

回梦涵, 黄凤兰, 倪枫, 等. 基于 NLP 技术对文本生成 BPMN 图[J]. 智能计算机与应用, 2024, 14(8): 11-18. DOI:10.20169/j.issn.2095-2163.240802

基于 NLP 技术对文本生成 BPMN 图

回梦涵, 黄凤兰, 倪枫, 刘姜, 李业勋

(上海理工大学 管理学院, 上海 200093)

摘要: 参考当今商业和组织环境, 流程管理是提效的核心。传统方法容易导致误解和模糊, 需求变更时, 手动更新和维护也非常困难。因此需要一种更形式化和自动化的方法, 来分析和验证需求确保正确性和一致性。本文提出了一种使用文本需求输入生成 BPMN 图的方法, 使用 NLP 处理文本需求以获取事实类型, 然后将事件映射到 BPMN 元素中, 通过基于电子表格的描述创建 BPMN 图。

关键词: 业务流程建模标注 2.0; 语义分析; 自然语言处理

中图分类号: TP311.52 **文献标志码:** A **文章编号:** 2095-2163(2024)08-0011-08

Generation of BPMN graphs for requirement texts based on NLP techniques

HUI Menghan, HUANG Fenglan, NI Feng, LIU Jiang, LI Yexun

(School of Business, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: With reference to today's business and organisational environment, process management is at the heart of efficiency gains. Traditional methods tend to lead to misunderstandings and ambiguities, and manual updating and maintenance is very difficult when requirements change. Therefore a more formal and automated approach is needed to analyse and validate requirements to ensure correctness and consistency. This paper proposes an approach to generate BPMN diagrams using textual requirement inputs, using NLP to process textual requirements to obtain fact types, and then mapping events to BPMN elements to create BPMN diagrams through spreadsheet-based descriptions.

Key words: business process modeling notation2.0; semantic analysis; natural language processing

0 引言

在目前的商业大环境中, 针对流程的管理已经成为提效、优化资源利用和提供卓越服务的关键因素。为了有效地管理和改进这些流程, 业界广泛采用了业务流程建模和标准化。其业务流程建模标注 2.0 (Business Process Modeling Notation 2.0, BPMN2.0) 为业务流程设计和流程实现之间的差距建立了一座标准化的桥梁^[1]。

在软件开发中, 正确理解和规格化系统的需求决定了软件的功能和行为。由于需求通常以自然语言的形式表达, 并涉及多个参与者和复杂的场景, 因

此存在理解和规格化的挑战。传统的需求工程方法往往依赖于人工分析和解释自然语言需求, 容易导致误解。当需求变更时, 手动更新和维护需求规格也非常困难。因此, 需要一种更形式化和自动化的方法来分析和验证需求, 以确保其正确性和一致性^[2]。而随着 ChatGPT 等人工智能聊天机器人的涌入, 互联网公司相继推出了基于深度神经网络的 NLP 服务和应用程序^[3]。

基于此背景, 本文提出使用自然语言处理 (Natural Language Processing, NLP) 技术生成 BPMN 图的方法, 以帮助解决需求理解和规格化的复杂性问题。该方法结合自然语言处理、形式化建模和可

基金项目: 国家自然科学基金 (12371508); 教育部产学研合作协同育人项目 (220603760210846); 上海市“大学生创新创业训练计划”资助项目 (SH2022072)。

作者简介: 回梦涵 (2000-), 女, 硕士研究生, 主要研究方向: Petri 网和业务流程管理; 黄凤兰 (1998-), 女, 硕士研究生, 主要研究方向: Petri 网和业务流程管理, 系统科学; 刘姜 (1983-), 女, 博士, 副教授, 主要研究方向: 复杂系统理论与方法, 符号代数计算; 李业勋 (2002-), 男, 本科生, 主要研究方向: 管理科学与工程。

通讯作者: 倪枫 (1982-), 男, 博士, 副教授, 主要研究方向: 系统科学, 系统分析与集成。Email: nifeng@usst.edu.cn

收稿日期: 2024-01-31

可视化技术,可以自动提取和分析用例,并生成可视化的图,以更好地理解 and 验证系统的需求。

本文的方法分为两个阶段:一是通过自然语言处理技术对文本需求进行分析,提取关键流程类型。文本解析将非结构化的文本数据转化为结构化的流程知识。其次,将流程类型映射到 BPMN 元素,利用电子表格方法生成直观的 BPMN 图,实现业务需求的可视化和结构化。

通过本文方法,可以将需求文本所描述的业务过程转化为 BPMN 图的可视化表示。生成的 BPMN 图将提供一个形象直观的视觉图像,用于展示和分析需求文本所描述的业务过程,帮助相关方更好地理解和沟通业务需求,以便进一步进行流程管理和优化。

1 理论及方法基础

1.1 业务流程建模标注

业务流程模型可以用文字表示为流程描述,这种多元化的表示方式为建模业务活动提供了多种选项^[4],为建模业务活动提供了不同的图表^[5]。

BPMN 提供了标准的符号和规范,使其能够以可视化的方式描述和呈现组织内的业务过程^[6]。其不仅提供了对业务流程的可视化描述,还在决策制定、流程自动化和跨功能沟通方面提供了重要的支持。业务架构是以业务为中心的框架,旨在明确企业的核心功能和架构,同时指导系统设计和技术实现,确保业务目标与技术战略的一致性^[7]。

从本质上讲,系统是动态的,组件系统之间的相互作用根据相应事件发生触发,进一步带动相应系统功能的执行^[8]。BPMN 模型主要用于对系统动态行为进行静态描述。

用自然语言描述的流程,有时可能包含误导性信息,可能会导致一系列错误^[9]。一个好的 BPMN 模型属性的形式化应该具备形式化属性,即满足结构良好^[10]、安全性^[11]和健全性^[12]。其提供了一种规范化的方式来描述和分析组织内的业务流程。通过使用 BPMN 模型,可以将业务活动和流程以可视化的方式进行建模和分析,从而更好地理解 and 改进组织的业务流程。

1.2 自然语言处理技术

自然语言处理(Natural Language Processing, NLP)是计算机科学、人工智能和语言学等学科交叉的前沿领域^[13]。其目标是让计算机能够理解、处理和生成自然语言,使计算机具有类似人类的语言交互和文本理解能力^[14]。预训练的模型在大量未标

记的语料库上进行训练,使模型能够从语料库中学习文本的最佳表示^[15]。在现实世界应用中得到广泛采用,包括情感分析^[16]和机器翻译^[17]。

NLP 技术通过将文本或语音转化为计算机能够理解的形式,使计算机能够处理和分析大量的自然语言数据。其主要目标是使计算机能理解和使用自然语言,从而为用户提供更好的交互体验和更高效的工作方式。

计算机用于处理自然语言文本,就需要词汇处理提取、识别、组织、关联单词和短语等过程。早期用于术语识别的词典由手工编制,或改编自现有的词典或列表,后来开发了自动工具,直接生成词典或者减少构建词典所需的人力。

词法分析(Lexical Analysis)是对词与词之间的联系进行分析,从而获得对文段局部表达信息的抓取,帮助研究者将输入文本或一系列字符打上词汇标签(Token)。在词法分析过程中,研究者尝试通过分析手段对每个单词进行识别并为其打上合适的词汇标签。最基础的标签标注方式是赋予每个单词一个词性标签(Part-Of-Speech, POS 标签)。

本文使用的分析工具 NLTK(Natural Language Toolkit)是在 NLP 领域中常使用的一个 Python 库,其可以处理预料库、分类文本、分析语言结构等多项操作,提供了一系列的 NLP 功能。其中包括分词、词性标注、命名实体识别、句法分析、依存句法分析、语义角色标注等。

2 自然语言处理与分析

从广义上讲,NLP 任务有效执行其操作所需的任何信息实体都被视为上下文。其可能包括部分注释的文本文档,或描述特定文本文档的任何元数据的可用性,或任何话语的存在也可以作为后续 NLP 任务的有用上下文^[13]。然而,现有的文字处理方法在有效建模和利用上下文时面临挑战,限制了存储、检索和利用上下文信息的能力。

文献[18-19]尝试生成与域无关和与任务无关的文本表示。这些系统通常依靠开放信息提取技术,来构建存储在基于 RDF 的三重存储中的文本表示。然而,其创建的文本表示更侧重于数据集成,而不是语言分析。本文提出的方法将需求文本拆分成词组,然后结合上下文对拆分单位进行情感与语义分析,解决了数据储存的冗余与词多义的问题。

首先对文字进行预处理,其中包括对需求文本进行预处理,以及语法分析、语义分析、文本分类与

聚类,以减少冗余和提取关键信息。随后,使用 NLP 技术中的文本分类方法,对需求文本进行分类。这些方法可以根据文本的特征和上下文信息,将需求文本归类为不同的事实类型。完成分析阶段后,系统将从文本中提取与业务过程相关的事实类型,确保生成的 BPMN 图符合真实需求文本的要求和意图。这些事实类型是后续生成 BPMN 图所需的基本信息。

2.1 文本预处理

首先进行关键词的提取。即将多个词组组成的文本需求进行拆分并提取出来,然后将其存储在数组列表中。该过程可以通过将文本需求分解为多个词组或短语,并将其存储在一个数组列表中来实现。然后进行标记化,将文本需求的原始文本分解成更小的单位,有助于理解文本的结构和语法。标记化可以通过将文本分解为句子,并将每个句子进一步分解为词语或标记来实现。

2.2 语法分析

2.2.1 自然语言处理与分析

语法分析包括词性标注和句法分析。词性标注是将文本中的每个词汇赋予一个特定的词性标签,使其可以更好的理解每个词汇在句子中的作用和语法功能,为后续的语法分析提供更准确的信息。句法分析是通过语法分析确定句子的基本结构,识别句子中的短语、从句和成分,并确定这些短语、从句和成份之间的关系,生成依存关系树,描述句子中的词汇之间的依存关系。

本文使用 NLTK 进行标记。并为文本中的每个标记分配适当的语义角色,以更好地理解句子中各个元素之间的语义关系。对于每个句子中的每个词语,包括词性、依存关系和命名实体信息都进行了标记。表 1 为分离后词性的标注规则。

对分割的词语进行标注后,通过表 2 的依存句法关系规则,可以对句子的依存关系进行解析,生成句法关系树。

2.2.2 文本类型提取

根据句子结构的不同,可以将句子分为 4 种类型:简单句、复杂句、复合结构和特殊句型^[20]。

简单句由主语和谓语构成,表达一个完整的思想。复合句由两个或多个简单句通过连接词或分句关系组合而成。复合句可以包括一个主句和一个或多个从句,从句可以是条件、目的、原因、结果等不同类型。并列句由两个或多个主句并列连接,可以表示并列关系、选择关系、对比关系等。特殊句型有倒装

句,其是一种对语序进行颠倒的句子结构,用于强调或满足特定语法要求。复杂句直接由从属分句表示。

表 1 词性标注集

Table 1 Lexical annotation set

标注	描述
JJ	形容词
MD	情态动词
NN	名词单数
NNS	名词复数
NNP	专有名词
PRP	人称代词
RB	副词
UH	感叹词
VB	动词原形
VBD	动词过去式
VBG	动名词或现在分词
VBN	动词过去分词
VBP	非第三人称单数的现在时
VBZ	第三人称单数的现在时
TO	正在

表 2 依存句法关系

Table 2 Dependent syntactic relations

关系类型	标注
NSUBJ	名词主语
NSUBJPASS	被动的名词主语
DOBJ	直接宾语
POBJ	介词的宾语
AMOD	修饰名词短语的形容修饰语
XCOMP	开放从句(缺主语的从句)补语
MARK	从句的主要成分

将输入的文本每一句分析归类为 3 种类型。一元句是只包含一个主语和一个谓语的简单句;二元句,有两个主语和两个谓语的句子;三元句或多元句为包含相应多个主语和谓语的句子。

2.3 语义分析

语义分析是对文本中的词汇和短语进行深入理解的过程。其中包括识别词义、句子的含义、上下文和语义关系。通过语义分析,可以更好地理解文本中的重要信息,以便进行后续的事实类型提取。在语义分析中,语义角色标签(Semantic Role Labeling, SRL)是一种自然语言处理常用的技术,用于识别句子中每个词或短语在句子中扮演的语义角色,这些语义角色表示词或短语在句子中扮演的不同角色或功能。通过为每个词或短语分配相应的语义角色,使用 SRL 可以更好地理解句子的结构和含义。

本文遵循角色基于 SBVR(Semantics of Business Vocabulary and Business Rules)^[1]元素,包括名词/一般概念、单个名词概念,动词概念属性、量化等。从结构文本中提取事件的规则如下:

规则 1 将结构文本分割成其组成的简单句,以便更好地理解每个句子。

规则 2 对于复合或复杂句子,将其分解成与句子中从句数量相等的简单句子以便更好地处理每个从句。

规则 3 对于全部倒装句,将主语移到谓语动词之前。对于部分倒装句,将助动词或情态动词移到主语之前。

规则 4 对于简单句,处理主语和谓语的复杂情况。如果主语由多个名词组成,用逗号或连词分隔,将句子分成多个新句子,与主语中的名词数量相等。如果谓语由多个动词组成,用逗号或连词分隔,将句子分成多个新句子,与谓语中的动词数量相等。

规则 5 识别多元句。识别主语+谓语对和主语+谓语+宾语对,并将其标识为一元事实类型句(identified as Unary Fact Type ;UFT)和二元事实类型句(Binary Fact Type ;BFT)。

规则 6 处理连接词和分离词的使用。连接词和分离词通常使用如‘和’、‘然而’、‘但是’、‘所以’、‘甚至’和‘即使’等词语来表示。在简单句子中,连接词连接各种语言单位,如名词与名词、动词与动词、从句与从句等,在简单句、复合句或复杂句中起到连接作用。

连接词也可以连接复合句或复杂句中的从句。分离词可以是包容性或排他性的,需要根据其词语表示的含义来进行推断。包容性分离词使用诸如‘或者’、‘除非’、‘和’等词语来表示,而排他性分离词用诸如‘如果’等词语来表示其推断。

2.4 事实类型识别

文本分析与自然语言处理的最后一步是识别事实类型,从文本中提取出与特定事件相关的信息。当系统从文本中提取与业务过程相关的事实类型后,按照分类对应 BPMN 的不同元素。事实类型通常是指描述业务中的对象、属性和关系的术语。这些事实类型是后续生成 BPMN 图所需的基本信息。

在 SBVR 1.5 中,事件由具有一个或多个动词概念角色的动词短语形成。这意味着动词短语可以具有不同的角色,通过识别动词短语及其对应的角色,可以确定事件的类型。文本需求可以包括一元事件类型、二元事件类型和具有 3 个或更多角色的事件类型。一元事件类型是指只涉及一个名词概念和一个动词概念的事件;二元事件类型是指涉及两个名词概念和一个动词概念的事件;而具有 3 个或更多角色的事件类型则涉及多个名词概念和一个动词概

念,并且每个名词概念都扮演不同的角色。

语法分析树的建立过程如图 1 所示。



图 1 NLP 分析举例

Fig. 1 Example of NLP analysis

3 电子表格生成

根据文本需求分析阶段得到的流程类型映射到 BPMN 元素上。映射过程需要依据规范和语义进行,以确保生成图能够准确地反映需求文本所描述的业务过程。

利用基于电子表格生成 BPMN 图的方法涉及一个基于 Python 的插件,通过映射后生成的表格可以生成 BPMN 图的结构和连接关系。

设:定义五元组 $C_0 = \{E, A, G, S, F\}$ 的 BPMN2.0 模型,其中:

(1) E 是事件集合, $E = (E^S \cup E^I \cup E^E) \wedge (E^S \cap E^I \cap E^E) = \emptyset$, 即事件集合 E 由开始事件集合 E^S 、结束事件集合 E^E 和中间事件集合 E^I 组成,其两两不相交。开始事件、结束事件和中间事件不区分每种事件中的具体类型,只关注事件的种类,且一个过程中只有一个开始事件和一个结束事件。

(2) A 是活动集合。活动可以分为原子任务 T 和子过程 S , 即 $A = T \cup S$ 。

(3) G 是网关集合, $G = (G^P \cup G^E) \wedge (G^P \cap G^E) = \emptyset$, 即网关集合 G 由并行网关集合 G^P 和排他网关集合 G^E 组成,且这两类网关集合无共同元素。其中,并行网关集合 G^P 包括并行分叉网关集合 G^{PF} 、并行汇总网关集 G^{PJ} 。这两个子集之间没有重叠元素,即 $(G^{PF} \cup G^{PJ}) \wedge (G^{PF} \cap G^{PJ}) = \emptyset$ 。基于数据的网关包括基于数据的排他网关集合 G^{DX} 、基于数据的合并网关集合 G^{DM} , 这两个子集之间没有重叠的元素,即 $(G^{DX} \cup G^{DM}) \wedge (G^{DX} \cap G^{DM}) = \emptyset$; 基于事件的网关包括基于数据的排他网关集合 G^{EE} 、基于事件的合并网关集合 G^{EM} , 这两个子集间没有重叠元素,即 $(G^{EE} \cup G^{EM}) \wedge (G^{EE} \cap G^{EM}) = \emptyset$ 。

(4) S 是流对象的集合, $S = (A \cup E \cup G) \wedge (A \cap E \cap G) = \emptyset$, 即流对象的集合由活动集合 A 、事件集合 E 和网关集合 G 组成,其两两之间没有交集。

(5) F 是连接关系的集合, $G = S \times S$, 对象与对

象之间依靠控制流来连接。

(6) 在编排中没有孤立的流对象。因此BPMN2.0模型是一个有向图,其中流对象作为节点,连接对象作为有向边。假设 x 和 y 都属于流对象的集合 S , 则 (x, y) 表示 x 与 y 之间通过顺序流进行连接。

(7) $M = \{model_1, model_2, \dots, model_n\}$ 。其中

$M_i(i = 1, 2, \dots, n)$ 为流程模型中的活动模块, n 等于活动模块的数量。

3.1 映射

映射用于将句子中的特定结构或词语映射到相应的 BPMN 元素,以捕捉业务过程中的准确信息。映射规则见表 3。

表 3 事件映射到 BPMN 元素的映射规则

Table 3 Mapping rules for event mapping to BPMN elements

序号	NL 描述	逻辑操作	BPMN 元素	BPMN 符号
1	过程文本标题		泳池	
2	事件类型主题		泳道	
3	事件		活动	
4	出现在‘如果’或者‘如果...那么...’中的事件	推导	异或网关	
5	出现在‘如果’或者‘如果...那么...’间的事件	连接	流元素	
6	出现在‘如果’或者‘如果...那么...’间的事件,但不是两者都出现	连接	排他网关	
7	数据		数据存储	

3.2 生成 BPMN 图

BPMN 模型由 BPMN 元素组成,包括对象、池、流元素等,表示和建模各种类型的业务过程。通过这些符号和元素,组织可以创建可视化的流程图,以便更好地理解、沟通和改进其业务流程。

对象可以是事件(Events)、活动(Activity)或者网关(Gateways);流元素包括序列流(Sequence

Flow)和消息流(Message Flow)。序列流用于连接流程的两个对象,并表示控制流(即排序)关系。

得到 SRL 结果后,通过二进制事实类型对事件生成基于电子表格的描述,并且通过一个基于 python 的插件 py_BPMN 直接生成 BPMN 图。BPMN 生成原理如图 2 所示。

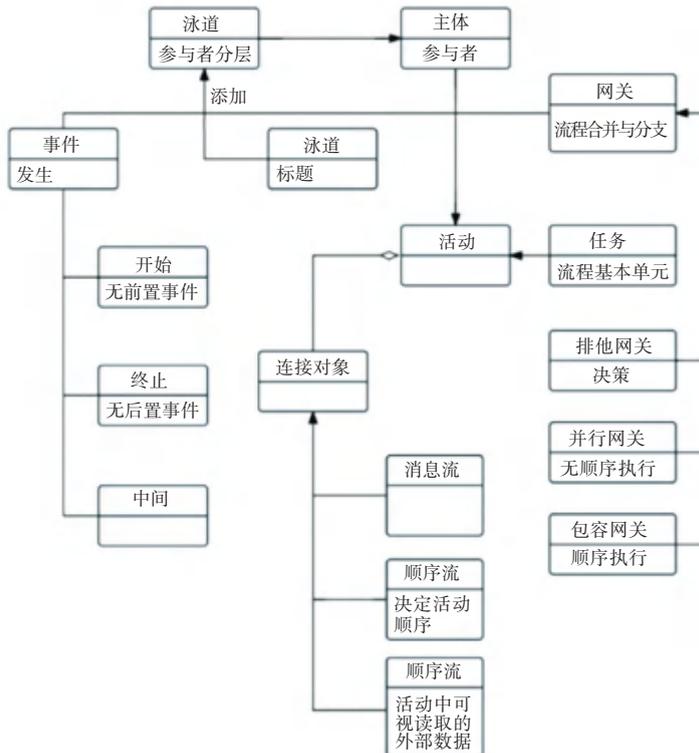


图 2 BPMN 模型对应关系

Fig. 2 BPMN model correspondence

将 SRL 结果中的事件和其对应的角色映射到电子表格中的行和列上,创建一个矩阵式的描述,其中每个单元格表示事件涉及的元素和其对应角色的二进制事实类型。生成基于电子表格的描述,可以更直观地呈现 SRL 结果。

4 案例研究

4.1 学生借书

对于事件“学生去图书馆借书,如果有库存,则去登记,否则返回。”可将其分割成:[学生][去][图书馆][借][书][,][如果][有][库存][,][则][去][登记][,][否则][返回][。];标记为:学生/NN;去/VBZ;图书馆/NN;借书/VBZ;。/;如果/RB;有/VBZ;库存/NN;,/标点;则/RB;去/VB;办理/VBZ;借书/NN;手续/NN;,/;否则/RB;返回/VBZ;。/。

解析完关系后,将创建一个词树结构来表示文本输入的结构,用于标识文本输入的各种句法结构之间的关系。

根据表 2 的规则,将会转变为活动,并使用相应的网关元素来处理“如果”、“否则”和“那么”的逻辑。这样,这个句子将在 BPMN 图中被表示为一个活动和相应的网关元素。句子的 SRL(语义角色标注)结果见表 4。在本文中,主要提取名词/单个概念和动词概念作为事实类型的主要组成部分。表 5 展示了从输入文本中提取的事实类型,以及利用一组规则得出的结果。

表 4 SRL 结果

Table 4 SRL results

SBVR 概念	计数	词汇
名词	4	学生、图书馆、库存、手续
动词	6	去、借书、有、去、办理、返回
特征	-	C1:状态接受 C2:状态拒绝
二进制事实类型		BFT1:学生要求借书 BFT2:管理员接收要求 BFT3:管理员请求库存 BFT4:信息库查询库存 BFT5:信息库分配信息状态 BFT6:管理员接收 BFT7:管理员反馈学生 BFT8:学生收到信息 BFT9:借书 BFT10:返回

表 5 中序号 1 到 6 是通过映射 BFT1 到 BFT6 获得的,而 7.1 是通过映射 BFT9 获得的活力,在条件为 C 的 IF 中;7.2 通过映射 BFT10 获得的活力,位于“否则”中。

表 5 输入的电子表格

Table 5 Input spreadsheet

序号	活动	条件	执行者	终止
0	开始			
1	借书(A)		学生	
2	接收要求(B)		老师	
3	请求库存(C)		老师	
4	查询库存(D)		信息库	
5	为库存分配状态(E)		信息库	
6	将信息发送给老师(F)		信息库	
7	接收信息(G)		老师	
7.1	办理借书手续(H)	C1	老师	
7.2	拒绝(I)	C2	老师	
8	传送消息给学生(J)		老师	
9	接受信息(K)		学生	
10	结束			是

参考表 3,创建基于电子表格的描述(如表 5)。得到以下内容:“BFT1 后 BFT2。BFT3 后 BFT4 后 BFT5 后 BFT6, BFT7 后 BFT8。如果 C1,则 BFT9;如果 C2,则 BFT10。”将事实类型映射到 BPMN 元素后,下一步是创建基于电子表格的描述。开始和结束事件在基于电子表格的说明中分别分配为“0”和“是”。

基于表 3,生成图 3 中的 BPMN 图。该过程首先根据功能过程的名称创建一个名为“学生借书”的池,以反映功能过程的名称。在“成员”列中,有“学生”、“信息库”和“老师”,因此需要创建 3 条通道。根据“成员”列,活动或事件发布在各自的通道中。活动可能涉及到数据存储或数据对象。

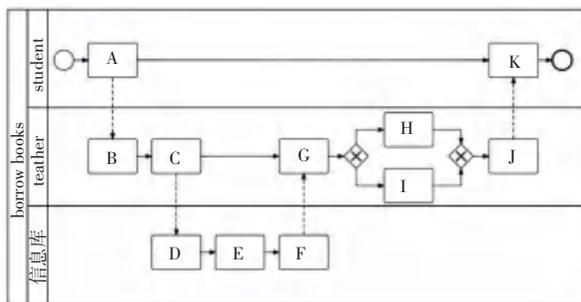


图 3 对应生成的 BPMN 图

Fig. 3 Corresponding generated BPMN diagram

4.2 推广

如图 4 所示。

表 6 为不同类型句子的示例,生成的 BPMN 图

表 6 推广句子
Table 6 Promotional sentences

序号	主题	自然语言文本
1	买饭	学生刷校园卡买饭
2	学生请假	学生提交请假申请,老师将申请记录提交到系统。
3	已毕业学生除名	教务从学生库中读取该学生信息,更新该学生的新状态到已毕业学生库中。
4	制作成绩	老师从成绩单中提取出学生书面成绩,综合平时成绩计算。如果总成绩大于等于 60,则及格,否则挂科。

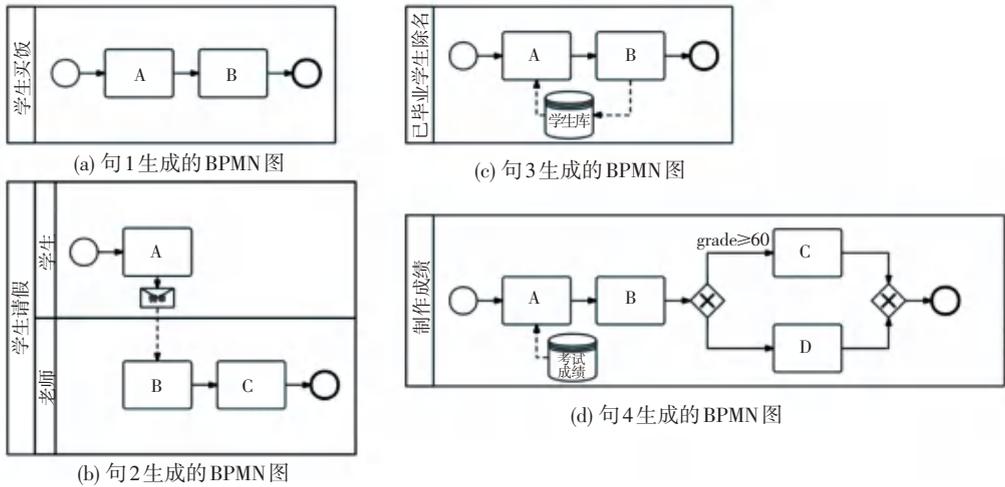


图 4 表 6 对应生成的 BPMN

Fig. 4 Table 6 corresponds to the generated BPMN

5 结束语

本文探索了自然语言文本中的关键信息,并提取出与业务流程相关的实体、动作和关系;通过一个基于 python 的 py_BPMN 插件将这些信息映射到 BPMN 图像中的任务、网关、流程线等元素上。最终,成功地实现了自然语言到 BPMN 图像的转换,并验证了该方法的有效性和准确性。

将自然语言文本转换为可视化的 BPMN 图像,提供了一种更直观、易理解的方式来描述和分析业务流程。对于企业流程管理和自动化的实践具有重要的意义,能够帮助企业更好地理解 and 优化其业务流程,提高工作效率和质量。

参考文献

[1] ROSING M V, WHITE S, CUMMINS F, et al. Business process model and notation - BPMN [J]. Complete Business Process Handbook, 2015, 95:433-457.

[2] 黄凤兰,倪枫,刘姜,等. 基于 HCPN 的复杂 BPMN 协作模型数据流建模与验证[J]. 计算机集成制造系统,2024,30(5):1754-1769.

[3] DAI L, MAO J, XU L, et al. SecNLP: An NLP classification model watermarking framework based on multi-task learning [J]. Computer Speech & Language, 2024, 86: 101606.

[4] GARCIA M T, NUNES M M, FANTINATO M, et al. BPMN-Sim: A multilevel structural similarity technique for BPMN process models[J]. Information Systems, 2023, 116: 102211.

[5] CORRADINI F, PETTINARI S, RE B, et al. A BPMN-driven framework for Multi-Robot System development[J]. Robotics and Autonomous Systems, 2023, 160: 104322.

[6] 王新康,倪枫,刘姜,等. 基于扩展 BPMN 的“家园互动”式儿童健康管理信息系统架构[J]. 智能计算机与应用,2022,12(10):189-199,202.

[7] 钟贤欣,倪枫,刘姜,等. 基于 ROADS 的面向场景业务架构建模方法[J]. 上海理工大学学报,2023,45(4):415-424.

[8] 黄凤兰,倪枫,刘姜,等. 基于 ROAD-CPN 业务架构的可执行建模方法[J]. 上海理工大学学报,2023,45(5):534-542.

[9] SUCHENIA A, POTEPA T, LIGŹA A, et al. Selected approaches towards taxonomy of business process anomalies[C]//Advances in Business ICT: New Ideas from Ongoing Research Cham: Springer, 2017: 65-85.

- [10] KIEPUSZEWSKI B, TER HOFSTEDE A H M, BUSSLERC J. On structured workflow modelling [C]//Proceedings of Advanced Information Systems Engineering: 12th International Conference. Stockholm, Sweden:IEEE,2000: 431-445.
- [11] ROZENBERG G, ENGELFRIET J. Elementary net systems [J]. Lecture Notes in Computer Science, 1998, 1491: 173-187.
- [12] VAN DER AALSTW M P. Process-oriented architectures for electronic commerce and interorganizational workflow [J]. Information Systems, 1999, 24(8): 639-671.
- [13] CAMBRIA E, WHITE B. Jumping NLP curves: A review of natural language processing research [J]. IEEE Computational Intelligence Magazine, 2014, 9(2): 48-57.
- [14] GAROUSI V, BAUER S, FELDERER M. NLP-assisted software testing: A systematic mapping of the literature [J]. Information and Software Technology, 2020, 126: 106321.
- [15] CHUNG S, MOON S, KIM J, et al. Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA) [J]. Automation in Construction, 2023, 154: 105020.
- [16] LI R, CHEN H, FENG F, et al. Dual graph convolutional networks for aspect-based sentiment analysis [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics .IEEE,2021: 6319-6329.
- [17] ZHANG T, KISHORE V, WU F, et al. Bertscore: Evaluating text Generation with BERT [C]//Proceedings of International Conference on Learning Representations.IEEE,2020: 26-30.
- [18] MARTINEZ-RODRIGUEZ J L, LOPEZ-AREVALO I, RIOS-ALVARADO A B. Openie-based approach for knowledge graph construction from text [J]. Expert Systems with Applications, 2018, 113: 339-355.
- [19] VOSSEN P, AGERRI R, ALDABE I, et al. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news [J]. Knowledge-Based Systems, 2016, 110: 60-85.
- [20] 孟佳营. 基于句法分析的关系分类和关系抽取方法研究 [D]. 杭州: 杭州电子科技大学, 2023.