $9 1
919    country western dancing,    $9 1
920    0-525    $9 1
920    3-499    $9 1
920    3-502    $9 1
921    dutton    $9 1
921    rowohlt    $9 1
921    scherz    $9 1
922    gw    $9 2
922    nyu    $9 1
940    eng    $9 1
940    ger    $9 2
942    18    $9 1
943    197x    $9 3
944    am    $9 3
950 11 oconnor, dick    $9 2
950 11 oconnor, dick    $d 1938    $9 1
999    1    $b 75014386 //r94    $2 DLC
999    1    $b n 94057411    $2 LoCNA
999    2    $b 780147766    $b 790425319    $2 DDB

```

图 3 : VIAF 记录

图 4：增强记录格式

90x 控制号

901	ISBN	\$a 子字段：ISBN 的数值部分 (没有校验码或分隔符)
902	ISSN	\$a 子字段：ISSN 的数值部分 (没有校验码或分隔符)
903	LCCN	\$a 子字段：LCCN 的数值部分 (没有校验码或分隔符)

91x 题名字段

910	245 中的题名	\$a 和 \$b 子字段
911	210 中的缩略题名 130 或 240 中的统 一题名	\$a 和 \$b 子字段
913	242 中的翻译题名	\$a 和 \$b 子字段
914	243 中的集合统一 题名	\$a 和 \$b 子字段
915	246 中的变异题名	所有子字段
916	规范记录统一题名	\$a 和 \$b 子字段
917	从名称/题名规范记录 100 字段 \$t 子字段抽取 从其他文字信息抽 取的题名	从名称/题名规范记录 100 字段 \$t 子字段抽取
919	各种附注或类似字段	各种附注或类似字段

92x 出版者字段

920	出版者号码	\$a 子字段：ISBN 中的出版者代码 \$a 子字段：260 字段 \$b 子字段或 533 字段 \$c 子 字段中的出版者名称
921	出版者名称	\$a 子字段：008 字段中的 出版者名称
922	出版地	出版国代码

93x 用法

930	名称用法	\$a 子字段：责任者说明中的名称形式，245 字段 \$c 子字段
-----	------	------------------------------------

94x 属性

940	语言	\$a 子字段：008 字段或 041 字段 \$a 子字段的语言代码
941	著者的角色	\$a 子字段：700 字段 \$e 和/或 \$4 子字段的的责任方式代码
942	NATC 主题	\$a 子字段：NATC 普查行号
943	出版年代 (10 年)	\$a 子字段：出版年代 (10 年)
944	格式	\$a 子字段：类型和书目级别(008/06-07)
945	大纲主题	习惯用法，参见 PND 的讨论

95x 合著者

950	个人著者	100 或 700 字段中的 \$a, \$b, \$c, \$d, \$q 子字段
951	团体著者	110 或 710 字段中的 \$a 子字段

96x 名称主题

- | | | |
|-----|------|--------------------------------------|
| 960 | 名称主题 | 600 字段中的 \$a, \$b, \$c, \$d, \$q 子字段 |
| 969 | 主题用法 | 文字“主题”表明规范标目用做主题，从 600 字段抽取。 |

99x 特殊字段

- | | | |
|-----|---------|-------------|
| | | \$a : 记录总数 |
| | | \$b : 记录控制号 |
| 999 | 关联的书目记录 | \$2 : 记录来源 |

中文翻译：顾犇（中国国家图书馆）
Chinese translation by Ben Gu (National Library of China)