

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE ENGENHARIA DE COMPUTAÇÃO

GUILHERME DE OLIVEIRA TASSINARI

**Descoberta de Informações de Processo de  
Negócio a partir de Dados da Plataforma de  
Gestão de Projetos Trello**

Monografia apresentada como requisito parcial  
para a obtenção do grau de Bacharel em  
Engenharia da Computação

Orientador: Prof. Dr. Lucineia Heloisa Thom

Porto Alegre  
2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões Mendes

Vice-Reitora: Prof<sup>a</sup>. Patrícia Helena Lucas Pranke

Pró-Reitora de Ensino: Prof<sup>a</sup>. Cíntia Inês Boll

Diretora do Instituto de Informática: Prof<sup>a</sup>. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Engenharia de Computação: Prof. Walter Fetter Lages

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“The good thing about science is that it’s true  
whether or not you believe in it.*

” — NEIL DEGRASSE TYSON

*“I don’t want to believe. I want to know.*

” — CARL SAGAN

*“Be ashamed to die until you have  
won some victory for humanity.*

” — HORACE MANN

## AGRADECIMENTOS

Em primeiro lugar quero agradecer aos meus pais, Nara e Sérgio, meus irmãos, Fabrício e Vinícius, meus dindos, Cristina e Marcelo e minha companheira, Caroline, por todo o amor, aprendizado, carinho e paciência nestes últimos 8 anos. Sem vocês eu não teria chegado tão longe e não seria um décimo do ser humano e profissional que sou hoje. Quero deixar aqui a minha homenagem ao meu tio, Valdir Picetti, “In Memoriam”. Muito obrigado por todos os churrascos e galetos de domingo em família, sentiremos sua falta. Muito obrigado, também, ao meu amigo mais antigo, Rafael Lopes, que esteve comigo nos melhores momentos da minha vida. Quero agradecer aos meus colegas de barra, Caroline e Renato, por serem os meus primeiros companheiros nesta longa empreitada. Ao time Enactus UFRGS, por se tornar a minha família dentro da universidade e proporcionar as lembranças desta jornada que guardarei pelo resto de minha vida. Não existem palavras capazes de externar o amor e carinho que tenho por todos vocês. Em especial, muito obrigado Filipe, Gustavo e Erik, por se tornarem os meus melhores companheiros dentro da UFRGS e por serem a minha rota de escape da rotina e do estresse no LoL e Valorant de toda a noite. À Empresa Júnior IDE, por acreditar na minha visão e por tornar a minha experiência na UFRGS completa. Mal posso esperar para vê-los como a maior EJ do estado. Aos meus colegas de estágio no Me Salva!, que desde então se tornaram meus amigos e minhas referências profissionais. Vocês são grande parte do profissional que me tornei, e por isso, muito obrigado. A todos os professores e demais colegas que conheci na universidade, obrigado pelas experiências e pelos conhecimentos que me proporcionaram nesta caminhada. Por fim, gostaria de agradecer a Profa. Dra. Lucinéia Heloisa Thom, por aceitar se tornar minha orientadora e me proporcionar a minha primeira experiência em pesquisa científica, sendo ao mesmo tempo firme, exigente e protetora nas suas orientações, revisões e críticas. Muito obrigado por todo o aprendizado e pela amizade.

## RESUMO

Processos de negócio são fundamentais em organizações. Através da execução de processos eficazes e de qualidade, as organizações são capazes de atingir resultados, tais como diminuição de custos, aumento da produtividade, entrega de mais valor aos seus clientes e, por consequência, destacarem-se frente a concorrência. Com foco em atingir tais resultados, o Gerenciamento de Processos de Negócio (*Business Process Management - BPM*) vêm sendo utilizado por diversas organizações para aprimorar e monitorar processos, através de projetos de BPM. Entretanto, o esforço e custo necessários para executar estes projetos, principalmente em grandes organizações, pode ser significativamente alto, especialmente na fase chamada descoberta de processos, em que analistas de processo devem realizar um conjunto de estudos de documentação, entrevistas e *workshops* para descobrir quais são e como são executados os processos na organização. Objetivando tornar esta fase mais ágil e eficaz, diversos estudos vêm sendo realizados para automatizar a extração de informações de processo de negócio, tanto de *logs* de eventos de sistemas, quanto de textos em linguagem natural (*e-mails*, documentações, manuais, etc) e até mesmo de plataformas de comunicação. Neste contexto, o presente trabalho propõe uma análise da plataforma de gestão de projetos Trello como fonte de informações de processo de negócio propensa à extrações automatizadas, em especial com o uso de técnicas de processamento de linguagem natural. Isto é realizado com a elaboração de definições capazes de extrair estas informações da plataforma e sua posterior implementação em um protótipo, que as aplica sobre dados da plataforma Trello e as transforma em texto em linguagem natural. Este texto, por sua vez, pode ser usado tanto por analistas de processo em projetos de BPM quanto por outras ferramentas de extração automatizada de informações de processo de negócio. Para avaliar as definições e o protótipo, um experimento foi realizado com a participação de analistas de processo, com o objetivo de comparar a capacidade de extração da abordagem frente a extração manual realizada em projetos de BPM.

**Palavras-chave:** Gerenciamento de processos de negócio. Processamento de linguagem natural. Trello. BPM. Descoberta de processos.

## **Process Discovery using the Trello project management platform data**

### **ABSTRACT**

Business processes are fundamental in organizations. Through the execution of effective and quality processes, organizations can achieve results, such as lower costs, increased productivity, delivering more value to their customers, and consequently, stand out in the face of the competition. With a focus on achieving such results, Business Process Management (BPM) has been used by several organizations to improve and monitor processes through BPM projects. However, the effort and cost required to carry out these projects, especially in large organizations, can be significantly high, especially in the phase called process discovery, where process analysts must carry out a set of documentation studies, interviews, and workshops to find out what the processes are and how they run in the organization. To make this phase more agile and effective, several studies have been carried out to automate the extraction of business process information, both from system's events logs and texts in natural language (e-mails, documentation, manuals, etc.) and even communication platforms. In this context, the present work proposes an analysis of the Trello project management platform as a source of business process information prone to automated extractions, especially with the use of natural language processing techniques. This is accomplished with the elaboration of definitions capable of extracting this information from the platform and its subsequent implementation in a prototype, which applies them to data from the Trello platform and transforms them into text in natural language. This text, in turn, can be used both by process analysts in BPM projects and by other tools for automated extraction of business process information. To evaluate the definitions and the prototype, an experiment was carried out with the participation of process analysts, to compare the extraction capacity of the approach compared to manual extraction performed in BPM projects.

**Keywords:** Business Process Management, Natural Language Processing, Trello, BPM, Process Discovery.

## LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
BPG	Business Process Graph
BPD	Business Process Diagram
BPM	Business Process Management
BPMN	Business Process Model and Notation
MVC	Model-View-Controller
NLP	Natural Language Processing
PLN	Processamento de Linguagem Natural
POS	Part-of-speech
RoR	Ruby on Rails
TI	Tecnologia da Informação

## LISTA DE FIGURAS

Figura 1.1	Visão geral da metodoliga aplicada neste trabalho.....	15
Figura 2.1	O Ciclo de vida de BPM de acordo com Dumas et al. (2018) .....	18
Figura 2.2	Principais elementos da BPMN 2.0 utilizados neste trabalho .....	21
Figura 2.3	Exemplo de modelo de processo em BPMN 2.0 .....	22
Figura 2.4	Exemplo de pipeline de PLN.....	23
Figura 2.5	Exemplo de classificação gramatical de uma sentença.....	25
Figura 2.6	Exemplo comparativo entre análise de constituintes e análise de dependências.....	26
Figura 2.7	Exemplo real de análise de dependências.....	26
Figura 2.8	Exemplo real de análise de constituintes .....	27
Figura 2.9	Arquitetura geral da Stanford CoreNLP toolkit .....	28
Figura 2.10	Exemplo de quadro da Trello contendo 4 listas.....	30
Figura 2.11	Exemplo de cartão da Trello .....	31
Figura 3.1	Exemplo de comentário de um cartão .....	43
Figura 3.2	Exemplo de uso de uma menção em um comentário .....	44
Figura 3.3	Exemplo de extração da definição de atividade representada por um cartão Trello a partir da fig. 2.11. ....	46
Figura 3.4	Exemplo de extração de sub-tarefas a partir de uma <i>checklist</i> utilizando exemplo da fig. 2.11.....	47
Figura 3.5	Exemplo de extração de tarefa, membros de processo e sub-tarefas a partir de um cartão da Trello utilizando exemplo da fig. 2.11 .....	48
Figura 3.6	Exemplos de descrições da Trello em edição com formatação, pós edição com formatação e em edição filtrado. ....	50
Figura 3.7	Exemplo visual da aplicação das definições 3, 6, 5 e 4, nesta ordem, ao cartão da fig. 2.11 .....	52
Figura 4.1	Visão em alto nível de abstração das etapas e artefatos do protótipo .....	56
Figura 4.2	Visão geral da arquitetura cliente servidor pelo protocolo HTTP .....	57
Figura 4.3	Modelo MVC aplicado a uma arquitetura cliente-servidor na <i>web</i> .....	58
Figura 4.4	Visão geral da arquitetura do protótipo desenvolvido .....	59
Figura 4.5	Diagrama de sequência em alto nível de abstração mostrando a interação entre usuário, navegador, protótipo e API da Trello durante a etapa de acesso às informações da Trello.....	60
Figura 4.6	Diagrama de classes da etapa de acesso às informações da Trello.....	61
Figura 4.7	Pipeline de pré-processamento de cartão Trello .....	62
Figura 4.8	Diagrama de sequência mostrando o fluxo de pré-processamento do cartão .....	64
Figura 4.9	Diagrama de classes dos pacotes e classes envolvidas na etapa de pré-processamento de cartão .....	67
Figura 4.10	Diagrama de classes dos pacotes e classes envolvidas na etapa de extração de informações de processo de negócio .....	70
Figura 4.11	Página contendo o formulário de credenciais da Trello .....	71
Figura 4.12	Página contendo os quadros do usuário.....	72
Figura 4.13	Página contendo os cartões do quadro escolhido pelo usuário.....	73
Figura 4.14	Página contendo o cartão pré-processado e o texto com informações de processo de negócio .....	74
Figura 5.1	Metodologia de avaliação .....	79

Figura 5.2	Questões demográficas dos formulários aplicados .....	80
Figura 5.3	Questões sobre a experiência em modelagem de processos de negócio dos formulários aplicados .....	81
Figura 5.4	Questões sobre a experiência no uso da plataforma Trello dos formulários aplicados .....	82
Figura 5.5	Introdução ao Trello exibida no primeiro questionário .....	83
Figura 5.6	Questões para a modelagem de um processo de negócio em um cartão Trello .....	84
Figura 5.7	Questões para a transformação de um cartão Trello em descrição de processo.....	85
Figura 5.8	Questão para modelagem de processo de negócio com base em texto/descrição de processo.....	86
Figura 5.9	Exemplo de BPG, elaborado a partir do BPD da fig. 2.3 .....	86
Figura 5.10	Cartões Trello elaborados a partir das respostas do primeiro questionário .....	88
Figura 5.11	Análise da sentença “ <i>Register claim</i> ” pela <i>Stanford CoreNLP</i> .....	89
Figura 5.12	Modelo de processo elaborado a partir da descrição de processo .....	91
Figura 5.13	Modelo de processo elaborado a partir do texto gerado pelo protótipo .....	91
Figura 5.14	Grafos de processo de negócio elaborados a partir dos modelos nas figuras 5.12 e 5.13.....	92
Figura 5.15	Tabela contendo os valores de similaridade sintática entre os nós dos grafos da figura 5.14 .....	93

## LISTA DE TABELAS

Tabela 2.1	Trabalhos relacionados .....	34
Tabela 3.1	Resumo dos axiomas elaborados a partir da análise da plataforma Trello - Parte 1 de 2 .....	38
Tabela 3.3	Resumo dos axiomas elaborados a partir da análise da plataforma Trello - Parte 2 de 2 .....	39
Tabela 3.5	Resumo das definições de identificação de elementos de processo a partir de cartões da plataforma Trello .....	51
Tabela 3.7	Exemplo de aplicação das definições sobre o cartão da fig. 2.11 .....	53
Tabela 5.1	Dados demográficos e de experiências individuais dos respondentes do primeiro questionário .....	87
Tabela 5.3	Dados demográficos e de experiências individuais dos respondentes do segundo questionário .....	95
Tabela 5.5	Dados demográficos e de experiências individuais dos respondentes do terceiro questionário do Grupo 1 .....	96

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>12</b>
<b>1.1 Objetivos</b> .....	<b>14</b>
<b>1.2 Metodologia de Pesquisa</b> .....	<b>14</b>
<b>1.3 Organização do texto</b> .....	<b>15</b>
<b>2 FUNDAMENTOS E TRABALHOS RELACIONADOS</b> .....	<b>16</b>
<b>2.1 Gerenciamento de Processos de Negócio</b> .....	<b>16</b>
2.1.1 Notação e Modelo de Processos de Negócio .....	20
<b>2.2 Processamento de linguagem natural</b> .....	<b>22</b>
2.2.1 O Pipeline de PLN .....	23
2.2.2 Stanford CoreNLP .....	27
<b>2.3 Gestão de projetos e plataforma Trello</b> .....	<b>29</b>
<b>2.4 Trabalhos relacionados</b> .....	<b>31</b>
<b>3 EXTRAÇÃO DE INFORMAÇÕES DE PROCESSO DE NEGÓCIO A PARTIR DA TRELLO</b> .....	<b>36</b>
<b>3.1 Análise dos dados presentes em um cartão da Trello</b> .....	<b>36</b>
3.1.1 Cartão .....	36
3.1.2 Nome de um cartão .....	37
3.1.3 Membros de um cartão.....	39
3.1.4 Descrição de um cartão.....	40
3.1.5 Comentários de um cartão .....	42
3.1.6 <i>Checklists</i> .....	43
3.1.7 Menções .....	44
<b>3.2 Métodos para a extração de informações de processo presentes em cartões Trello</b> .....	<b>45</b>
3.2.1 Definição de atividade de processo de negócio a partir de nome e membros e cartão.....	45
3.2.2 Sub-tarefas a partir de <i>checklists</i> .....	47
3.2.3 Informações de processo em textos de linguagem natural.....	49
<b>3.3 Resumo</b> .....	<b>53</b>
<b>4 PROTÓTIPO</b> .....	<b>55</b>
<b>4.1 Desenvolvimento e arquitetura</b> .....	<b>56</b>
<b>4.2 Acesso às informações da Trello</b> .....	<b>59</b>
<b>4.3 Pré-processamento de um cartão da Trello</b> .....	<b>62</b>
<b>4.4 Extração de informações de processo de negócio a partir de um cartão da Trello</b> .....	<b>66</b>
4.4.1 Extração de definição de atividade de cartão .....	68
4.4.2 Extração de sub-tarefas a partir de <i>checklists</i> .....	71
4.4.3 Extração de atividades a partir de texto .....	72
4.4.4 Geração de texto com informações de processo de negócio.....	72
<b>4.5 Exemplo de uso do protótipo</b> .....	<b>73</b>
<b>5 RESULTADOS, AVALIAÇÃO E DISCUSSÃO</b> .....	<b>76</b>
<b>5.1 Metodologia</b> .....	<b>77</b>
<b>5.2 Resultados</b> .....	<b>83</b>
<b>5.3 Discussão</b> .....	<b>92</b>
<b>6 CONCLUSÃO</b> .....	<b>97</b>
<b>6.1 Contribuições</b> .....	<b>97</b>
<b>6.2 Trabalhos futuros</b> .....	<b>98</b>
<b>REFERÊNCIAS</b> .....	<b>100</b>

## 1 INTRODUÇÃO

Processos de negócio são tudo aquilo que organizações, com fins lucrativos ou não, executam quando entregam um serviço ou produto aos seus clientes (DUMAS et al., 2018). Organizações podem se tornar mais competitivas através da elaboração e execução de melhores processos, tanto os internos, como contratações de funcionários, quanto aqueles diretamente ligados à como seus consumidores percebem a qualidade de seus serviços (DUMAS et al., 2018). Para se manter competitivas, organizações têm utilizado técnicas para gerenciar seus processos e, dessa forma, vêm atingindo resultados significativos em termos de custos, agilidade, qualidade e satisfação de clientes (ROSEMANN; BROCKE, 2010).

O Gerenciamento de Processos de Negócio (*Business Process Management - BPM*) é a arte e ciência de supervisionar como o trabalho é realizado para garantir resultados consistentes e aproveitar oportunidades de melhoria. Esta supervisão é aplicada através do Ciclo de Vida de BPM, que define fases (Identificação de processos, Descoberta de processos, Análise de processos, Redesenho de processos, Implementação de processos, Monitoramento de processos) a serem seguidas durante a aplicação de um projeto de BPM. Dentre estas fases destaca-se a *Descoberta de processos*, na qual a equipe de analistas responsável pela aplicação do projeto de BPM documenta e modela os processos identificados conforme são executados naquele momento (DUMAS et al., 2018).

Na aplicação de técnicas para a análise e aprimoramento de processos de negócio, a fase de descoberta de processos é conhecida por ser complexa e demorada devido às inúmeras entrevistas, workshops, pesquisas e análises de documentos que a compõe (AA et al., 2018). Tal coleta de informações é, geralmente, realizada através de entrevistas com especialistas do domínio (*Descoberta baseada em entrevistas*), análises de documentos relacionados ao processo (*Descoberta baseada em evidências*) e/ou oficinas com diversos especialistas de domínio e facilitadas por analistas de processos (*Descoberta baseada em oficinas*). De acordo com Verner (2004), a abordagem que muitos projetos de BPM utilizam para realizar a descoberta de processos é inadequada por assumirem que o processo é pré-conhecido, o que não ocorre na maioria das grandes organizações.

Com o intuito de tornar a descoberta de processos mais ágil e eficaz, trabalhos vêm sendo elaborados nas áreas de mineração de processos (ex.: Aalst et al. (2011)) e no uso de técnicas de Processamento de Linguagem Natural (PLN) (ex.: Silva (2019)),

Rosa (2020), Silva (2019), Friedrich, Mendling and Puhmann (2011), Ferreira, Thom and Fantinato (2017)).

Dentre os trabalhos com foco em PLN, destacam-se os trabalhos de Friedrich, Mendling and Puhmann (2011) e Silva (2019), que propõe técnicas para gerar de forma automática modelos de processos de negócio e respectivas descrições de processo a partir de textos em linguagem natural. Embora ambos os trabalhos extraíam informações de processo a partir de textos em linguagem natural, como descrições de processos e e-mails, tais textos compõe uma minoria das informações disponíveis no contexto de uma organização. De acordo com relatório de Merrill Lynch & Co. (1998), aproximadamente 80% dos dados presentes em uma organização estão em formatos não estruturados como modelos, páginas *web*, apresentações de *slides*, conversas pela internet e etc. Têm-se então que, no contexto da aplicação de PLN em projetos de BPM, que outras fontes de dados, além de textos em linguagem natural, podem ser utilizadas para se auxiliar na fase de descoberta de processos de negócio.

Uma das fontes de dados que podem ser utilizadas são plataformas de comunicação e gestão utilizadas no contexto de organizações. Em seu trabalho, Silva (2019) analisa a plataforma de comunicação empresarial *Slack*<sup>1</sup> como fonte de informações de processos de negócio. Utilizando técnicas de PLN, Silva (2019) cria definições capazes de identificar informações de processos de negócio e extraí-las em formato de texto.

Neste contexto, o presente trabalho estabelece a seguinte hipótese: *A plataforma de gestão de projetos Trello*<sup>2</sup> *pode servir como fonte de informações de processo de negócio*, isto é, seus dados podem ser úteis para a fase de descoberta de processos de negócio. Para verificar esta hipótese, este trabalho analisa a plataforma Trello e suas funcionalidades, estabelecendo relações entre estas e conceitos de PLN e BPM na forma de axiomas. Com base nestes axiomas são elaboradas definições capazes de extrair estas informações na forma de texto em linguagem natural. Estes textos, por sua vez, podem ser utilizados por analistas de processo na descoberta de processos em projetos de BPM. Para validar as definições, um questionário foi aplicado com a participação de analistas de processos. O objetivo do questionário é comparar as informações extraídas pela abordagem proposta no presente trabalho com aquelas extraídas pelos analistas participantes, avaliando a correção das extraídas pela abordagem deste trabalho em relação àquelas extraídas pelos analistas de processo.

---

<sup>1</sup><<https://slack.com/>>

<sup>2</sup><<https://trello.com/>>

## 1.1 Objetivos

Considerando a hipótese estabelecida, o objetivo geral deste trabalho é evidenciar que a plataforma Trello possui informações de processos de negócio disponíveis em seus dados. Como objetivos específicos tem-se:

- i) criar um conjunto de definições capazes de extrair informações de processo da plataforma Trello na forma de texto;
- ii) desenvolver um protótipo que sirva como prova de conceito para estas definições.

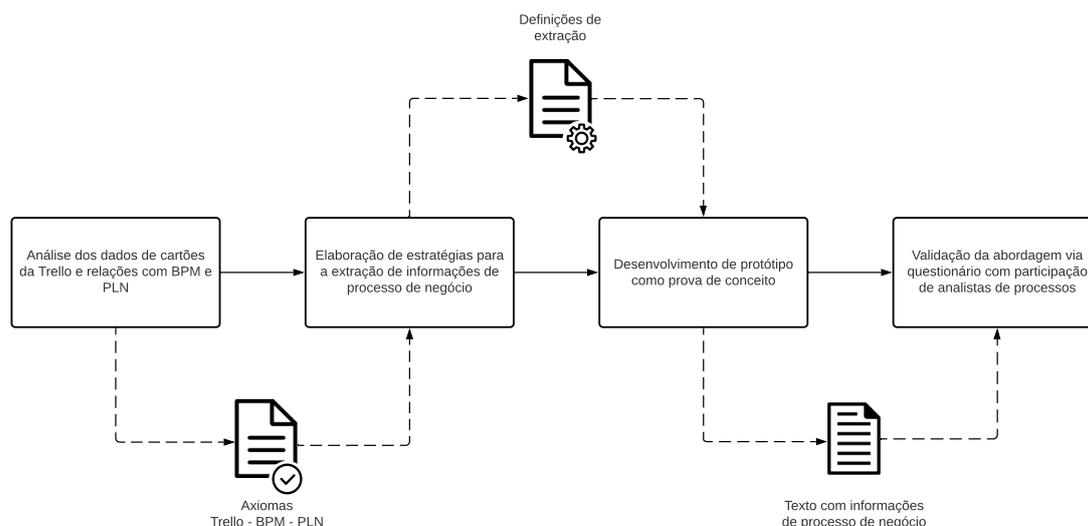
## 1.2 Metodologia de Pesquisa

A metodologia de pesquisa adotada neste trabalho inclui a análise da Plataforma Trello, a formalização de axiomas, o desenvolvimento de um protótipo para prova de conceitos e a aplicação de um questionário de avaliação da abordagem. A metodologia é aplicada da seguinte forma: Primeiro, é feita a análise da Plataforma Trello tendo como foco os dados e funcionalidades disponíveis no contexto de um cartão Trello. Estes dados e funcionalidades são relacionados com conceitos de BPM e PLN a partir de sua documentação oficial e trabalhos que analisem a plataforma. Estas relações são formalizadas em um conjunto de axiomas que servem como entrada para a etapa seguinte, onde são elaboradas definições que habilitam a extração de informações de processo de negócio a partir de cartões da Trello e sua transformação em textos em linguagem natural.

Como prova de conceito, um protótipo em software é desenvolvido, capaz de acessar informações da plataforma Trello através de sua API e aplicar as definições de extração, tendo por saída o texto contendo informações de processo de negócio.

Para avaliar a abordagem, um conjunto de questionários foi aplicado com a participação de analistas de processos, que elaboram descrições e modelos de processos de negócio a partir de cartões da Trello. As descrições e modelos são comparadas com aquelas geradas a partir do protótipo em termos de sua correção. Uma visão geral da metodologia é apresentada na fig. 1.1.

Figura 1.1 – Visão geral da metodologia aplicada neste trabalho



Fonte: Os autores

### 1.3 Organização do texto

O presente trabalho está organizado como segue. O capítulo 2 apresenta os conceitos fundamentais para o correto entendimento do restante do trabalho. Discute, ainda, trabalhos relacionados ao tema. O capítulo 3 traz a metodologia de pesquisa utilizada neste estudo, assim como a análise dos dados disponibilizados pela plataforma Trello, estabelecendo relações entre estes dados e conceitos de BPM com o objetivo de elaborar definições capazes de extrair informações de processos de negócio. O capítulo 4 apresenta o desenvolvimento do protótipo e a aplicação das definições na prática, detalhando a arquitetura do protótipo e o *pipeline* de PLN utilizado. No capítulo 5 são apresentados os resultados de questionário com a participação de analistas de processos para a validação da abordagem proposta. Finalmente, no capítulo 6 o estudo é concluído com discussões acerca da abordagem e propondo pontos de melhoria e ideias para trabalhos futuros que possam estender e complementar a abordagem proposta.

## 2 FUNDAMENTOS E TRABALHOS RELACIONADOS

Este capítulo traz os conceitos fundamentais necessários para o entendimento do restante do trabalho. Primeiro, a seção 2.1 apresenta conceitos inerentes a BPM. Em particular a seção conceitualiza processo de negócio, que elementos compõe um processo de negócio e como é aplicado um projeto de BPM através do Ciclo de Vida de BPM. A seção também define e elenca os principais elementos da Notação e Modelo de Processos de Negócio (*BPMN*). A seção 2.2 descreve PLN, elencando suas principais aplicações, práticas mais comuns em projetos de PLN e as principais ferramentas utilizadas neste trabalho. A seção 2.3 detalha a plataforma Trello e seu funcionamento, descrevendo seus principais elementos e funcionalidades a partir de sua documentação oficial, exemplos de quadros disponibilizados pela plataforma e estudos relacionados. Finalmente, a 2.4 apresenta e discute os principais trabalhos relacionados ao presente estudo, suas contribuições e como este estudo visa contribuir com o tema.

### 2.1 Gerenciamento de Processos de Negocio

Um processo de negócio é composto por uma cadeia de eventos, atividades e pontos de decisão que, quando executados, agregam valor à organização que a executou e a seus clientes (DUMAS et al., 2018).

Um processo envolve também um conjunto de atores (pessoas, organizações ou sistemas), nos quais destacam-se os *participantes do processo*, que são os atores que executam suas atividades no dia-a-dia do processo (DUMAS et al., 2018).

Uma *atividade* é uma ação no contexto do processo composta por passos que são executados do início ao fim da atividade. Quando uma atividade é simples e pode ser vista como uma unidade única de trabalho, ela é chamada de *arefa*.

Um *evento* é um acontecimento atômico, isto é, sem duração, que geralmente desencadeia uma série de atividades. Já um *ponto de decisão* é um momento no processo em que uma decisão deve ser tomada e que influenciará a maneira com o que o restante da execução do processo (DUMAS et al., 2018) se dará.

No contexto de pontos de decisão, é importante definirmos os seus três tipos principais: *Decisões exclusivas*, *decisões paralelas* e *decisões inclusivas*. Uma decisão exclusiva ocorre quando a decisão tomada influencia o processo de maneira que apenas uma dentre um conjunto de alternativas de atividades seja executada no seguimento do pro-

cesso. Uma decisão paralela ocorre quando a decisão tem por consequência a execução de um conjunto de atividades em paralelo, devido à independência dessas atividades entre si. Por fim, uma decisão inclusiva ocorre quando a decisão acarreta na execução opcional de uma ou mais das atividades alternativas (DUMAS et al., 2018).

Para exemplificar os elementos que compõe um processo, considere a descrição de processo a seguir, escrita por Dumas et al. (2018) e traduzida pelos autores deste trabalho. No contexto da descrição, temos como exemplo de eventos: “*reclamação é registrada*”. Como exemplos de atividades: “*redige uma recomendação de solução*” e “*classificar a reclamação*”. Como exemplos de pontos de decisão: “*Se a reclamação estiver marcada como "Não OK", [...]. Se a reclamação for "OK", [...].*”.

### **Texto 1 : Descrição de processo de solução de reclamação**

1 *Depois que uma reclamação é registrada, ela é examinada por um oficial de re-*  
 2 *clamações que então redige uma recomendação de solução. Esta recomendação é,*  
 3 *então, verificada por um oficial sênior de reclamações, o qual pode classificar a re-*  
 4 *clamação como "OK" ou "Não OK". Se a reclamação estiver marcada como "Não*  
 5 *OK", esta será retornada ao oficial de reclamações e a recomendação é refeita. Se a*  
 6 *reclamação for "OK", o processo de solução de reclamações continua.*

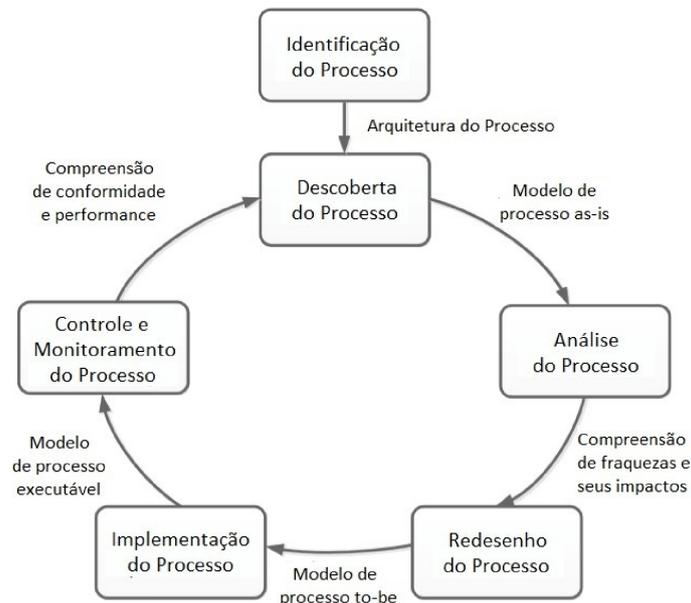
BPM é uma disciplina que consolida e integra as ferramentas, técnicas, métodos e metodologias desenvolvidas nas últimas décadas para prover o suporte a todas as fases do ciclo de vida de um processo de negócio (ROSEMANN; BROCKE, 2010; WESKE, 2007). Utilizando este conhecimento profissionais desenham, analisam, executam e monitoram processos para que estes atinjam os objetivos da organização, no que é denominado *projeto de BPM* (DUMAS et al., 2018).

Algumas das principais características que tornam BPM importante em contextos atuais são a sua interdisciplinaridade, envolvendo diversas áreas em suas aplicações, em especial a *Tecnologia da Informação (IT)*, e o fato de ser tanto ativamente aplicada quanto ativamente estudada, isto é, é uma disciplina em constante aplicação e evolução (DUMAS et al., 2018).

Um projeto de BPM é aplicado a partir da perspectiva do *Ciclo de vida de BPM*. O Ciclo de vida de BPM é composto por 6 fases, representadas visualmente na Fig. 2.1 e detalhadas a seguir, na ordem de sua execução. Na primeira fase, denominada *identificação de processo*, são delimitados quais processos serão aprimorados no projeto de

BPM sendo iniciado. Para isto, um problema presente na organização é escolhido para ser solucionado e os processos relacionados ao mesmo são identificados (DUMAS et al., 2018). Estes mesmos processos são, então, detalhados em termos de seus escopos (onde começam e onde terminam) e em termos de suas relações. Estas relações podem ser de diversos tipos. Uma *relação hierárquica*, por exemplo, ocorre quando um processo é parte de outro. Já uma *relação sequencial* ocorre quando um processo deve acontecer antes ou depois do outro. Ao fim da fase de identificação de processo, têm-se como resultado uma *arquitetura de processos* da organização. Uma arquitetura de processos é uma visão geral e organizada dos processos existentes na organização e as relações que possuem entre eles (POLANČIČ et al., 2020).

Figura 2.1 – O Ciclo de vida de BPM de acordo com Dumas et al. (2018)



Fonte: Dumas et al. (2018), traduzido por Silva (2019)

Após, na fase de *descoberta de processo*, cada processo identificado é detalhado e modelado de acordo com seu estado atual, resultando em um modelo denominado modelo *as-is*. Para que a modelagem possa ser feita é fundamental coletar informações acerca do processo em questão (DUMAS et al., 2018). Tal coleta de informações é, geralmente, realizada através de entrevistas com especialistas do domínio (*descoberta baseada em entrevistas*), análises de documentos relacionados ao processo (*descoberta baseada em evidências*) e/ou oficinas com diversos especialistas de domínio e facilitadas por analistas de processos (*descoberta baseada em oficinas*) (DUMAS et al., 2018).

A fase de descoberta de processo é especialmente importante no contexto deste trabalho. De acordo com Verner (2004), os métodos geralmente utilizados para a descoberta de processos são inadequados, visto que assumem que a organização já tenha um

conhecimento prévio, acurado e detalhado do funcionamento do processo em questão, o que geralmente não ocorre. Verner (2004) cita também que a maior parte do conhecimento acerca dos processos em uma organização estão, geralmente, descentralizados e não formalizados.

Na fase de *análise de processo* os modelos *as-is* criados na fase de descoberta de Processo são analisados em busca de problemas e oportunidades de melhorias. Esta análise é feita utilizando-se de diversas técnicas qualitativas (ex.: análise de valor agregado, análise de causa raiz) e quantitativas (ex.: análise de filas, utilização de indicadores de performance) para identificar e listar elementos indesejados no processo, como desperdícios e ineficácias. Após listados, estes elementos são documentados, estruturados e priorizados de acordo seu impacto no processo e esforço estimado para solução (DUMAS et al., 2018).

Na fase de *redesenho de processos* são propostas mudanças capazes de solucionar os problemas identificados na fase de análise de processos. Cada mudança proposta é analisada usando técnicas de análise de processos e comparada com outras mudanças em termos de métricas de performance. Por fim, as melhores mudanças são escolhidas e combinadas em um novo modelo de processo, denominado modelo *to-be*, que servirá como base para a etapa de implementação de processo (DUMAS et al., 2018).

Na fase de *implementação de processo* as mudanças propostas pelo modelo *to-be* são preparadas e implementadas através da automação de processos por sistemas de TI e de alterações na maneira com que todos os participantes do processo trabalham. Estas alterações geralmente são realizadas, principalmente, com treinamentos dos participantes do processo acerca do novo modo de trabalho (DUMAS et al., 2018).

Por fim, na fase de *monitoramento e controle de processo* o novo processo é acompanhado e mensurado através da coleta de dados sobre seu funcionamento. Estes dados fornecem métricas de performance que são analisadas para se identificar o quão bem o processo está sendo executado em relação aos seus objetivos. Esta análise pode indicar a necessidade de ajustes no processo e até mesmo novos problemas em processos adjacentes. Nestes casos, um novo ciclo de BPM poderá ser iniciado a partir da fase de descoberta de processo (DUMAS et al., 2018).

### 2.1.1 Notação e Modelo de Processos de Negócio

Durante a execução de um projeto de BPM, muitos modelos podem ser gerados para representar visualmente os processos e seus elementos, como os modelos *as-is* e *to-be*. O ato de elaborar modelos de processo é chamado *Modelagem de Processos de Negócio* e é definido como a atividade de representar de forma gráfica processos de negócio de uma organização, possibilitando a análise e aprimoramento do processo. Com a popularização do uso de BPM e Modelagem de Processos de Negócio em organizações, cada vez mais tornou-se necessário estabelecer uma linguagem de modelagem padrão para estas atividades. No presente, o estado da arte em linguagens para modelagem de processos é a *Notação e Modelo de Processos de Negócio (Business Process Model and Notation - BPMN)*<sup>1</sup> (CHINOSI; TROMBETTA, 2012).

A BPMN é a notação padrão para a elaboração de *Diagramas de Processos de Negócio (BPD)* (ROSING et al., 2015; OMG, 2011), mais especificamente a BPMN 2.0, a versão mais recente da notação. Seu objetivo é fornecer uma notação formal, porém compreensível o suficiente para os diversos envolvidos em processos de negócios, desde analistas de processos até os participantes do processo (CHINOSI; TROMBETTA, 2012).

A BPMN 2.0 é composta por uma biblioteca de elementos notacionais disponíveis para utilização em modelos de processo. Esta biblioteca é extensa e a maioria de seus elementos não são relevantes para o escopo do presente trabalho. Por isso serão introduzidos apenas os elementos notacionais relevantes. Os elementos da biblioteca notacional da BPMN 2.0 são categorizados em cinco grupos: (i) Objetos de Fluxo, (ii) Dados, (iii) Objetos Conectivos, (iv) Raias de Piscina e (v) Artefatos (OMG, 2011). Neste trabalho utilizaremos determinados Objetos de Fluxo, Objetos Conectivos e Raias de Piscina. Para referência visual, a fig. 2.2 apresenta os elementos notacionais da BPMN 2.0 descritos neste capítulo.

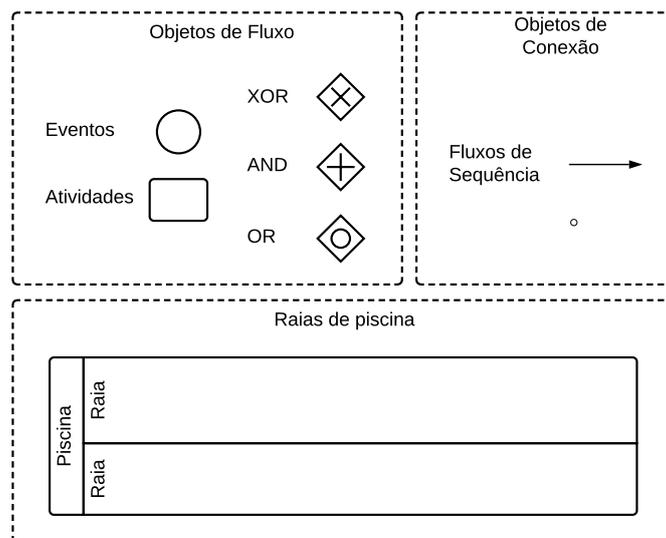
Na categoria Objetos de Conexão, *Fluxos de sequência* são os arcos com flechas em uma de suas extremidades que representam uma relação temporal entre elementos de um modelo de processo. Um *caminho* pode ser definido como um conjunto de objetos de fluxo conectados sequencialmente através de fluxos de sequência.

Na categoria de Objetos de Fluxo, *atividades* e *eventos* correspondem às mesmas definições de BPM para estes elementos de processo. Atividades são representadas por um retângulo com bordas arredondadas contendo no seu interior um rótulo nomeando

---

<sup>1</sup><<https://www.omg.org/bpmn/>>

Figura 2.2 – Principais elementos da BPMN 2.0 utilizados neste trabalho



Fonte: Os autores, adaptado de Silva (2019)

a atividade que este representa. Por convenção, rótulos de atividades são geralmente escritos na forma *verbo + objeto*, por exemplo “Examinar reclamação”, por ser uma forma considerada menos ambígua (MENDLING; REIJERS; AALST, 2010; DUMAS et al., 2018). Eventos são representados por uma circunferência, com seu rótulo geralmente posicionado abaixo desta. Por convenção, rótulos de eventos geralmente são escritos na forma *objeto + verbo no passado*, por exemplo “Reclamação registrada”. Vale ressaltar que o início e o fim de um processo são vistos como eventos e representados pelo seu respectivo elemento notacional em BPMN, com pequenas variações. O evento de início de processo é representado por uma circunferência com bordas finas, enquanto o evento de fim é representado com uma borda grossa (DUMAS et al., 2018).

Ainda na categoria de Objetos de Fluxo, *Gateways* são responsáveis por controlar divergências (separação) e convergências (junção) de fluxos de sequência, levando a diferentes caminhos ou unindo diferentes caminhos. Existem seis tipos de *Gateways* que diferem entre si tanto em lógica quanto em representação notacional. Os principais são: (i) *Gateway Exclusivo (XOR)*, representado visualmente por um losango com um “X” em seu centro, onde apenas um dos caminhos disponíveis poderá ser executado após a tomada de decisão, representando uma decisão exclusiva; (ii) *Gateway Paralelo (AND)*, representado visualmente por um losango com um “+” em seu centro, no qual todos os caminhos possíveis devem ser executados, representando uma decisão paralela; e *Gateway Inclusivo (OR)*, representado visualmente por um losango com um círculo em seu centro,

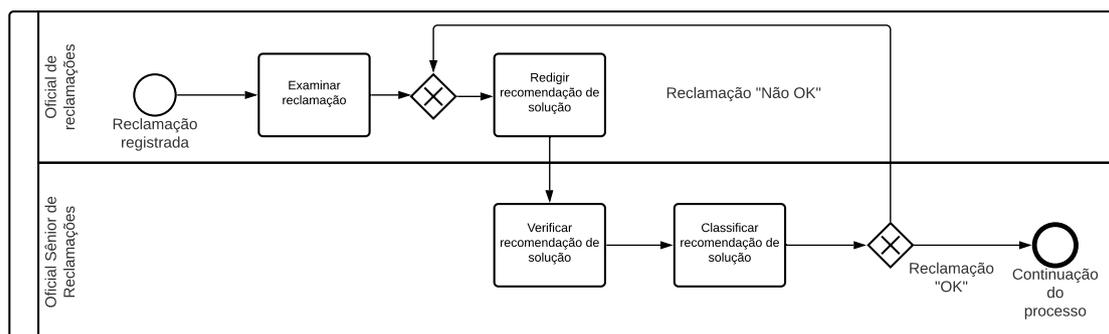
no qual pelo menos um dos caminhos possíveis deve ser executado, representando uma decisão inclusiva (SILVA, 2019).

Cada *gateway* deve ser utilizado em dois casos: quando há a divergência de caminhos e quando estes mesmos caminhos devem ser convergidos. Por convenção, sempre que há um *gateway* de divergência do tipo XOR, os fluxos de sequência que dele saem devem possuir rótulos descrevendo a condição pela qual aquele caminho é executado (DUMAS et al., 2018).

Na categoria Raias de Piscina, uma *Piscina* representa uma classe de recursos ou equipes participantes do processo de negócio, enquanto *Raias* são utilizadas para particionar piscinas em sub-recursos e recursos únicos, de maneira a organizar e categorizar atividades dentro de uma piscina (DUMAS et al., 2018). Geralmente raias são utilizadas para representar cargos e departamentos internos à organização ou sistemas (SILVA, 2019). Tanto raias quanto piscinas são representadas visualmente por retângulos que englobam atividades, eventos, gateways e dados. Geralmente são utilizados na horizontal, com seus rótulos exibidos no lado esquerdo de maneira vertical. Nos casos em que uma piscina é sub-dividida em raias, os rótulos são separados por linhas.

Para exemplificar o uso da BPMN na prática, a Fig 2.3 é um exemplo simples de modelagem de processo utilizando a BPMN 2.0 para modelar a descrição de processo utilizada como exemplo anteriormente, na seção 1.

Figura 2.3 – Exemplo de modelo de processo em BPMN 2.0



Fonte: Os autores

## 2.2 Processamento de linguagem natural

*Linguagem natural* são linguagens que se desenvolveram naturalmente, sem planejamento ou modelagem (LYONS, 1991). São as linguagens naturais que nos possi-

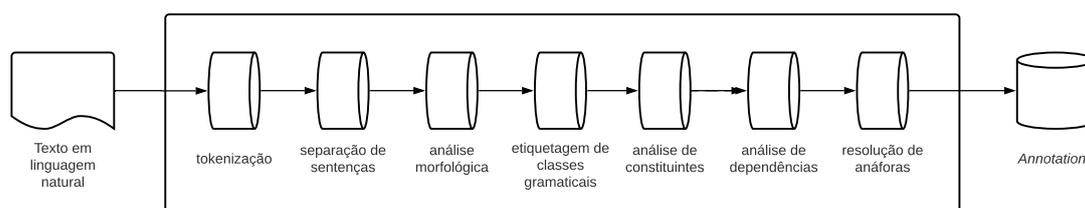
bilitam expressar nossa vontade, trocar pensamentos e documentar nosso conhecimento em fontes escritas (LEOPOLD, 2013). As linguagens utilizadas por seres humanos para sua comunicação, seja escrita ou falada, como Inglês, Hindi e Português, são exemplos bem conhecidos de linguagens naturais, visto que sua evolução ocorre com a passagem de conhecimento de geração em geração e cuja definição com regras explícitas é complexa (BIRD; KLEIN; LOPER, 2009).

Neste contexto, PLN é uma área da Ciência da Computação que fornece métodos computacionais para analisar, modelar, entender e extrair informações de linguagens naturais (BIRD; KLEIN; LOPER, 2009). Devido ao seu uso multidisciplinar, PLN tem se tornado cada vez mais presente no cotidiano, como assistentes pessoais em Smartphones (FRANZ; MILCH, 2002; VAJJALA et al., 2020) e combate a notícias falsas na internet (DALE, 2017). Em BPM, técnicas de PLN vêm sendo utilizadas com diversos fins, como a geração de modelos e extração de informações de processos de negócio a partir de textos.

### 2.2.1 O Pipeline de PLN

Tradicionalmente, aplicações em PLN são vistas da perspectiva de um conjunto de estágios cujas saídas servem como entradas para os estágios seguintes, possibilitando análises e resultados mais complexos e de mais alto nível (INDURKHYA; DAMERAU, 2010). De fato, qualquer aplicação prática de PLN deve executar diversas sub-tarefas e a possibilidade de se utilizarem diversos algoritmos para uma mesma tarefa torna a construção de *pipelines* de PLN uma prática bastante utilizada (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011). Um exemplo de *pipeline* é exibido na Fig. 2.4, que remete a um dos possíveis *pipelines* utilizados pela ferramenta *Stanford CoreNLP* (MANNING et al., 2014).

Figura 2.4 – Exemplo de pipeline de PLN.



Fonte: Os autores. Adaptado do *website* da ferramenta Stanford CoreNLP

Cada aplicação em PLN tem objetivo próprio, podendo utilizar ou não diferentes técnicas para atingir este objetivo, o que acarreta em *pipelines* com diferentes estágios para diferentes aplicações, apesar de muitas compartilharem estágios iniciais. No contexto deste trabalho os principais estágios utilizados são: (i) tokenização, (ii) separação de sentenças, (iii) análise morfológica, (iv) etiquetamento de classes gramaticais (na língua inglesa, “*part-of-speech tagging*”), (vi) análise de constituintes (na língua inglesa, “*constituency parsing*”) e (viii) resolução de anáfora (na língua inglesa, “*anaphora resolution*”).

*Tokenização* se refere à separação de um texto em suas palavras constituintes. A maneira com que o processo de tokenização é realizado depende da língua sendo analisada. No caso deste trabalho, em que a língua analisada é a inglesa, a tokenização é geralmente realizada identificando-se os elementos identificadores de fronteiras entre palavras, como o espaço em branco, e realizando a separação da sentença nestes identificadores (INDURKHYA; DAMERAU, 2010). Cada elemento identificado recebe o nome *token*. Considerando, por exemplo, a frase “*After a claim is registered, it is examined.*”, teríamos como tokens: “*after*”, “*a*”, “*claim*”, “*is*”, “*registered*”, “*,*”, “*it*”, “*is*”, “*examined*” e “*.*”. É importante ressaltar que não apenas o espaço em branco serve como identificador, e que nem só palavras são identificadas como tokens. Pontuações como “*,*” também são tokens identificados por esta técnica.

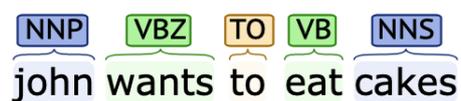
*Separação de sentenças* refere-se à separação de um texto nas sentenças que o compõe. Assim como na Tokenização, a separação de sentenças costuma utilizar identificadores de fronteiras entre sentenças, que em muitas línguas são pontuações específicas, como “*.*”, “*!*” e “*?*” (INDURKHYA; DAMERAU, 2010). De acordo com Indurkha and Damerau (2010), as etapas de Tokenização e Separação de sentenças são cruciais para qualquer aplicação de PLN.

A *Análise morfológica* analisa o texto no nível de uma palavra, preocupando-se com a sua estrutura e variações. Uma das tarefas básicas da análise morfológica é a identificação de variantes morfológicas de uma palavra. O verbo inglês *DELIVER*, por exemplo, possui um conjunto de variações como {*delivers, deliver, delivering, delivered*}. A palavra original de um conjunto de variações, isto é, a forma original da palavra e que geralmente aparece nos dicionários, é chamada *lema* (na língua inglesa, *lemma*). Quando utilizamos técnicas de PLN para realizar a análise morfológica de uma palavra e encontrar o seu lema, chamamos este processo de *lematização* (na língua inglesa, *lemmatization*).

Um exemplo de uso de lematização seria trocarmos uma variação do verbo *deliver* em um texto, como *delivering*, pelo seu lema *deliver* (INDURKHAYA; DAMERAU, 2010).

Classes gramaticais (na língua inglesa, *part-of-speech* ou *POS*) são categorias atribuídas a palavras de acordo com seu comportamento gramatical. O *etiquetamento de classes gramaticais* é considerado um sub-processo da análise morfológica e seu objetivo é classificar cada palavra de uma sentença, atribuindo a ela uma classe gramatical de acordo com um conjunto de etiquetas. Apesar de haverem diferenças entre conjuntos de etiquetas, é consenso entre linguistas que as classes gramaticais *verbo*, *substantivo* e *adjetivo* são a base de qualquer conjunto (INDURKHAYA; DAMERAU, 2010). No contexto deste trabalho utilizaremos o conjunto de etiquetas do Penn Treebank Project<sup>2</sup>, pois ele é o conjunto utilizado pela ferramenta de PLN sobre a qual o protótipo foi construído, a Stanford CoreNLP (MANNING et al., 2014).

Figura 2.5 – Exemplo de classificação gramatical de uma sentença



Fonte: Os Autores, utilizando a ferramenta Stanford CoreNLP

A Fig. 2.5 mostra um exemplo de classificação gramatical utilizando a ferramenta Stanford CoreNLP (MANNING et al., 2014) para a frase da língua inglesa “*john wants to eat cakes*”, retirada de Indurkhya and Damerau (2010). Nela “*john*” e “*cakes*” são identificadas como substantivos (NNP), enquanto “*wants*” é identificada como um verbo na terceira pessoa (VBZ) e “*eat*” é identificada como verbo em sua forma base (VB). Por fim, “*to*” é identificado pela sua própria etiqueta gramatical (TO).

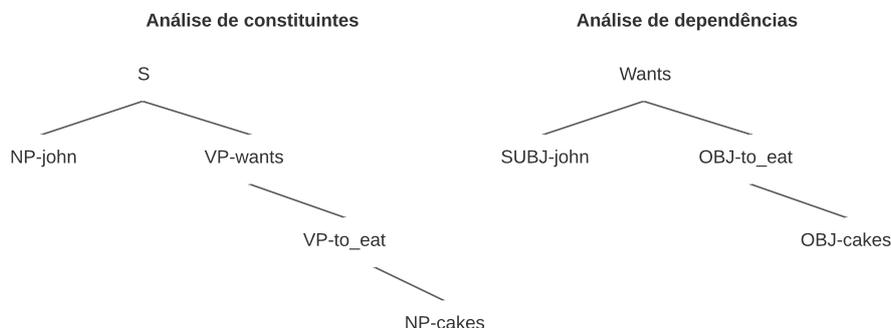
A *Análise de constituintes* e *Análise de dependências* são semelhantes entre si e muitas vezes concorrem pela escolha do desenvolvedor quando do desenvolvimento de uma aplicação em PLN. Ambas analisam a estrutura de uma sentença e a organizam em uma estrutura no formato de grafo. A diferença é que a análise de constituintes organiza seu grafo de acordo com a classificação das orações que a compõe. Já a análise de dependências organiza a sentença de acordo com as relações gramaticais entre as palavras (INDURKHAYA; DAMERAU, 2010).

A Fig. 2.6 traz um exemplo simplificado de análise constituinte e de dependência retirado de Indurkhya and Damerau (2010) para a mesma frase utilizada no exemplo de classificação gramatical. No caso da análise de constituintes a sentença é dividida em uma oração substantiva (na língua inglesa, *Noun-phrase*), simbolizada pelo *NP* em *VP*-

<sup>2</sup><[https://www.ling.upenn.edu/courses/Fall\\_2003/ling001/penn\\_treebank\\_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)>

*john*, seguida de duas orações verbais (na língua inglesa, *Verb-phrase*) em *VP-wants* e *VP-to\_eat*, e terminando em uma nova oração substantiva, *NP-cakes*.

Figura 2.6 – Exemplo comparativo entre análise de constituintes e análise de dependências



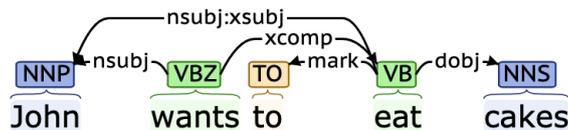
Fonte: Os autores, adaptado de exemplo de Indurkha and Damerau (2010)

Já na análise de dependências a sentença é dividida de acordo com as relações gramaticais entre as palavras. Tomando, por exemplo, o verbo principal, *Wants*, o substantivo *john* é seu sujeito e é assim marcado pelo prefixo *SUBJ* em *SUBJ-john*. Já *to eat* e *cakes* são marcados como objetos do verbo principal pelo prefixo *OBJ* em *OBJ-to\_eat* e *OBJ-cakes*.

Entretanto, análises de constituintes e dependências reais costumam ter como saída grafos maiores e com as informações melhor distribuídas entre seus nodos. Para exemplificar um uso real, as fig. 2.7 e 2.8 exibem, respectivamente, um exemplo de análise de dependência e análise constituinte para a mesma frase utilizada no exemplo da fig. 2.6. Estas, porém, foram geradas pela ferramenta *Stanford CoreNLP*.

Na análise de dependências, exibida na fig. 2.7, podemos observar as relações gramaticais entre as palavras, como “*John*” sendo o sujeito (*subj*) de “*wants*” e “*cakes*” o objeto de “*eat*”. Já na análise de constituintes, na fig. 2.6, não temos as relações entre as palavras, mas sim a classificação gramatical destas e das sentenças, como a palavra “*John*” que é classificada como um substantivo próprio (NNP) em uma frase nominal (NP).

Figura 2.7 – Exemplo real de análise de dependências

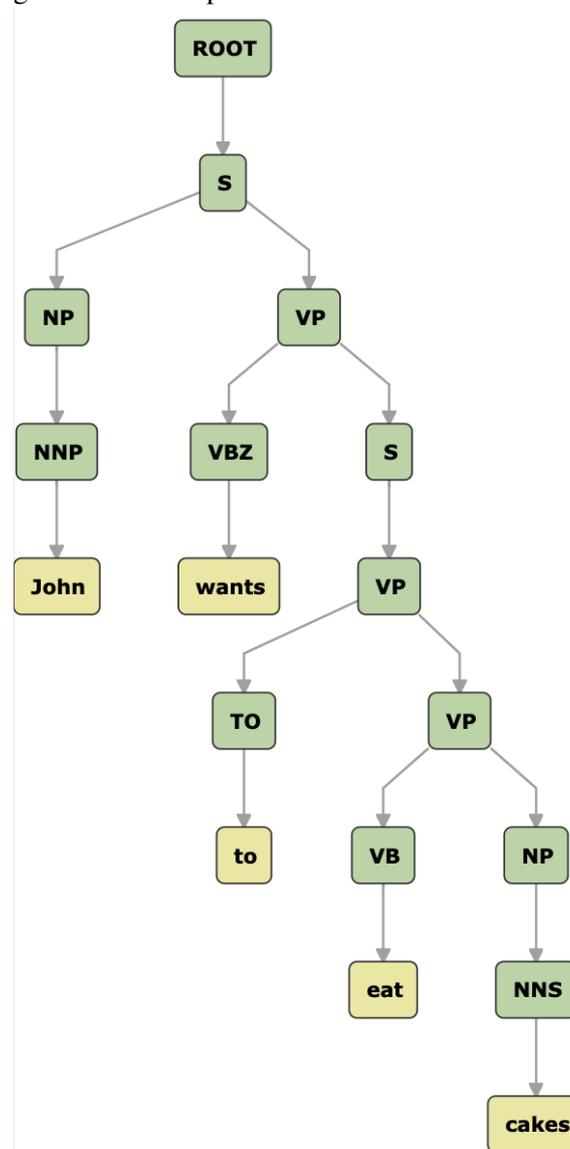


Fonte: Os autores, utilizando a ferramenta *Stanford CoreNLP* (MANNING et al., 2014)

Por fim, *Resolução de anáfora* é uma técnica em PLN utilizada para determinar a palavra original à qual um pronome se refere (MITKOV, 1999). Tomando, por exemplo,

a frase “Depois que uma reclamação é registrada, ela é examinada por um oficial de reclamações”, a resolução de anáfora permite que a aplicação identifique que o pronome “ela” se refere à palavra “reclamação”.

Figura 2.8 – Exemplo real de análise de constituintes



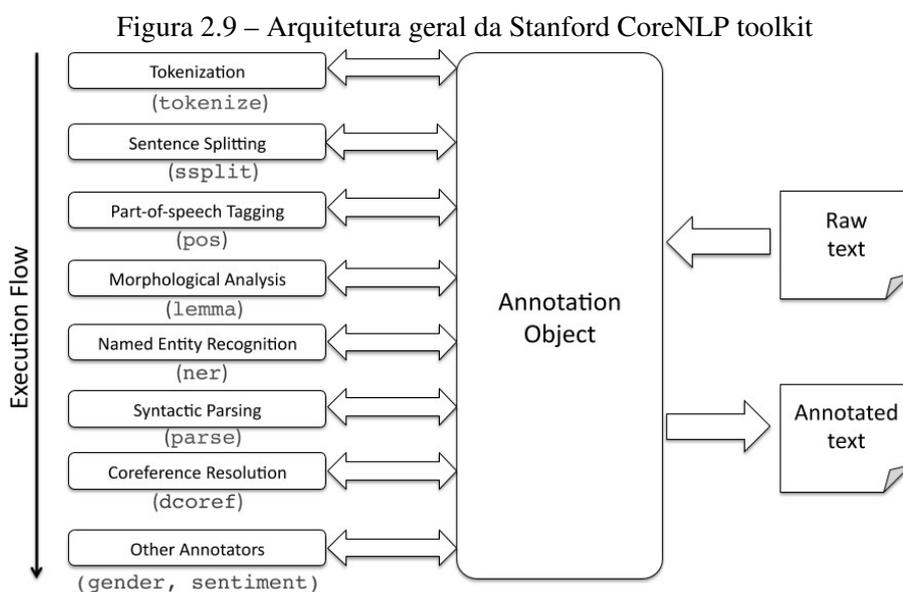
Fonte: Os autores, utilizando a ferramenta Stanford CoreNLP (MANNING et al., 2014)

### 2.2.2 Stanford CoreNLP

Dentre as diversas ferramentas para o desenvolvimento de aplicações em PLN, a *Stanford CoreNLP toolkit* é uma das mais utilizadas, tanto em ambientes comerciais quanto em pesquisas acadêmicas. A ferramenta é construída com base em conceito de *Annotators*, que são pequenas aplicações capazes de analisar e adicionar informações a

um texto. Uma das vantagens de se utilizar a Stanford CoreNLP é a facilidade de se criar pipelines customizados utilizando a sua API em Java. Com esta API é possível definir quais estágios do pipeline deseja-se utilizar e, assim, criar um pipeline de PLN específico para os objetivos da aplicação em questão (MANNING et al., 2014). Vale ressaltar que existem algumas dependências entre certos estágios do pipeline, o que torna obrigatório o uso de um estágio caso se queira utilizar outro que o tenha como dependência.

A Fig. 2.9 representa uma visão geral da arquitetura da Stanford CoreNLP de acordo com Manning et al. (2014). O pipeline de execução, representado pelo conjunto de caixas contendo *Annotators*, como *Tokenization* e *Sentence splitting*, exibe alguns dos estágios de pipeline disponíveis pela ferramenta. No contexto deste trabalho os mais importantes são (i) *tokenize*, responsável pela tokenização; (ii) *ssplit*, responsável pela separação de sentenças; (iii) *pos*, responsável pelo etiquetamento de classes gramaticais; (iv) *lemma*, responsável pela análise morfológica e variações morfológicas; (v) *parse* e *depparse*, responsáveis pela análise de constituintes e análise de dependências; e (vi) *dcoref*, responsável pela resolução de anáfora.



Fonte: Manning et al. (2014)

Um dos principais objetivos do desenvolvimento do Stanford CoreNLP toolkit era tornar simples de configurar e executar pipelines de processamento. Apesar de inicialmente escrito como uma API em Java, iniciativas em código aberto possibilitaram seu uso em outras linguagens de programação como Ruby<sup>3</sup> e Python<sup>4</sup> (MANNING et al., 2014).

<sup>3</sup><<https://github.com/louismullie/stanford-core-nlp>>

<sup>4</sup><<https://github.com/stanfordnlp/python-stanford-corenlp>>

### 2.3 Gestão de projetos e plataforma Trello

Projeto é um esforço temporário empreendido para criar um produto, serviço ou resultado único (The Project Management Institute, 2017). Gestão de projetos é a aplicação de conhecimentos, habilidades, ferramentas e técnicas a atividades de projetos para atingir os objetivos do projeto (The Project Management Institute, 2017). Dentre as ferramentas mais utilizadas por organizações estão os softwares de gestão de projetos (WHITE; FORTUNE, 2002), cuja utilização é estimada em 90% (LIBERATORE; POLLACK-JOHNSON, 2003).

Um exemplo de software de gestão de projetos é a *Trello*<sup>5</sup>, uma plataforma *web* que permite a mais de 50 milhões de usuários ao redor do mundo gerenciar projetos de maneira fácil, flexível e visual (TRELLO, 2020b). A plataforma está disponível em três versões: gratuita, *Business Class* e *Enterprise*. As versões *Business Class* e *Enterprise* são pagas e oferecem um conjunto expandido de funcionalidades, segurança e suporte (KAUR, 2018; SOARES, 2017).

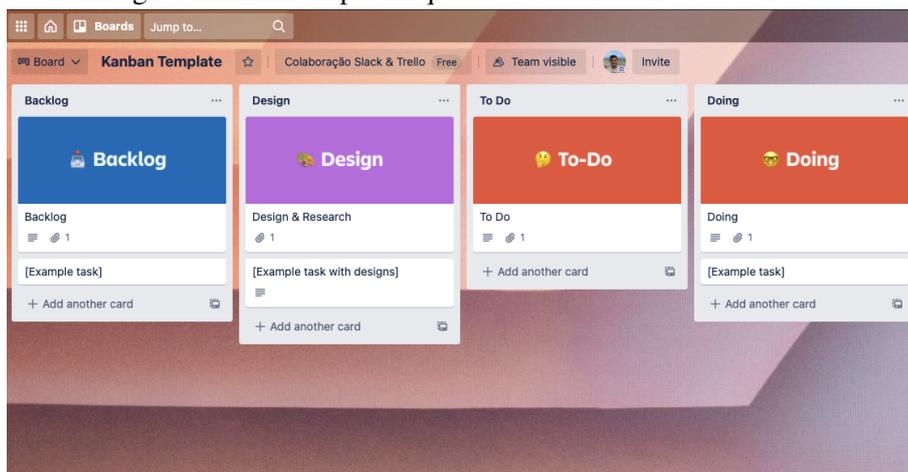
A Trello se inspira na metodologia de gestão KanBan para seu funcionamento. Criada pela Toyota na década de 60, a metodologia Kanban utilizava quadros e cartões para realizar o controle de estoques e produção de empresa (SILVA; ANASTÁCIO, 2019). Em quadros Kanban é possível dividir todos passos necessários para se completar um projeto em cartões contendo uma descrição da tarefa a ser feita, quem é responsável pela sua realização e outras informações (KAUR, 2018). Uma das principais vantagens do uso desta metodologia é a visibilidade que esta dá ao fluxo de informações (SILVA; ANASTÁCIO, 2019).

Da mesma maneira que a metodologia Kanban, a Trello também se organiza com quadros e cartões para permitir a gestão de projetos. Após cadastrar-se na plataforma, um usuário pode criar um número ilimitado de quadros (JOHNSON, 2017). Quadros são compostos por um conjunto de listas, ordenadas visualmente uma ao lado da outra e que podem ser movidas e posicionadas livremente pela tela. Estas listas, por sua vez, são compostas por um conjunto de cartões, ordenados visualmente na vertical e que também podem ser posicionados livremente na tela, desde que estejam dentro de uma lista (KAUR, 2018). A Fig. 2.10 exibe um exemplo de quadro da Trello, contendo 4 listas: “*Backlog*”, “*Design*”, “*To Do*” e “*Doing*”. Todas as listas contém dois cartões, com exceção da lista “*To Do*”, que contém apenas um.

---

<sup>5</sup><<https://trello.com/>>

Figura 2.10 – Exemplo de quadro da Trello contendo 4 listas



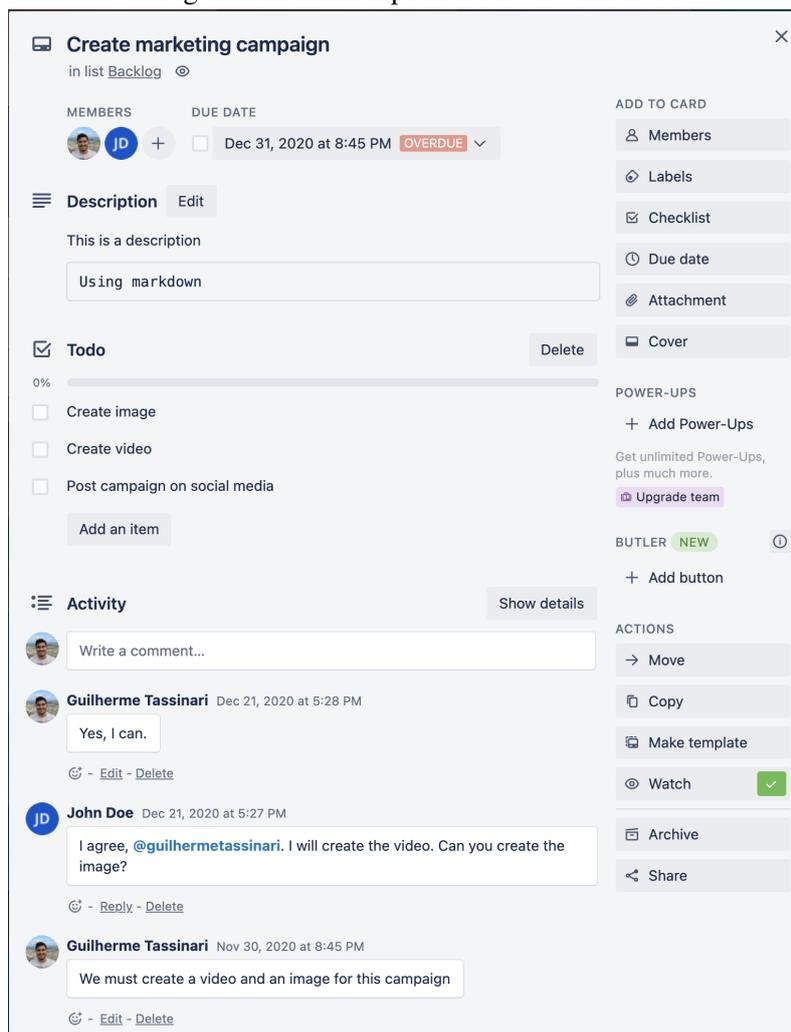
Fonte: Trello Kanban Template<sup>6</sup>

A maior parte das informações da Trello estão contidas dentro de um cartão. Um cartão contém, dentre outras informações, um nome, uma descrição, um ou mais usuários responsáveis, um conjunto de *checklists* contendo sub-tarefas relacionadas ao cartão e comentários deixados pelos usuários com acesso ao cartão. Algumas destas informações possuem peculiaridades e funcionalidades relacionadas que valem ser ressaltadas.

Enquanto o nome de um cartão é composto por um texto simples, a descrição e os comentários possuem funcionalidades que oferecem mais liberdade de escrita aos usuários. Em ambos é possível utilizar a linguagem de marcação *Markdown* (GRUBER, 2004) para formatar os textos (TRELLO, 2020a). Além disso, é possível mencionar outros usuários através de menções (TRELLO, 2021b) utilizando o formato “@ + nome de usuário”, o que irá notificar o usuário mencionado e permitir que outros usuários acessem o perfil deste ao clicar na menção.

Na Fig. 2.11 é exibido um exemplo de cartão Trello. O nome do cartão é exibido no topo deste, onde é possível ler “*Create marketing campaign*”. Abaixo do nome temos os usuários responsáveis pelo cartão, representados pelas suas fotos de perfil ou iniciais de seus nomes. Em seguida temos a descrição do cartão utilizando a linguagem *Markdown* junto com texto não formatado. Logo abaixo temos uma checklist de nome “*Todo*” com três sub-tarefas: “*Create image*”, “*Create video*” e “*Post campaign on social media*”. Por fim temos a seção “*Activity*” onde são exibidos os comentários deixados pelos usuários, com destaque para menção “@*guilhermetassinari*” ao usuário Guilherme Tassinari no comentário do usuário John Doe.

Figura 2.11 – Exemplo de cartão da Trello



Fonte: Os autores

## 2.4 Trabalhos relacionados

Não é novidade que a descoberta de processos é uma tarefa longa e custosa, principalmente em grandes organizações, devido aos métodos manuais necessários para sua execução, como entrevistas e workshops (DUMAS et al., 2018; AA et al., 2018). Sabendo desta dificuldade, diversos trabalhos vêm sendo realizados com o objetivo de tornar a descoberta de processos mais ágil e eficaz. Muitos destes esforços se concentram em propor abordagens para automatizar etapas deste processo, como a geração de modelos de processo, utilizando técnicas de PLN.

Nesta seção revisamos os principais estudos com foco no uso de PLN em BPM e respectivas contribuições para a fase de descoberta de processos. Apesar de haverem outros estudos com estas temáticas, esta fora do escopo deste trabalho realizar uma listagem exaustiva destes, portanto, nos centramos na revisão dos principais trabalhos relacionados

ao presente estudo. Estes trabalhos estão listados na tabela 2.1, ordenados conforme seu ano de publicação e são discutidos nos próximos parágrafos.

Para a melhor compreensão da relação entre os trabalhos discutidos, abordamos, tais trabalhos, na seguinte ordem:

- (i) estudos que estabelecem as bases e os desafios do uso de PLN no contexto de processos de negócio (Aa et al. (2018) e Leopold (2013));
- (ii) trabalhos que estabelecem estratégias para a análise e extração de informações de processos de negócio utilizando PLN (Ferreira, Thom and Fantinato (2017));
- (iii) trabalhos que aplicam estas estratégias em um nível básico de extração (Silva (2019) e Rosa (2020)) e;
- (iv) trabalhos mais complexos, que unem diversas técnicas em PLN com o intuito de se ter uma solução específica do contexto de BPM (Friedrich, Mendling and Puhmann (2011) com modelos de processos e Silva (2019) com descrições de processo consistentes).

Concluimos a seção com uma breve discussão do presente trabalho e de sua contribuição no contexto dos estudos apresentados.

O trabalho de Leopold (2013) possui diversas contribuições para o estudo do uso de PLN em BPM. Dentre eles destacam-se técnicas propostas para a anotação de rótulos em modelos de processo de acordo com o estilo de rotulação, detecção e correção de violações de diretrizes de escrita em modelos de processo e geração de textos em linguagem natural a partir de modelos de processo. Entretanto, no contexto deste trabalho, as observações de Leopold (2013) quanto aos papéis e a importância de PLN no Ciclo de Vida de BPM são especialmente importantes. Leopold (2013) identifica utilidades do uso de PLN em todas as fases do Ciclo de Vida de BPM, com destaque para a fase de Descoberta de processos, onde argumenta que as técnicas frequentemente utilizadas nesta fase são baseadas na análise de linguagem natural, como entrevistas e análises de documentos. Leopold (2013) também cita que PLN pode ser usado na fase de Identificação de processos para automatizar a busca por atividades chave e relações entre processos, tarefa esta que é geralmente realizada também pela análise de documentos de maneira manual e custosa.

Em seu trabalho, Aa et al. (2018) corrobora com as observações de Leopold (2013), inclusive citando-o em seus argumentos. Aa et al. (2018) coloca que BPM, em suas análises clássicas, frequentemente não considera o processamento automático de dados em texto como alternativa. No entanto, técnicas como esta vêm sendo destacadas em

diversos estudos recentes e baseados em PLN. Aa et al. (2018) argumenta que ferramentas como a *Stanford Core* (MANNING et al., 2014) podem ser utilizadas para se extrair diversas informações de processo a partir de textos e documentos e aponta que técnicas relacionados ao uso de PLN na fase de descoberta de processos podem ser aplicadas em diferentes granularidades, como a transformação de descrições de processo em modelos de processos e a anotação e inferência de textos, dentre outros.

Em seu estudo, Ferreira, Thom and Fantinato (2017) propõe um conjunto de 33 regras capazes de identificar elementos de processo em textos em linguagem natural. As 33 regras estão divididas de acordo com os elementos de processos que são capazes de identificar, sendo estas atividades, eventos, *gateways* paralelos, *gateways* exclusivos e raias de piscina. Cada regra é composta por um conjunto ordenado de elementos de textos, que podem variar entre classes gramaticais, relações gramaticais, palavras específicas (palavras chave) e até mesmo outras regras.

Entrando em estudos mais práticos, o recente trabalho de Rosa (2020) utiliza o trabalho de Friedrich, Mendling and Puhmann (2011) como base para desenvolver um conjunto de algoritmos capazes de identificar elementos de processos em textos em linguagem natural e marcá-los visualmente. A abordagem propõe diretrizes de escrita de descrições de processos que terão melhores resultados quando utilizadas com os algoritmos. O processo de identificação inclui etapas de tratamento do texto de entrada, extração de elementos de processo, identificação destes elementos dentro do texto e sua posterior marcação visual.

A maior parte dos estudos apresentados utilizam técnicas de PLN para identificar elementos de processo em textos em linguagem natural, mas o trabalho de Silva (2019) se diferencia ao utilizar técnicas semelhantes para extrair elementos de dados da plataforma de comunicação empresarial *Slack*<sup>7</sup>. Silva (2019) argumenta que informações de processo não estão presentes apenas em documentos e textos, mas também em sistemas de informação utilizados em organizações. Logs de eventos destes sistemas costuma ser utilizados por técnicas como Mineração de Processos (AALST et al., 2011; AALST, 2012) na busca por informações de processo, mas outros dados podem ser utilizados para a aplicação de técnicas de PLN e extração de informações complementares.

Silva (2019) estabelece conjuntos de definições que pré-processam e extraem informações de processo a partir de conversas entre funcionários e metadados relacionados presentes na Slack. Estas conversas em muito se assemelham com comentários em car-

---

<sup>7</sup><https://slack.com/>

Tabela 2.1 – Trabalhos relacionados

<i>Autores</i>	<i>Tópico de estudo</i>
Friedrich, Mendling and Puhlmann (2011)	Geração automática de modelos de processo a partir de textos em linguagem natural.
Leopold (2013)	Fundamentos e técnicas para a aplicação de PLN na análise de documentos em busca da descoberta de atividades, decisões críticas e relações entre processos.
Ferreira, Thom and Fantinato (2017)	Identificação de elementos de processo em textos em linguagem natural.
Aa et al. (2018)	Discussão acerca dos desafios, vantagens e limitações da aplicação de PLN em BPM.
Silva (2019)	Geração de textos orientados a processos a partir de textos em linguagem natural.
Silva (2019)	Análise da plataforma <i>Slack</i> como fonte de informações de processos de negócio.
Rosa (2020)	Abordagem para anotação e marcação visual de elementos de processo em descrições de processo.

Fonte: Os Autores

tões Trello e, por isso, algumas definições de Silva (2019) foram reutilizadas ou adaptadas para o contexto da Trello, como o ajuste de pontuações. Quando comparada à abordagem de Silva (2019), o presente trabalho se aprofunda ao estabelecer primeiro uma base teórica na forma de axiomas e por aproveitar o maior conjunto de dados presentes na Trello fora os comentários, além de considerar a presença de textos formatados em campos de texto, algo que não foi explorado pelo trabalho de Silva (2019), apesar de a Slack utilizar a mesma linguagem de formatação.

Por fim, partindo para os trabalhos mais complexos, Silva (2019) elabora uma arquitetura orientada a serviços que possibilita a análise de descrições de processo em língua natural inglesa e sua transformação no que chama de “*soundness process description*”, ou “*descrição de processo consistente*”, em tradução livre. Uma descrição de processo consistente, conforme Silva (2019) refere-se a um processo consistente, no qual

há pelo menos um caminho de término, que ao fim do processo nenhuma atividade ainda esteja em execução e que não há caminhos que nunca são executados.

O trabalho de Friedrich, Mendling and Puhmann (2011) é considerado o estado da arte quando o assunto é a extração automática de modelos de processos a partir de textos em linguagem natural. Utilizando-se da *Stanford Parser*, ferramenta predecessora a *Stanford CoreNLP*, e um conjunto de outras técnicas e ferramentas de PLN, incluindo seu próprio algoritmo de resolução de anáfora, os autores propõe uma abordagem capaz de identificar elementos de processo e suas relações e posteriormente gerar modelos de processo a partir destes elementos. A abordagem, quando avaliada frente a um conjunto de 47 pares de textos e modelos de processo, atingiu uma média de aproximadamente 77% de correção nos modelos gerados.

Observando os estudos apresentados e suas relações é possível percebermos a tendência da utilização de textos em linguagem natural como entrada de diversos trabalhos ((FRIEDRICH; MENDLING; PUHLMANN, 2011; ROSA, 2020; SILVA, 2019)), o que é razoável dadas as técnicas clássicas de descoberta de processos, fortemente baseadas em linguagem natural, e a alta disponibilidade de recursos em texto no contexto de organizações, como *e-mails*, manuais, descrições de processo e páginas *web*.

Entretanto, informações de processo também estão presentes em outras fontes de informação como sistemas de informação utilizados no contexto da organização. Estes sistemas já vêm sendo utilizados para a descoberta de processos através de mineração de processos e suas análises de logs de evento (AA et al., 2018), mas são passíveis, também, de serem analisadas por técnicas de PLN, como no trabalho de Silva (2019) e sua análise da plataforma Slack.

Neste contexto, o presente trabalho busca contribuir ao analisar a plataforma Trello, identificando quais elementos de processo podem ser extraídos de seus dados. O presente trabalho propõe uma abordagem para a extração dessas informações e sua transformação em texto passível de ser utilizado por analistas de processo em projetos de BPM e como entrada em trabalhos como os de Friedrich, Mendling and Puhmann (2011), Silva et al. (2018) e Rosa (2020), contribuindo para as etapas de identificação de processo e, especialmente, descoberta de processo.

### **3 EXTRAÇÃO DE INFORMAÇÕES DE PROCESSO DE NEGÓCIO A PARTIR DA TRELLO**

Neste capítulo são analisadas as funcionalidades presentes em cartões da Trello, com foco na busca de informações de processos de negócio e maneiras de extrair estas informações. A seção 3.1 detalha os principais dados presentes em cartões da Trello, abordando como são utilizados e qual é o uso padrão e sugerido pela documentação oficial, bem como em outros estudos e análises da plataforma. Também são estabelecidas relações entre estes padrões de uso, fundamentos de processos de negócios e conceitos de linguagem natural, com o objetivo de elaborar axiomas que permitam uma maior compreensão destas relações e suas consequências. Na seção 3.2, a partir das relações estabelecidas, são criadas definições que possibilitem a extração das informações de processo de maneira algorítmica e sua posterior consolidação em texto em linguagem natural.

#### **3.1 Análise dos dados presentes em um cartão da Trello**

Para que possamos identificar que informações de processo estão presentes na plataforma Trello, precisamos analisar os dados disponibilizados por esta e estabelecer relações com fundamentos de BPM e PLN. Isto pode ser feito com um estudo da plataforma, sua documentação e de trabalhos que a citem, procurando por relações entre a definição das suas funcionalidades e conceitos de BPM e PLN.

Apesar de a Trello possuir diversas estruturas (equipes, quadros, listas, cartões e etc), esta análise se limitará ao estudo de cartões Trello como estruturas individuais. O principal motivo para esta escolha é o fato de que, dentre todas as estruturas da Trello, cartões são aquelas que concentram a maior parte das informações da plataforma, em especial devido à alta interação entre usuários através de suas descrições, comentários e menções.

##### **3.1.1 Cartão**

Muitos dos guias e trabalhos que citam a plataforma Trello definem cartões como comumente utilizados para representar tarefas e ideias (TRELLO, 2021e). Kaur (2018), por exemplo, coloca que a metodologia KanBan, na qual a Trello se baseia, possibilita

gerenciar tarefas futuras, presentes e os responsáveis por tais tarefas. Ainda mais, Kaur (2018) também define a Trello como uma plataforma organizada entorno de um quadro virtual onde é possível visualizar e gerenciar tarefas. Um dos guias de primeiros passos da Trello, inclusive, recomenda criar um cartão para cada tarefa que precisa ser executada (TRELLO, 2021c). A relação entre cartões e tarefas é um padrão que se em diversos outros trabalhos, como Johnson (2017), Soares (2017), e na documentação da plataforma (TRELLO, 2020b). Levando-se em consideração este padrão, podemos assumir assumir que cartões representam atividades ou tarefas no contexto de BPM, o que nos leva ao axioma 1.

**Axioma 1** *Cartões geralmente representam atividades e/ou tarefas de processo de negócio.*

### 3.1.2 Nome de um cartão

A documentação oficial da Trello não fornece uma definição ou boas práticas para nomes de cartões. Entretanto, ela inclui exemplos de nomes de cartões, além de quadros exemplo que podem ser reutilizados por usuários como *templates* para criarem seus próprios quadros de maneira mais ágil. Analisando estes exemplos podemos inferir padrões de uso de nomes de cartão e estabelecer relações com BPM.

No seu guia de como criar um quadro e adicionar cartões a este, a documentação da Trello fornece dois exemplos de nomes de cartões: “*Pick up the dry cleaning*” e “*Write a blog post*” (TRELLO, 2021c). Em uma breve análise destes exemplos é possível perceber o formato de texto “*verbo + objeto*”, que é citado por Dumas et al. (2018) como padrão de uso para rótulos de atividades. Também, apesar de não termos o sujeito da frase, a forma se assemelha à regra 1 de Ferreira, Thom and Fantinato (2017) para identificação de atividades de processo de negócio em textos em linguagem natural (“*sujeito + verbo + objeto*”).

Também contribuem para a análise os quadros exemplo fornecidos e divulgados pela Trello<sup>1</sup> para auxiliar e inspirar seus usuários no uso da plataforma. Os quadros são muito diversos e exibem diferentes padrões de escrita para nomes de cartões. No entanto, especialmente em quadros na categoria gestão de projetos<sup>2</sup>, como o *Kanban Workflow*<sup>3</sup>

<sup>1</sup><<https://trello.com/templates>>

<sup>2</sup><<https://trello.com/templates/project-management>>

<sup>3</sup><<https://trello.com/b/4wddd1zf/kanban-workflow>>

e *Team Tasks*<sup>4</sup>, é possível observar o uso dos padrões descritos por Dumas et al. (2018) e Ferreira, Thom and Fantinato (2017). Levando-se em consideração o axioma 1 e os padrões de escrita para nomes de cartões, podemos assumir que, nos casos em que o nome de cartão estiver no formato definido por Dumas et al. (2018), este expressa um rótulo da atividade representada pelo cartão, o que nos leva ao axioma 2.

**Axioma 2** *Um nome de cartão é, geralmente, escrito na forma “verbo + objeto” e expressa um rótulo da atividade/tarefa de processo de negócio representada pelo cartão.*

Tabela 3.1 – Resumo dos axiomas elaborados a partir da análise da plataforma Trello - Parte 1 de 2

<i>Nº</i>	<i>Axioma</i>
1	Cartões geralmente representam atividades e/ou tarefas.
2	Um nome de cartão é, geralmente, escrito na forma “ <i>verbo + objeto</i> ” e expressa um rótulo da atividade/tarefa representada pelo cartão.
3	Os membros atribuídos a um cartão são os participantes de processo responsáveis pela execução da tarefa representada pelo cartão.
4	Como texto livre em linguagem natural, a descrição de um cartão pode conter diversas informações de processo em seu conteúdo.
5	Pronomes pessoais na 1ª pessoa do plural na descrição do cartão se referem, geralmente, aos membros deste mesmo cartão.
6	Em descrições de cartões da Trello, a formatação em <i>Markdown</i> para alterações visuais, como negrito e itálico, não altera o conteúdo do texto, que pode ser analisado como texto em linguagem natural na busca por informações de processo de negócio.
7	Em descrições de cartões da Trello, seções de código em <i>Markdown</i> , por não se caracterizarem como texto em linguagem natural, fogem ao escopo do presente estudo e devem ser ignoradas em suas análises.
8	Em descrições de cartões da Trello, <i>links</i> rotulados devem ter apenas os seus rótulos analisados por informações de processo de negócio, visto que o <i>link</i> não se caracteriza como texto em linguagem natural, fugindo ao escopo do presente estudo.
9	Como texto livre em linguagem natural, comentários de um cartão podem conter diversas informações de processo em seu conteúdo.
10	Pronomes pessoais na 1ª pessoa do plural em comentários se referem, geralmente, aos membros do cartão em questão.

Fonte: Os Autores

<sup>4</sup><<https://trello.com/b/I8X0Ed9y/team-tasks>>

Tabela 3.3 – Resumo dos axiomas elaborados a partir da análise da plataforma Trello - Parte 2 de 2

<i>Nº</i>	<i>Axioma</i>
11	Em comentários de cartões da Trello, a formatação em <i>Markdown</i> para alterações visuais, como negrito e itálico, não altera o conteúdo do texto, que pode ser analisado como texto em linguagem natural na busca por informações de processo de negócio.
12	Em comentários de cartões da Trello, seções de código em <i>Markdown</i> , por não se caracterizarem como texto em linguagem natural, fogem ao escopo do presente estudo e devem ser ignoradas em suas análises.
13	Em comentários de cartões da Trello, <i>links</i> rotulados devem ter apenas os seus rótulos analisados por informações de processo de negócio, visto que o <i>link</i> não se caracteriza como texto em linguagem natural, fugindo ao escopo do presente estudo.
14	Pronomes pessoais na 1ª pessoa do singular em comentários se referem, geralmente, ao autor deste mesmo comentário.
15	<i>Checklists</i> de um cartão são conjuntos sub-tarefas da atividade do cartão, que devem ser executadas para que a atividade do cartão seja completa. Estas sub-tarefas se relacionam entre si ou por sequencialidade ou por paralelismo.
16	Uma menção em comentários ou descrições pode ser interpretada como uma anáfora para o membro mencionado.

Fonte: Os Autores

### 3.1.3 Membros de um cartão

Uma das práticas da metodologia Kanban é a atribuição de membros da organização às tarefas que devem ser executadas para que o projeto tenha continuidade (KAUR, 2018). Na Trello também é possível atribuir membros do quadro a cartões, fazendo com que sejam notificados de diversas mudanças que ocorrerem no contexto dos cartões a que forem adicionados (SOARES, 2017).

A ação de atribuir membros é, inclusive, descrita na documentação da Trello como “atribuir membros a tarefas” (TRELLO, 2021c). É razoável, a partir destas informações, assumirmos que os membros atribuídos à um cartão da Trello sejam os responsáveis pela execução da tarefa representada por este cartão. Dumas et al. (2018) define que os atores de um processo responsáveis pela execução de suas atividades são chamados *participantes de processo*. A partir desta definição de Dumas et al. (2018) e da relação de membros atribuídos a um cartão com a tarefa representada por este cartão, podemos assumir que

estes membros são os participantes de processo responsáveis pela atividade do cartão, o que nos leva ao axioma 3.

**Axioma 3** *Os membros atribuídos a um cartão são os participantes de processo responsáveis pela execução da tarefa representada pelo cartão.*

### 3.1.4 Descrição de um cartão

A descrição de um cartão é um espaço em texto para o usuário adicionar informações diversas acerca do cartão, como links para *websites* e até mesmo direcionamentos passo-a-passo de como executar a tarefa que este representa (TRELLO, 2021d). Como texto livre em linguagem natural, a descrição de um cartão pode conter diversas informações acerca do processo de negócio em execução. Supondo, por exemplo, que a descrição do cartão contenha uma descrição de processo, teríamos não apenas um elemento de processo, mas sim um conjunto de elementos passíveis de extração.

Assim, um texto em linguagem natural, como a descrição de um cartão não pode ser mapeado para um único elemento de processo, visto que em seu conteúdo podem ser encontradas diversas informações de processo, até mesmo uma descrição de processo completa. Tanto isto é verdade que trabalhos, tais como Ferreira, Thom and Fantinato (2017), Rosa (2020) Silva (2019) e Friedrich, Mendling and Puhlmann (2011) propõe abordagens para a extração de não apenas diversos elementos de processo a partir de textos em linguagem natural, como também a extração de modelos de processo e geração de descrições de processo consistentes. Logo, um mapeamento da descrição de um cartão para um único elemento de processo não se faz correto, o que nos leva ao axioma 4

**Axioma 4** *Como texto livre em linguagem natural, a descrição de um cartão pode conter diversas informações de processo de negócio em seu conteúdo.*

Contudo, como dado de um cartão Trello, a descrição está suscetível a relações com o próprio cartão e outros dados deste. Estas relações podem permitir a utilização de outros dados presentes em um cartão para enriquecer as informações presentes na descrição do cartão. De acordo com o axioma 3, os membros de um cartão são aqueles responsáveis por executar as tarefas relacionados ao cartão e que devem trabalhar em conjunto para garantir a execução destas tarefas. Portanto, podemos assumir que pronomes pessoais na 1ª pessoa do plural, como “nós” (na língua inglesa, “we”), presentes na descrição

do cartão, geralmente, se referem aos participantes de processo diretamente relacionados com a atividade sendo descrita, isto é, os membros deste cartão, o que nos leva ao axioma 5.

**Axioma 5** *Pronomes pessoais na 1ª pessoa do plural na descrição do cartão se referem, geralmente, aos membros deste mesmo cartão.*

Outra peculiaridade da descrição de cartões da Trello é a possibilidade da utilização de linguagem da marcação *Markdown* (GRUBER, 2004) para a formatação de seu texto. De acordo com a documentação da Trello, os usos mais comuns da formatação em *Markdown* Trello (2020a) são para inserir textos em itálico, negrito e tachado e de códigos de programação e *links* rotulados. Esta formatação se dá através da colocação de marcadores da linguagem *Markdown* que delimitam como o texto entre os marcadores deve ser formatado. Por exemplo, para inserir um texto negrito deve-se adicionar “\*\*” antes e depois do texto a ser formatado (ex.: “\*\*este texto está em negrito\*\*”). Seções de código também funcionam de maneira similar, sendo delimitadas pelos marcadores “` `” ou “`”.

Já no caso de *links* rotulados, a formatação é composta por dois marcadores utilizados em sequência, com o primeiro sendo utilizado para demarcar o rótulo e o segundo sendo utilizado para demarcar o *link*. Por exemplo, para criar um *link* rotulado com o rótulo “Este é um link para a Trello” que aponta para o *link* “https://trello.com”, deve-se utilizar o marcador “[ ]” para o rótulo e “( )” para o link, da seguinte forma: “[Este é um link para a Trello](https://trello.com)”.

Com exceção de códigos e *links* rotulados, o restante da formatação em *Markdown* tem apenas a função de alterar o texto visualmente, mantendo seu conteúdo. Nestes casos, a formatação em *Markdown* não adiciona informações de processo, mas o conteúdo do texto formatado pode conter tais informações, o que os torna úteis para a descoberta de processos.

No caso de códigos de programação, estes já não se classificam como texto em linguagem natural, fugindo do escopo do estudo deste trabalho. Já para *links* rotulados, os *links* em si também não se classificam como texto em linguagem natural, mas seu rótulo, exibido ainda no fluxo do descrição, faz parte do texto, assim como o conteúdo de marcações visuais. Portanto, o rótulo pode, também, ser útil para descoberta de processos e deve ser analisado. Tais observações formam os axiomas 6, 7 e 8.

**Axioma 6** *Em descrições de cartões da Trello, a formatação em Markdown para alterações visuais, como negrito e itálico, não altera o conteúdo do texto, que pode ser analisado como texto em linguagem natural na busca por informações de processo de negócio.*

**Axioma 7** *Em descrições de cartões da Trello, seções de código em Markdown, por não se caracterizarem como texto em linguagem natural, fogem ao escopo do presente estudo e devem ser ignoradas em suas análises.*

**Axioma 8** *Em descrições de cartões da Trello, links rotulados devem ter apenas os seus rótulos analisados por informações de processo de negócio, visto que o link não se caracteriza como texto em linguagem natural, fugindo ao escopo do presente estudo.*

### 3.1.5 Comentários de um cartão

Comentários de um cartão são apenas mais uma maneira de adicionar informações ao contexto de um cartão (TRELLO, 2021c). Comentários permitem que os membros de um cartão interajam através de conversas em texto (TRELLO, 2021b). Em seu formato, comentários são similares a descrições. Ambos são textos em linguagem natural e possuem as mesmas opções de formatação de texto. Neste sentido, podemos criar axiomas para comentários correspondentes aos axiomas 4, 5, 6, 7 e 8, de descrição de cartões, nos levando aos axiomas 9, 10, 11, 12 e 13.

**Axioma 9** *Como texto livre em linguagem natural, comentários de um cartão podem conter diversas informações de processo em seu conteúdo.*

**Axioma 10** *Pronomes pessoais na 1ª pessoa do plural em comentários se referem, geralmente, aos membros do cartão em questão.*

**Axioma 11** *Em comentários de cartões da Trello, a formatação em Markdown para alterações visuais, como negrito e itálico, não altera o conteúdo do texto, que pode ser analisado como texto em linguagem natural na busca por informações de processo de negócio.*

**Axioma 12** *Em comentários de cartões da Trello, seções de código em Markdown, por não se caracterizarem como texto em linguagem natural, fogem ao escopo do presente estudo e devem ser ignoradas em suas análises.*

**Axioma 13** *Em comentários de cartões da Trello, links rotulados devem ter apenas os seus rótulos analisados por informações de processo de negócio, visto que o link não se caracteriza como texto em linguagem natural, fugindo ao escopo do presente estudo.*

Apesar de similares, uma das principais diferenças entre um comentário e uma descrição é a presença de um autor no contexto do comentário. A descrição de um cartão não possui um membro diretamente atrelado à sua escrita, enquanto todo o comentário possui, de maneira visível, uma referência ao membro que o escreveu, como exibido na fig. 3.1, onde o comentário é acompanhado pelo nome de seu autor acima de seu conteúdo, no caso, “Guilherme Tassinari”.

Figura 3.1 – Exemplo de comentário de um cartão



Fonte: Os autores

Dado que comentários podem ser interpretados como conversas entre seus autores, é razoável assumirmos que pronomes pessoais na 1ª pessoa do singular, como “eu” (na língua inglesa, “I”), geralmente, se referem ao autor do comentário, nos levando ao axioma 14.

**Axioma 14** *Pronomes pessoais na 1ª pessoa do singular em comentários se referem, geralmente, ao autor deste mesmo comentário.*

### 3.1.6 Checklists

A Trello define *checklists* como uma maneira de gerenciar sub-tarefas em um cartão (TRELLO, 2021a). Cada *checklist* é composta por um conjunto de itens, cada item representando uma sub-tarefa. É possível converter um item de uma sub-tarefa em um novo cartão e, em atualizações recentes, foi adicionada a funcionalidade de atribuir um membro do cartão a um item de uma *checklist*.

A partir da definição da Trello de *checklists* e da possibilidade de transformar itens de uma *checklist* em cartões, em conjunto com o axioma 1, podemos assumir que, no contexto de BPM, uma *checklist* representa uma lista de sub-tarefas no contexto de um processo ou atividade. Apesar de a documentação, através da nomenclatura “sub-tarefa”, nos permitir inferir uma relação hierárquica entre a atividade do cartão e os itens

da *checklist*, esta não possibilita descobrir a relação dos itens entre si. Considerando que todos os itens da devem ser executadas para a execução completa da atividade do cartão, a relação entre estes itens só pode ser sequencial ou paralela. Esta ambiguidade entre paralelismo e sequencialidade de atividades se assemelha à relatada por Friedrich, Mending and Puhlmann (2011) e Silva et al. (2018) em relação ao uso de conjunções aditivas (“and”) em textos em linguagem natural, que também pode indicar tanto sequencialidade quanto paralelismo entre atividades. Portanto, podemos definir que itens de *checklists* indicam sub-tarefas do da atividade do cartão, relacionadas entre si ou por sequencialidade ou por paralelismo.

**Axioma 15** *Checklists de um cartão são conjuntos sub-tarefas da atividade do cartão, que devem ser executadas para que a atividade do cartão seja completa. Estas sub-tarefas se relacionam entre si ou por sequencialidade ou por paralelismo.*

### 3.1.7 Menções

Em campos de texto da Trello, como comentários e a descrição de um cartão, é possível utilizar a funcionalidade de menções para referenciar e notificar membros de um cartão. Para mencionar um membro deve-se adicionar ao texto sendo escrito um termo no formato “@ + nome de usuário”, onde “nome de usuário” é um identificador único do membro em questão (TRELLO, 2021b). Um exemplo de menção pode ser visto na fig. 3.2, onde o usuário de nome “Guilherme Tassinari” menciona o membro com o nome de usuário “claimsofficer”. Dado que uma menção é uma referência a um membro do cartão e que membros de cartões são participantes do processo, como definido no axioma 3, podemos interpretar menções como anáforas que referenciam estes membros, nos levando ao axioma 16.

Figura 3.2 – Exemplo de uso de uma menção em um comentário



Fonte: Os autores

**Axioma 16** *Uma menção em comentários ou descrições pode ser interpretada como uma anáfora para o membro mencionado.*

As tabelas 3.1 e 3.3 exibem um resumo de todos os axiomas elaborados pela análise da plataforma Trello e suas relações com conceitos de BPM e PLN.

### **3.2 Métodos para a extração de informações de processo presentes em cartões Trello**

Realizadas as análises sobre as funcionalidades presentes na Trello e suas relações com conceitos de PLN e BPM, pode-se construir técnicas e algoritmos para extrair estas informações em formato de texto em linguagem natural. Nas subseções seguintes serão propostas técnicas que possibilitam estas extrações, incluindo exemplos de aplicações destas.

#### **3.2.1 Definição de atividade de processo de negócio a partir de nome e membros e cartão**

De acordo com a regra 1 de Ferreira, Thom and Fantinato (2017) para a identificação de atividades em textos em linguagem natural, atividades são geralmente representadas em texto pela estrutura gramatical “*sujeito + verbo + objeto*”. Assim, caso tenha-se as informações que representam o sujeito, verbo e objeto, isto é, respectivamente, os atores, a ação que realizam e o objeto sobre a qual a ação é executada no contexto da atividade, é possível gerar a estrutura “*sujeito + verbo + objeto*” que representa a atividade em texto a partir destas informações.

No contexto da Trello, de acordo com o axioma 1, sabe-se que cartões geralmente representam atividades e/ou tarefas. Ainda mais, dado o axioma 2, sabe-se que nomes de cartões são geralmente escritos como rótulos de atividades, na forma “*verbo + objeto*”. Em conjunto, estas duas definições nos permitem assumir que, quando na forma “*verbo + objeto*”, o nome de um cartão já satisfaz a parte final da estrutura “*sujeito + verbo + objeto*” definida por Ferreira, Thom and Fantinato (2017), restando apenas a parte inicial, “*sujeito*”.

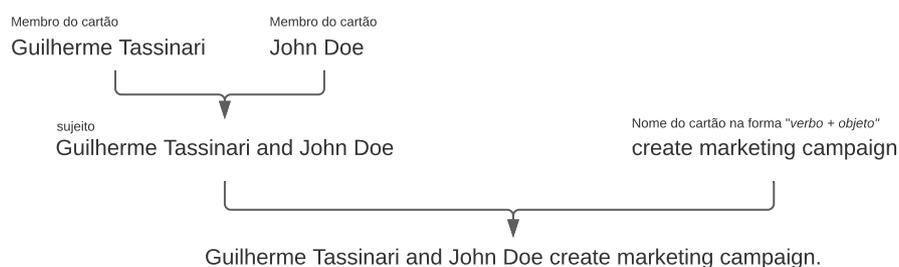
Dado que o sujeito representa os atores que executam a atividade definida pela estrutura “*sujeito + verbo + objeto*”, pelo axioma 3, sabe-se que os participantes de processo responsáveis pela execução da tarefa de um cartão são representados pelos membros atribuídos à este cartão. Assim, podemos definir o “*sujeito*” da estrutura “*sujeito + verbo + objeto*” como sendo a união dos nomes dos membros de um cartão. Desta forma, tem-

se todas as informações necessárias para se elaborar a estrutura textual “*sujeito + verbo + objeto*” que define a atividade representada pelo cartão.

Para exemplificar a geração desta forma textual, tomemos o exemplo de cartão Trello apresentado na fig. 2.11. Primeiro deve-se analisar o nome do cartão para verificar se este obedece à forma “*verbo + objeto*”. Neste caso, dado o nome de cartão “*Create marketing campaign*”, é possível observamos que este obedece à forma “*verbo + objeto*”, dado que que “*Create*” é traduzido para o verbo “criar” e “*marketing campaign*” é traduzido para o objeto “campanha de marketing”.

Após, deve-se unir os nomes dos membros do cartão por uma conjunção, formando uma única frase que sirva como sujeito para o texto a ser gerado. No caso, têm-se como membros do cartão os usuários de nomes “Guilherme Tassinari” e “John Doe”. Unindo-os pela conjunção “and”, têm-se a frase “*Guilherme Tassinari and John Doe*”. Por fim, unindo-se o sujeito, o verbo e o objeto identificados, têm-se a sentença “*Guilherme Tassinari and John Doe create marketing campaign*” que, em tradução livre, significa “Guilherme Tassinari e John Doe criam campanha de marketing”. A fig. 3.3 mostra de maneira visual a elaboração da sentença a partir dos dados do cartão da Trello, que é definida formalmente pela definição 1.

Figura 3.3 – Exemplo de extração da definição de atividade representada por um cartão Trello a partir da fig. 2.11.



Fonte: Os autores

**Definição 1 (Atividade a partir de nome e membros de cartão)** *Quando o nome de um cartão se encontra no formato “verbo + objeto”, é possível definir textualmente a atividade representada por este cartão através da estrutura “sujeito + verbo + objeto” de Ferreira, Thom and Fantinato (2017), utilizando a união dos nomes dos membros do cartão como “sujeito” e o nome do cartão como “verbo + objeto”.*

### 3.2.2 Sub-tarefas a partir de *checklists*

Conforme o axioma 15, *checklists* em cartões da Trello podem ser vistas como conjuntos de sub-tarefas, que devem ser executadas para que a atividade do cartão seja completa. Estes itens podem ser de execução paralela ou sequencial, mas não é possível definir a relação pelos dados presentes no cartão.

Dado que Friedrich, Mendling and Puhlmann (2011) e Silva et al. (2018) relatam que sentenças contendo atividades unidas por conjunções aditivas (“*and*”) podem representar tanto sequencialidade quanto paralelismo, que são as relações possíveis para itens da *checklist*, podemos tomar estas formas como equivalentes. Portanto, os itens de uma *checklist* podem ser descritos em texto por seus rótulos unidos por conjunções aditivas. Tomando, por exemplo, a *checklist* presente na fig. 2.11, pode-se descrevê-la através da sentença “*Create image and create video and post campaign on social media*”.

Figura 3.4 – Exemplo de extração de sub-tarefas a partir de uma *checklist* utilizando exemplo da fig. 2.11



Fonte: Os autores

Contudo, como sentença isolada, o resultado da utilização desta técnica não fornece contexto acerca da atividade do cartão em si ou daqueles que devem executá-la. Isto é, a sentença não contextualiza estas sub-tarefas em relação ao processo/atividade em que estão contidas. É possível perceber esta falta de contextualização ao analisar a sentença e perceber que não há nenhuma referência à atividade do cartão ou a quem são os participantes do processo responsáveis por executar estas sub-tarefas. No caso da utilização desta sentença por um analista de processo, de maneira isolada, este seria incapaz de inferir o contexto ao qual as sub-tarefas estão inseridas.

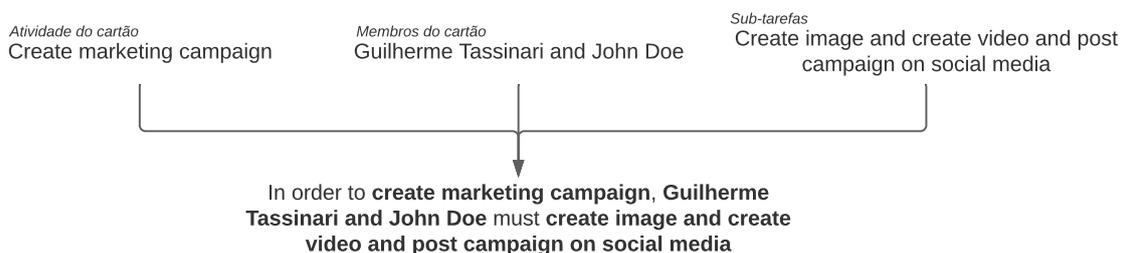
Entretanto, como visto na definição 1, têm-se acesso a informações de contexto, como o nome da atividade do cartão e os participantes de processo, o que torna possível sua adição à sentença de forma a melhor contextualizar e oferecer informações mais precisas. De maneira mais prática, pode-se utilizar o *template*

*Para executar a <atividade do cartão>, <membros do cartão> devem <sub-tarefas>*

onde <atividade do cartão> é a atividade identificada pela definição 1 a partir do nome do cartão, <membros do cartão> são os nomes dos participantes de processo por conjunção, também como utilizado na definição 1 e <sub-tarefas> refere-se à frase formada a partir dos itens da *checklist* em questão e exemplificado na fig. 3.3. A elaboração desta sentença tem por objetivo expressar o contexto completo da *checklist*, possibilitando seu uso tanto por analistas de processos quanto por ferramentas de análise automática. A definição formal desta sentença é descrita na definição 2.

Tomando as sentenças formadas nas figuras 3.3 e 3.4 como base, teríamos a sentença “*In order to create marketing campaign, Guilherme Tassinari and John Doe must Create image and create video and post campaign on social media*” (“Para criar campanha de marketing, Guilherme Tassinari e John Doe devem criar uma imagem e criar um vídeo e publicar a campanha em mídias sociais”, em tradução livre). A fig. 3.5 exhibe visualmente este exemplo.

Figura 3.5 – Exemplo de extração de tarefa, membros de processo e sub-tarefas a partir de um cartão da Trello utilizando exemplo da fig. 2.11



Fonte: Os autores

**Definição 2 (Sub-tarefas a partir de *checklist* de um cartão)** *Dada uma checklist de um cartão da Trello, pode-se utilizar o template “Para executar a <atividade do cartão>, <membros do cartão> devem <sub-tarefas>” para expressar a atividade do cartão, seus participantes de processo e sub-tarefas representada pela checklist em questão.*

### 3.2.3 Informações de processo em textos de linguagem natural

Como definido pelos axiomas 4 e 9, comentários e descrições de cartões Trello são textos em linguagem natural. A identificação e extração de elementos de processo a partir de textos em linguagem natural vêm sendo o foco de diversos estudos envolvendo PLN e BPM, como os de Friedrich, Mendling and Puhmann (2011), Silva (2019), Ferreira, Thom and Fantinato (2017) e Rosa (2020). A existência destes estudos indica que extrair elementos de processo de textos em linguagem natural não é uma tarefa trivial e, por isto, foge ao escopo do presente trabalho apresentar novas técnicas com este fim.

Portanto, ao invés de propor técnicas para a extração de informações de processo a partir de textos em linguagem natural, o presente trabalho propõe técnicas para realizar o processamento e enriquecimento dos dados em texto presentes em um cartão da Trello, transformando-os em texto mais completo em termos de informações de processos de negócio e que possam ser utilizados tanto por analistas de negócios quanto por ferramentas criadas por trabalhos relacionados para a extração automática de elementos de processo.

Tomando os axiomas 6, 7, 8, 11, 12 e 13, observa-se que a formatação de comentários e descrições utilizando *Markdown* deve ser filtrada para se ter um texto em linguagem natural que possa ser útil para projetos de BPM. De acordo com os axiomas, formatações em *Markdown* que alterem apenas visualmente o texto devem ser retiradas, mantendo apenas o seu conteúdo disponível para análise. Já formatações de *links*, compostas por um rótulo e um *link*, devem ser filtradas de maneira que apenas o rótulo permaneça no texto. Por fim, seções de código em *Markdown* devem ser completamente removidas do texto em questão. Estas definições estão expressas na definição 3.

**Definição 3 (Filtragem de formatação *Markdown*)** *Em cartões Trello, textos formatados utilizando Markdown devem ser filtrados de acordo com o tipo de formatação. Formatações visuais, como itálico, negrito e tachado, devem ser filtradas para retirar os marcadores de Markdown correspondentes. Links devem manter apenas os seus rótulos. Seções de código devem ser removidas.*

A fig. 3.6 exibe três textos de descrição da Trello. O primeiro exibe o texto em edição, puro, com elementos de formatação *Markdown*. O segundo texto é o mesmo que o primeiro, porém, após editado e salvo, exibindo a visualização da formatação *Markdown*. Por último, um texto filtrado, exemplificando como os textos anteriores seriam após filtrados pelas observações definidas nos axiomas.

Figura 3.6 – Exemplos de descrições da Trello em edição com formatação, pós edição com formatação e em edição filtrado.



Fonte: Os autores

Considerando os axiomas 5 e 10, que dizem que pronomes na 1ª pessoa do plural em descrições e comentários geralmente se ferem aos membros do cartão, é direto assumir que se possa realizar algo similar a Resolução de Anáfora ao substituímos estes pronomes pelos nomes dos membros do cartão. Considere, por exemplo, o último comentário do cartão exemplo da fig. 2.11, “*We must create a video and an image for this campaign*”. Dado que “*we*” (na língua portuguesa, “*nós*”) é o pronome na 1ª pessoa do plural, podemos substituí-lo pelo nome dos membros no cartão, isto é “*Guilherme Tassinari and John Doe*”, alterando a sentença para “*Guilherme Tassinari and John Doe must create a video and an image for this campaign*” (na língua portuguesa, “*Guilherme Tassinari e John Doe devem criar um vídeo e uma imagem para esta campanha*”). Isto adiciona maior contexto ao texto, permitindo seu uso isolado, fora da plataforma Trello, em projetos de BPM. A definição 4 oficializa esta técnica.

**Definição 4 (Substituição de pronomes da 3ª pessoa do plural)** *Pronomes da 3ª pessoa do plural presentes em comentários e descrições de cartões podem ser substituídos pelos nomes dos membros do cartão unidos por conjunções.*

No mesmo contexto da substituição de pronomes da 3ª pessoa, o axioma 14 estabelece que pronomes da 1ª pessoa do singular em comentários se referem ao autor do comentário. Nestes casos, também é possível realizar algo similar a Resolução de Anáfora, substituindo estes pronomes pelo nome do autor do comentário. Tomando, por exemplo, o comentário de John Doe na fig. 2.11, “*I agree, @guilhermetassinari. I will create the video. Can you create the image?*”, pode-se substituir o pronome “*I*” (na língua portuguesa, “Eu”) pelo nome do autor do comentário, John Doe, resultando na sentença “*John Doe agree, @guilhermetassinari. John Doe will create the video. Can you create the image?*” (na língua portuguesa, “John Doe concorda, @guilhermetassinari. John Doe vai criar o vídeo. Você pode criar a imagem?”). A definição 5 oficializa esta técnica.

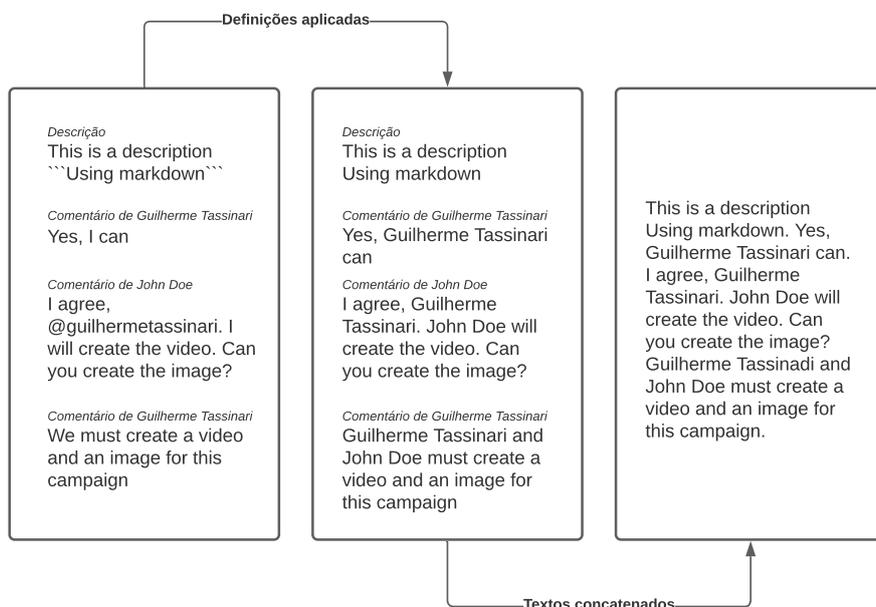
Tabela 3.5 – Resumo das definições de identificação de elementos de processo a partir de cartões da plataforma Trello

Nº	Definição
1	Quando o nome de um cartão se encontra no formato “verbo + objeto”, é possível definir textualmente a atividade representada por este cartão através da estrutura “sujeito + verbo + objeto” de Ferreira, Thom and Fantinato (2017), utilizando a união dos nomes dos membros do cartão como “sujeito” e o nome do cartão como “verbo + objeto”.
2	Dada uma <i>checklist</i> de um cartão da Trello, pode-se utilizar o <i>template</i> “Para executar a <atividade do cartão>, <membros do cartão> devem <sub-tarefas>” para expressar a atividade do cartão, seus participantes de processo e sub-tarefas representada pela <i>checklist</i> em questão.
3	Em cartões Trello, textos formatados utilizando <i>Markdown</i> devem ser filtrados de acordo com o tipo de formatação. Formatações visuais, como itálico, negrito e tachado, devem ser filtradas para retirar os marcadores de <i>Markdown</i> correspondentes. <i>Links</i> devem manter apenas os seus rótulos. Seções de código devem ser removidas.
4	Pronomes da 3ª pessoa do plural presentes em comentários e descrições de cartões podem ser substituídos pelos nomes dos membros do cartão unidos por conjunções.
5	Pronomes da 1ª pessoa do singular presentes em comentários de cartões podem ser substituídos pelo nome do autor do comentário.
6	Uma menção em um comentário ou descrição pode ser substituída pelo nome do usuário referenciado.

Fonte: Os Autores

**Definição 5 (Substituição de pronomes da 1ª pessoa do singular)** *Pronomes da 1ª pessoa do singular presentes em comentários de cartões podem ser substituídos pelo nome do autor do comentário.*

Figura 3.7 – Exemplo visual da aplicação das definições 3, 6, 5 e 4, nesta ordem, ao cartão da fig. 2.11



Fonte: Os autores

Finalmente, menções também podem ser interpretadas como anáforas no contexto de cartões da Trello, como definido pelo axioma 16. Assim, menções em comentários e descrições podem ser substituídas pelo nome do membro que está sendo mencionado. Tomando novamente por exemplo o comentário de John Doe na fig. 2.11, pode-se substituir a menção “@guilhermetassinari” pelo nome do membro representado por este nome de usuário, no caso, “Guilherme Tassinari” resultando na sentença “*I agree, Guilherme Tassinari. I will create the video. Can you create the image?*” (na língua portuguesa, “Eu concordo, Guilherme Tassinari. Eu vou criar o vídeo. Você pode criar a imagem?”). A definição 6 oficializa esta técnica.

**Definição 6 (Substituição de menções)** *Uma menção em um comentário ou descrição pode ser substituída pelo nome do usuário referenciado.*

Todas as definições criadas até agora foram definidas independentes uma das outras e, portanto, podem ser aplicadas de maneira isolada. Entretanto, seu real valor se mostra ao aplicarmos todas estas definições para se extrair um texto próximo da linguagem natural a partir de um cartão da Trello. Tomemos, mais uma vez, o cartão da figura

2.11 para exemplificarmos. A tabela 3.7 mostra o resultado da aplicação destas definições ao cartão em questão. As definições 1 e 2 são aplicadas individualmente, pois cada uma possui como saída uma sentença específica contendo informações de processos de negócio.

Já as definições 3, 6, 5 e 4 são aplicadas seguidas uma das outras, nesta mesma ordem, sobre a descrição e os comentários do cartão e seu resultado é então consolidado em um único texto formado por estes. Para melhor visualização das definições aplicadas à descrição e comentários, a fig. 3.7 exibe essa aplicação de maneira visual, dividindo-a em etapas, onde a primeira contém os textos originais, a segunda, a direita, exibe os textos após a aplicação das definições e, por fim, os textos são concatenados e exibidos na última etapa, resultando no texto exibido na tabela 3.7.

Tabela 3.7 – Exemplo de aplicação das definições sobre o cartão da fig. 2.11

<i>Definições aplicadas</i>	<i>Dados sobre o qual a definição é aplicada</i>	<i>Resultado</i>
1	Nome e membros do cartão	<i>“Guilherme Tassinari and John Doe create marketing campaign.”</i>
2	Nome, membros e checklist do cartão	<i>“In order to create marketing campaign Guilherme Tassinari and John Doe must create image, create video and post campaign on social media.”</i>
3, 6, 5 e 4	Descrição e comentários	<i>“This is a description Using markdown. Yes, Guilherme Tassinari can. I agree, Guilherme Tassinari. John Doe will create the video. Can you create the image? Guilherme Tassinari and John Doe must create a video and an image for this campaign”</i>

Fonte: Os Autores

### 3.3 Resumo

Neste capítulo foram analisadas as principais funcionalidades presentes em um cartão Trello, explicando seu funcionamento e estabelecendo relações destas funcionali-

dades com conceitos de PLN e BPM. A partir destas relações foram definidos axiomas que estabelecem bases para a elaboração de definições capazes de identificar informações de processo em um cartão da Trello e extraí-las em formato de texto em linguagem natural. Este texto, por sua vez, por ser em linguagem natural, pode ser utilizado por analistas de processos em projetos de BPM ou em ferramentas desenvolvidas para analisar automaticamente textos em linguagem natural na busca por informações de processo.

Através da transformação de dados previamente separados em um único texto, como realizado nas definições 1 e 2, com nome de cartão, membros de cartão e *checklists*, podemos agilizar a análise dos analistas de processos ao fornecer não mais um conjunto de dados não estruturados na forma de um cartão Trello, mas sim um texto em linguagem natural contendo apenas as informações de processo de negócio contidas no cartão.

No caso das definições 3, 6, 5 e 4, apesar de não identificarem as informações de processo, estas podem auxiliar o analista removendo símbolos, marcações e dados que não são úteis para a descoberta de processos e complementando o texto com informações de processo contidas outros dados do cartão.

Por fim, a transformação dos dados do cartão Trello em texto em linguagem natural torna desnecessário que o analista de processo conheça a plataforma e seu funcionamento. Uma vez extraído o texto, o analista pode se ater a utilizá-lo como fonte de informações, sem a necessidade de estudar a plataforma e sua documentação para compreender como elementos de processo podem estar contidos nela.

## 4 PROTÓTIPO

Este capítulo apresenta o protótipo<sup>1</sup> com o objetivo de servir como prova de conceito para as definições de extração propostas neste trabalho. Através da aplicação de técnicas de PLN e das definições de extração, o protótipo, construído como uma plataforma para a Internet, permite a usuários da plataforma Trello navegar entre quadros e cartões até encontrar um cartão que deseje processar e extrair dele informações de processo de negócio. A partir desta escolha, a plataforma aplica uma série de técnicas para o pré-processamento do cartão e, após, aplica as definições de extração, exibindo por fim ao usuário o cartão pré-processado e o texto com as informações de processo de negócio extraídas.

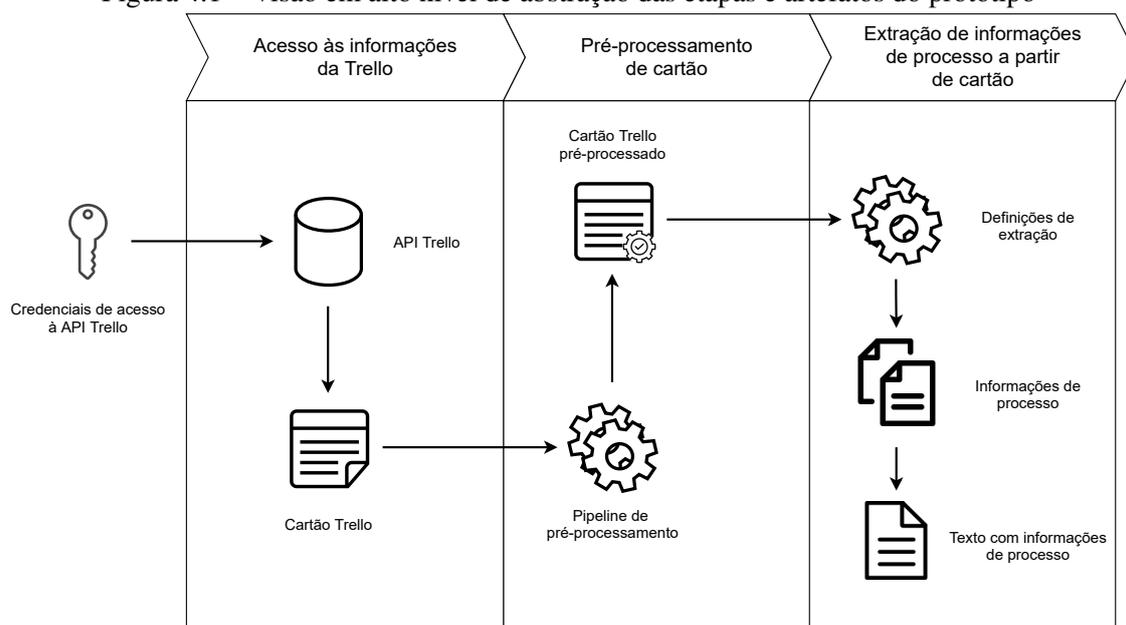
As seções neste capítulo detalham o desenvolvimento e o funcionamento do protótipo. A 4.1 apresenta as principais decisões tomadas em relação à arquitetura e desenvolvimento do protótipo, além da escolha de tecnologias, como linguagem de programação e ferramentas de código aberto. Após, as três etapas que compõe a execução do protótipo são apresentadas, detalhando seu funcionamento, arquitetura e interação entre o usuário e o protótipo, bem como dos componentes internos da plataforma. A fig. 4.1 mostra a visão em alto nível de abstração do protótipo, exibindo as etapas e os principais artefatos utilizados e gerados em cada uma destas.

A seção 4.2 traz a primeira etapa do protótipo, em que o usuário navega entre quadros e cartões até escolher o cartão desejado. A seção detalha o passo-a-passo do usuário na utilização do protótipo e os artefatos de software utilizados, bem como as relações entre estes. A seção 4.3 descreve a segunda etapa, em que o cartão é pré-processado por um *pipeline* de pré-processamento composto por diversos estágios. Na seção 4.4 a etapa de extração é descrita, na qual o cartão, já pré-processado, tem suas informações de processo extraídas através da aplicação das definições de extração e da aplicação de uma das regras de identificação de elementos de processo em textos em linguagem natural definida por Ferreira, Thom and Fantinato (2017) (“sujeito + verbo + objeto”). Por último, a seção 4.5, exemplifica o uso do protótipo, com imagem das páginas utilizadas e descrições detalhadas do que compõe cada uma e como ela deve ser usada.

---

<sup>1</sup><<https://github.com/guitassinari/bpm-trello>>

Figura 4.1 – Visão em alto nível de abstração das etapas e artefatos do protótipo



Fonte: Os autores.

#### 4.1 Desenvolvimento e arquitetura

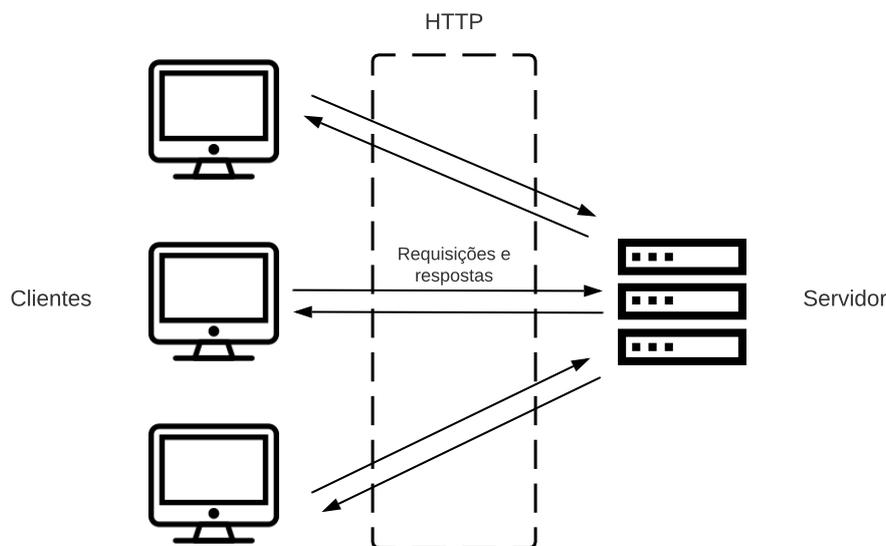
A arquitetura do protótipo foi escolhida tendo em vista dois fatores: (i) agilidade de desenvolvimento e (ii) disponibilidade de ferramentas e suas respectivas documentações. Dada a experiência prévia dos autores com o desenvolvimento de plataformas para a Internet, foi escolhida a arquitetura cliente-servidor pela Internet, na qual o cliente é um sistema que realiza requisições ao servidor. Este, por sua vez, recebe estas requisições, processa informações e responde as requisições do cliente (OLUWATOSIN, 2014).

No caso deste protótipo, cliente e servidor se comunicam pelo protocolo HTTP, onde o cliente é o navegador de Internet utilizado pelo usuário, enquanto o servidor é a aplicação hospedada em uma máquina e disponível na Internet. Esta arquitetura é exibida visualmente na fig. 4.2

Escolhida a arquitetura geral do protótipo, deve-se escolher as tecnologias utilizadas para o seu desenvolvimento, como a linguagem de programação e as ferramentas disponibilizadas para esta linguagem. Ainda levando em consideração a experiência prévia dos autores com o desenvolvimento de plataformas para a Internet, com foco em se ter uma maior agilidade de desenvolvimento, a linguagem de programação *Ruby*<sup>2</sup> foi a escolhida.

<sup>2</sup><<https://www.ruby-lang.org/pt/>>

Figura 4.2 – Visão geral da arquitetura cliente servidor pelo protocolo HTTP



Fonte: Os autores

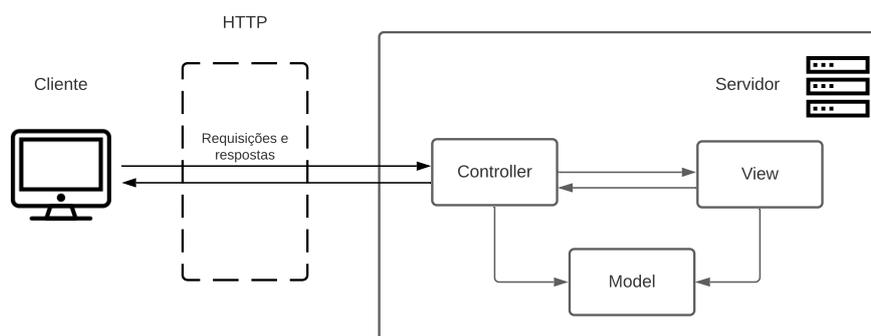
Foi relevante também para a escolha da linguagem a existência de ferramentas para a construção de plataformas para a Internet no modelo cliente-servidor e para a implementação de *pipelines* em PLN. No contexto do desenvolvimento do modelo cliente-servidor, o *framework Ruby on Rails*<sup>3</sup> (RoR) foi escolhido para utilização, também devido à experiência prévia dos autores.

O *Ruby on Rails* é um *framework* para o desenvolvimento de aplicações para a *web* que implementa o modelo modelo-visão-controlador (MVC) (Bächle; Kirchberg, 2007). O modelo MVC divide a aplicação em três partes: o modelo, a visão e o controlador, responsáveis, respectivamente, pelo processamento de dados, saída de dados e entrada de dados (QURESHI; SABIR, 2014). Em arquiteturas cliente-servidor pela internet o controlador é responsável por receber as requisições dos clientes, buscar os modelos necessários para o processamento da requisição e envia-las à visão para que seus dados sejam exibidos. A visão, por sua vez é, enviada como resposta à requisição do cliente e é alterada pelo modelo quando seus dados são alterados. Uma vez enviada para o cliente, a visão pode realizar requisições ao controlador para atualizar dados do modelo (QURESHI; SABIR, 2014). A fig. 4.3 exhibe as relações entre os componentes do modelo MVC quando aplicados em uma arquitetura cliente-servidor disponível na *web*.

No contexto de ferramentas PLN, devido ao seu uso por diversos trabalhos envolvendo BPM e PLN, como Silva et al. (2018), Rosa (2020) e, principalmente, Friedrich, Mendling and Puhmann (2011), a ferramenta *Stanford CoreNLP* (MANNING et

<sup>3</sup><<https://rubyonrails.org/>>

Figura 4.3 – Modelo MVC aplicado a uma arquitetura cliente-servidor na *web*



Fonte: Os autores. Adaptado de Qureshi and Sabir (2014).

al., 2014) foi escolhida para a implementação de *pipelines* em PLN. Devido à sua escrita na linguagem Java<sup>4</sup>, a utilização da *Stanford CoreNLP* na linguagem Ruby está condicionada a uma integração entre a linguagem e a ferramenta. Tal integração foi desenvolvida e disponibilizada na Internet como código aberto por Louis Mulli<sup>5</sup> e também é utilizada neste trabalho.

Outra necessidade específica deste protótipo é o acesso aos dados do usuário na plataforma Trello. A plataforma Trello disponibiliza uma API (*Application Programming Interface*) *web* para que terceiros possam ter acesso aos seus dados. Com o objetivo de facilitar o acesso a esta API com a linguagem Ruby, Jeremy Tregunna desenvolveu a biblioteca *ruby-trello*<sup>6</sup> que encapsula os chamadas à API em classes Ruby e é utilizada no protótipo.

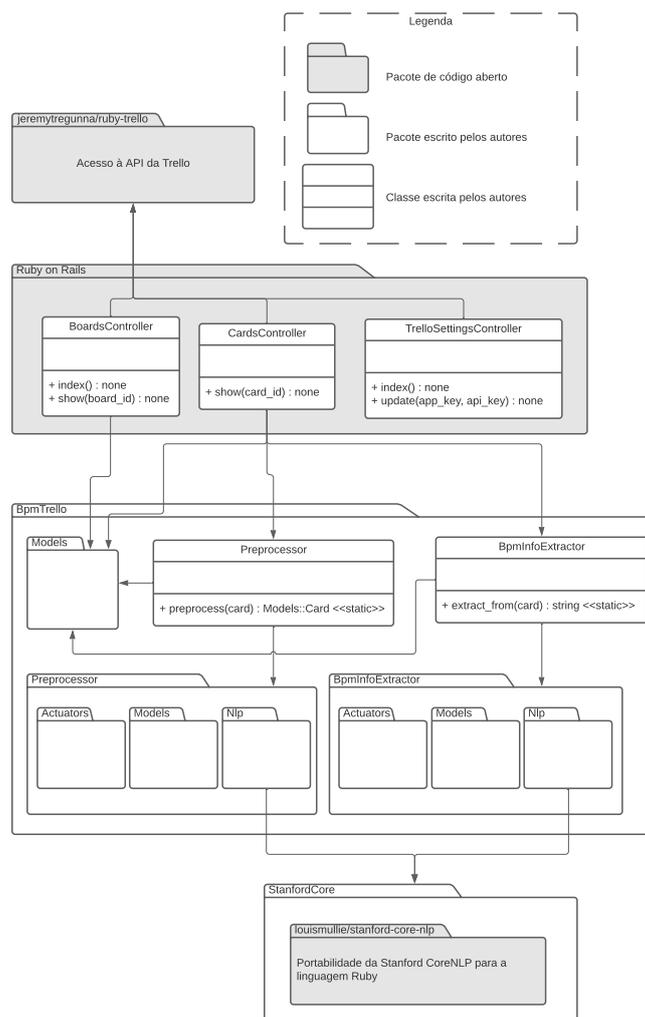
Com todas as ferramentas e a arquitetura escolhida, o protótipo foi desenvolvido conforme exibido no diagrama de pacotes da fig. 4.4, que apresenta uma visão geral, não detalhada, dos pacotes do protótipo. O diagrama mostra as três bibliotecas/*frameworks* de código aberto utilizadas no contexto do protótipo (*ruby-trello*, *stanford-core-nlp* e *Ruby on Rails*). O pacote *BpmTrello* contém toda a lógica para o pré-processamento e extração de informações de processo de negócio de um cartão da Trello. Os controladores *BoardsController*, *CardsController* e *TrelloSettingsController*, apesar de estarem inseridas no contexto do *Ruby on Rails*, foram escritas pelos autores e são o ponto de entrada do navegador do usuário com a aplicação.

<sup>4</sup><<https://docs.oracle.com/javase/8/docs/technotes/guides/language/index.html>>

<sup>5</sup><<https://github.com/louismullie/stanford-core-nlp>>

<sup>6</sup><<https://github.com/jeremytregunna/ruby-trello>>

Figura 4.4 – Visão geral da arquitetura do protótipo desenvolvido



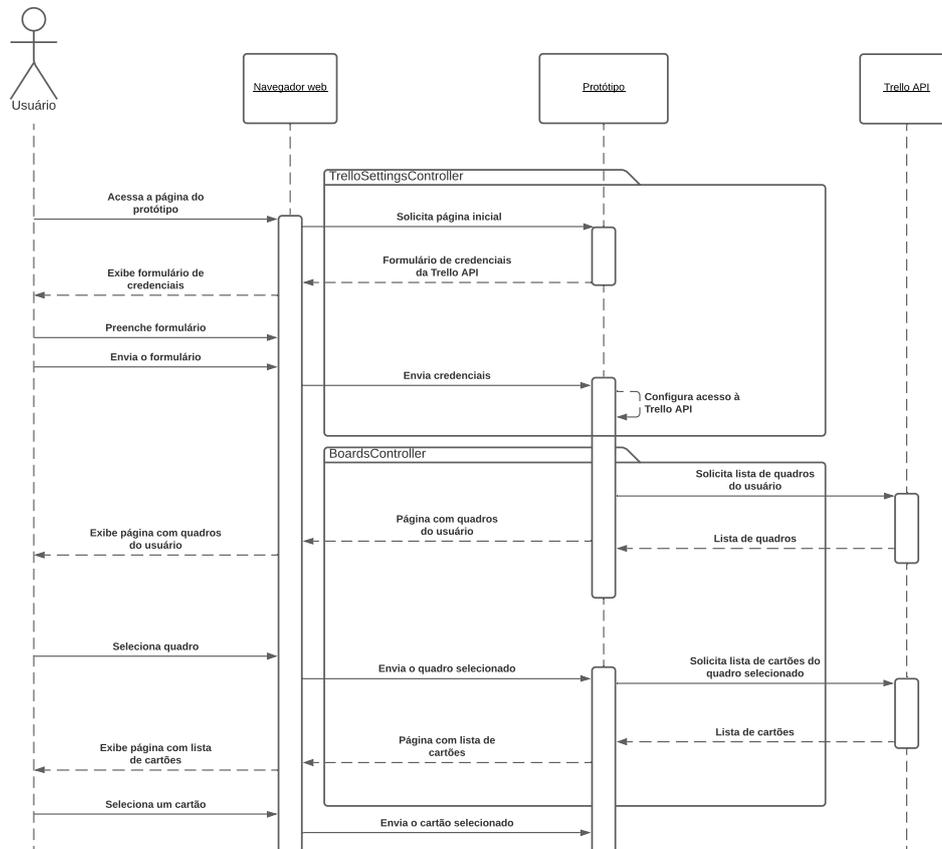
Fonte: Os autores.

## 4.2 Acesso às informações da Trello

A etapa de acesso às informações da Trello é composta pelo fornecimento das credenciais de API da conta Trello do usuário ao protótipo, o acesso do protótipo aos dados do usuário e a navegação deste até o cartão do qual deseja extrair informações de processo de negócio. Todo o fluxo desta primeira etapa de uso do protótipo é representada pelo diagrama de sequência da fig. 4.5, onde a interação em alto nível entre usuário, navegador *web*, protótipo e API da Trello é apresentada. O diagrama também separa as interações realizadas pela *TrelloSettingsController* e *BoardsController*.

Para que se possa ter acesso aos quadros e cartões de um usuário da Trello é necessário que este usuário crie duas credenciais, a chave de API e o *token* de API, e às forneça a aplicação, permitindo que esta acesse seus dados na plataforma. Estas creden-

Figura 4.5 – Diagrama de sequência em alto nível de abstração mostrando a interação entre usuário, navegador, protótipo e API da Trello durante a etapa de acesso às informações da Trello



Fonte: Os autores.

ciais podem ser geradas na conta do usuário na Trello, como explicado na documentação oficial da API<sup>7</sup>. O fornecimento das credenciais ao protótipo é realizado através de um formulário *web* que deve ser preenchido pelo usuário.

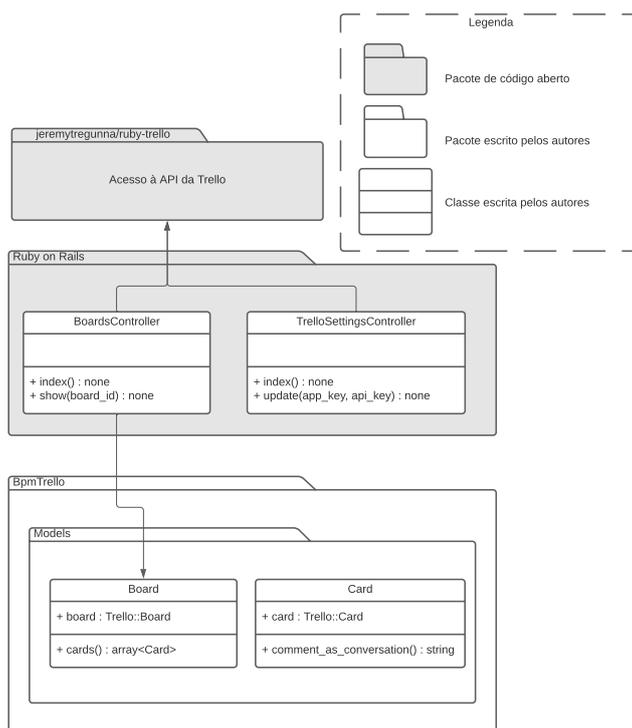
Uma vez preenchido e terminado, o formulário envia as credenciais à aplicação através da *TrelloSettingsController*. Esta então configura a biblioteca *ruby-trello* com estas credenciais para autenticar a aplicação junto à API da Trello, permitindo que o protótipo tenha acesso a dados do usuário, quadros, cartões e suas informações, incluindo nome do cartão, membros do cartão, descrição, comentários e *checklists*. Após, a *TrelloSettingsController* redireciona o usuário para a página contendo a lista de quadros, que é gerada pela *BoardsController*.

Tendo configurado a biblioteca *ruby-trello* com as credenciais necessárias, a *BoardsController* a utiliza para acessar a API da Trello e buscar todos os quadros da conta do usuário. Estes são exibidos em tela para que o usuário escolha aquele em que está o cartão que deseja processar.

<sup>7</sup><<https://developer.atlassian.com/cloud/trello/guides/rest-api/api-introduction/>>

Uma vez escolhido o quadro, a *BoardsController* mais uma vez utiliza a biblioteca *ruby-trello* para acessar a API e buscar todos os cartões existentes no quadro escolhido. Apesar de a biblioteca fornecer classes para cada entidade da Trello, como *RubyTrello::Card* (cartão) e *RubyTrello::Board* (quadro), foram criadas classes no contexto do pacote *BpmTrello* que recebem as classes da biblioteca como entrada e as encapsulam, possibilitando a criação de métodos mais abstratos e complexos para uso no restante do protótipo. Estas classes estão no pacote *BpmTrello::Models* e são exibidas de maneira detalhada na fig. 4.6. No contexto da *BoardsController*, a classe *BpmTrello::Models::Board* é utilizada para encapsular a classe *RubyTrello::Board* e fornece o método *cards*, utilizado para buscar a lista de cartões do quadro escolhido.

Figura 4.6 – Diagrama de classes da etapa de acesso às informações da Trello



Fonte: Os autores.

Tendo-se a lista de cartões do quadro, estes são exibidos para o usuário, que deve escolher aquele que deseja processar. Ao escolher o cartão, o processamento deste é iniciado, começando pela etapa de pré-processamento, detalhada a seguir.

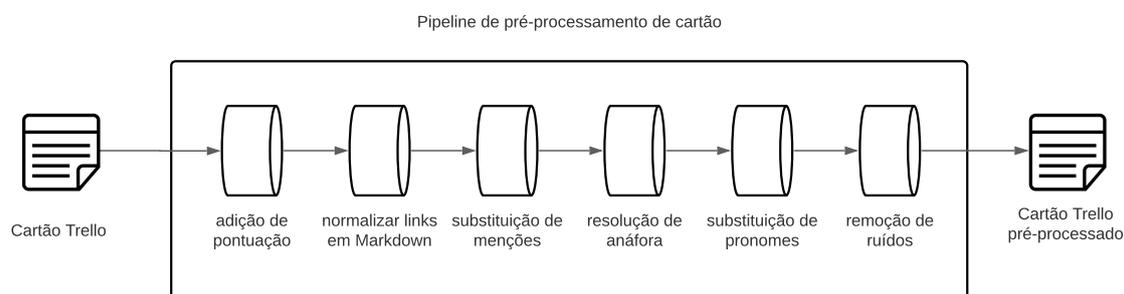
### 4.3 Pré-processamento de um cartão da Trello

Após a escolha do cartão, ocorre o pré-processamento do cartão, em que são aplicados os primeiros axiomas e definições elaboradas anteriormente e respectivamente resumidos nas tabelas 3.1, 3.3 e 3.5. Em especial, nesta etapa são aplicadas as definições relacionadas com comentários e descrições (definições 3, 4, 5 e 6), por sua natureza como texto em linguagem natural. Além disso, a etapa também realiza alterações para aprimorar a análise feita pela *Stanford CoreNLP*, como a adição de pontuação.

O pré-processamento de um cartão é dividido em um conjunto ordenado de estágios que chamamos de *pipeline de pré-processamento*. Cada estágio é responsável por uma alteração específica no cartão. O estágio de “adição de pontuação”, por exemplo, é responsável por adicionar pontos finais (“.”) ao nome, descrição e cada comentário do cartão. A figura 4.7 exibe os estágios do *pipeline* de pré-processamento e a ordem em que são executados.

A ordem de execução dos estágios é importante, pois as alterações de um estágio podem impactar as alterações dos estágios anteriores, de maneira a torná-los mais ou menos eficientes. Por exemplo, o estágio de “substituição de menções” substitui menções em descrições e comentários pelos nomes dos usuários mencionados. Quando executada antes do estágio “resolução de anáfora”, em que a *Stanford CoreNLP* é utilizada para substituir anáforas pelo termo da sua referência original a qual ela se refere, esta pode possibilitar um maior número de resoluções ao substituir as menções, que são termos que não fazem parte da língua inglesa, por pronomes pessoais.

Figura 4.7 – Pipeline de pré-processamento de cartão Trello



Fonte: Os autores.

Para aplicar todas estas alterações foi criado, no contexto do desenvolvimento do protótipo, o pacote *Preprocessor*, que é dividido em três sub-pacotes: (i) *Actuators*, que contém as classes responsáveis pelos estágios de pré-processamento, (ii) *Models*, que contém classes que encapsulam os dados de um cartão Trello, como nome, descrição,

comentários e *checklists*, para possibilitar a sua modificação sem afetar os dados reais da plataforma, e (iii) *Nlp*, que contém classes responsáveis por fornecer técnicas complexas de PLN para os estágios utilizando a *Stanford CoreNLP*.

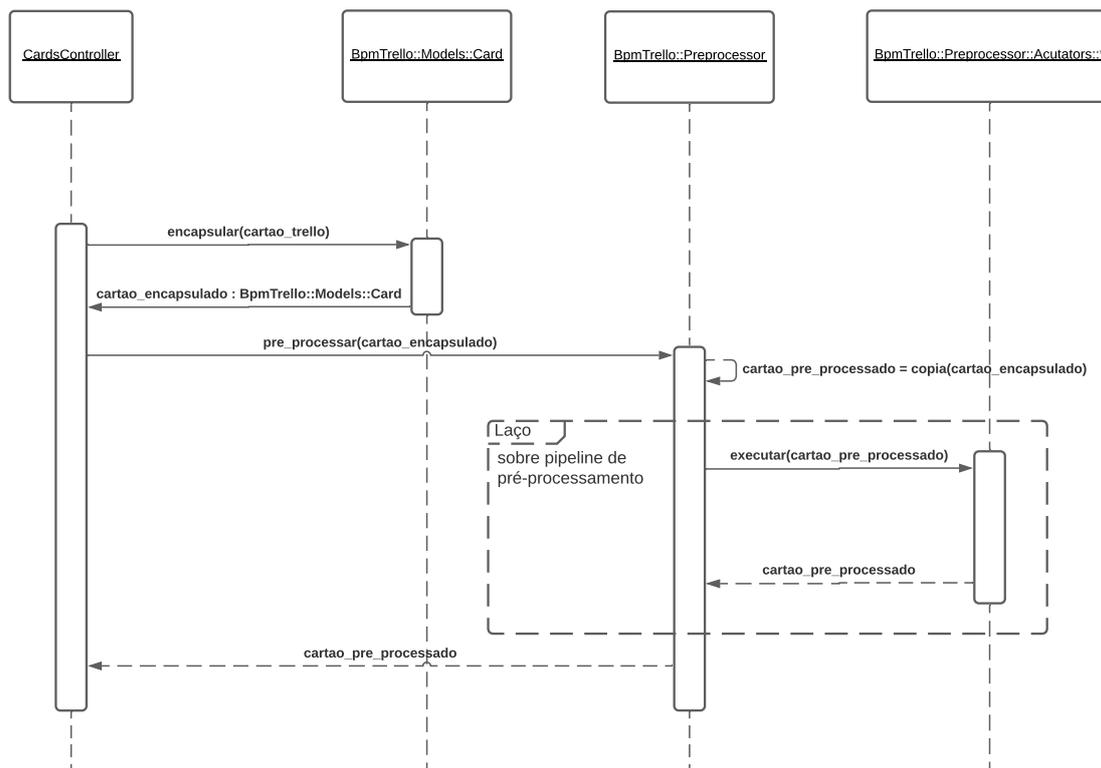
A fig. 4.9 exibe um diagrama de classes contendo os pacotes e classes utilizados na etapa de pré-processamento, bem como as suas relações de uso e extensão. Já a fig. 4.8, dado que a etapa de pré-processamento ocorre logo após a escolha do cartão pelo usuário, mostra a continuação direta do diagrama de sequência da fig. 4.5, representando a interação entre o controlador *CardsController* e as demais classes envolvidas no pré-processamento do cartão.

As classes do pacote *Actuators* são as mais importantes para a etapa de pré-processamento, pois são nelas que se encontram as lógicas para a aplicação de cada estágio do *pipeline*. Cada uma destas classes estende uma classe base, a *Base*, que define a interface que todas devem obedecer. No caso, essa interface define que todas as classes que a utilizam devem receber como entrada um objeto do tipo *Preprocessor::Models::Card* e implementar um método *run*, que também retorna um objeto desta mesma entidade. Isto permite que os estágios sejam encadeados no *pipeline*, com a entrada de um servindo como entrada para o seguinte. Nos próximos parágrafos desta seção são detalhados cada um dos estágios, na ordem em que são executados no *pipeline* de pré-processamento, definindo quais são as alterações que realizam, como estas são realizadas e porque o estágio é executado nesta ordem.

O primeiro estágio é o de *adição de pontuação*, realizado pela classe *Periodifier*. Este estágio adiciona pontos finais (“.”) ao fim do nome, descrição e comentários de um cartão. Isto é feito para facilitar a análise dos textos pela ferramenta *Stanford CoreNLP*, principalmente na separação de sentenças, visto que etapas de *pipelines* de PLN, principalmente tokenização e separação de sentenças, utilizam estas pontuações como identificadores para aplicar seus algoritmos (INDURKHYA; DAMERAU, 2010). A execução correta destes estágios é de suma importância visto que servem como base crucial para qualquer aplicação de PLN (INDURKHYA; DAMERAU, 2010).

O segundo estágio, *normalizar links em Markdown*, implementado pela classe *MarkdownLinksNormalizer*, executa a parte da definição 3 que trata de *links*, removendo todos os links em formato *Markdown* presentes em descrições e comentários do cartão, deixando no texto apenas seus rótulos. Isto é feito utilizando-se expressões regulares para encontrar padrões de *links* em *Markdown* e removê-las do texto, mantendo apenas o rótulo do *link*.

Figura 4.8 – Diagrama de sequência mostrando o fluxo de pré-processamento do cartão



Fonte: Os autores.

O terceiro estágio, *substituição de menções*, implementado pela classe *Mention-Substitution*, executa a definição 6, substituindo todas as menções presentes em um cartão pelo nome do respectivo usuário. Isto é feito utilizando expressões regulares para encontrar seções do texto no formato de menções (isto é, “@ + nome-de-usuário”). Em seguida, todas as menções encontradas são utilizadas para buscar pelo usuário mencionado na API da Trello usando seu nome-de-usuário (isto é, o texto da menção). Uma vez encontrado o usuário, a menção é substituída pelo seu nome.

O quarto estágio, *resolução de anáfora*, implementado pela classe *AnaphoraResolver*, utiliza a ferramenta de resolução de anáfora da *Stanford CoreNLP* para encontrar co-referências em descrições e comentários, substituindo-as pela referência principal. Como comentários podem ser interpretados como conversas, é natural que hajam co-referências entre comentários de usuários diferentes. Por isso, para que a ferramenta tenha um maior contexto quando da resolução de anáforas, comentários são concatenados em um texto único antes da execução da resolução e divididos novamente logo após, já com as anáforas substituídas.

É importante ressaltar que o estágio de substituição de menções é executado antes do estágio de resolução de anáforas. Isto é feito para remover as menções do texto, pois

as mesmas não fazem parte da sintaxe da língua inglesa, fazendo com que a análise da ferramenta *Stanford CoreNLP* seja realizada em um texto o mais próximo possível do inglês gramaticalmente correto, que é a língua suportada pela ferramenta e utilizada no protótipo.

O quinto estágio, *substituição de pronomes*, implementado pela classe *Pronouns-Substitution*, executa as definições 5 e 4, substituindo pronomes da 3ª pessoa do plural pelos nomes dos membros do cartão em descrições e comentários, e os de 1ª pessoa do singular pelo nome do autor do comentário em comentários. Isto é feito utilizando-se expressões regulares para encontrar pronomes como “*I*” e “*We*” no texto e, em seguida, substituí-los pelo nome correspondente (isto é, nome do comentário ou nome dos membros do cartão, respectivamente).

O último estágio, *remoção de ruídos*, implementado pela classe *NoiseRemover* é utilizado para filtrar seções do texto que não possuem informações de processo de negócio e que, portanto, não são úteis para analistas de processos ou ferramentas que automatizam a descoberta de processos. Neste sentido, têm-se dois tipos de seções que não agregam para a descoberta de processo: seções em *Markdown* e sentenças sem informações de processo de negócio.

De acordo com a definição 3, seções de código em *Markdown* podem ser ignoradas na análise por informações de processo de negócio. Por isso, a classe *NoiseRemover*, utilizando-se de expressões regulares, remove todo e qualquer texto que esteja entre marcadores de código em *Markdown* (isto é, “````” ou “```”). Já formatações de texto, como negrito, itálico e tachado, não possuem todo o seu conteúdo removido, mas apenas os seus marcadores, visto que o conteúdo ainda faz parte do texto em linguagem natural e pode ser analisado, como definido também na definição 3. Para isto, novamente expressões regulares são utilizadas para encontrar estes marcadores (isto é, “**\*\*\***”, “*\**” e “~”), que são removidos, mas mantendo o conteúdo textual presente entre eles.

Já quanto a sentenças sem informações de processo de negócio, (SILVA et al., 2018) define *ruído* (na língua inglesa, *noise*) como “sentenças que não fazem parte de uma descrição de processo” e que diminuem a eficácia da identificação e descoberta de processos.

Ainda mais, Ferreira, Thom and Fantinato (2017) define um conjunto de pré-requisitos para que um texto em linguagem natural possa ser utilizado em sua abordagem semi-automática de identificação de elementos de processo de negócio. Em particular, seu trabalho requer que o texto de entrada não possua sentenças interrogativas, que são con-

sideradas irrelevantes para a extração de informações de processo de negócio e, portanto, podem ser consideradas como ruído. Também é importante observar que todas as regras de identificação de elementos de processo definidas por Ferreira, Thom and Fantinato (2017) possuem verbos em sua composição, portanto, é possível definir que sentenças sem verbos não constituem uma informação de processo de negócio e, portanto, também podem ser consideradas ruídos.

Portanto, o estágio de remoção de ruídos remove de descrição e comentários sentenças que sejam interrogativas ou que não possuam verbos em sua composição, visto que não são úteis para a extração de informações de processo de negócio. Após sua execução, o cartão pré-processado é passado adiante para a etapa de extração de informações de processo de negócio

Sentenças são identificadas como interrogativas quando terminam pelo símbolo de interrogação “?”. Já sentenças sem verbos são identificadas quando nenhum de seus *tokens*, identificados pela ferramenta *tokenize* da *Stanford CoreNLP*, possui uma classe gramatical verbal.

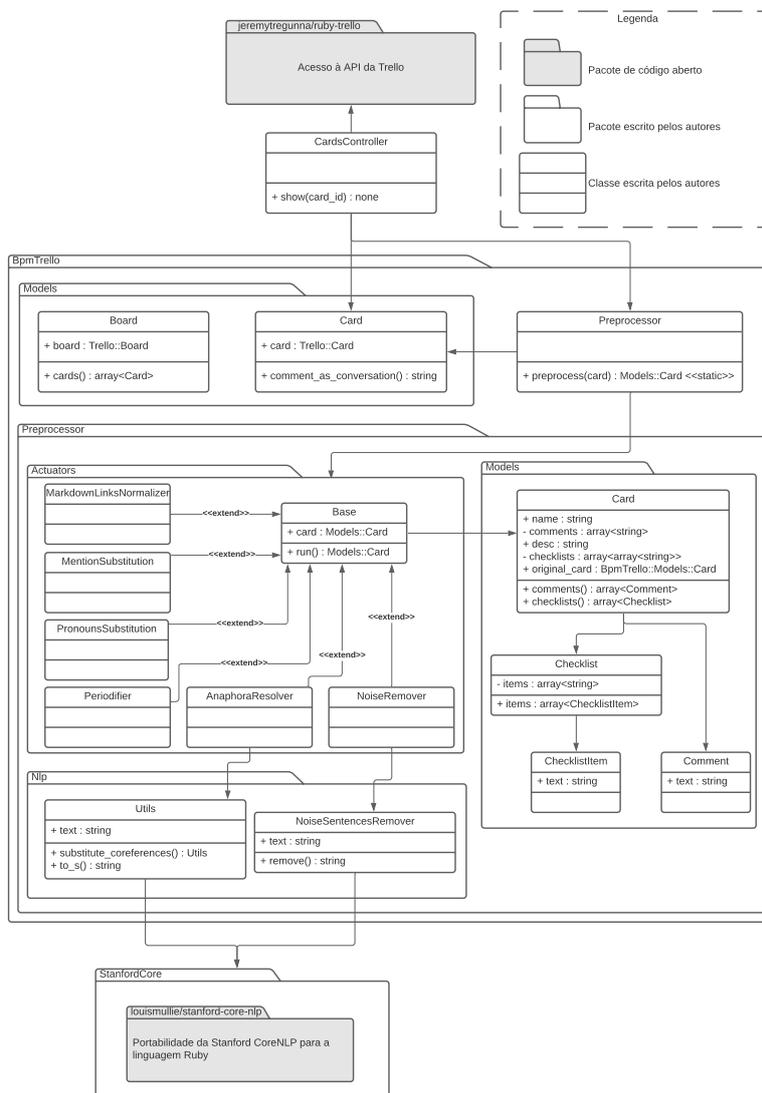
#### **4.4 Extração de informações de processo de negócio a partir de um cartão da Trello**

A terceira e última etapa do protótipo envolve a extração de informações de processo de negócio do cartão pré-processado e sua transformação em um único texto em linguagem natural. As extrações são feitas utilizando as definições de extração, em especial as definições 1 e 2.

As demais definições não tem como foco a extração direta de informações de processo de negócio, mas sim o processamento e enriquecimento dos textos de descrições e comentários com foco em seu uso por analistas de processos e ferramentas de extração automática de informações de processo de negócio. Para evidenciar a utilidade destas definições, foi implementado no protótipo uma versão simples da identificação de elementos de processo de negócio utilizando a regra 1 de Ferreira, Thom and Fantinato (2017) para a identificação de atividades em textos em linguagem natural.

Esta etapa em muito se assemelha com a etapa de pré-processamento, principalmente por ser uma continuação direta do seu contexto. Os algoritmos aqui detalhados são executados ainda no contexto da *CardsController*, logo após encerrada a etapa de pré-processamento. Depois que ambas as etapas são encerradas, isto é, o cartão foi completa-

Figura 4.9 – Diagrama de classes dos pacotes e classes envolvidas na etapa de pré-processamento de cartão



Fonte: Os autores.

mente processado, tanto o cartão pré-processado quanto o texto contendo as informações de processo são exibidas ao usuário.

Os pacotes e classes também são semelhantes entre as etapas, com os pacotes *Actuators*, *Models* e *Nlp* mantendo seus propósitos, porém com classes diferentes, dadas as necessidades distintas das duas etapas. A fig. 4.10 exhibe o diagrama de classes dos pacotes e classes envolvidos nesta etapa.

#### 4.4.1 Extração de definição de atividade de cartão

A aplicação da definição 1 se dá a partir da classe *TaskDefinition*. Para compreender seu funcionamento, se faz necessário, primeiramente, entender o algoritmo utilizado pela classe *Nlp::ActivitiesExtractor*, responsável por extrair atividades de textos em linguagem natural.

Para extrair atividades a partir de um texto foi utilizada a regra 1 para a identificação de atividades definida por Ferreira, Thom and Fantinato (2017). De acordo com a regra, atividades podem ser encontradas em textos pelo padrão “sujeito + verbo + objeto”. Desta forma, foi elaborado um algoritmo para buscar por seções do texto que estejam neste padrão e extraí-las como atividades no contexto do processo de negócio. O Algoritmo 1 apresenta o pseudocódigo da extração.

---

**Algoritmo 1:** Busca por atividades de processo de negócio em texto
 

---

**Entrada:** *texto* : texto em linguagem natural

**Saída:** *atividades* : lista de atividades de processo de negócio encontradas no texto

*sentencas*  $\leftarrow$  StanfordCore.anotar(*texto*);

*sentencas\_para\_analisar*  $\leftarrow$  [];

**para** *sentenca* em *sentencas* **faça**

**se** *sentenca.tem\_verbo* **então**

*sentencas\_para\_analisar*  $\leftarrow$

*sentencas\_para\_analisar*  $\cup$  [*sentenca*];

**fim**

**fim**

*atividades*  $\leftarrow$  [];

**para** *sentenca* em *sentencas\_para\_analisar* **faça**

**para** *verbo* em *sentenca* **faça**

*sujeitos*  $\leftarrow$  encontrar\_sujeitos(*verbo*, *sentenca*);

*objetos*  $\leftarrow$  encontrar\_objetos(*verbo*, *sentenca*);

*atividades\_and*  $\leftarrow$  atividades\_relacionadas\_and(*verbo*, *sentenca*);

*atividades\_or*  $\leftarrow$  atividades\_relacionadas\_or(*verbo*, *sentenca*);

*atividade*  $\leftarrow$  Atividade(*sujeitos*, *verbo*, *objetos*, *atividades\_and*,  
            *atividades\_or*);

*atividades*  $\leftarrow$  *atividades*  $\cup$  [*atividade*];

**fim**

**fim**

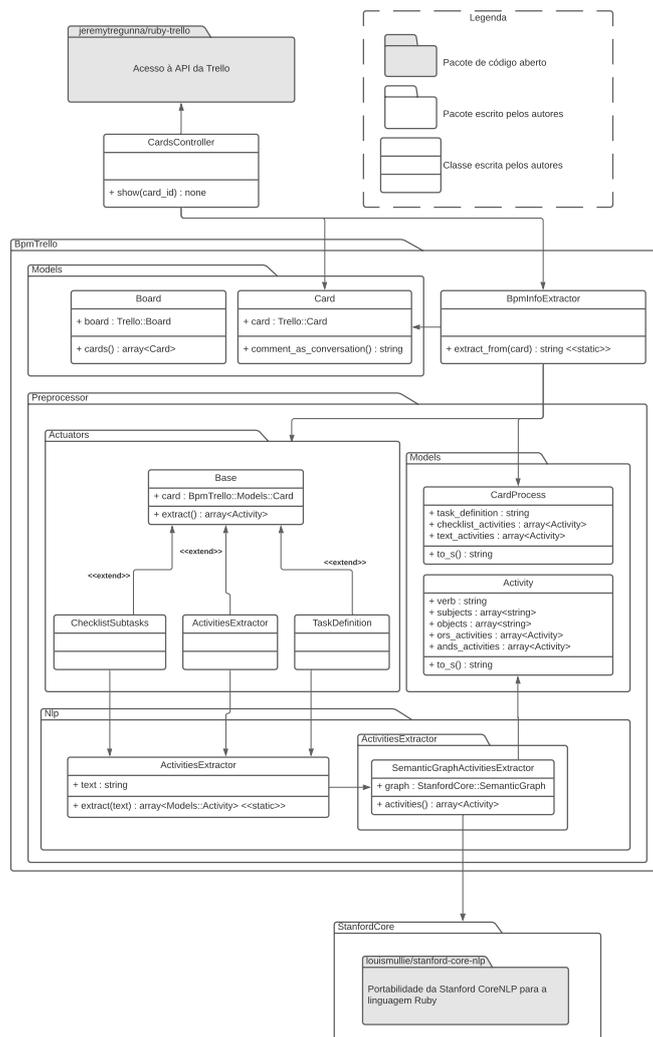
**retorna** *atividades*

---

O algoritmo começa utilizando a ferramenta *ssplit* da *Stanford CoreNLP* para separar o texto em suas respectivas sentenças e, então, itera sobre cada sentença, removendo da análise aquelas que não possuem nenhum verbo. Isto é feito utilizando a ferramenta *pos* da *Stanford CoreNLP*, que adiciona anotações contendo a classe gramatical de cada *token* da sentença. Caso nenhum *token* da sentença possua uma classe gramatical verbal, esta é removida da análise na busca por atividades no texto.

Após, cada uma das sentenças restantes é analisada na busca por atividades. Utilizando a ferramenta *depparse* da *Stanford CoreNLP*, é realizada a análise de dependências de cada sentença, buscando os sujeitos e objetos relacionados com cada verbo da sen-

Figura 4.10 – Diagrama de classes dos pacotes e classes envolvidas na etapa de extração de informações de processo de negócio



Fonte: Os autores.

tença em análise. Além disso, considerando a prática de Silva (2019) na identificação de atividades na plataforma *Slack*, através das dependências são encontradas, também, atividades relacionadas por conjunções *and* e *or*, que podem configurar decisões paralelas, exclusivas ou inclusivas.

Cada conjunto de verbos, sujeitos e objetos relacionados constituem uma atividade identificada. Estas atividades são representadas no protótipo pela classe *Activity*. Cada instância do tipo *Activity* pode ter um conjunto de outras instâncias *Activity* relacionadas, representando as outras atividades relacionadas com esta, seja por conjunções *and* ou *or*.

Apesar de utilizar o formato “sujeito + verbo + objeto”, O algoritmo não se limita a encontrar apenas atividades que contenham sujeito e objeto, visto que alguns verbos podem não ter um sujeito relacionado, o que torna impossível identificar o papel ou ator

Figura 4.11 – Página contendo o formulário de credenciais da Trello

Fonte: Os autores.

que executa a atividade. Isto possibilita seu uso para identificar o padrão “verbo + objeto” de rótulo de atividade no nome do cartão, que é necessário para a implementação da definição 1.

Assim, a classe *TaskDefinition* utiliza a *Nlp::ActivitiesExtractor* para analisar o nome do cartão, buscando por uma atividade no formato “verbo + objeto”. Quando encontrada, a atividade é combinada com os nomes dos membros do cartão, que são adicionados como seus sujeitos, resultando em uma instância de *Activity* que representa a atividade do cartão de acordo com a definição 1.

#### 4.4.2 Extração de sub-tarefas a partir de *checklists*

A definição 2 estabelece que *checklists* podem ser interpretadas como sub-tarefas da atividade representada pelo cartão. Ainda mais, a definição propõe um *template* que expressa esta relação na forma de texto.

Dentre as informações necessárias para a elaboração do texto a partir do *template*, a atividade do cartão e os membros do cartão já foram extraídas na extração da definição da atividade do cartão, restando apenas a extração das sub-tarefas contidas em *checklists*.

Para extrair as sub-tarefas de uma *checklist*, a classe *ChecklistSubtasks* utiliza a classe *Nlp::ActivitiesExtractor*, aplicando seu algoritmo em cada item da *checklist*, resultando em uma lista de atividades, uma para cada item. Isto é, o algoritmo 1, previamente utilizado para extrair a definição de atividade do nome do cartão, é também utilizado para extrair as atividades presentes em cada item de uma *checklist*. As atividades identificadas são retornadas pela classe e serão posteriormente utilizadas na elaboração da sentença de acordo com o *template* definido.

Figura 4.12 – Página contendo os quadros do usuário



Fonte: Os autores.

#### 4.4.3 Extração de atividades a partir de texto

Como textos em linguagem natural podem conter inúmeras informações de processo de negócio, como definido pelos axiomas 4 e 9, a extração de toda e qualquer informação presente nestes textos é complexa o suficiente para se tornar estudo de diversos trabalhos, como os de Friedrich, Mendling and Puhmann (2011) e Silva et al. (2018).

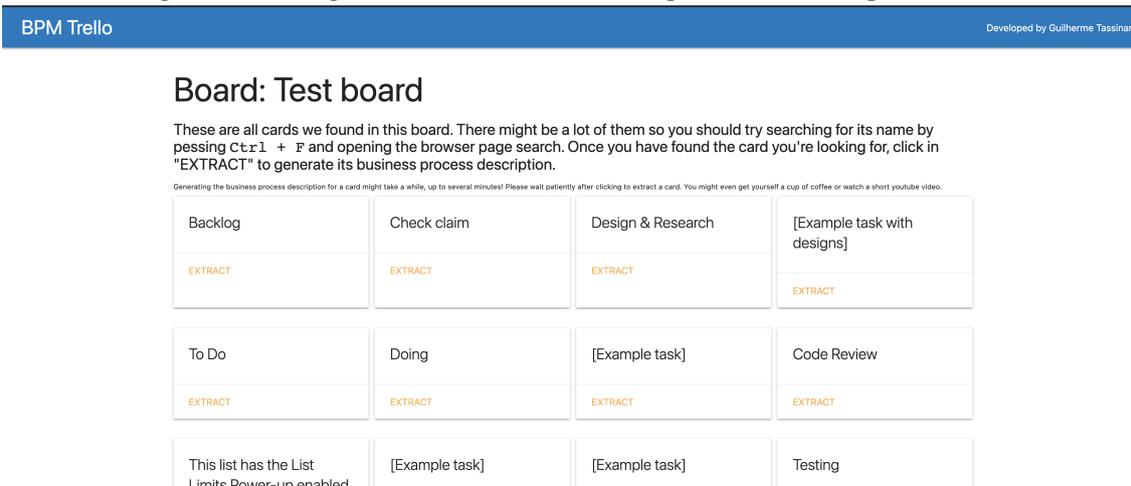
Com estes argumentos, as definições 3, 4, 5 e 6 foram elaboradas com o intuito de processar estes textos, enriquecendo-os com informações relevantes para a descoberta de processos de negócio. Ainda mais, o último estágio da etapa de pré-processamento foi elaborado com o intuito de remover destes textos sentenças que não se provem úteis para a descoberta de processos de negócio. Por fim, têm-se um conjunto de textos pré-processados contendo apenas sentenças consideradas úteis para a descoberta de processos de negócio.

Para evidenciar a utilidade deste processamento, o protótipo aplica o algoritmo 1, que implementa a regra 1 de identificação de atividades, definida por Ferreira, Thom and Fantinato (2017), nos textos pré-preprocessados de descrição e comentários do cartão, extraindo assim uma lista de atividades de processo de negócio, que também serão adicionados ao texto final contendo informações de processo de negócio.

#### 4.4.4 Geração de texto com informações de processo de negócio

Após executadas todas as definições, têm-se um conjunto de atividades composta pela definição da atividade do cartão, as sub-tarefas identificadas nas *checklists* do cartão e as atividades identificadas nas descrições e comentários do cartão. Todas estas atividades

Figura 4.13 – Página contendo os cartões do quadro escolhido pelo usuário



Fonte: Os autores.

são, por fim, passadas à classe *CardProcess*, que é responsável por combiná-las em um único texto contendo todas as informações de processo identificadas.

O texto começa com a concatenação da definição da atividade do cartão. Em seguida, é adicionada a sentença elaborada pelo *template* da definição 2, utilizando-se os nomes membros do cartão e as atividades identificadas nas *checklists* do cartão. Por último, as atividades identificadas na descrição e comentários são adicionadas ao fim do texto como “outras atividades” no contexto do processo de negócio, visto que não se tem informações suficientes para inferir relações entre elas e a atividade do cartão ou as sub-tarefas de *checklists*.

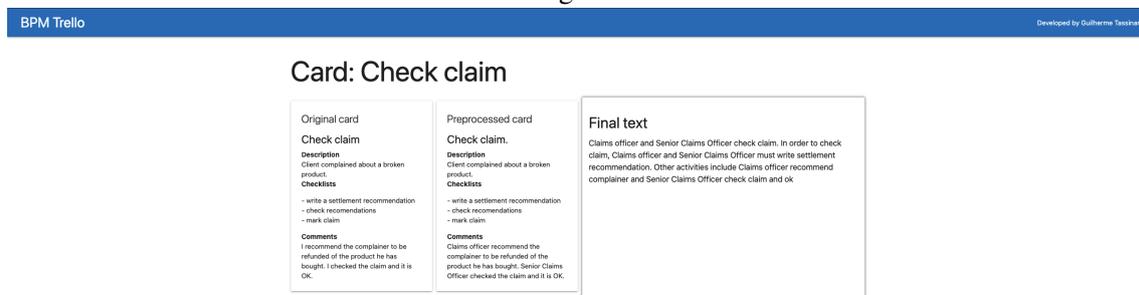
#### 4.5 Exemplo de uso do protótipo

Esta seção demonstra o uso do protótipo. Isto é feito através de imagens (*print screens* do protótipo) acompanhadas de uma descrição, passo-a-passo, da navegação pelas páginas, desde a etapa de entrada das credenciais da Trello até a exibição dos textos extraídos do cartão.

Ao acessar o protótipo por um navegador web, a primeira página exibida é a que contém o formulário para o envio das credenciais de acesso à API Trello (fig. 4.11). A página é composta por um cabeçalho, em azul, contendo o título do projeto, “BPM Trello”, e um *link* que credita o desenvolvimento da plataforma à um dos autores do trabalho e

redireciona o usuário para o perfil do autor na plataforma GitHub<sup>8</sup>. Este cabeçalho é exibido em todas as páginas da plataforma.

Figura 4.14 – Página contendo o cartão pré-processado e o texto com informações de processo de negócio



Fonte: Os autores.

O formulário de envio de credenciais é composto por um título, um texto contendo descrições sobre como conseguir as credenciais na plataforma Trello, incluindo um *link* para a página da documentação da Trello que ensina o usuário a gerar as credenciais. Há também um pequeno texto avisando o usuário de que as credenciais não são armazenadas e apenas são conhecidas pelo protótipo durante a seção corrente.

Logo abaixo estão as entradas do formulário, compostas por dois campos de texto simples, onde devem ser inseridas as credenciais *APP KEY* e *API KEY* ou *API TOKEN*. Após inseridas as credenciais o usuário deve clicar no botão “*SUBMIT*” para enviá-las ao servidor.

Após inseridas e enviadas as credenciais de acesso à API da Trello, o usuário é redirecionado para a página de quadros (fig. 4.12). Nesta página são exibidos todos os quadros do usuário ao qual o protótipo teve acesso. A página começa com um título que cumprimenta o usuário, chamando-o pelo seu nome cadastrado na plataforma Trello. Abaixo do título há um pequeno texto explicando que esta página exibe uma lista de quadros do usuário e que, ao clicar em um dos quadros, o usuário poderá ver os cartões deste.

Clicando em quadro, o usuário é redirecionado para a página de cartões. A página exibe todos os cartões encontrados pelo quadro selecionado e é composta por um título, exibindo o nome do quadro escolhido, seguido por um texto orientando o usuário a escolher um cartão, clicando no botão com texto “*EXTRACT*”, para que este seja processado na busca por informações de processo de negócio. O texto também indica ao usuário que,

<sup>8</sup><<http://github.com/>>

devido ao possível grande número de cartões disponíveis, este pode buscar pelo nome do cartão utilizando a funcionalidade de busca do navegador.

Abaixo do texto de orientação há um pequeno texto avisando o usuário de que, após escolhido o cartão, o processamento deste pode demorar vários minutos e que o usuário deve ter paciência durante esse período, aguardando o término do processamento e o carregamento da próxima página. Em seguida, todos os cartões do quadro são listados na forma de retângulos brancos contendo o nome do cartão e um botão com o título “*EXTRACT*”, que o usuário deve clicar para processar o cartão.

Após selecionado o cartão, é iniciado o pré-processamento e a extração de informações de processo de negócio. O processamento pode demorar vários minutos. Após terminado, é exibida ao usuário a página contendo um título com o nome do cartão, os conteúdos principais do cartão original e do cartão pré-processado, e o texto contendo as informações de processo de negócio (fig. 4.14).

## 5 RESULTADOS, AVALIAÇÃO E DISCUSSÃO

Este capítulo avalia e discute os resultados obtidos na aplicação da abordagem e, respectivo protótipo desenvolvido no contexto deste trabalho, ressaltando pontos positivos, negativos e possíveis melhorias.

Para avaliar os resultados da abordagem e do protótipo desenvolvido, precisamos relembrar a hipótese do trabalho e seus objetivos. O presente estudo têm como hipótese que *a plataforma de gestão de projetos Trello pode servir como fonte de informações de processo de negócio*, em especial no contexto da etapa de descoberta de processos.

Para validar esta hipótese, (i) *criou-se um conjunto de definições capazes de extrair informações de processo da plataforma Trello na forma de texto* e (ii) *foi desenvolvido um protótipo como prova de conceito para estas definições*.

Portanto, para avaliar os resultados deste trabalho, deve-se avaliar se os artefatos de texto gerados pelo protótipo, uma vez aplicadas as definições de extração, possuem informações de processo de negócio relevantes para a descoberta de processos de negócio e são passíveis de uso por analistas de processo.

Com este intuito, a abordagem de avaliação utiliza um conjunto de questionários, aplicados com a participação de analistas de processo, com foco em identificar se o texto gerado pelo protótipo é passível de ser utilizado para a modelagem manual de processos de negócio, como geralmente é feito em projetos de BPM.

Para avaliar a utilidade do texto gerado pelo protótipo para analistas de processo em projetos de BPM, conduzimos um estudo de caso onde, através de um conjunto de questionários, analistas de processo realizam um conjunto de análises e transformações a partir de uma descrição de processo, passando por um cartão Trello e terminando em um modelo de processo utilizando-se a BPMN 2.0. Em paralelo, a mesma abordagem é aplicada com a utilização do protótipo, que também gera um modelo de processo utilizando a BPMN 2.0. Por fim, ambos os modelos gerados são comparados em termos de suas Similaridades Estruturais (DIJKMAN et al., 2011).

O objetivo do estudo é usar como referência as informações de processo que analistas de processos são capazes de identificar a partir de um cartão Trello. Comparando ambos os modelos gerados pelos questionários, isto é, aquele elaborado a partir da descrição de processo elaborada por analistas e aquele elaborado a partir do texto gerado pelo protótipo, pode-se avaliar o quanto das informações de processo disponíveis em um cartão Trello o protótipo foi capaz de identificar.

Este capítulo está organizado como segue. A seção 5.1 explica a metodologia de avaliação, explicando e exibindo em detalhes os formulários aplicados, quais informações foram solicitadas em cada um, em que ordem estes foram aplicados e que medidas foram tomadas para se evitarem vieses durante o processo de avaliação. Além disso, também é resumida a técnica de comparação de modelos de processo através da Similaridade Estrutural, definida por Dijkman et al. (2011). A seção 5.2 apresenta os resultados dos questionários, detalhando as respostas coletadas e apresentando os dados demográficos dos participantes e as similaridades entre os modelos. Por fim, a seção 5.3 discute os resultados da avaliação, apresentando pontos positivos e negativos, analisando pontos de falha e propondo soluções e melhorias para a abordagem e o protótipo.

## 5.1 Metodologia

A metodologia de avaliação é composta por um conjunto de questionários que foram aplicados de maneira consecutiva, com a resposta de um questionário servindo como entrada para o questionário seguinte. O objetivo disto é que diferentes analistas de processos participem em diferentes estágios, sem ter conhecimento prévio do que foi feito nos estágios anteriores, prevenindo a existência de possíveis vieses.

No total foram elaborados 4 questionários, aplicados no decorrer de três estágios, com cada questionário sendo respondido por uma analista de processo diferente dos demais. A fig. 5.1 exibe uma visão geral dos estágios e questionários aplicados, bem como os artefatos recebidos e gerados em cada um deles. O objetivo com a aplicação deste conjunto de questionários é (i) simular o caminho de um processo quando implementado com a utilização da Trello e (ii) simular a utilização de dados da plataforma Trello por analistas de processo em uma etapa de descoberta de processos.

Para simular a implementação de um processo de negócio na plataforma Trello, o primeiro questionário solicita aos respondentes que, dada uma descrição de processo (fornecida no questionário), estes a transformem em um cartão da Trello utilizando nome de cartão, membros do cartão, descrição, data de entrega, *checklists* e comentários.

No questionário são pedidos também dados demográficos (idade, principal ocupação, setor da sua principal ocupação), dados sobre a experiência do respondente com modelagem de processos de negócio e também sobre a sua experiência com a plataforma Trello (anos de experiência de uso, quando foi a última vez que modelou um processo de

negócio/usou a plataforma Trello e autoavaliação do seu nível de experiência em modelagem/Trello).

Foi colocada também uma breve introdução à Trello e às suas principais funcionalidades, com imagens de exemplo para quadros, listas e cartões. O propósito é garantir que, mesmo quando o analista de processo não tem nenhuma experiência prévia com a plataforma, este tenha o mínimo de conhecimento necessário para transformar a descrição em um cartão Trello.

As figuras 5.2, 5.3 e 5.4 exibem, respectivamente, as questões demográficas, de experiência em modelagem de processo de negócio e de experiência no uso da plataforma Trello utilizadas nos questionários. Já a fig. 5.5 exibe a seção de introdução ao Trello colocada no primeiro questionário, contendo uma breve explicação sobre a plataforma e imagens de um quadro, uma lista e um cartão da Trello.

Na última seção do primeiro questionário é solicitado ao respondente que, considerando a descrição de processo exibida, este “modele” o processo utilizando o cartão Trello, procurando expressar da melhor maneira possível as informações do processo nos dados do cartão.

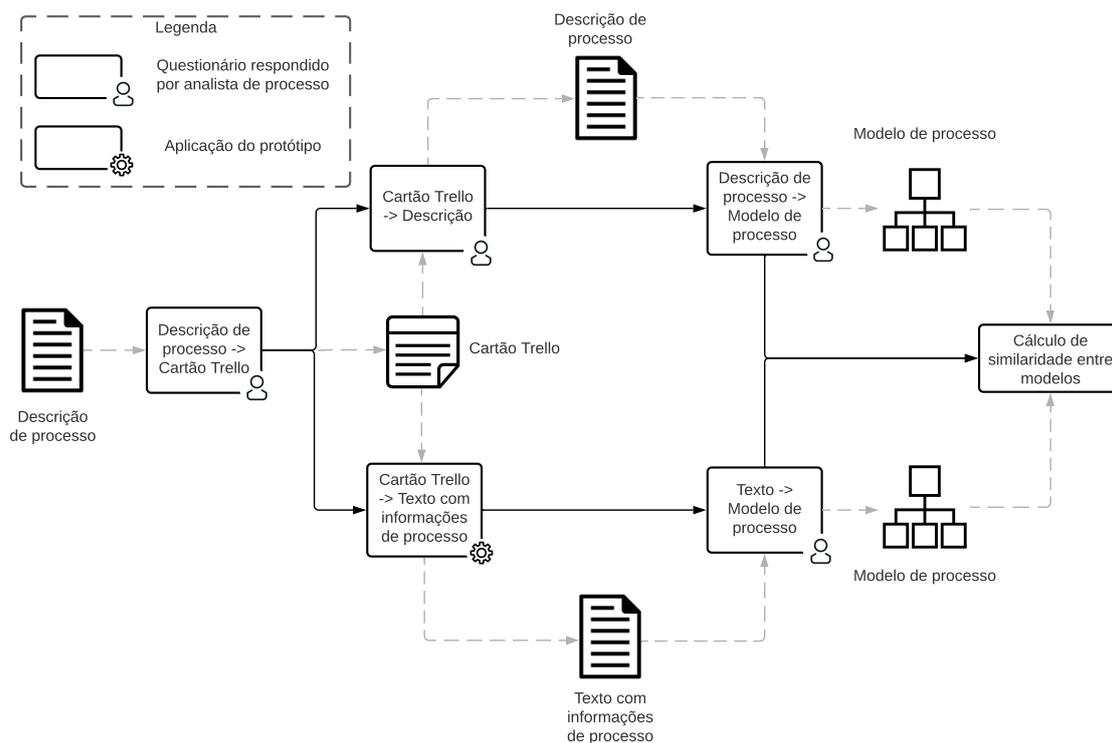
Para evitar que o respondente tenha que criar uma conta na Trello para responder ao questionário, foram fornecidas instruções a respeito do formato com que cada campo deveria ser preenchido. Por exemplo, no campo de *checklists*, foi solicitado que o respondente escreva uma *checklist* por linha, começando pelo título da *checklist*, seguido dos seus itens, todos separados por “;”. A fig. 5.6 exibe esta seção, contendo todas as questões necessárias para que o respondente modele o processo em um cartão Trello.

Após aplicado o primeiro formulário, têm-se as informações necessárias para se criar um cartão Trello baseado na descrição de processo fornecida. O segundo questionário é feito com base neste cartão Trello e tem por objetivo simular a análise feita por um analista de processo para descobrir informações de processo de negócio a partir do cartão.

Da sua elaboração, o segundo questionário é muito similar ao primeiro, mantendo as questões demográficas (fig. 5.2), de experiência em modelagem de processo (fig. 5.3) e experiência com a plataforma Trello (fig. 5.4). Já a seção de introdução à Trello é mantida parcialmente, mantendo-se apenas as informações e imagens sobre cartões da Trello.

A seção final é onde está a maior diferença, já que o respondente deve, agora, realizar o inverso do primeiro questionário, isto é, ele deve transformar um cartão Trello em uma descrição de processo. Para a elaboração deste questionário, utilizamos as informações das respostas do primeiro questionário para elaborar cartões Trello. Em seguida,

Figura 5.1 – Metodologia de avaliação



Fonte: Os autores.

realizamos capturas de tela destes cartões e as colocamos no segundo questionário, pedindo ao respondente que, considerando o cartão exibido, escrevesse uma descrição de processo que melhor representasse o que estava sendo expresso no cartão. A fig. 5.7 exibe esta última seção de um dos questionários elaborados.

Em paralelo com a aplicação do segundo questionário, o cartão Trello elaborado é também processado pelo protótipo, gerando o texto com informações de processo de negócio de acordo com a aplicação das definições elaboradas neste estudo. Tendo-se então a descrição de processo da resposta do segundo questionário e o texto com informações de processo de negócio gerado pelo protótipo, aplica-se o terceiro questionário, de maneira paralela, a dois analistas de processo. Um dos analistas recebe o questionário contendo a descrição de processo (resposta do segundo questionário), o outro recebe o questionário com o texto com informações de processo (gerado pelo protótipo). Ambos devem, com base em seus respectivos textos, elaborar um modelo de processo de negócio.

O terceiro questionário não possui as questões envolvendo a experiência do respondente com a plataforma Trello, visto que este conhecimento não se faz necessário neste questionário, mas mantém as questões demográficas (fig. 5.2) e de experiência em modelagem de processo (fig. 5.3). A seção de introdução à Trello também é removida

Figura 5.2 – Questões demográficas dos formulários aplicados

Seção 1 de 3

### Process modeling on Trello cards

This questionnaire aims to collect data that enable the validation of an approach to extract business process information from cards of the Trello platform.

In this questionnaire, general information about the respondent will be collected.

When answering this questionnaire, the respondent allows these data to be used for the research in question and disclosed anonymously.

#### Respondent general information

Descrição (opcional)

What is your age? \*

- 18 years or younger
- 18-24 years
- 25-34 years
- 35-44 years
- 45-54 years
- 55 years or older

Are you connected to the public, academic or private sector? \*

- Academic (eg.: Professor, researcher, student...)
- Private (eg.: Process Analyst, Software Engineer ...)
- Public (eg.: Public profession in federal, state or municipal institution)
- Outros...

What is your main occupation? \*

- Graduation student
- Masters student
- PhD student
- Process analyst
- Researcher
- Professor
- Outros...

Fonte: Os autores.

pelo mesmo motivo. A última seção é composta pelo texto com informações de processo de negócio (ou descrição de processo) e pede que o respondente insira, como anexo, uma imagem contendo um modelo de processo de negócio modelado com base no texto exibido. A fig. 5.8 exhibe esta última seção.

Finalmente, após aplicados todos os questionários, têm-se dois modelos de processo: um baseado na descrição de processo elaborada por um analista de processo sobre o cartão Trello e outro baseado no texto com informações de processo gerado pelo protótipo sobre o mesmo cartão. Pra avaliar o quanto o modelo elaborado sobre o texto do protótipo se aproxima daquele elaborado por analistas de processos, é calculada a sua similaridade utilizando a técnica *Similaridade Estrutural* definida por Dijkman et al. (2011).

Figura 5.3 – Questões sobre a experiência em modelagem de processos de negócio dos formulários aplicados

About your experience with business process modeling  
Descrição (opcional)

How many years of experience do you have in business process modeling? \*

None  
 Less than 2 years  
 Between 2 and 4 years  
 More than 4 years

When was the last time you modeled a process? \*

I have never modeled a process  
 This week  
 Last week  
 Last month  
 Last 6 months or latter

...

How do you consider your level of experience in business process modeling? \*

Very low    1    2    3    4    5    Very high

Fonte: Os autores.

O cálculo de similaridade estrutural realiza uma comparação estrutural entre dois modelos de processo de acordo com a *distância de edição entre grafos*, isto é, o número de alterações necessárias para transformar um grafo em outro.

Para realizar o cálculo deve-se, primeiro, transformar ambos os modelos em grafos de processos de negócio (*Business Process Graph* ou BPG), onde cada nó do grafo representa um elemento de processo de negócio (atividade, evento, decisão) e arestas representam os fluxos de sequência entre elementos do modelo. Cada nó do grafo é composto, também, por um conjunto de atributos, que caracterizam a representação deste elemento no modelo de processo de negócio, como o seu rótulo, por exemplo.

É importante ressaltar que existem outros atributos, como o tipo do nó ou a raia a qual o elemento pertencia no modelo de processo. No contexto deste trabalho, considerarmos apenas o atributo rótulo é suficiente. A fig. 5.9 exhibe um exemplo de BPG, elaborado a partir do modelo de processo da fig. 2.3 e contendo apenas o atributo rótulo.

Tendo-se os BPGs dos dois modelos de processo a serem comparados, deve-se calcular o número de edições necessárias para se transformar um grafo no outro, isto é, a distância de edição dos grafos (DIJKMAN et al., 2011). São consideradas alterações (i) a remoção/adição de nós, (ii) a remoção/adição de arestas e (iii) a similaridade entre nós semelhantes dos dois grafos.

Figura 5.4 – Questões sobre a experiência no uso da plataforma Trello dos formulários aplicados

About your experience with the Trello platform

Descrição (opcional)

How many years of experience do you have with the Trello platform use? \*

None

Less than 2 years

Between 2 and 4 years

More than 4 years

!!!

When was the last time you used the Trello platform? \*

I have never used it

This week

Last week

Last month

Last 6 months or latter

How do you consider your level of experience using the Trello platform? \*

Very low    1    2    3    4    5    Very high

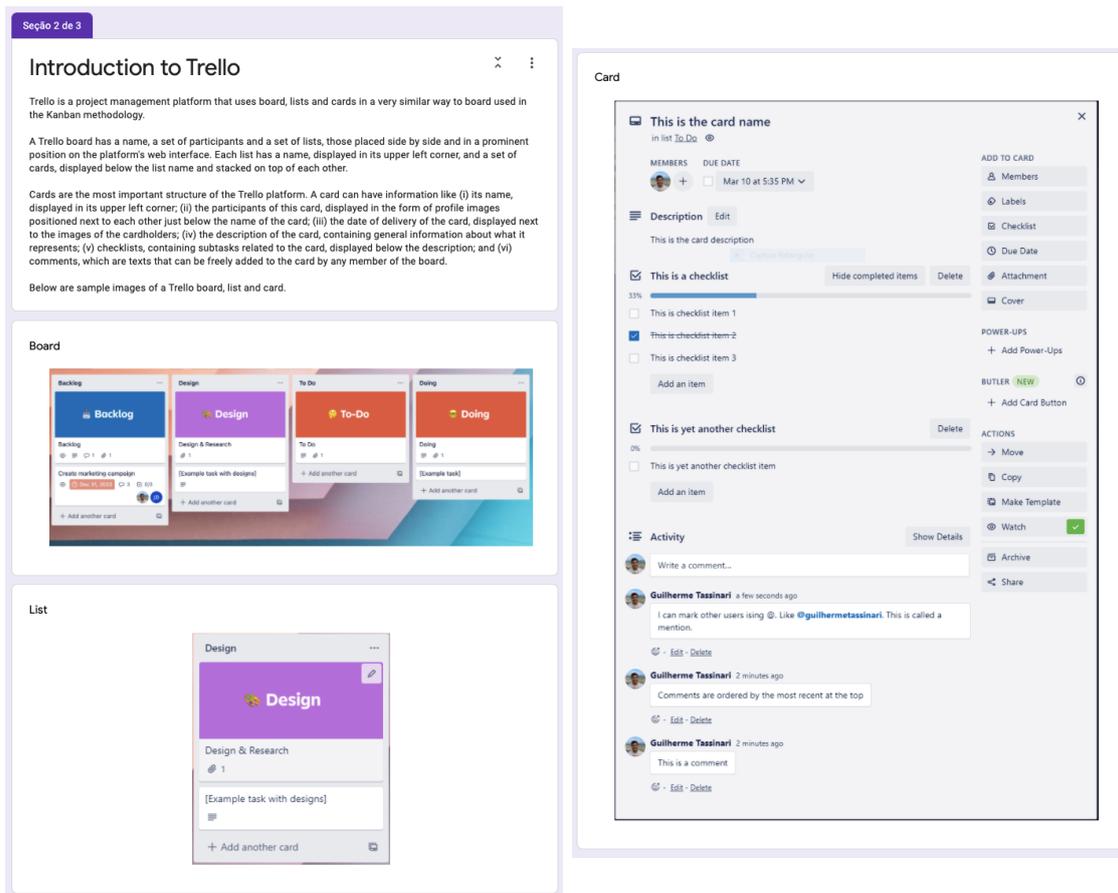
Fonte: Os autores.

O primeiro passo para o cálculo é identificarmos nós entre os dois grafos que possam ser considerados semelhantes, chamados *nós mapeados*. Isto é feito calculando-se a similaridade sintática entre todos os nós dos dois BPGs e considerando como mapeados todos aqueles que atingirem um limiar mínimo de similaridade. A *similaridade sintática* de dois nós é calculada com base na distância de edição (ou distância de Levensthein) entre os rótulos destes, isto é, quantas alterações atômicas são necessárias para transformar um rótulo em outro (DIJKMAN et al., 2011).

Tendo-se o conjunto de nós mapeados, todos os nós restantes são considerados como alterações necessárias para a transformação de um BPG em outro. Já as arestas, só não são consideradas alterações aquelas que ligarem nós mapeados em ambos os grafos (DIJKMAN et al., 2011).

O resultado final dos cálculos de similaridade estrutural e similaridade sintática é um número entre 0 e 1, onde 0 significa que os itens comparados são completamente diferentes e 1 significa que os itens são idênticos (DIJKMAN et al., 2011). Para o limiar mínimo de similaridade sintática entre nós consideramos o mesmo utilizando como exemplo por Dijkman et al. (2011), 0,5. Isto é, para serem considerados mapeados, dois nós não podem alterar mais da metade de seus rótulos para serem transformados um no outro.

Figura 5.5 – Introdução ao Trello exibida no primeiro questionário



Fonte: Os autores.

Considerando dois BPGs, A e B, sendo  $nalt$  o conjunto de nós não mapeados,  $aalt$  o conjunto de arestas consideradas alterações,  $nmap$  os pares de nós mapeados,  $tn$  o número total de nós em ambos os grafos e  $ta$  o número total de arestas em ambos os grafos, a similaridade estrutural  $Sim$  entre os BPGs A e B, baseada no cálculo de similaridade sintática  $SimSint$ , é dada por

$$Sim(A, B) = 1 - media\left(\frac{|nalt|}{tn}, \frac{|aalt|}{ta}, 2\left(\frac{\sum_{n \in nmap} 1 - SimSint(n)}{tn - |nalt|}\right)\right)$$

## 5.2 Resultados

Foram realizadas duas rodadas de questionários, em paralelo, com grupos diferentes de analistas de processo, totalizando 8 respostas a questionários. Ambos os grupos responderam os questionários com a mesma descrição de processo inicial, a versão origi-

Figura 5.6 – Questões para a modelagem de um processo de negócio em um cartão Trello

Seção 3 de 3

## Process modeling on a Trello card

Given the process description below, fill in the questionnaire questions trying to represent the process described, simulating the use of a Trello card for this representation. At the end of this section, the example image of the Trello card shown previously is repeated, for reference purposes.

**Process description**

After a claim is registered, it is examined by a claims officer who then writes a settlement recommendation. This recommendation is then checked by a senior claims officer who may mark the claim as "OK" or "Not OK". If the claim is marked as "Not OK", it is sent back to the claims officer and the recommendation is repeated. If the claim is "OK", the claim handling process proceeds.

**Card name \***

Texto de resposta curta

**Card members**

Names or positions, separated by commas

Texto de resposta curta

**Card due date**

Mês, dia, ano 

**Card description**

Texto de resposta longa

**Card checklists**

One line for each checklist, starting with the checklist title followed by its items, all separated with semi-colons. For example: "This is a checklist title; this is checklist item 1; this is checklist item 2; this is checklist item 3"

Texto de resposta longa

**Card comments**

One line per comment, starting with the name of the participant who writes the comment, followed by a colon and the content of the comment. For example: "Guilherme Tassinari: this is a comment"

Texto de resposta longa

Fonte: Os autores.

nal da descrição de processo traduzida na seção 1 e retirada de Dumas et al. (2018). Segue a versão original, em inglês, utilizada nos questionários.

### Texto 2 : Descrição de processo de solução de reclamação (original)

1 *After a claim is registered, it is examined by a claims officer who then writes a set-*  
 2 *tlement recommendation. This recommendation is then checked by a senior claims*  
 3 *officer who may mark the claim as "OK" or "Not OK". If the claim is marked as*  
 4 *"Not OK", it is sent back to the claims officer and the recommendation is repeated.*  
 5 *If the claim is "OK", the claim handling process proceeds.*

Figura 5.7 – Questões para a transformação de um cartão Trello em descrição de processo

Seção 3 de 3

## Process description from Trello card

Given the Trello card below, fill in the questionnaire question creating a process description that describes the process shown in the Trello card.

Here are the Trello card member names represented by the images with their initials:  
 - CO: Claims Officer  
 - SO: Senior Claims Officer

Trello card

**Check claim**  
 in list Backlog

MEMBERS: CO, SO

DUE DATE: Mar 17 at 12:00 AM (OVERDUE)

**Description** Edit

After a claim is registered, it is examined by a claims officer who then writes a settlement recommendation. This recommendation is then checked by a senior claims officer who may mark the claim as "OK" or "Not OK". If the claim is marked as "Not OK", it is sent back to the claims officer and the recommendation is repeated. If the claim is "OK", the claim handling process proceeds.

**Tasklist** Delete

0%

- write a settlement recommendation
- check recommendations
- mark claim

Add an item

**Activity** Show details

Write a comment...

**ADD TO CARD**

- Members
- Labels
- Checklist
- Due date
- Attachment
- Cover

**POWER-UPS**

- + Add Power-Ups
- Get unlimited Power-Ups, plus much more.
- Upgrade team

**BUTLER**

- + Add button

**ACTIONS**

- Move
- Copy
- Make template
- Watch
- Archive
- Share

**Process description \***

Texto de resposta longa

Fonte: Os autores.

No primeiro questionário os participantes tinham de, com base na descrição de processo fornecida, elaborar um cartão Trello que representasse o processo descrito. Como são poucos os participantes em todo o processo de avaliação, exibir gráficos para demonstrar as informações demográficas tem muita utilidade. Ainda mais, saber dos níveis individuais de experiência dos respondentes em cada questionário é importante para contextualização dos resultados e a análise destes. Por isso, mostramos os dados demográficos de experiência na forma de tabelas, uma para cada questionário. A do primeiro questionário é exibida na tabela 5.1, com os cartões Trello elaborados a partir das respostas exibidos na fig 5.10.

Figura 5.8 – Questão para modelagem de processo de negócio com base em texto/descrição de processo

Seção 2 de 2

### Process modelling

Given the process description below, build a business process model for it and upload it as a image.

**Process description**

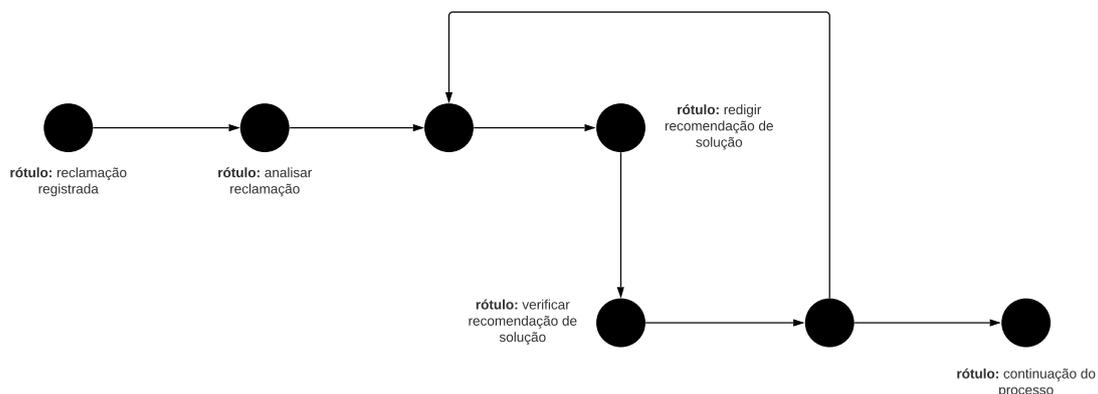
Process: Check Claim  
 Rules involved: Claims Officer (CO) and Senior Claims Officer (SO)  
 Process description:  
 The process starts when a claim is registered.  
 The CO should "exam the claim," and after that, "write a settlement recommendation."  
 The "settlement recommendation" goes to the SO;  
 The SO should "check the settlement recommendation," and after that, "mark the claim".  
 If the claim is marked as "Not OK", the SO should "send the settlement recommendation back" to the CO.  
 If the claim is marked as "OK", the process "check claim" is finished.

**Process model image \***

Adicionar arquivo

Fonte: Os autores.

Figura 5.9 – Exemplo de BPG, elaborado a partir do BPD da fig. 2.3



Fonte: Os autores.

Ao analisarmos a resposta do Grupo 1 é possível perceber a presença de alguns dos axiomas definidos no presente estudo, mesmo sem nenhum tipo de indicação sobre as possíveis relações entre os dados do cartão e conceitos de BPM ter sido fornecida aos respondentes do questionário. Por exemplo, os atores presentes na descrição de processo, *Claims Officer* e *Senior Claims Officer*, foram colocados como membros do cartão, como definido pelo axioma 3. Outras presenças são: nome do cartão como tarefa no formato “verbo + objeto” (axioma 2), descrição do cartão contendo diversas informações de processo, visto que a descrição de processo completa foi inserida no campo (axioma 4) e sub-tarefas colocadas na *checklist* (axioma 15).

Já a resposta do Grupo 2, quando comparada à do Grupo 1, possui bem menos informações no cartão Trello, que foi composto apenas por nome do cartão, descrição e membros do cartão. Acreditamos que o principal motivo para a discrepância entre os car-

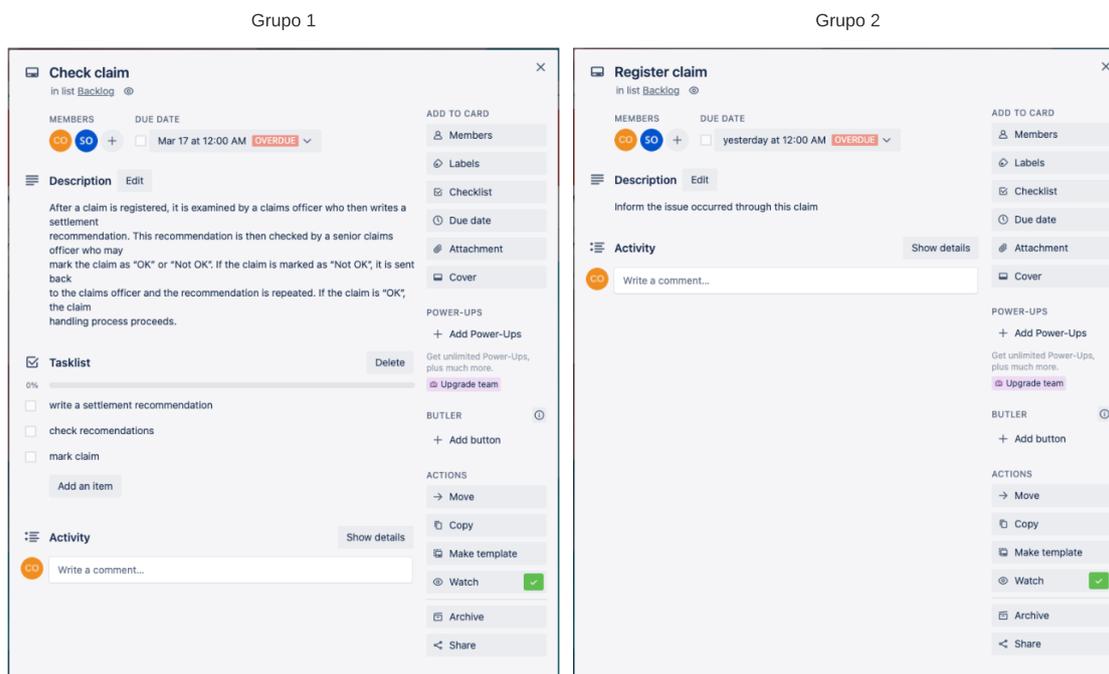
tões é o fato de o respondente do Grupo 2, apesar de possuir experiência em modelagem de processo, não possuir experiência alguma no uso da plataforma Trello, o que pode ter se refletido na maneira com que dispôs as informações no cartão. Isto também pode indicar que a seção de introdução à Trello (fig. 5.5) tenha sido insuficiente em fornecer os conhecimentos necessários para o preenchimento do cartão. Ainda assim, os axiomas 2 e 3 também se fazem presentes neste cartão.

Tabela 5.1 – Dados demográficos e de experiências individuais dos respondentes do primeiro questionário

<i>Questão</i>	<i>Grupo 1</i>	<i>Grupo 2</i>
Qual é a sua idade?	25-34 anos	25-34 anos
Você está conectado ao setor público, acadêmico ou privado?	Acadêmico	Privado
Qual é a sua principal ocupação?	Estudante de mestrado	Estudante de doutorado
Quantos anos de experiência você possui em modelagem de processos de negócio?	Entre 2 e 4 anos	Mais de 4 anos
Quando foi a última vez que você modelou um processo?	No último mês	No último mês
Como você considera o seu nível de experiência em modelagem de processos de negócio?	4 (alto)	4 (alto)
Quantos anos de experiência você possui com a plataforma Trello?	Menos de 2 anos	Nenhum
Quando foi a última vez que você utilizou a plataforma Trello?	No último mês	Eu nunca usei
Como você considera o seu nível de experiência em modelagem de processos de negócio	3 (médio)	1 (muito baixo)

Fonte: Os Autores

Figura 5.10 – Cartões Trello elaborados a partir das respostas do primeiro questionário



Fonte: Os autores.

No segundo questionário os respondentes tinham de, considerando os dados presentes no cartão extraído no primeiro questionário, elaborar uma descrição de processo que representasse o cartão em questão. Os dados demográficos dos participantes são exibidos na tabela 5.3 e as descrições de processo elaboradas são exibidas a seguir. É possível perceber a diferença na quantidade de informações entre os dois textos, que pode ser uma consequência da diferença de informações presentes nos cartões elaborados no primeiro questionário.

### Texto 3 : Descrição de processo elaborada a partir do cartão Trello do Grupo 1

- 1 *Process: Check Claim*
- 2 *Rules involved: Claims Officer (CO) and Senior Claims Officer (SO)*
- 3 *Process description:*
- 4 *The process starts when a claim is registered.*
- 5 *The CO should "exam the claim,"and after that, "write a settlement recommenda-*
- 6 *tion."*
- 7 *The "settlement recommendation" goes to the SO;*
- 8 *The SO should "check the settlement recommendation,"and after that, "mark the*
- 9 *claim".*
- 10 *If the claim is marked as "Not OK", the SO should "send the settlement recommen-*
- 11 *dation back"to the CO.*
- 12 *If the claim is marked as "OK", the process "check claim" is finished.*

ambiguidades da linguagem natural e dificuldades da sua análise pela ferramenta de PLN, *Stanford CoreNLP*.

### Texto 5 : Texto extraído pelo protótipo a partir do cartão Trello do Grupo 1

4 *Claims officer and Senior Claims Officer check claim. In order to check claim, Claims*  
 5 *officer and Senior Claims Officer must write settlement recommendation. Other ac-*  
 6 *tivities include who write settlement recommendation, who mark claim and claim*  
 7 *handle process proceeds.*

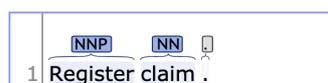
Por exemplo, o nome do cartão elaborado pelo Grupo 2, “*Register claim*”, quando analisado pela ferramenta *Stanford CoreNLP*, é identificado como um substantivo composto (5.11), ou seja, “*Register*” é classificado gramaticalmente como um substantivo, e não um verbo, o que descarta a utilização do nome do cartão como uma atividade (axioma 2 e definição 1), visto que a sentença não possui verbo. Contribuindo com esta hipótese, quando investigado o porquê de apenas o item *write a settlement recommendation* ter sido identificado pela aplicação da definição 2, descobriu-se que os itens *check recommendation* e *mark claim* também foram interpretados como substantivos compostos.

Outro exemplo é a adição de “*who*” como sujeito das atividades identificadas na descrição do cartão Trello do Grupo 1 (texto 5). Quando da investigação utilizando a *Stanford CoreNLP*, identificamos que “*who*” realmente é interpretado como o sujeito das atividades pela relação “*nsubj*”, enquanto os reais atores (*Claims Officer* e *Senior Claims Officer*) são interpretados com a relação *dep* com as palavras “*who*”, relação esta que é utilizada em casos em que a ferramenta não é capaz de encontrar uma relação mais precisa entre as palavras (MARNEFFE; MANNING, 2008).

Entretanto, estas investigações também não descartam a possibilidade de os axiomas e definições elaborados estarem incorretos ou não serem suficientes para a extração correta das informações de processo de negócio.

Figura 5.11 – Análise da sentença “*Register claim*” pela *Stanford CoreNLP*

#### Part-of-Speech:



#### Basic Dependencies:



Fonte: Os autores.

Devido à ausência do texto gerado pelo protótipo a partir do cartão Trello do Grupo 2, a continuação do processo de avaliação deste não pode ser realizado, uma vez que não se teria modelo de processo de negócio para comparação.

Já para o Grupo 1, utilizando a descrição de processo elaborada pelo segundo questionário e o texto com informações de processo de negócio gerado pelo protótipo, foram elaborados dois questionários, um para cada texto, solicitando que os respondentes elaborassem modelos de processo de negócio considerando o texto exibido. Os dados demográficos dos respondentes destes questionários são exibidos na tabela 5.5. O modelo elaborado a partir da descrição de processo é exibido na fig. 5.12 e o elaborado a partir do texto gerado pelo protótipo é exibido na fig. 5.13.

Tendo-se ambos os modelos de processo, podemos realizar a comparação entre eles através do cálculo de Similaridade Estrutural de Dijkman et al. (2011). Primeiro os modelos foram transformados em BPGs. Para facilitar a compreensão, cada nó de cada BPG foi nomeado por um par alfa-numérico, onde pares  $A_i$  ( $A_1, A_2, \dots, A_6$ ) se referem aos nós do BPG elaborado a partir da descrição de processo e pares  $B_i$  ( $B_1, B_2, \dots, B_6$ ) se referem aos nós do BPG elaborado a partir do texto gerado pelo protótipo. A figura 5.14 exibe os BPGs gerados. O grafo de cima é aquele elaborado a partir do modelo da descrição de processo (fig. 5.12) e o de baixo é o elaborado a partir do modelo do texto gerado pelo protótipo (fig. 5.13). Todos os nós que representam elementos com rótulos têm o seu rótulo exibido junto a si.

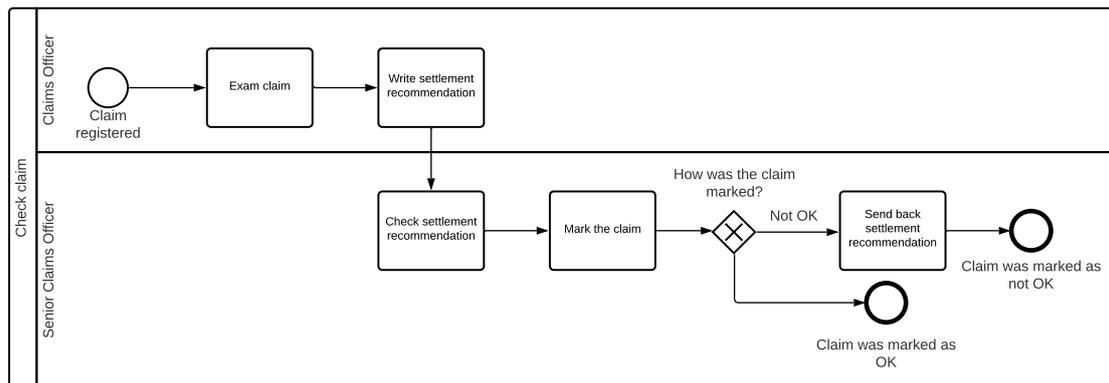
Com os BPGs elaborados, calculamos a similaridade sintática de cada par de nós dos dois grafos (DIJKMAN et al., 2011). O valor calculado para cada está exposto na tabela exposta pela fig. 5.15. Os valores são coloridos de acordo com o nível de similaridade, com tons vermelhos significando maior diferença e tons verdes significando maior similaridade.

Analisando a tabela levando-se em consideração o limiar mínimo de similaridade considerado para o mapeamento de nós (0, 5), percebe-se que os pares com similaridade suficiente são  $(A_3, B_3)$ ,  $(A_3, B_4)$ ,  $(A_4, B_3)$ ,  $(A_4, B_4)$ ,  $(A_8, B_3)$  e  $(A_8, B_4)$ . Dado que o mapeamento de nós deve ser um mapeamento injetivo parcial (DIJKMAN et al., 2011), não se pode ter um nó  $B_i$  mapeado por dois ou mais outros nós  $A_i$ . Por isso, mapeamos os nós de acordo com os maiores valores de similaridade, neste caso os pares  $(A_3, B_3)$  e  $(A_4, B_4)$ . Na figura 5.14 estes mapeamentos são representados pelas flechas tracejadas entre os nós mapeados.

Com o mapeamento de nós realizado, podemos definir quais nós e arestas devem ser removidas/inseridas e, com isso, calcularmos a Similaridade Estrutural entre os dois grafos. Neste caso, para transformar um grafo em outro, são necessárias 11 remoções/inserções de nós ( $A1, A2, A5, A6, A7, A8, A9, B1, B2, B5, B6$ ) e 14 arestas (todas). Para os nós mapeados, a sua contribuição para a distância de edição dos dois gráficos é dada por  $2(1 - 0,78 + 1 - 0,65)$ , cujo valor é 1,14. Também são necessários o número total de nós dos dois grafos (15) e o número total de arestas (14). Assim, têm-se a similaridade entre ambos os grafos pelo cálculo:

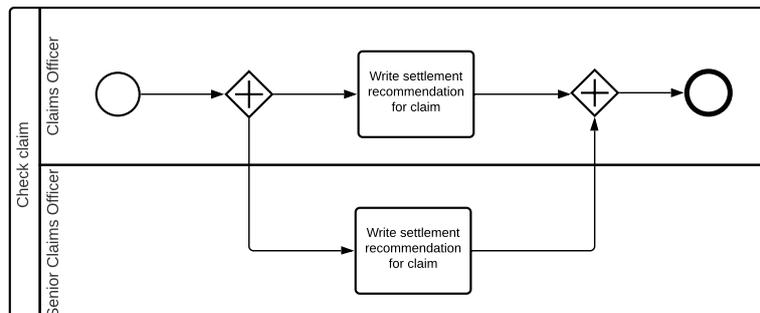
$$Sim(A, B) = 1 - media\left(\frac{13}{15}, \frac{14}{14}, \frac{1,14}{15 - 13}\right) = 0,19$$

Figura 5.12 – Modelo de processo elaborado a partir da descrição de processo



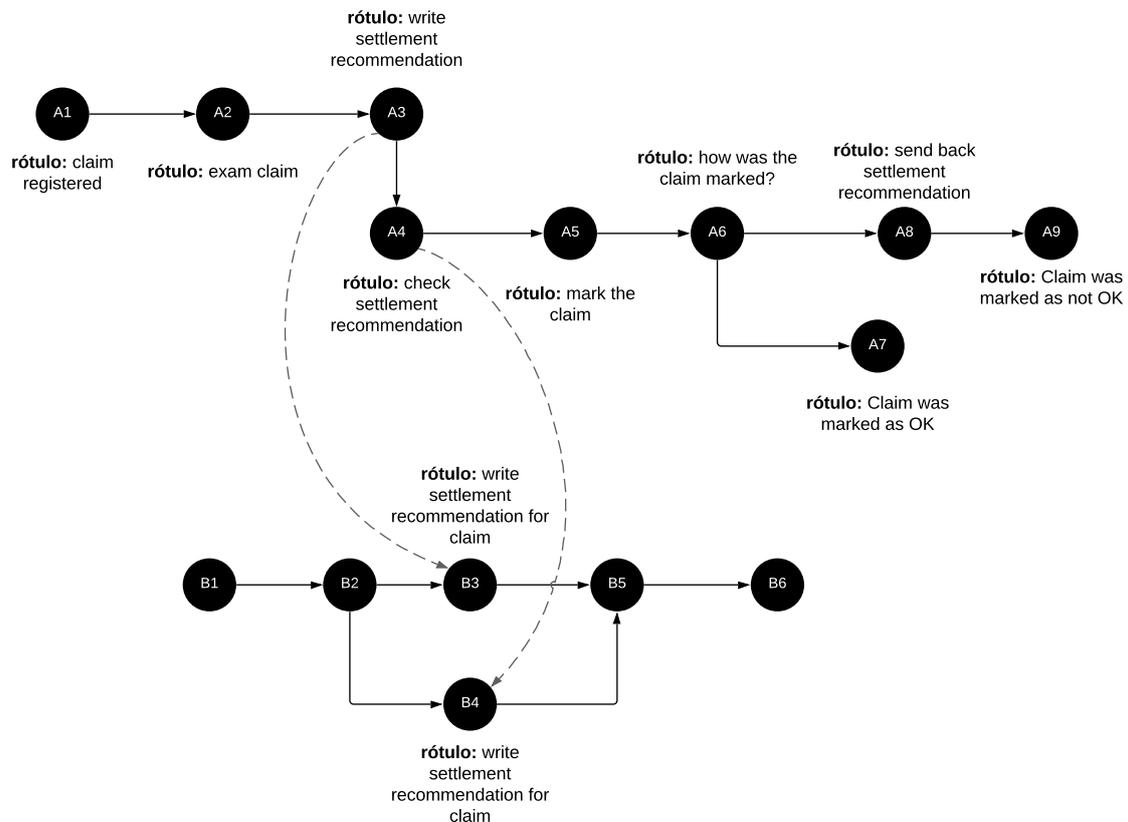
Fonte: Os autores.

Figura 5.13 – Modelo de processo elaborado a partir do texto gerado pelo protótipo



Fonte: Os autores.

Figura 5.14 – Grafos de processo de negócio elaborados a partir dos modelos nas figuras 5.12 e 5.13



Fonte: Os autores.

### 5.3 Discussão

Ao avaliarmos os resultados obtidos pelas definições e protótipo na avaliação realizada, está claro que ainda há muito trabalho a ser feito para se ter uma abordagem definitiva para a extração de informações de processos de negócio a partir da Trello, principalmente na implementação prática das definições.

Os primeiros questionários, de transformação da descrição de processo em cartão Trello e depois o processo inverso, principalmente as respostas do Grupo 1, se mostraram promissoras, com os cartões e a descrição de processo evidenciando a eficácia de alguns axiomas definidos neste estudo. Entretanto, é evidente que uma avaliação mais extensa do que um único estudo de caso precisa ser realizada, visto que axiomas como o 6, 7, 8, 11, 12 e 13, relativas ao uso de *Markdown*, não puderam ser avaliados.

Ainda nos primeiros questionários, a falta de conhecimento e experiência do respondente do Grupo 2 com a plataforma Trello no segundo questionário se mostrou um possível fator determinante na falha dos questionários e validações dos estágios seguintes.

Figura 5.15 – Tabela contendo os valores de similaridade sintática entre os nós dos grafos da figura 5.14

	B1	B2	B3	B4	B5	B6
A1	0,00	0,00	0,38	0,38	0,00	0,00
A2	0,00	0,00	0,32	0,32	0,00	0,00
A3	0,00	0,00	0,78	0,78	0,00	0,00
A4	0,00	0,00	0,65	0,65	0,00	0,00
A5	0,00	0,00	0,22	0,22	0,00	0,00
A6	0,00	0,00	0,19	0,19	0,00	0,00
A7	0,00	0,00	0,16	0,16	0,00	0,00
A8	0,00	0,00	0,57	0,57	0,00	0,00
A9	0,00	0,00	0,19	0,19	0,00	0,00

Fonte: Os autores.

A falta de informações de processo de negócio no cartão Trello elaborado pelo respondente, somada às ambiguidades da linguagem natural que acabaram por tornar a análise da *Stanford CoreNLP* não tão eficaz quanto desejada, acarretou em um texto raso, ou seja, nenhuma informação de processo de negócio extraída, comprometendo a etapa de modelagem e comparação de modelos.

A análise dos textos pelo protótipo, utilizando a *Stanford CoreNLP*, também se mostrou um fator determinante no processo de avaliação, principalmente no segundo e terceiro questionário. A interpretação de sentenças no formato “verbo + objeto” como substantivos compostos (ex.: “*Register claim*”, “*check recomendations*” e “*mark claim*”) prejudicou a extração de informações de processo de negócio e a geração do texto contendo estas informações, deixando de adicionar atividades que poderiam acarretar em elementos de processo nos modelos elaborados no terceiro questionário, o que poderia aumentar consideravelmente a similaridade entre os modelos. O mesmo pode ser dito das relações do tipo “*det*” interpretadas pela *Stanford CoreNLP* entre as palavras “*who*” e os atores do processo na descrição do cartão. Tivessem sido consideradas estas relações no

protótipo ou tivessem as relações de sujeito, “*nsubj*”, sido classificadas diretamente aos atores, ainda mais elementos de processo poderiam ter sido adicionados aos modelos.

De uma maneira geral, apesar das ressalvas encontrados durante o estudo de caso e da baixa similaridade entre os modelos gerados pelo Grupo 1 (apenas 0, 19), os axiomas e definições elaborados no estudo parecem promissores, principalmente pelas evidências de uso dos axiomas pelos analistas de processo, mesmo com sua pouca experiência de uso na plataforma Trello e sem qualquer estudo prévio das possíveis relações dos dados do cartão com conceitos de BPM e PLN.

Como principal ponto negativo, o protótipo se mostrou parcialmente eficaz na aplicação das definições, principalmente por não lidar com todas as ambiguidades e possibilidades de textos em linguagem natural e, conseqüentemente, das relações que a *Stanford CoreNLP* pode interpretar em sua análise. Uma implementação mais abrangente, estendendo a aplicação das definições e considerando uma maior variabilidade de casos especiais de texto, tem potencial de aumentar significativamente a eficácia da extração de informações de processo de negócio.

Finalmente, um único estudo de caso, apesar de útil para se evidenciar a eficácia dos axiomas, definições e do protótipo, não é suficiente para se ter uma validação definitiva. Considerando uma possível continuação e extensão do presente trabalho, se faz necessária uma validação mais extensa, com maior diversidade de descrições de processo, capazes de fornecer dados para validação de todos os axiomas e definições, com um maior número de participantes e, idealmente, com uso de quadros e cartões Trello utilizados em organizações reais.

Tabela 5.3 – Dados demográficos e de experiências individuais dos respondentes do segundo questionário

<i>Questão</i>	<i>Grupo 1</i>	<i>Grupo 2</i>
Qual é a sua idade?	25-34 anos	25-34 anos
Você está conectado ao setor público, acadêmico ou privado?	Outro: Acadêmico e Privado	Acadêmico
Qual é a sua principal ocupação?	Outro: Analista de sistemas e estudante de mestrado	Estudante de graduação
Quantos anos de experiência você possui em modelagem de processos de negócio?	Entre 2 e 4 anos	Entre 2 e 4 anos
Quando foi a última vez que você modelou um processo?	No último mês	Na última semana
Como você considera o seu nível de experiência em modelagem de processos de negócio?	3 (médio)	4 (alto)
Quantos anos de experiência você possui com a plataforma Trello?	Menos de 2 anos	Menos de 2 anos
Quando foi a última vez que você utilizou a plataforma Trello?	No último mês	Nos últimos 6 meses ou antes
Como você considera o seu nível de experiência em modelagem de processos de negócio	2 (baixo)	2 (baixo)

Fonte: Os Autores

Tabela 5.5 – Dados demográficos e de experiências individuais dos respondentes do terceiro questionário do Grupo 1

<i>Questão</i>	<i>Descrição de processo</i>	<i>Texto do protótipo</i>
Qual é a sua idade?	25-34 anos	35-44 anos
Você está conectado ao setor público, acadêmico ou privado?	Acadêmico	Privado
Qual é a sua principal ocupação?	Estudante de doutorado	Analista de processo
Quantos anos de experiência você possui em modelagem de processos de negócio?	Mais de 4 anos	Mais de 4 anos
Quando foi a última vez que você modelou um processo?	Esta semana	Esta semana
Como você considera o seu nível de experiência em modelagem de processos de negócio?	5 (muito alto)	54 (muito alto)

Fonte: Os Autores

## 6 CONCLUSÃO

O presente trabalho, levando em consideração o contexto da fase de descoberta de processos de negócio, que é frequentemente caracterizada como custosa e demorada, buscou somar-se a diversos outros esforços e estudos que têm por foco a automatização desta fase, em especial àqueles que utilizam PLN para extrair informações de processo de negócio de textos em linguagem natural.

Tendo em vista o crescente uso de plataformas de gestão de projetos em organizações, este trabalho teve como objeto de estudo a plataforma de gestão de projetos Trello e definiu como hipótese que esta pode servir como fonte de informações de processo de negócio para fase de descoberta de processos.

Para avaliar esta hipótese, as funcionalidades e dados presentes na plataforma, em especial em cartões, foram analisados e relacionados com conceitos de BPM e PLN, com foco em identificar possíveis elementos de processo presentes em seu conteúdo (seção 3.1). Como resultado dessa análise, foi elaborado um conjunto de 16 axiomas que estabelecem estas relações e fornecem a base para 6 definições de extração (seção 3.2) que buscam, através de transformações sobre os dados da plataforma Trello, extrair informações de processo de negócio em formato de texto.

Para pôr em prática os axiomas e definições, um protótipo em software foi desenvolvido (capítulo 4), possibilitando que o usuário navegue por seus quadros e cartões Trello, selecione um cartão e aplique as definições de extração, tendo por resultado um texto em linguagem natural contendo as informações de processo extraídas.

Por fim, com foco na avaliação dos resultados obtidos pelas definições e pelo protótipo, um questionário foi aplicado com a participação de analistas de processo (capítulo 5). O objetivo do questionário foi possibilitar a comparação entre um modelo de processo gerado a partir de uma descrição de processo elaborada com base em um cartão Trello e outro modelo, gerado a partir do texto gerado pelo protótipo.

### 6.1 Contribuições

A principal contribuição do presente trabalho está nas evidências de que plataformas digitais, através da análise de seus dados utilizando técnicas PLN, podem se tornar úteis para a extração de informações de processos de negócio e a automatização da fase de descoberta de processos. Tais evidências começaram a ser coletadas pelo trabalho de

Silva (2019), em sua análise da plataforma Slack, que indicou como possíveis trabalhos futuros a análise de outras plataformas, dentre elas a Trello.

Apesar da avaliação não extensa e da disparidade entre os modelos comparados, a identificação do uso dos axiomas pelos analistas de processo, quando da elaboração do cartão Trello, evidencia que a abordagem pode ser promissora e obter melhores resultados, principalmente com a realização de ajustes na sua implementação.

Ainda mais, com a existência de inúmeras plataformas à Trello e que também utilizam estruturas inspiradas na metodologia KanBan, como Jira<sup>1</sup>, Clubhouse<sup>2</sup>, GitHub Projects<sup>3</sup> e ZenHub<sup>4</sup>, trabalhos futuros podem analisar estas ferramentas e utilizar como base os axiomas e definições elaborados no presente estudo, que podem ser reutilizados ou modificados de acordo com o contexto da ferramenta em questão. Para isto, estas ferramentas devem conter cartões com, pelo menos, os mesmos campos utilizados neste estudo: título/nome do cartão, descrição, usuários atribuídos e comentários.

## 6.2 Trabalhos futuros

O presente trabalho, como mostrado pela avaliação de seus resultados (capítulo 5), possui aspectos que podem ser melhorados, tal como sua metodologia de avaliação. Para se validar a eficácia da abordagem e protótipo desenvolvidos, se faz necessário um processo de validação mais extenso, contendo um maior número de participantes e uma maior diversidade de dados para estudo, principalmente dados reais, retirados de quadros e cartões Trello utilizados em organizações.

Futuras validações podem, ainda, utilizar ferramentas de extração de elementos de processo de negócio desenvolvidas por outros estudos, como as de Rosa (2020) e Silva (2019), para validar a utilidade das informações extraídas pelas definições quando utilizadas nestas ferramentas.

Do ponto de vista dos axiomas e definições, vale ressaltar que apenas as funcionalidades presentes em cartões Trello foram consideradas e, ainda assim, nem todas foram inclusas. Portanto, é natural que trabalhos futuros considerando a plataforma Trello incluam outros dados e funcionalidades. Um exemplo seria a análise de nomes de listas e datas de entrega de cartões.

---

<sup>1</sup><<https://www.atlassian.com/br/software/jira>>

<sup>2</sup><<https://clubhouse.io/>>

<sup>3</sup><<https://github.com/features/project-management/>>

<sup>4</sup><<https://www.zenhub.com/>>

Por fim, existem inúmeras outras plataformas digitais passíveis de análise por outros trabalhos, também com foco na extração de informações de processo de negócio com técnicas de PLN. Tais plataformas não se limitam às de comunicação e gestão de projetos, como Slack, Trello, mas também plataformas de gestão de recursos humanos (Namely<sup>5</sup>, Gupy<sup>6</sup>), controle de vendas (Pipedrive<sup>7</sup>, Salesforce<sup>8</sup>) e outras categorias.

---

<sup>5</sup><<https://www.namely.com/>>

<sup>6</sup><<https://www.gupy.io/>>

<sup>7</sup><<https://www.pipedrive.com/>>

<sup>8</sup><<https://www.salesforce.com/>>

## REFERÊNCIAS

- AA, H. van der et al. Challenges and opportunities of applying natural language processing in business process management. In: . [S.l.: s.n.], 2018. p. 2791–2801.
- AALST, W. et al. Process mining manifesto. In: . [S.l.: s.n.], 2011. v. 99, p. 169–194.
- AALST, W. V. D. Process mining. **Communications of the ACM**, ACM New York, NY, USA, v. 55, n. 8, p. 76–83, 2012.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2009. ISBN 0596516495.
- Bächle, M.; Kirchberg, P. Ruby on rails. **IEEE Software**, v. 24, n. 6, p. 105–108, 2007.
- CHINOSI, M.; TROMBETTA, A. Bpmn: An introduction to the standard. **Computer Standards & Interfaces**, v. 34, n. 1, p. 124 – 134, 2012. ISSN 0920-5489. Available from Internet: <<http://www.sciencedirect.com/science/article/pii/S0920548911000766>>.
- DALE, R. Nlp in a post-truth world. **Natural Language Engineering**, Cambridge University Press, v. 23, n. 2, p. 319–324, 2017.
- DIJKMAN, R. et al. Similarity of business process models: Metrics and evaluation. **Information Systems**, v. 36, n. 2, p. 498–516, 2011. ISSN 0306-4379. Special Issue: Semantic Integration of Data, Multimedia, and Services. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S0306437910001006>>. Accessed in: 04/2021.
- DUMAS, M. et al. **Fundamentals of Business Process Management**. Berlin: Springer-Verlag GmbH Germany, part of Springer Nature, 2018.
- FERREIRA, R. C. B.; THOM, L. H.; FANTINATO, M. A semi-automatic approach to identify business process elements in natural language texts. In: . [S.l.: s.n.], 2017.
- FRANZ, A.; MILCH, B. Searching the web by voice. In: **Proceedings of the 19th International Conference on Computational Linguistics (COLING)**. [s.n.], 2002. p. 1213–1217. Available from Internet: <<https://research.google.com/archive/webbyvoice.html>>.
- FRIEDRICH, F.; MENDLING, J.; PUHLMANN, F. Process model generation from natural language text. In: . [S.l.: s.n.], 2011. p. 482–496. ISBN 978-3-319-98176-5.
- GRUBER, J. 2004. Available from Internet: <<https://daringfireball.net/projects/markdown/>>. Accessed in: 03/2021.
- INDURKHYA, N.; DAMERAU, F. J. **Handbook of natural language processing**. [S.l.]: CRC Press, 2010.
- JOHNSON, H. A. Trello. **Journal of the Medical Library Association : JMLA**, p. 209–211, 2017.
- KAUR, A. App review: Trello. **Journal of Hospital Librarianship**, 2018.

- LEOPOLD, H. **Natural language in business process models**. Thesis (PhD) — Springer, 2013.
- LIBERATORE, M.; POLLACK-JOHNSON, B. Factors influencing the usage and selection of project management software. **IEEE Transactions on Engineering Management**, v. 50, 05 2003.
- LYONS, J. **Natural Language and Universal Grammar: Essays in Linguistic Theory**. [S.l.]: Cambridge University Press, 1991.
- MANNING, C. D. et al. The stanford corenlp natural language processing toolkit. In: **ACL (System Demonstrations)**. The Association for Computer Linguistics, 2014. p. 55–60. ISBN 978-1-941643-00-6. Available from Internet: <<http://dblp.uni-trier.de/db/conf/acl/acl2014-d.html#ManningSBFBM14>>.
- MARNEFFE, M.-C. de; MANNING, C. D. Stanford typed dependencies manual. sep. 2008. Available from Internet: <[http://nlp.stanford.edu/software/dependencies\\_manual.pdf](http://nlp.stanford.edu/software/dependencies_manual.pdf)>. Accessed in: 03/2021.
- MENDLING, J.; REIJERS, H.; AALST, W. Seven process modeling guidelines (7pmg). **Information and Software Technology**, v. 52, p. 127–136, 02 2010.
- Merril Lynch & Co. Enterprise information portals. 1998. Available from Internet: <[https://web.archive.org/web/20110724175845/http://ikt.hia.no/perep/eip\\_ind.pdf](https://web.archive.org/web/20110724175845/http://ikt.hia.no/perep/eip_ind.pdf)>. Accessed in: 02/2021.
- MITKOV, R. **Anaphora resolution: the state of the art**. [S.l.]: Citeseer, 1999.
- NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, v. 18, n. 5, p. 544–551, 09 2011. ISSN 1067-5027. Available from Internet: <<https://doi.org/10.1136/amiajnl-2011-000464>>.
- OLUWATOSIN, H. S. Client-server model. **IOSRJ Comput. Eng**, Citeseer, v. 16, n. 1, p. 2278–8727, 2014.
- OMG. Business Process Model and Notation (BPMN), Version 2.0. January 2011. Available from Internet: <<http://www.omg.org/spec/BPMN/2.0>>.
- POLANČIČ, G. et al. An empirical investigation of the intuitiveness of process landscape designs. In: NURCAN, S. et al. (Ed.). **Enterprise, Business-Process and Information Systems Modeling**. Cham: Springer International Publishing, 2020. p. 209–223. ISBN 978-3-030-49418-6.
- QURESHI, M. R. J.; SABIR, F. A comparison of model view controller and model view presenter. 2014.
- ROSA, L. S. A visual approach for identification and annotation of business process elements in process descriptions. 2020. Available from Internet: <<http://hdl.handle.net/10183/218643>>. Accessed in: 12/2020.
- ROSEMANN, M.; BROCKE, J. V. **Handbook on Business Process Management 1**. [S.l.]: Springer-Verlag Berlin Heidelberg, 2010.

ROSING, M. v. et al. Business process model and notation (bpmn). In: \_\_\_\_\_. [S.l.: s.n.], 2015. v. 1, p. 429–453. ISBN 978-0-12-799959-3.

SILVA, D. B. Análise da plataforma de comunicação slack para descoberta de informações de processos de negócio. 2019. Available from Internet: <<http://hdl.handle.net/10183/198634>>. Accessed in: 03/2021.

SILVA, J. B. d.; ANASTÁCIO, F. A. d. M. Método kanban como ferramenta de controle de gestão. 2019. ISSN 1981-1179.

SILVA, T. S. An architecture for enabling business process-oriented text generation. 2019. Available from Internet: <<http://hdl.handle.net/10183/194331>>.

SILVA, T. S. et al. Empirical analysis of sentence templates and ambiguity issues for business process descriptions. In: PANETTO, H. et al. (Ed.). **On the Move to Meaningful Internet Systems. OTM 2018 Conferences**. Cham: Springer International Publishing, 2018. p. 279–297. ISBN 978-3-030-02610-3.

SOARES, M. M. Análise comparativa de ferramentas utilizadas para kanban. 2017. Available from Internet: <<https://lume.ufrgs.br/handle/10183/168978>>.

The Project Management Institute. **Um Guia do Conhecimento em Gerenciamento de Projetos (Guia PMBOK)**. [S.l.: s.n.], 2017. ISBN 9781628253900.

TRELLO. **How to format your text in Trello**. 2020. Available from Internet: <<https://help.trello.com/article/821-using-markdown-in-trello>>. Accessed in: 03/2021.

TRELLO. **What is Trello?** 2020. Available from Internet: <<https://trello.com/en-US/about>>. Accessed in: 03/2021.

TRELLO. **Adding checklists to cards**. 2021. Available from Internet: <<https://help.trello.com/article/737-adding-checklists-to-cards>>. Accessed in: 03/2021.

TRELLO. **Commenting on a card**. 2021. Available from Internet: <<https://help.trello.com/article/765-commenting-on-cards>>. Accessed in: 03/2021.

TRELLO. **Create a Board**. 2021. Available from Internet: <<https://trello.com/guide/create-a-board>>. Accessed in: 03/2021.

TRELLO. **Feature deep dive**. 2021. Available from Internet: <<https://trello.com/guide/feature-deep-dive>>. Accessed in: 03/2021.

TRELLO. **Trello 101**. 2021. Available from Internet: <<https://trello.com/guide/trello-101>>. Accessed in: 03/2021.

VAJJALA, S. et al. **Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems**. O'Reilly Media, Incorporated, 2020. ISBN 9781492054054. Available from Internet: <<https://books.google.com.br/books?id=G40jyWEACAAJ>>.

VERNER, L. Bpm: The promise and the challenge. 2004. Available from Internet: <<https://dl.acm.org/doi/10.1145/984458.984503>>.

WESKE, M. **Business Process Management: Concepts, Languages, Architectures.** Berlin, Heidelberg: Springer-Verlag, 2007. ISBN 3540735216.

WHITE, D.; FORTUNE, J. Current practice in project management — an empirical study. **International Journal of Project Management**, v. 20, n. 1, p. 1 – 11, 2002. ISSN 0263-7863. Available from Internet: <<http://www.sciencedirect.com/science/article/pii/S0263786300000296>>.