业务流编排中基于RESTful的半自动派生

摘要

企业为了向其他组织提供合作，为其提供复杂的产品和服务市场。信息促进了不同组织之间的这种合作与协调，获得了很好的技术分享.BPMN流程编排是一种建模语言，用于指定业务级别上不同组织之间的信息和服务交换。最近，REST体系结构风格在网络上提供服务的使用激增，但很少有系统的工程方法来设计它们的协作。在本文中，我们通过定义用于从过程编排中推导出RESTful编排的半自动方法，以全面的方式解决这个缺陷。该方法基于自然语言分析技术来从过程编排中的文本信息导出相互作用。所提出的方法在效率方面进行了评估，从而导致网络工程师仅对所有生成的RESTful交互作用的约10％进行干预。

关键字:业务流程编排 RESTful编排 自然语言分析

# 1 引言

传统上，BPM的研究集中在组织的内部流程上。 更复杂的服务趋势将BPM扩展到多个流程之间交互的视角。 这种由信息技术支持的交互需要标准模型，如BPMN [1]，所有参与者都可以理解。 特别是，BPMN业务流程编排指定了两个或多个参与者之间的交互以及这些交互在业务级别发生的顺序。 在技术层面上，REST [2]正日益成为Web上提供服务的首选架构风格，从而导致RESTful API的主流开发。

涉及具体网络交互的RESTful API的设计应基于通过编排图建模的业务交互。但是，将业务交互降低到RESTful交互级别具有挑战性。编排图的设计者通常是业务流程领域的专家，并且不具备软件开发的知识。网络工程师在业务流程编排方面也是如此，例如，在设计URI时，他们通常与实现的角度太接近，这使得RESTful API的客户端更难以理解URI [3]。在[4]中描述了解决这个问题的第一个概念，然而，概念评分和评估存在局限性。

在本文中，我们以全面的方式解决了这一研究缺陷，并提出了一种将标准BPMN编排图作为输入的方法，并将其作为输出生成RESTful编排[5]。 RESTful编排是一种非常有用的[6,7]基于BPMN的建模语言，专门针对不同业务参与者之间的RESTful交互规范。我们的方法基于自然语言处理技术，该技术使用编排任务的文本描述来映射到具有相应REST URI的最合适的REST动词。我们在研究原型中实现了我们的方法，并将其应用于来自不同领域的一组编排图。派生的REST请求也由REST专家进行了评估，确认了我们方法的有用性。创建的RESTful编排用于派生有助于开发REST API的代码框架。应用我们的方法意味着所有参与者都使用相同的REST接口生成逻辑，这有助于更好地理解，维护和演进参与者的RESTful API。

这种方法扩展了[4]中的工作，提供了一个增强型标签分析方法和一个更大的评估数据集的扩展评估。这种扩展可以在正确生成RESTful交互的过程中产生更高的效率，其中仅需要大约10％的时间来进行网络工程师干预。

本文组织如下。第2节介绍了编排图和RESTful编排图。这些概念通过一个运行的例子来说明。第3部分介绍了我们的派生RESTful编排图的半自动方法。第4节讨论了用户评估的设置和结果。第5节提供了第Sect节前的相关工作。 6结束本文并介绍了未来的工作。

# 2 预备

本节通过示例的帮助简要描述编排图。 此外，在引入RESTful编排图的概念之前，将解释REST架构风格。

## 2.1 编排图

BPMN 2.0 [1]中介绍的业务流程编排图是一种建模语言，它侧重于描述两个或更多参与者之间的交互作用，他们通常是企业主角，例如企业，客户或组织。 与业务流程模型相比，编排图抽象出参与者的内部流程，并指定参与者之间交换消息的顺序。

图1描绘了一个编排图的例子。 该图描述了参与提交，审核和组织过程的不同参与者之间的相互作用，以安排科学会议。 会议中的一些主要利益相关者包括组织者，作者和评论者。 该图描述了这三位参与者之间的相互作用，这些参与者从发表电话会议（CFP）开始，最终结束，确认纸质出版物。

为了促进这些互动，参与者使用了一个评论管理系统（RMS），在我们的案例中，这个系统受到了http://easychair.org的启发。 RMS负责在整个协作过程中协调这三名参与者。

如图1所示，编排图的主要元素是编排任务（图形描绘为圆角矩形）。它代表两个参与者之间的消息交换。发起消息交换的参与者称为发起者，而另一个参与者称为接收者。返回消息是可选的，可以从接收者发送到发起者。为了以图形方式区分发起者和接收者，后者总是以灰色突出显示。这同样适用于启动和返回消息，尽管消息不需要以图形方式描述。

编排图定义了进行交互的顺序。编排任务具有通过顺序流建模的顺序依赖性。序列流，事件和网关以与常规BPMN业务流程模型中类似的方式使用。但是，在编排图中只能使用事件和网关的一个子集。图1中的事件是开始事件，计时器事件最后期限和三个结束事件：拒绝论文;短论文下降;论文接受发表。

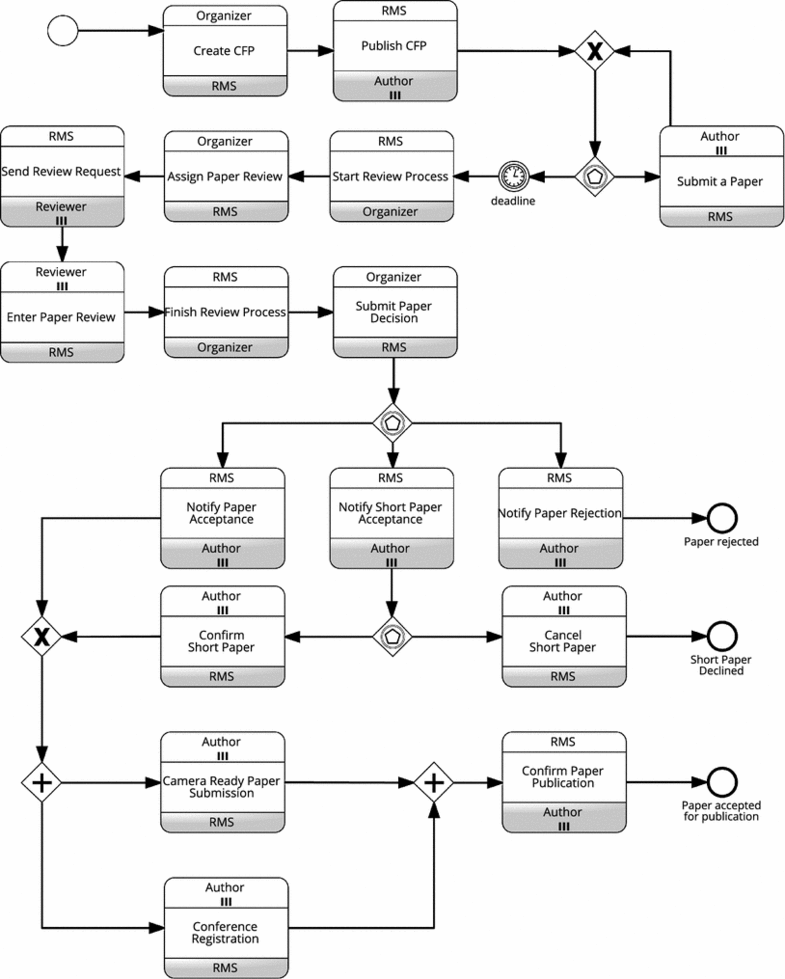


图1 提交论文和审查管理的编排图

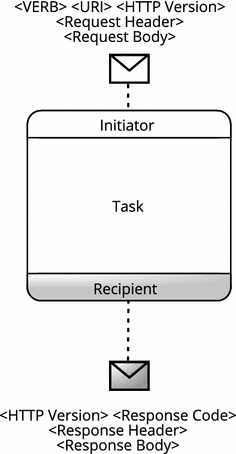


图2 RESTful编排中编排任务的注释[5]

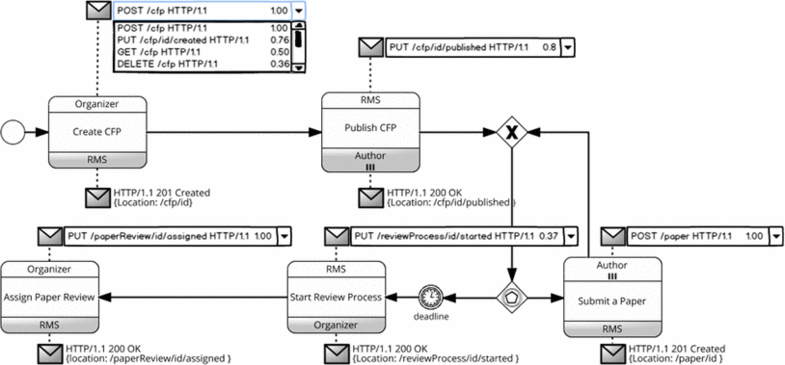


图3 RMS生成的RESTful编排的一部分

## 2.2 RESTful编排图

REST架构风格[2]越来越多地用于开发RESTful Web服务。其架构限制有助于提高可扩展性和可移植性。在几乎所有情况下，REST都使用HTTP协议作为不同参与者之间交互的手段。交互是通过对资源使用标准HTTP动词（GET，POST，PUT，DELETE）来实现的。驻留在服务器上的资源通过URL全局唯一地标识。客户可以通过这些REST动词改变其状态。由于无状态约束，服务器不需要记住之前与客户端的交互以了解客户端的请求。

RMS例子的假设是所有参与者都是RESTful服务，即他们通过发送REST调用来相互交互。在简单的浏览器设置中，组织者，审阅者和作者是RMS的用户。在这个例子中，我们假设他们也提供了一个RESTful API。将这些服务视为RMS移动应用程序，其中RMS可根据用户角色推送通知[8]，例如，向审阅者通知分配给审阅的论文，将论文决定发送给作者，通知组织者何时提交所有审阅。

业务流程可以用来模拟[9]中提出的参与RESTful会话的参与者的内部行为。但是，当涉及到多个参与者之间的交互时，重要的是关注全球视角，以便捕捉到公共资源的状态以及与这些资源的交互。

为此，Nikaj等人[5]介绍RESTful编排图 - 一个轻量级的BPMN编排图，增加了REST细节。这些细节包括表示RESTful交互的编排任务的注释，称为RESTful任务，如图1中的Submit paper decision编排任务。这是通过将两个编排任务消息分别改进为REST请求和一个REST响应，如图2所示。图3提供了一个RESTful编排图的摘录，模拟组织者，RMS和作者之间的RESTful交互。但是，负责使用REST符号增强编排任务的人员必须了解编排的业务方面和RESTful交互的实现方面。这个问题在我们的论文中提出，提出了一种从业务流程编排中派生RESTful编排的半自动方法。

# 3 RESTful编排的半自动生成

本节介绍我们用于生成RESTful编排的半自动方法。 第3.1节讨论了该方法的相关概念。 第3.2节花费在上一节中，侧重于编排特定的标签风格。 然后第3.3节将解释如何使用这些概念来标识在编排任务标签中表示的REST请求的类型。 3.4节展示了如何通过RESTful信息最终丰富编排任务。

## 3.1 基础

本小节从编排图的正式说明开始，作为我们方法的核心工件。 我们使用编排图的所有元素来生成RESTful编排图。 我们认为编排图是一个元组

C =（N，S，P，L，label），使得：

-N =T∪E∪G是一组节点。 T，E和G是成对相互排斥的;

T是一系列的编排任务;

E是一组事件;

G是一组网关;

S⊆N×N是一组序列流;

P是一组参与者;

L是一组自然语言文本标签;

label：T↦L是一个将文本标签分配给编排任务的功能。

正如第2节所指出的，编排任务可以表示单个消息交换或两个消息交换（即发送消息和回复消息）。 在我们的方法中，我们将每个案例映射到单个REST式交互，因为发起者向收件人发出REST请求。 由于根据BPMN规范[1]，可选的第二条消息是返回消息，因此我们不认为它是新的REST请求，而是来自HTTP中嵌入的接收方（REST术语中的服务器）的响应体（见图2）。 此外，我们观察到，受自然语言处理影响的编排任务标签与其相应的业务流程发送任务相比，更接近于接收任务。 这意味着编排任务标签不提供有关如何命名第二个REST请求的信息。

为了处理标签的文本信息，有必要以结构化的方式访问这些信息。 作为一个起点，我们观察到，如上所述，编排任务被类似地标记为活动，通常指业务过程模型中的相应发送任务[1]。 因此，我们假设编排任务的每个标签都包含以下组件：一个动作和一个应用动作的业务对象[10]。 例如，考虑图1中的提交文章决策的标签。它包含要提交的操作和业务对象文章决策。 重要的是要注意，这些组件可以用不同的语法变体进行通信。 例如，标签相机准备好的文章提交通过使用表达要提交的动作的名词在不同的语法结构中传达动作。

为了独立于语法标签结构，我们依赖Leopold等人的标签注释方法[11]，它确定了相当程度的准确性（平均精确度：91％，平均召回率：90.6％）的行为和业务对象。 考虑到l = label（t）∈L是任意编排任务的标签，并且分别考虑WV和WN来描述所有动词和名词的集合，我们参考l的动作和业务对象如下：

α：L↦WV是一项将动作分配给编排任务标签的功能;

β：L↦WN是将业务对象分配给编排任务标签的函数。

作为一个例子，考虑从图1中提交文章决策标记的编排任务。根据之前的概念化，该行动由α（提交文章决策）=提交并且业务对象由β给出（提交文章决策 ）=文章决定。 我们将在下面使用这些标签组件来派生各自的REST请求，并从编排任务的文本标签中生成RESTful注释。

## 3.2 编排特定的标签分析

在前述的标签处理方法中，我们观察到关于编排任务和业务流程任务的相似性的假设并不总是成立。由于编排任务代表跨组织发送的消息，因此标签倾向于包含发送单词或类似的单词，如enter，submit，mail，它描述了从一个编排参与者到另一个参与者的消息传递。这些单词后跟单个或多个单词，例如，order, invoice, request, application request, paper submission, short paper acceptance。由于所有REST请求都是发送消息，因此send这个词表示对任何特定REST动词没有语义偏好。例如，可以发送用于检索资源表示的GET请求或用于删除资源的DELETE请求。因此，我们将注意力集中在包含单词send（或类似单词）的标签的其余部分，以寻找一些额外的语义信息，这些信息可以指示标签对特定REST动词的倾向。

在检查标签是否包含发送动作或类似内容之前，我们需要确定可用于编排中的动词，以表示正在发送的消息。 表1中提供了同义词和语义上类似的发送动词。下一节将介绍我们填充此集的方式。

表1 REST动词的同义词词组和发送

|  |  |  |
| --- | --- | --- |
| 动词 | 描述 | 同义词词组 |
| POST | 在服务器上创建新资源 | SynPOSTSynPOST = {{create, request, [produce, make, ...] } |
| PUT | 编辑存在的资源 | SynPUTSynPUT = {{confirm, edit, accept, send [support, redact, ...] } |
| GET | 从服务器检索现有资源 |  |
| DELETE | 删除现有资源 |  |
| Send | 从一个编排参与者向另一个编排参与者发送消息 |  |

如果标签有一个作为SynSend元素的动作，我们将分析标签的其余部分。我们对其余部分重新应用标签注释器（来自上一小节）以标识新的操作和业务对象。例如，标签发送审阅请求，通知拒收文章，通知短纸接受被重新分析为审查请求，拒收文章和短纸接受。应用相同的批注方法将产生以下操作 - 业务对象对：请求审查;拒绝纸;接受短纸。

但是，如果没有新的行为和业务对象，我们会像处理其他任何动词一样对待这个标签。例如，在标签发送评论中，我们有α（发送评论）=发送和β（发送评论）=评论。在这种情况下，我们将发送映射到REST动词PUT以表示发送的资源处于“已发送”状态（参见3.4节）。其余部分不需要是单个单词，例如，在发送短文件的标签中，分析剩余的短文件不会产生新的业务对象。再次在这种情况下，我们将发送映射到REST动词PUT。

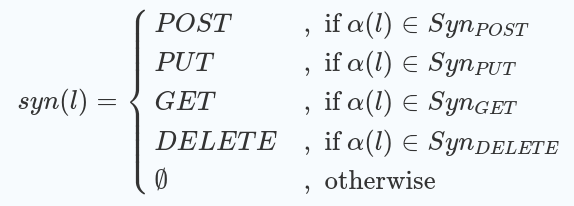
剩余分析提供了更多的语义信息，这是生成REST请求所需的，如以下两节所述。

## 3.3 通过自然语言分析派生REST动词

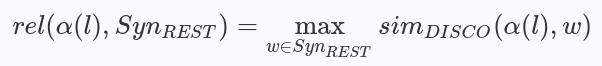
通过自然语言分析推导REST请求的总体思路是基于这样的假设，即编排任务标签提供所有相关信息。 具体而言，我们关注标签的行为，因为它们描述了必须执行哪些特定活动以及这些活动如何影响系统。 REST动词派生应用两个步骤。 第一步将相应的编排任务标签的动作与反映不同REST动词含义的同义词进行比较。 第二步涉及编排任务标签和同义词的动作的语言相似性分析，以防标签的动作与任何同义词不完全匹配。 在下文中，我们将进一步详细讨论这两个步骤。

首先，我们需要一组同义词，然后才能进行派生。一个挑战是，REST动词与特定的技术含义相关，并不一定与单词的语言含义相符。例如，REST动词POST指示服务器创建一个新的可区分资源，而要发布的动词通常描述在公告板上公布新闻的行为。因此，有必要在技术意义上定义反映POST意义的一组同义词。为此，我们向REST专家询问最接近REST动词含义的自然语言动词。这个过程的结果如表1所示。例如，专家一致认为POST的含义最好反映在动词创建或请求中。由于识别出的动词可能无法捕捉到语言的所有变化，我们进一步考虑可能从计算词典中提取的附加同义词，如WordNet [12]。例如，POST动词也可能与动词产生或制作有关。其他的例子可以用边框中的前一个表格检索。同样，我们使用相同的方法来推导表1中的发送同义词集。

同义词分析步骤调查编排任务标签的动作是否等于REST动词的同义词之一。 如果此条件评估为true，我们将相应的REST动词映射到编排任务。 否则，不会将REST动词映射到此任务。 作为一个例子，考虑编排任务创建CFP和确认短文。 第一项任务将映射到POST，因为其创建的操作是SynPOST集合的同义词的成员。 第二个任务将映射到PUT，因为它的确认操作是集合SynPUT的成员。 这个逻辑由以下函数表示，假设REST是REST动词的集合：



在同义词分析步骤未能将REST动词分配给编排任务的情况下，相似性分析步骤用作回退策略。 在这种情况下，有必要找到与动作最相关的REST动词。 因此，有必要确定一个动作与同义词的关联性。 在我们的方法中，我们使用语义相似性的概念（例如，参见[13,14,15]）来量化这种相关性。 我们利用DISCO字词相似度工具[16]的分布相似性，用simDISCO表示，因为它比现有的相似性度量更好[17]。 给定编排任务标签l，其动作α（l）和任意REST动词SynREST的同义词集合，编排任务标签和同义词REST集合的动作的相关性如下给出：



作为一个例子，我们考虑编排任务输入图1中的文章评论。由于输入的动作不是REST动词同义词集的成员，因此我们确定它与每个同义词集的相关性。 使用二阶分布相似度，我们收到以下相关性值：rel（enter，SynPOST）= 0.48，rel（enter，SynPUT）= 0.92，rel（enter，SynGET）= 0.92，rel（enter，SynDELETE）= 0.55。

最后，我们考虑所有的相关性分数来为给定的编排任务标签派生最合适的REST动词。 在这种情况下，我们假设最高关联度分数反映了给定编排任务中最适合的REST动词。 因此，我们将这个REST动词分配给最高的相关性分数。 然而，可能出现这样的情况，即多个相关性分数相等，从而导致REST动词的多于一个分配，强调用户选择正确的REST动词的必要性。 在形式上，我们描述相似性分析步骤如下：



作为一个例子，再次考虑编排任务输入文章评论及其相关性分数。由于PUT和GET的得分相等，相似性分析策略将两个REST动词PUT和GET分配给编排任务。

标签的行为不是唯一可以定义适当的REST动词的实体。我们也考虑确定的业务对象。 REST调用中PUT和POST之间的区别在于服务器处理资源的不同方式。当使用POST时，服务器将自行创建一个分配标识符的资源。但是，使用PUT时，客户端必须在请求中指定资源的标识符。例如，请求POST / paper后面跟着一个response / papers / id，其中id由服务器选择。否则，作者应该知道他或她想要修改或创建的论文的ID，当他或她请求PUT / paper / id时。

我们观察到，POST和PUT之间的这种内在差异可以分别在不定冠词和定冠词的自然语言概念中找到。在这些情况下，动作映射到PUT，我们会对英文不定冠词“a”或“an”的存在进行额外检查。如果事情变成这样，我们将这个行为映射到POST。例如，提交图1中的文章任务被映射到POST / paper，因为客户端没有引用特定的文章。

以下部分将解释如何使用各自标识的REST动词为编排任务生成RESTful请求。

## 3.4 生成REST请求

生成REST请求的任务涉及生成唯一的资源标识符（URI），以解释如何通过HTTP访问资源。为了生成URI，我们将其生成视为语言生成问题，该问题使用编排任务的可用信息和上一步中的REST动词派生。许多语言生成系统采用三步流水线方法，首先确定所需的句子信息，然后计划表达这些信息，然后将其转化为正确的句子[18]。与这些系统相比，我们不需要完全灵活的方法，因为最终的链接遵循常规结构[5]。因此，我们使用基于模板的方法[19,20,21]来生成REST URI。具体来说，我们使用编排任务和上一步中的REST动词并选择相应的链接模板。之后，我们为模板填充必要的信息，即编排任务标签的操作和业务对象。必须注意的是，有些情况下，当请求派生显示多于一个REST动词或需要用户更正链接的REST动词时，编排与其他REST链接无关。

表2 基于REST请求的链接模板

|  |  |
| --- | --- |
| 链接模板 | 例子 |
| POST /<β(l)<β(l)><<HTTP Version> | POST /CFP HTTP/1.1 |
| PUT /<β(l)><β(l)>/id/<Past Participle of α>α> <HTTP Version> | PUT /paperReview/id/entered HTTP/1.1 |
| GET /<β(l)><β(l)>/id <HTTP Version> | GET /paperReview/id HTTP/1.1 |
| DELETE /<β(l)><β(l)>/id <HTTP Version> | DELETE /shortPaper/id HTTP/1.1 |

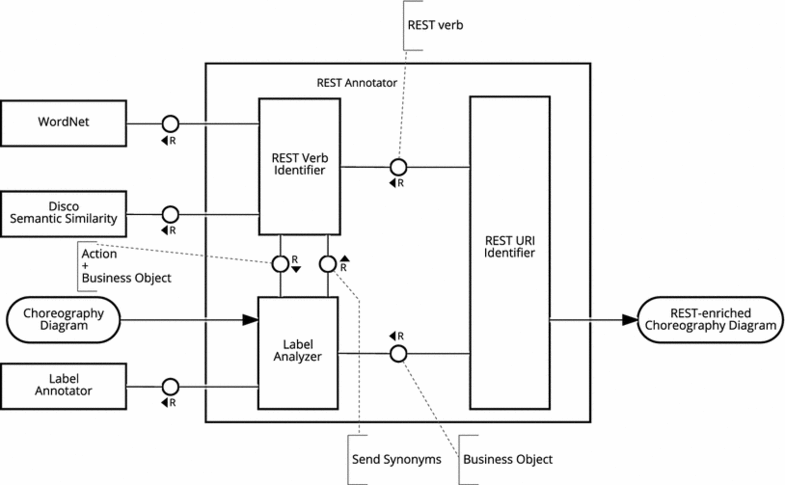


图4 REST Annotator软件体系结构

表2显示了不同REST动词的链接模板，并给出了从图1的编排任务创建的示例。模板强调编排任务标签（β（l））的业务对象对REST链接扮演着重要角色，它类似于需要由REST动词解决的服务器资源。因此，我们将业务对象与唯一标识符关联在一起。如果特定资源的状态必须更改，链接还会解释其状态如何随REST动词而变化。这个变化是由一个编排任务标签的过去分词表达的。

我们的方法的最终输出是一个RESTful编排。图3是一个模型，描绘了将我们的方法应用到运行示例中生成的RESTful编排图的摘录。在这个图中，我们展示了REST工程师如何与生成的RESTful编排进行交互。 REST工程师提供所有四个生成的链接（每个REST动词一个），根据匹配得分（1为最好，0为差）进行排名。根据选择，假定交互始终有效,HTTP响应会自动生成。

# 4 评估

本节介绍我们的评估。 首先，我们解释我们原型实现的架构。 然后，我们从实践中呈现一组172个编排图的推导步骤的准确性结果。

## 4.1 评估配置

为了评估我们的方法，我们开发了一个名为REST Annotator的工具。 REST Annotator的体系结构在图4中被描述为FMC图[22]。 REST Annotator将一组编排图作为输入，并输出一组REST丰富的编排图。 该工具使用三个外部组件：Leopold等人的标签注释器。 [11]，WordNet [12]以及DISCO工具[16]的分布相似性组件。 构成该工具的主要组件由三个子组件组成：标签分析器，REST动词标识符和REST URI标识符。

标签分析器负责从模型中提取所有标签，并在标签注释器的帮助下对其进行分析。 后者用于表示编排任务标签的操作和业务对象。 标签分析器将动作和每个标签的业务对象映射到REST动词标识符和REST URI标识符组件。 REST动词标识符组件需要由标签分析器提供的操作以及WordNet的同义词，类似于各自的REST动词。 如果未找到同义词，则该组件需要动作和来自Disco语义相似性组件的REST动词的同义词集之间的语义相似性分数。 一旦识别出动作与每个REST动词的语义关系，就将REST动词及其各自的分数传递给REST URI标识符组件。 这个组件产生一组充满REST注解的编排图。

此外，标签分析器在提供最终操作和业务对象之前，需要应用3.2节中所述的方法。 为此，它需要SynSend，它由REST动词标识符生成。 另外，后者需要前者的业务对象来检查是否存在不定冠词。 正如在3.3节中解释的那样，不定冠词的存在可以改变REST动词标识符的输出。

作为评估数据，我们使用BPM 学者倡导的编排图。 该计划提供了来自不同领域的丰富的流程模型。 总的来说，我们检索了424个BPMN编排图。 由于这些图表是由专家和非专家一起创建的，因此有必要清理数据。 我们采用以下清理标准：

1.仅限英语的图表我们仅包含带英文文字标签的图表。 这个标准是必要的，因为大多数自然语言分析组件仅支持英语。

1. 语法正确的图表排除具有关于BPMN 2.0编排图规范的语法错误的图表。

关于评估程序，REST专家必须对每个编排任务进行三步评估：（a）标签的语法正确性，（b）REST动词生成的适当性，以及（c） 生成的REST URI的适应性。 在（a）成立的情况下，评估者还必须检查具有最佳匹配分数的识别的REST动词是否适当。 如果（b）成立，评估者必须检查生成的URI是否合适。

然而，评估有其自身的局限性。 单一评估者的存在可能会在确定正确匹配时施加主观性。 为了缓解这个问题，评估人员重复了几次检查程序，同时参考开发RESTful API的常见做法。 这样可以避免像相同标签 - RESTful请求对的不一致验证等错误，并有助于解决非明显的错误配对。

## 4.2 评估结果

本节讨论表3中汇总的结果。172个模型总共包含1213个编排任务标签。 模型大小（根据编排任务计数）为1到26，平均每个模型约7个编排任务。 从这些标签中，864个标签（71.14％）在语法上是正确的标签。 在下面的讨论中，我们只关注那些句法正确的标签，并讨论动词识别和链接生成在这些情况下如何执行。

动词识别策略在772个标签中识别出正确的REST动词，相当于所有句法正确标签的89.35％。在这些标签中，我们进一步区分已经用同义词策略和相似策略识别的动词。同义词策略能够在332个标签中导出正确的REST动词，而相似性策略则为450个编排标签导出正确的REST动词。结果强调了对REST动词相似性识别策略的需求。最明显的REST动词是PUT，因为它是在参与者改变业务对象状态的编排上下文中预期的，例如，订单被发送，接受，交付和支付。 POST发现139例，其中8例使用不定冠词。 DELETE标识的数量较少也反映了在REST环境中使用DELETE的罕见情况，这是由于资源通常存档或保存在特定状态而不是故意删除。

表3 用户评估的定量结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

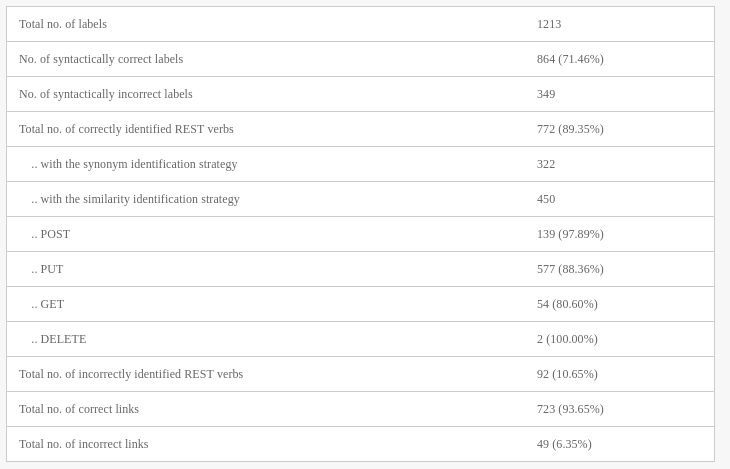


表4 来自图1的相应RESTful任务的REST请求结果

