

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

DIOGO BORGES DA SILVA

**Análise da plataforma de comunicação
slack para descoberta de informações de
processos de negócio**

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

Orientador: Profa. Dra. Lucinéia Heloisa Thom

Porto Alegre
2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

Pró-Reitor de Graduação: Prof. Wladimir Pinheiro do Nascimento

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Engenharia de Computação: Prof. André Inácio Reis

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

*“Conhecimento não é aquilo que você sabe,
mas o que você faz com aquilo que você sabe.”*

— ALDOUS HUXLEY

AGRADECIMENTOS

Inicio meus agradecimentos pelos meus pais, Cilon e Odete, minhas irmãs, Natáche e Nicole, e minha sobrinha Maria, por todo amor, apoio, incentivo e paciência durante todos esses anos de graduação. Que mesmo distantes fisicamente, estiveram perto todos os dias, por meio de ligações e mensagens, tornando suportáveis todos os momentos difíceis que passei. Agradeço meus amigos e ex colegas de Santa Catarina, por todos os bons momentos que passamos e também por todo esforço em conjunto para superar as dificuldades. Foram anos de muito crescimento e também de alegrias junto deles. Meu muito obrigado também aos amigos do Sport Club PampaCats, em especial aos meninos do Pampahand, que fizeram meu último ano bem movimentado, porém cheio de momentos alegres e descontraídos. Aprendi e aprendo muito com eles. Aos demais amigos e familiares, agradeço as conversas, conselhos e todos os bons momentos que tivemos. A vida é mais leve com eles. Por fim, agradeço aos professores do instituto de informática, por todo o conhecimento e desenvolvimento pessoal e profissional que me proporcionaram. Em especial, quero agradecer a Profa. Dra. Lucinéia Heloisa Thom, por ter aceitado ser minha orientadora e me proporcionar tamanho auxílio para o desenvolvimento deste trabalho, mesmo distante. Meu muito obrigado por todas as revisões, críticas sempre construtivas, paciência e amizade.

RESUMO

Ter os processos bem mapeados e disponíveis para acesso é muito relevante para as organizações, uma vez que com essas informações é possível diminuir custos operacionais e melhorar o uso de recursos, por exemplo. Essas informações sobre processos podem estar nos mais variados locais e, com o desenvolvimento de novas ferramentas sociais, supõe-se que estejam presentes também neste tipo de rede de comunicação. No caso de ferramentas sociais, esse tipo de informação precisa ser extraída da plataforma social e, por estar em linguagem natural, precisa ser transformada, para melhor se adequar a um texto de descrição de processos, ou seja, um texto produzido com o intuito de descrever quais processos são realizados e sobretudo a maneira como são realizados, mostrando suas etapas de forma sequencial. Este trabalho apresenta uma abordagem de extração de informações referentes á processos, advindas da plataforma *Slack*, cuja linguagem é próxima de conversas informais. Após a extração, deve-se também pré-processar as informações, com o intuito de gerar-se um documento mais próximo de uma descrição de processos, em língua Inglesa, de forma sequencial. Para tanto, neste trabalho utiliza-se métodos de processamento de linguagem natural, como segmentação e marcação de texto, para analisar os diálogos obtidos da plataforma *Slack* e reestruturá-los para o padrão *sujeito + verbo + objeto*. Com isso, obtém-se um arquivo de saída composto de textos estruturados, similares a descrição de um processo, para ser utilizado em ferramentas de análise e modelagem de processos, por exemplo.

Palavras-chave: Extração de informação. Processo de Negócio. Plataformas sociais. Slack. BPM.

ABSTRACT

Analysis of the slack communication platform for business process information discovery

Having the processes well mapped and available for access is very relevant for companies, once this information makes it possible to decrease operating costs and improve the use of resources, for example. This information can be in the most varied places and, with the development of new social tools, it is assumed that they are also present in this type of communication network. In the case of social tools, this type of information needs to be extracted from the social platform and, because it is in natural language, needs to be transformed, to better fit a process description text, that is, a text produced with the intention of describe what processes are performed and, especially, the way they are performed, showing their steps sequentially. This work presents an approach to extracting process information from the Slack platform, whose language is closer to informal conversations. After the extraction, it is also necessary to pre-process the information, in order to generate a document closer to a description of processes, in English language, sequentially. In order to do this, in this work natural language processing methods, like segmentation and text tagging, are used to analyze the dialogs obtained from the Slack platform and to restructure them to the subject + verb + object pattern. Thus, we obtain an output file composed of structured texts, similar to the description of a process, for use in process analysis and modeling tools, for example.

Keywords: Information Extraction, Business Process, Social Plataforms, Slack , BPM.

LISTA DE ABREVIATURAS E SIGLAS

BPM	Business Process Management
BPMN	Business Process Management Notation
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
MVC	Model-View-Controller
NLP	Natural Language Processing
POS	Part Of Speech
SOA	Service-Oriented Architecture

LISTA DE FIGURAS

Figura 1.1	Estrutura geral do trabalho	14
Figura 2.1	Exemplo do processo de divulgação de conteúdo representado em BPMN	16
Figura 2.2	O ciclo de vida do BPM de acordo com Dumas et al. (2018)	18
Figura 2.3	Abordagem de Silva and Thom (2019)	20
Figura 3.1	Arquitetura MVC para uma aplicação Web.....	24
Figura 4.1	Exibição dos canais na plataforma Slack	26
Figura 4.2	Criação do token de acesso à plataforma Slack.....	27
Figura 4.3	Token gerado automaticamente	27
Figura 4.4	Retorno do método de listagem dos canais	29
Figura 4.5	Arquivo com o histórico estruturado do canal.....	30
Figura 4.6	Método de marcação para Stanford Parser	36
Figura 5.1	Conversas simuladas na plataforma Slack.....	39
Figura 5.2	Tela de Cadastro	40
Figura 5.3	Tela Inicial	41
Figura 5.4	Tela de Escolha do Canal de Comunicação.....	41
Figura 5.5	Tela de extração e processamento das informações.....	42
Figura 5.6	Arquivo final com as sentenças processadas	43
Figura A.1	Preenchimento dos campos obrigatórios	52
Figura A.2	Registro realizado	52
Figura A.3	Tela Inicial com os registros criados	52
Figura A.4	Preenchimento dos campos obrigatórios para edição.....	53
Figura A.5	Registro Modificado com sucesso	53
Figura A.6	Tela Inicial com registros realizados.....	53
Figura A.7	Registro deletado com sucesso	54
Figura A.8	Tela Inicial sem o registro deletado	54
Figura A.9	Token não reconhecido	54
Figura A.10	Comunicação com a <i>API Slack</i> e obtenção de informação.....	55
Figura A.11	Processamento de canal sem mensagens	56
Figura A.12	teste em canal com mensagens	56
Figura A.13	Arquivo de saída	57
Figura A.14	Conteúdo do arquivo de saída.....	57

LISTA DE TABELAS

Tabela 2.1	Visão geral dos trabalhos relacionados a BPM e NLP	22
Tabela 3.1	Técnicas Utilizadas no presente trabalho.	24
Tabela 4.1	Métodos da API de extração.....	28
Tabela 4.2	Definições do primeiro nível de processamento.....	33
Tabela 4.3	Definições do segundo nível de processamento	34
Tabela 4.4	Exemplos de aplicação das definições.....	35
Tabela 4.5	Lista de marcação do Stanford POS Tagger	37
Tabela 5.1	Casos de teste da integração das unidades.....	44
Tabela 5.2	Teste caixa-branca de acordo com Ricca and Tonella (2001)	45

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Motivação.....	11
1.2 Objetivos e Metodologia de Pesquisa	13
1.3 Organização do Texto	14
2 FUNDAMENTOS E TRABALHOS RELACIONADOS	15
2.1 Gerenciamento de Processos de Negócio	15
2.2 Processamento de linguagem	17
2.3 Trabalhos Relacionados.....	19
2.4 Resumo do Capítulo.....	21
3 TECNOLOGIAS UTILIZADAS NA ABORDAGEM DESENVOLVIDA	23
3.1 Resumo do Capítulo.....	25
4 EXTRAÇÃO DE INFORMAÇÕES DE PROCESSO A PARTIR DA PLATAFORMA SLACK	26
4.1 Passo 1: Configurando um <i>Token</i> de Acesso.....	27
4.2 Passo 2: Módulo de Extração.....	28
4.3 Passo 3: Pré-processamento das Informações.....	31
4.4 Abordagem de Silva and Thom (2019).....	31
4.5 Módulo de Processamento.....	32
4.5.1 Primeiro Nível de Processamento	32
4.5.2 Segundo nível de processamento	33
4.6 Resumo do Capítulo.....	37
5 DEMONSTRAÇÃO E VERIFICAÇÃO DA APLICAÇÃO DESENVOLVIDA ...	39
5.1 Utilização da ferramenta desenvolvida	40
5.2 Teste e verificação da abordagem	42
5.3 Resumo do Capítulo.....	45
6 CONCLUSÕES	46
6.1 Contribuições e limitações da ferramenta desenvolvida	47
6.2 Trabalhos Futuros.....	48
REFERÊNCIAS	49
APÊNDICE A — TESTES DA FERRAMENTA DESENVOLVIDA	52

1 INTRODUÇÃO

De acordo com Dumas et al. (2018), processo de negócio é o que uma empresa faz sempre que entrega um produto ou presta um serviço. Mais ainda, um processo deve ser visto como parte essencial para a existência de uma organização, uma vez que para fornecer algum produto ou serviço, deve exercer atividades de forma coordenada e estruturada, envolvendo assim os processos.

Os processos em uma organização devem ser gerenciados e coordenados para que sejam atingidos os objetivos específicos dessa organização, tal como manufaturar um produto ou prestar um serviço a um cliente. Para isso, são utilizados conceitos de Gerenciamento de Processos de Negócio (*Business Process Management - BPM*).

O método tradicional de descoberta de processos em organizações é baseado em técnica de entrevista com especialistas de domínio, ou seja, alguém na organização que conhece os processos realizados em determinado departamento, por exemplo. De acordo com Aalst, van der (2011), esse método funciona bem para a identificação de processos-chave, ou seja, os principais processos executados na organização, porém pode não ser adequado quando o especialista de domínio não possui conhecimento ou experiência com os processos. Afirma ainda que processos *periféricos*, ou seja, aqueles processos que se afastam do raciocínio do especialista de domínio, são difíceis de serem contemplados com essa técnica. Além disso, há um grande volume de dados presentes em variadas fontes (ex.: formulários, *emails*, páginas *web*, documentos corporativos), principalmente em meios digitais, com informações relacionadas às atividades e como elas se relacionam dentro das organizações. Somado a isso, tem-se o desenvolvimento acelerado de tecnologias de processamento de dados (ex.: inteligência artificial, mineração de dados, processamento de linguagem natural), o que torna muito atrativa a utilização de métodos computacionais para a descoberta de processos.

1.1 Motivação

De acordo com Anderson (2016) e Leonard, Huysman and Steinfield (2013), no contexto empresarial, as tecnologias sociais estão sendo utilizadas para comunicação interna e externa entre funcionários. Os autores afirmam que plataformas como *Facebook*, *Twitter*, *Instagram* e *Slack*, são exemplos de tecnologias sociais, utilizadas nas organizações para este tipo de comunicação e que essas tecnologias podem ter relevantes implica-

ções para a comunicação das organizações no que tange o compartilhamento de conhecimento, gerenciamento e construção de capital social.

De acordo com Robertson and Kee (2017), a facilidade de acesso a informações, principalmente sobre os membros da organização e suas tarefas de trabalho pode melhorar a comunicação e a colaboração entre colegas de equipe. Especialmente para novos funcionários e/ou colaboradores que não estão em tempo integral na organização e por isso possuem conhecimento limitado sobre algumas atividades a eles atribuídas. Conforme Robertson and Kee (2017), o uso de *wiki's* para projetos de equipe ou grupos no *Facebook* permitem manter os colegas atualizados sobre o status de seus projetos. Além disso, essas informações são ainda mais interessantes quando são considerados os benefícios para novos funcionários ou para funcionários que não participam em tempo integral, desconhecendo partes das rotinas.

Conforme Aalst, van der (2011), a maioria dos dados armazenados no universo digital é desestruturada e as organizações têm problemas em lidar com grandes volumes de dados. O autor afirma ainda que um dos principais desafios das organizações atuais é extrair informações e valor dos dados armazenados em seus sistemas de informação. Além disso, os dados estão armazenados nas mais diversas fontes e estruturas textuais, dificultando a descoberta de informações acerca de elementos de processos, os quais se tem interesse no presente trabalho.

As plataformas sociais se mostram fontes interessantes a serem analisadas para a descoberta de informações de processos, principalmente de elementos de processos. Entretanto, nestas plataformas, são utilizadas formas informais de conversação e troca de mensagens. Esse tipo de linguagem pode representar uma dificuldade para a extração de informações, pois os textos que geralmente são analisados na literatura, quando se deseja descobrir processos (ver seção 2.3) são descrições de como estes processos acontecem, de forma sequencial, escritos por algum especialista do domínio dos processos. Assim, para transformar estas conversas e modo de comunicação em conhecimento, faz-se necessário extrair os dados da plataforma e pré-processá-los, a fim de obter-se um texto descritivo e sequencial.

Portanto, tecnologias sociais, por viabilizarem a troca de informações sobre o negócio, podem ser uma das fontes de informação de processos de negócio executados em organizações. Informações sobre elementos de processo, tais como atividades, subprocessos e seus participantes podem estar presentes no texto referente as trocas de mensagens.

1.2 Objetivos e Metodologia de Pesquisa

Uma ferramenta utilizada para manter o tipo de comunicação, a qual é estudada no presente trabalho, é a *Slack*. De acordo com Slack (2019), ela é uma plataforma lançada em 2014, cujo objetivo é facilitar a colaboração e comunicação no setor empresarial, ou seja, foi desenvolvida para ser utilizada por organizações para gerenciamento de times, projetos e trabalhos específicos. Ela possui espaços de trabalho, *workspaces*, que podem ser vistos como uma divisão de departamento ou de projetos mais amplos em uma organização, onde é possível a criação de diversos canais de comunicação, que podem representar processos separados. Estes canais possuem uma estrutura de comunicação muito semelhante a plataformas de troca de mensagens e redes sociais, porém a *Slack* foi desenvolvida como uma solução para que as organizações possam gerenciar seus times e atividades.

O objetivo principal deste trabalho é apresentar uma abordagem para extração de informações referentes à processos, advindas de plataformas heterogêneas, com foco em ferramentas cuja linguagem se aproxima mais de conversas informais, como é o caso da *Slack*.

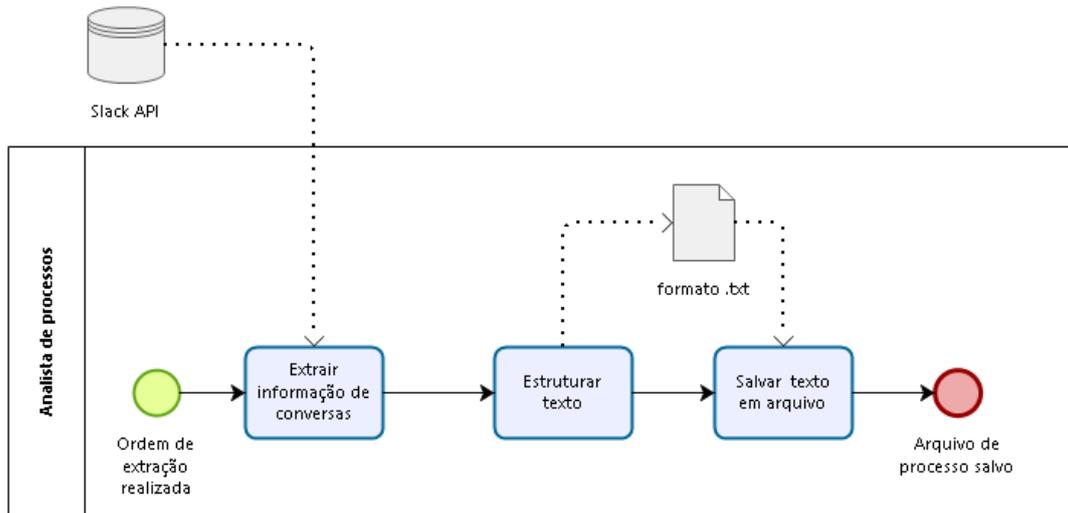
Para atingir este objetivo, a seguinte metodologia de pesquisa é usada: Primeiramente, pesquisou-se trabalhos na literatura que abordassem os temas de extração de dados, processamento de linguagem natural e BPM. Em seguida, fez-se uma análise da *API* fornecida pela *Slack* para descobrir os métodos necessários para interfacear uma comunicação com a plataforma e assim fazer a extração de dados. Além disso, percebeu-se a necessidade de um processamento das informações obtidas por meio da plataforma, para adequar as conversas desta, linguagem informal, para um formato mais próximo de uma descrição textual de processos, em língua inglesa, de forma corrida e sequencial.

Assim, este trabalho foi dividido em duas partes: Um módulos de extração de informação, onde é exposto como as informações são obtidas por meio da ferramenta *Slack* e um módulo de processamento e estruturação das informações, onde são abordados os formatos de texto que se deseja trabalhar, além de transformações necessárias na estrutura das conversas para que se obtenha tais formatos. Uma visão geral está representada na Figura 1.1.

Com este estudo será possível processar os dados provenientes desta fonte, por meio de Processamento de Linguagem Natural (*Natural Language Processing - NLP*), e usá-los, em trabalhos futuros, para a descoberta de processos em organizações e a cria-

ção de modelos de processos, em Notação e Modelo de Processo de Negócio (*Business Process Model and Notation - BPMN*).

Figura 1.1: Estrutura geral do trabalho



Fonte: Os Autores

1.3 Organização do Texto

O restante deste trabalho está organizado da seguinte forma: O Capítulo 2 apresenta os fundamentos necessários para o entendimento deste trabalho. Além disso, discute trabalhos relacionados a abordagem apresentada neste trabalho. O Capítulo 3 apresenta e justifica as tecnologias usadas no desenvolvimento da abordagem. O Capítulo 4 apresenta a ferramenta desenvolvida neste trabalho para a extração de informações de uma plataforma social conhecida como *Slack*, bem como processamentos e transformações que devem ser executados nos dados dela extraídos. O capítulo 5 apresenta a utilização da ferramenta desenvolvida e uma metodologia de teste e verificação desta. Por fim, o capítulo 6 traz as conclusões e perspectivas futuras.

2 FUNDAMENTOS E TRABALHOS RELACIONADOS

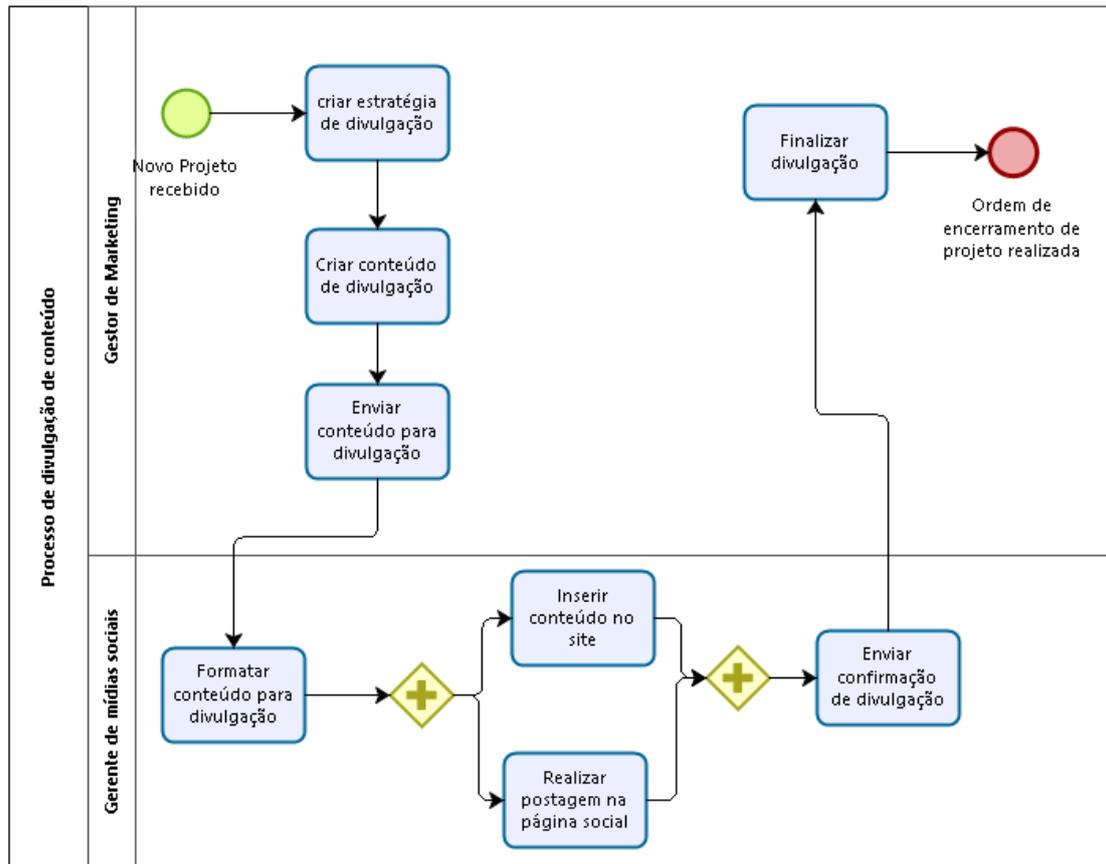
O presente capítulo apresenta os fundamentos em BPM, plataformas sociais, em particular a plataforma *Slack*, e Processamento de Linguagem Natural. Além disso, o capítulo discute trabalhos relacionados a abordagem apresentada neste trabalho.

2.1 Gerenciamento de Processos de Negócio

Conforme Dumas et al. (2018), um processo de negócio engloba eventos, atividades e tarefas. Um *evento* é um acontecimento, sem duração, de forma atômica, como a chegada de um *email*, por exemplo. Já *atividades* e *tarefas* são instâncias de trabalho, que levam certo tempo, como a inspeção de algum equipamento, por exemplo. A Figura 2.1 apresenta um exemplo de processo de negócio, representado através da BPMN. De acordo com The Object Management Group (2011), a BPMN é usada por quem projeta, gerencia e realiza processos de negócios, permitindo que os diagramas sejam traduzidos em componentes de processos de software, por meio de uma notação simples e que independe de qualquer ambiente de implementação específico. De acordo com Aagesen and Krogstie (2015) e Geiger, Neugebauer and Vorndran (2017), a BPMN é amplamente utilizado na indústria e suportado por grande parte das ferramentas de modelagem de processos do mercado. Segundo International Organization for Standardization (2013) a notação se tornou um padrão e foi adotada como uma ISO em 2013.

A Figura 2.1 apresenta um exemplo de processo, divulgação de conteúdo, representado segundo a BPMN. Este processo envolve dois participantes, representados por *Gestor de Marketing* e *Gerente de mídias sociais*, que desempenham as atividades necessárias para que o processo seja bem executado. O processo inicia com o recebimento de um novo projeto, denotado pelo primeiro círculo. Após, o gestor de marketing deve realizar, nesta sequência, as atividades de criar estratégia de divulgação, criar conteúdo de divulgação e enviar o conteúdo de divulgação para o gerente de mídias sociais. O gerente de mídias sociais deve formatar o conteúdo para divulgação e então deve, em paralelo, inserir o conteúdo no site e realizar uma postagem com o conteúdo também na página social. As duas últimas tarefas do processo devem ser feitas paralelamente, pois há um controle de fluxo, também chamado de *gateway*, que representa caminhos alternativos de um processo. Essa divisão de fluxo é denotada pelo losango com sinal de mais, indicando paralelização. Após isso, o gerente deve enviar uma confirmação de divulgação para o

Figura 2.1: Exemplo do processo de divulgação de conteúdo representado em BPMN



Fonte: Os Autores

gestor de marketing, que irá então finalizar a divulgação e encerrar o projeto. Esta informação é indicada pelo segundo círculo, com bordas mais espessas, que indica fim de processo.

Conforme Dumas Dumas et al. (2018), BPM é a arte e a ciência de supervisionar como o trabalho é realizado em uma organização para garantir resultados consistentes e aproveitar oportunidades de melhoria. Os autores classificam o ciclo de vida de BPM de acordo com 6 fases principais: Identificação de Processos, Descoberta de Processos, Análise de Processos, Redesenho de Processos, Implementação de Processos e Monitoramento de Processos. Este ciclo de vida é mostrado na Figura 2.2.

Na fase de *Identificação de Processos* é definido um problema do negócio e os processos relevantes para este problema são identificados, delimitados e relacionados entre si. O resultado dessa etapa é uma arquitetura de processo que fornece uma visão geral dos processos em uma organização e seus relacionamentos.

Na fase de *Descoberta de Processos*, o estado atual de cada um dos processos relevantes é documentado, geralmente na forma de um ou vários modelos de processos,

também conhecidos como modelo *As-is*.

Na fase de *Análise de Processos*, as questões associadas ao processo *As-is* são identificadas, documentadas e, sempre que possível, quantificadas usando medidas de desempenho. A saída dessa fase é uma coleção estruturada de problemas, que são tipicamente priorizados em termos de impacto e, às vezes, também em termos do esforço estimado necessário para resolvê-los.

Na fase de *Redesenho de Processos*, também denominada *melhoria de processo*, o objetivo é identificar mudanças no processo que ajudariam a resolver os problemas identificados na fase de análise de processos. Permite ainda que a organização cumpra seus objetivos de desempenho. A saída desta fase é tipicamente um modelo de processo futuro, também chamado *To-be*, que serve como base para a próxima fase, pois é a documentação em diagrama do processo mapeado que se quer implementar.

Na fase de *Implementação de Processos*, as mudanças necessárias para transformar o modelo *As-is* para o modelo melhorado do processo, *To-be*, são preparadas e executadas. A implementação do processo abrange dois aspectos: gerenciamento de mudanças organizacionais, atividades necessárias para mudar a maneira de trabalhar de todos os participantes envolvidos no processo, e automação de processos, desenvolvimento e implantação de sistemas de TI (ou versões aprimoradas de sistemas de TI existentes) que suportam o processo futuro.

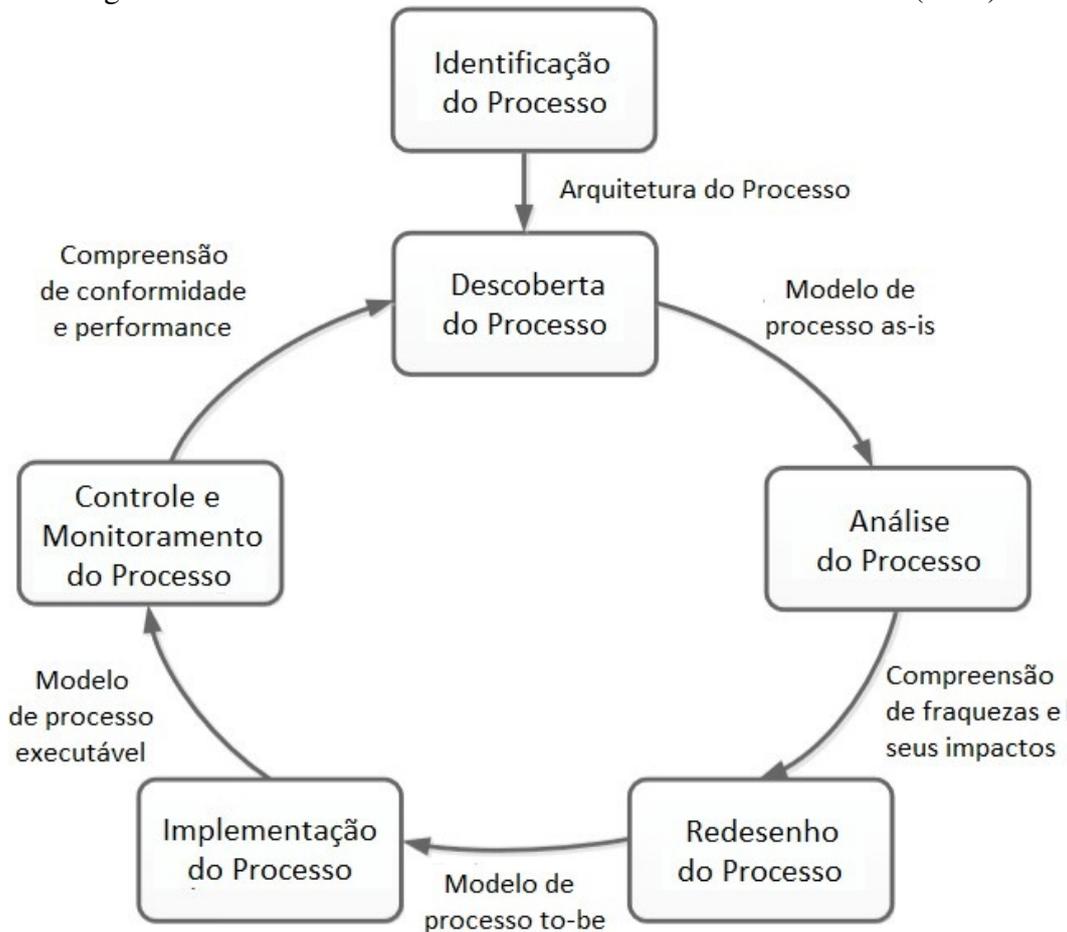
Na fase de *Monitoramento de Processos*, uma vez que o processo reprojeto está em execução, dados relevantes são coletados e analisados para determinar o desempenho do processo com relação a suas medidas de desempenho e objetivos de desempenho. São identificados gargalos, erros recorrentes ou desvios em relação ao comportamento pretendido e são realizadas ações corretivas. Novas questões podem surgir, no mesmo ou em outros processos, exigindo que o ciclo seja repetido de forma contínua.

No presente trabalho, é dada prioridade para as fases de Identificação e Descoberta de Processos, que compõem as duas primeiras etapas do ciclo de vida de BPM.

2.2 Processamento de linguagem

De acordo com Bird, Klein and Loper (2009), *linguagem natural* é a linguagem usada para comunicação cotidiana por humanos, como a comunicação no idioma inglês, ou português, por exemplo. Neste contexto, NLP se refere a todo tipo de manipulação computacional da linguagem natural. Segundo os autores, vai desde uma simples conta-

Figura 2.2: O ciclo de vida do BPM de acordo com Dumas et al. (2018)



Fonte: Dumas et al. (2018) - tradução nossa

gem de frequência de palavras em um texto até a compreensão de expressões humanas completas, com o intuito de obter informações relevantes para aplicações específicas.

Segundo Indurkha and Damerau (2010), segmentação de texto implica em converter um texto em suas palavras e sentenças componentes. As palavras são então limitadas, ou seja, os pontos em que uma palavra termina e outra começa são bem definidos. As palavras assim identificadas são frequentemente referidas como *tokens*, e a segmentação de palavras também é conhecida como tokenização. Por exemplo, a sentença *The marketing director must update the data base* é tokenizada da seguinte forma: ['The', 'marketing', 'director', 'must', 'update', 'the', 'data', 'base'].

Além da tokenização, utiliza-se uma técnica de marcação, também chamada de *taggeamento*, cujo objetivo é a obtenção das formas básicas de palavras (*part of speech - POS*), que trata de termos com funções semelhantes dentro de sentenças, como divisão de verbos, adjetivos, pronomes, ou ainda marcação das estruturas de dependências de palavras em uma sentença. A sentença anterior é marcada da seguinte forma: *The_DT marketing_NN director_NN must_MD update_VB the_DT data_NNS base_NN*. Desta forma,

sabe-se a qual classe gramatical cada palavra pertence dentro do contexto de determinada sentença, por meio dos elementos de marcação, como verbo, *_VB*, ou um substantivo, *_NN*, por exemplo. O uso desta técnica é necessária para fazer a reestruturação de sentenças, no capítulo 4.

Neste trabalho são utilizadas as ferramentas de processamento de linguagem do *Stanford Core Natural Language Processing*, em especial *Stanford Parser*, acessadas por meio de uma interface PHP desenvolvida em Schur (2019). A utilização dessas ferramentas é mostrada no Capítulo 4.

2.3 Trabalhos Relacionados

Diversos trabalhos com a proposta de modelagem de processos de negócios por meio de ferramentas que tratam texto em linguagem natural vêm sendo desenvolvidos.

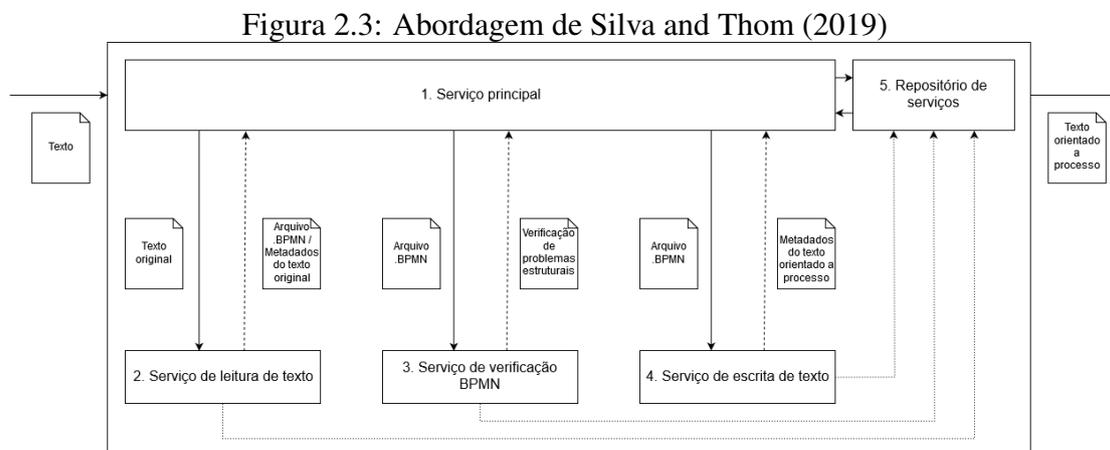
Friedrich (2010) e Friedrich, Mendling and Puhmann (2011) trazem essa discussão e apresentam um modelo para transformar textos em linguagem natural em modelos de processos. A abordagem proposta pelos autores tem 77% de sucesso em relação ao modelo gerado, quando comparado com o modelo esperado.

Leopold (2013) também traz discussões relacionadas com a proposta apresentada no presente trabalho. O autor afirma que na etapa de identificação de processos, pode-se utilizar NLP para analisar documentos em linguagem natural objetivando descobrir atividades executadas com frequência, decisões críticas e inter-relações entre processos. Supondo-se que existam informações de processos de uma organização nas conversas entre colegas em redes sociais, supõe-se também que seja possível, portanto, a utilização de técnicas de NLP para descoberta e análise de processos também para esse tipo de fonte de informação. Além disso, afirma também que atividades ou até mesmo modelos inteiros de processos podem ser sintetizados automaticamente com uso de tais técnicas, quando aplicadas em documentos de entrevista com especialista de domínio ou em relatórios e e-mails, por exemplo.

Outros trabalhos mais atuais também trazem a discussão do uso de NLP, como é o caso de Aa et al. (2018). Segundo os autores, NLP, juntamente com a evolução do *big data* e inteligência artificial, atingiu maturidade suficiente para ser utilizada em aplicações diversas, inclusive aquelas relacionadas à processos de negócio das organizações. Em seu trabalho são discutidas as vantagens e limitações da utilização de técnicas de NLP em aplicações relacionadas a gerenciamento de processos. Embora, sejam apresentadas algu-

mas dificuldades, tais como tratamento de ambiguidade e/ou necessidade de adaptação a domínios específicos, é reforçada a ideia de que com NLP é possível extrair informações sobre processos, com intuito de modelar esses processos.

Com base nos trabalhos de Friedrich (2010), Friedrich, Mendling and Puhmann (2011) e Silva et al. (2018) foi possível o desenvolvimento de uma ferramenta orientada a serviços, em pesquisa conduzida por Silva and Thom (2019), cuja estrutura está representada na Figura 2.3. Neste contexto, foram criados módulos que oferecem serviço de *text writer*, um serviço recebe como entrada um arquivo .BPMN e produz como saída um arquivo JSON contendo um texto com metadados. *bpmn verification*, serviço que recebe como entrada um arquivo .BPMN e produz como saída um arquivo JSON, contendo mensagens de verificação. Por último, oferece também um módulo de *text reader*. Este módulo recebe como entrada um arquivo texto “.txt”, descrevendo o processo e fornece como saída um arquivo .BPMN. Este último é de fundamental importância, pois o que apresenta-se no presente trabalho é uma extensão do módulo *text reader*, uma vez que este módulo ainda não é capaz de lidar com informações advindas de plataformas de comunicação, tal como a *slack*.



Fonte: Silva and Thom (2019)

Como mostrado na Figura 2.3, a ferramenta proposta por Silva and Thom (2019) recebe como entrada um arquivo com texto em linguagem natural, cujo conteúdo é a descrição de um processo. Porém, esta ferramenta não extrai informações de outras plataformas. Assim, para que se torne possível o processamento de informações de outras fontes, é necessário o desenvolvimento de módulos complementares, tais como os desenvolvidos neste trabalho. A tabela 2.1 mostra uma visão geral dos trabalhos relacionados a BPM e NLP utilizados como fundamentação do presente trabalho.

2.4 Resumo do Capítulo

Neste capítulo foram apresentados os conceitos fundamentais para o entendimento do presente trabalho e também trabalhos relacionados a abordagem apresentada. Definiu-se BPM e as fases de seu ciclo de vida, demonstrando-se o interesse nas duas primeiras fases, Identificação e Descoberta de Processos.

Conceituou-se também linguagem natural e citou-se as ferramentas de processamento de linguagem do *Stanford Core Natural Language Processing*, em especial *Stanford Parser*.

Finalmente, foram discutidos os trabalhos de Friedrich (2010) e Friedrich, Mending and Puhmann (2011), onde os autores lidam com transformação de textos em linguagem natural para modelos de processo de negócio. Citou-se também os trabalhos de Leopold (2013), que trata de NLP para análise de documentos em linguagem natural para descoberta de elementos de processos, o trabalho de Aa et al. (2018), que discute as vantagens e limitações do uso de NLP em BPM e também os trabalhos desenvolvidos por Silva et al. (2018) e Silva and Thom (2019), de fundamental relevância para o presente trabalho, uma vez que esse configura-se uma ferramenta complementar à arquitetura dos autores.

Tabela 2.1: Visão geral dos trabalhos relacionados a BPM e NLP

<i>Ano</i>	<i>Autor</i>	<i>Título</i>	<i>Abordagem</i>
2010	Friedrich (2010)	Automated generation of business process models from natural language input.	Processamento de linguagem natural para geração de modelo de processos de negócio.
2011	Friedrich, Mendling and Puhlmann (2011)	Process Model Generation from Natural Language Text	Transformação de textos de linguagem natural para modelos de processos.
2013	Leopold (2013)	Natural language in business process models: Theoretical Foundations, Techniques, and Applications.	Fundamentos e técnicas de NLP para análise de documentos em linguagem natural e descoberta de atividades executadas, decisões críticas e inter-relações entre processos.
2018	Aa et al. (2018)	Challenges and Opportunities of Applying Natural Language Processing in Business Process Management.	Discussão das vantagens e limitações de técnica de NLP em aplicações relacionadas a gerenciamento de processos. Extração de informações sobre processos, com intuito de modelar estes processos
2018	Silva et al. (2018)	Empirical Analysis of Sentence Templates and Ambiguity Issues for Business Process Descriptions.	Análise empírica de descrições de processos de negócios, a fim de encontrar modelos de frases recorrentes e filtrá-los para problemas de ambigüidade. Geração de descrições de processos mais adequadas para analistas de processo e especialistas em domínio.
2019	Silva and Thom (2019)	An architecture for enabling business process-oriented text generation	Desenvolvimento de uma arquitetura baseada em Service Oriented Architecture (SOA) na qual as etapas necessárias para gerar o texto orientado a processos de negócio são endereçadas a serviços individuais. Esta arquitetura recebe um arquivo com texto em linguagem natural, cujo conteúdo é a descrição de um processo e retorna um arquivo com “texto orientado à processo”.

3 TECNOLOGIAS UTILIZADAS NA ABORDAGEM DESENVOLVIDA

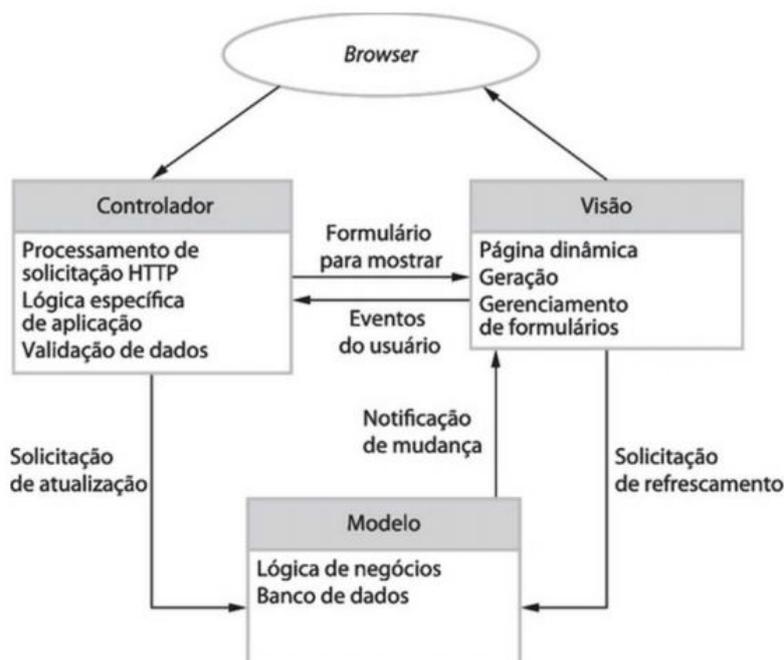
Para este trabalho, utilizou-se a linguagem de programação PHP. Um acrônimo recursivo para *PHP: Hypertext Preprocessor*) que segundo The PHP Group (2019) é uma linguagem de *script open source* de uso geral, muito utilizada e especialmente adequada para o desenvolvimento *web*. A linguagem PHP pode ser embutida dentro do HTML, que é a linguagem de marcação utilizada para a construção de páginas *web* e interpretada pelos navegadores. Integrado ao *PHP*, utilizou-se também *MySql* para configuração de banco de dados.

Como padrão de arquitetura, utilizou-se *Model-View-Controller (MVC)*. Conforme Sommerville et al. (2011), o objetivo do padrão MVC é separar a apresentação e a interação dos dados, estruturando o sistema em três componentes lógicos que interagem entre si (ver Figura 3.1). São eles: *Modelo*, cuja função é gerenciar o sistema de dados e as operações associadas a esses dados. *Visão*, usada para definir e gerenciar como os dados são apresentados ao usuário e; *Controlador*, cuja função é gerenciar a interação do usuário e passar essas interações para a Visão e o Modelo.

Ainda de acordo com Sommerville et al. (2011), a vantagem desse modelo é que ele permite que os dados sejam alterados de forma independente de sua representação, e vice-versa, ou seja, possibilita a reescrita da interface gráfica de usuário (*Graphical User Interface - GUI*) ou do *Controller* sem alterar o modelo. Além disso, facilita a manutenção e adição de recursos e o reaproveitamento de código. Este padrão de arquitetura é adequado para este trabalho devido a simplicidade para a divisão dos módulos do projeto, visando futuramente a adição de códigos de extração de dados de outras plataformas, além da *Slack*.

De acordo com Manning et al. (2014), o *Stanford CoreNLP* é um conjunto de softwares que provêm métodos de análise de linguagem natural para utilização em aplicações com esta devida finalidade. A ferramenta dá suporte a alguns idiomas como alemão, francês e espanhol. Entretanto, é utilizado o inglês no presente trabalho. Assim, a partir de um texto de entrada neste idioma, pode-se obter a POS de cada sentença. Por exemplo, na frase *The manager approves the report*, tem-se que as ocorrências da palavra *the* é um determinativo, ou seja, acompanha um substantivo fazendo referencia a ele. Já as palavras *manager* e *report* são substantivos, enquanto *approves* é um verbo que se relaciona com os substantivos anteriores. Sendo *manager* o sujeito que executa a ação e *report* representa o objeto direto que sofre a ação. Em especial, utiliza-se o *Stanford Parser*,

Figura 3.1: Arquitetura MVC para uma aplicação Web



Fonte: Sommerville et al. (2011)

uma das ferramentas que compõem o *Stanford Core NLP*, que possibilita trabalhar com as estruturas gramaticais de sentenças, bem como marcação de elementos gramaticais e suas dependências.

O *API stanford-nlp-tagger* é uma API PHP para *Stanford Natural Language Processing library*, desenvolvida por Schur (2019). Esta API faz interface com os módulos de *POSTagger*, usado para marcação de POS e estruturas gramaticais (ex.: verbos, adjetivos, substantivos) e *CRFClassifier*, utilizado para fazer reconhecimento e classificação de entidades (ex.: lugar, organização, pessoa, entre outros). Além disso, carrega automaticamente os pacotes necessários para o uso integrado do *Stanford Parser* e detecta o idioma do texto. É utilizada no presente trabalho para acessar as ferramentas do *Stanford Parser*.

A tabela 3.1 apresenta um resumo das tecnologias utilizadas neste trabalho.

Tabela 3.1: Tecnologias Utilizadas no presente trabalho.

<i>Objetivo</i>	<i>tecnologia</i>
Linguagem de programação	PHP
Banco de dados	MySQL
Padrão de arquitetura	model - view - controller (MVC)
Processamento de linguagem natural	Stanford CoreNLP
API para ferramentas de processamento de linguagem	API stanford-nlp-tagger

Fonte: Os Autores

3.1 Resumo do Capítulo

Este capítulo discutiu as tecnologias que foram utilizadas para o desenvolvimento deste trabalho. Utilizou-se *PHP* como linguagem de programação, pois é uma linguagem muito bem difundida para o desenvolvimento de aplicações *web* e *MySQL* para banco de dados. Como padrão de arquitetura, utilizou-se MVC. Já para a manipulação de textos em linguagem natural, fez-se uso das ferramentas do *Stanford CoreNLP*, em especial o *Stanford Parser*, cujas funcionalidades são acessadas neste trabalho por meio de uma interface, chamada de *API stanford-nlp-tagger*, desenvolvida por Schur (2019).

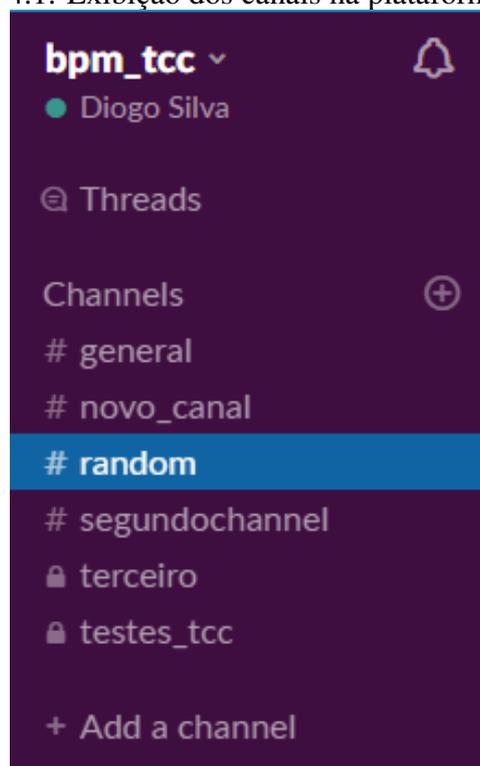
4 EXTRAÇÃO DE INFORMAÇÕES DE PROCESSO A PARTIR DA PLATAFORMA *SLACK*

Extrair informações da plataforma *Slack*, tal como conteúdo das conversas, remetente, canais de troca de mensagens, entre outras, é uma das principais tarefas destes trabalhos. Entretanto, é necessário entender a plataforma, e principalmente a API de comunicação fornecida, uma vez que cada ferramenta possui uma maneira própria de tratar os dados.

Na ferramenta, as conversas são separadas em canais, que se assemelham a grupos de comunicação *online*. São a partir destes canais que se deseja aplicar a extração de informação, com propósitos de descobrir elementos de processo, caso existam nas conversas. A Figura 4.1 mostra como estão dispostos os canais na *Slack*, bem como sua privacidade, ou seja, os canais *terceiro* e *testes_tcc* são privados, uma vez que o símbolo mais a esquerda, em forma de cadeado, indica esta informação. Já os demais canais criados são públicos, uma vez que o símbolo # denota esta informação.

As próximas seções, mostram os principais passos seguidos neste trabalho para a obtenção das informações que se deseja analisar e processar.

Figura 4.1: Exibição dos canais na plataforma Slack



Fonte: Slack (2019)

4.1 Passo 1: Configurando um *Token* de Acesso

Um *token* é como uma senha de acesso, utilizado para integrar aplicações customizáveis com os métodos da *API* da plataforma *Slack*. Ele permite identificar informações pessoais do usuário, como nome e equipe, solicitar dados sobre canais, mensagens, membros da equipe de trabalho, entre outras funcionalidades disponibilizadas pela plataforma.

Assim, o primeiro passo para ter acesso às informações dos canais é a criação do *token* de acesso, que será utilizado pela aplicação desenvolvida neste trabalho para garantir a identificação e acesso seguro na plataforma. Para isso, deve-se acessar <<https://api.slack.com/custom-integrations/legacy-tokens>>, conforme Figura 4.2 e clicar sobre *create token*. Logo em seguida, será criado, de forma automática, o token requisitado, que será mostrado na tela para cópia, conforme Figura 4.3.

Figura 4.2: Criação do token de acesso à plataforma Slack

Use this tool to generate legacy tokens.

⚠ Legacy tokens are just for you. Never share legacy tokens with other users or applications. Do not publish Legacy tokens in public code repositories. [Review token safety tips.](#)

By creating a test API token, you agree to the [Slack API Terms of Service](#).

Workspace	User	Token	
ufrgs	diogo_bsilva	none	Create token

Fonte: Os Autores

Figura 4.3: Token gerado automaticamente

Workspace	User	Token	
ufrgs	diogo_bsilva	xoyp-519965414144-52070286	Re-issue token
bpm_tcc	dibsilva	xoyp-468116830608-46868296	Re-issue token

Fonte: Os Autores

4.2 Passo 2: Módulo de Extração

De acordo com API... (2019), a *API Slack* possui quatro métodos importantes para a extração das informações das conversas realizadas na plataforma, como o texto e os membros do *chat*, por exemplo:

O *groups.list*, através do qual é possível listar todos os canais privados de comunicação de determinado usuário, com o intuito de obter acesso ao *id* que deve ser passado como parâmetro para o próximo método, ou *channel.list* para canais públicos.

O *conversations.history*, com o qual é possível obter o arquivo *JSON* com as conversas de determinado canal privado, escolhido previamente, ou *channels.history* para canais públicos.

Estes métodos estão apresentados na tabela 4.1.

Tabela 4.1: Métodos da API de extração

<i>Método</i>	<i>Tipo de canal</i>	<i>Parâmetros</i>	<i>Função</i>
groups.list	privado	token	listar canais privados e obter id interno dos canais
channel.list	público	token	listar canais públicos e obter id interno dos canais
conversations.history	privado	token e id do canal	obter conversas de canais privados
channels.history	público	token e id do canal	obter conversas de canais públicos

Fonte: Os Autores

Conforme API... (2019), as informações das conversas realizadas na plataforma podem ser acessadas por meio de uma requisição de dados com protocolo *HTTP*. Como retorno dessa requisição, recebe-se um arquivo *JSON*, estruturado conforme as funcionalidades que a plataforma disponibiliza.

Utiliza-se o método *groups.list* da *API Slack*, <(https://slack.com/api/groups.list?token=x)>, em conjunto com um *token* de validação, o qual deve ser criado para cada usuário no site da ferramenta para obter-se informações dos canais privados, como mostrado na Figura 4.4. Com esse método, é possível capturar os “ids” dos canais, a fim de

escolher de qual canal será extraído o histórico de conversas e eventos.

De posse do id do canal que se deseja, utiliza-se o método `conversations.history` <<https://slack.com/api/conversations.history?token=x&channel=GDSL57AG4>>, fornecendo o token de validação de usuário (`token=x`) e o id do canal (`channel=GDSL57AG4`, neste caso), resultando na obtenção do arquivo que contém o histórico do canal escolhido, de forma estruturada, como mostra a Figura 4.5.

Figura 4.4: Retorno do método de listagem dos canais

```
1 {
2   "ok": true,
3   "groups": [
4     {
5       "id": "GDSL57AG4",
6       "name": "testes_tcc",
7       "is_group": true,
8       "created": 1541004842,
9       "creator": "UDSL2TVFW",
10      "is_archived": false,
11      "name_normalized": "testes_tcc",
12      "is_mpim": false,
13      "members": [
14        "UDSL2TVFW"
15      ],
16      "topic": {
17        "value": "",
18        "creator": "",
19        "last_set": 0
20      },
21      "purpose": {
22        "value": "tcc",
23        "creator": "UDSL2TVFW",
24        "last_set": 1541004843
25      },
26      "priority": 0
27    }
28  ]
29 }
```

Fonte: Os Autores

A Figura 4.4 exibe o retorno do canal `testes_tcc`, `name` na linha 6. Neste docu-

mento, as informações mais importantes para a extração são: *id* na linha 5, pois com este parâmetro é possível requisitar as informações do canal desejado, *creator* na linha 9 e *members* na linha 13, pois com estes parâmetros é possível distinguir os participantes do *chat* e possivelmente dos processos que venham a existir referentes a este canal.

Figura 4.5: Arquivo com o histórico estruturado do canal

```

1 {
2   "ok": true,
3   "messages": [
4     {
5       "type": "message",
6       "user": "UDSL2TVFW",
7       "text": "another one",
8       "client_msg_id": "080e2c78-f4aa-4e0c-9f9d-d7e8ec6c4824",
9       "ts": "1541011012.001400"
10    },
11    {
12      "type": "message",
13      "user": "UDSL2TVFW",
14      "text": "trying another message, second test",
15      "client_msg_id": "269d45cb-2ddd-41d3-b4a7-f0abd46e6b98",
16      "ts": "1541004918.001200"
17    },
18    {
19      "type": "message",
20      "user": "UDSL2TVFW",
21      "text": "first test message",
22      "client_msg_id": "93419245-aa2b-45fc-8e75-f6c8ea7e4866",
23      "ts": "1541004881.000700"
24    },
25    {
26      "user": "UDSL2TVFW",
27      "purpose": "tcc",
28      "text": "<@UDSL2TVFW> set the channel's purpose: tcc",
29      "type": "message",
30      "subtype": "group_purpose",
31      "ts": "1541004843.000300"
32    },
33    {
34      "user": "UDSL2TVFW",
35      "text": "<@UDSL2TVFW> has joined the group",
36      "type": "message",
37      "subtype": "group_join",
38      "ts": "1541004842.000200"
39    }
40  ],}

```

A Figura 4.5 exibe todas as mensagens enviadas no canal: três mensagens neste caso, linhas 5, 12 e 19. Além disso, outras informações importantes são mostradas, como o usuário que enviou determinada mensagem, linhas 6, 13 e 20, o texto da mensagem, linhas 7, 14 e 21, e informações de eventos que ocorreram, como nas linhas 34 a 38.

4.3 Passo 3: Pré-processamento das Informações

Conforme Bordignon et al. (2018), ferramentas de NLP, são capazes de diminuir o tempo e o custo na elicitação de modelos de processo e facilitar a extração de informações de fontes textuais, como documentos das organizações ou ainda, redes sociais. Isso é evidenciado também em Aa et al. (2018), onde os autores sugerem que há uma grande quantidade de fontes textuais em linguagem natural nas organizações, o que torna interessante o uso de técnicas de NLP para obter-se informações sobre os processos dessas organizações. Assim, após a extração das conversas da plataforma *Slack*, é necessário fazer uma filtragem, ou processamento, para a geração de um texto descritivo mais próximo do que se espera segundo a abordagem de (SILVA; THOM, 2019).

O restante deste capítulo discute, portanto, a criação de definições de processamento das informações empregadas neste trabalho, fazendo a relação entre o que está sendo desenvolvido e a arquitetura criada por Silva and Thom (2019).

4.4 Abordagem de Silva and Thom (2019)

De acordo com Silva and Thom (2019) um texto orientado a processos de negócio é definido como um texto estruturado, capaz de manter o máximo de informações relacionadas ao processo de negócio e capaz de verificar a qualidade do processo em relação ao BPMN 2.0 e em relação à *soundness*. *Soundness*, neste contexto diz respeito à processos bem formulados e que não possuam atividades que nunca serão executadas, por exemplo.

Em Silva and Thom (2019), é desenvolvida uma arquitetura baseada em *Service Oriented Architecture (SOA)* na qual as etapas necessárias para gerar o texto orientado a processos de negócio são endereçadas a serviços individuais. Esta arquitetura recebe um arquivo com texto em linguagem natural, cujo conteúdo é a descrição de um processo e retorna, como saída um arquivo com texto orientado à processo.

A arquitetura não faz a extração de informações de plataformas sociais, tal como

a *Slack*. Além disso, o texto utilizado como entrada no serviço de leitura (*text reader*) possui algumas especificações importantes. Este deve ser uma descrição de um processo, em idioma inglês, com sentenças gramaticalmente corretas e sequenciais, divididas em parágrafos. Além disso, não deve conter sentenças de questionamento, ou seja, perguntas.

Em Silva and Thom (2019), os dados de entrada do modelo se referem a textos em linguagem natural que são armazenados em formato .txt, podendo ser obtidos de diferentes fontes, tais como: formulários, normas, dados de eventos de sistemas de informação, mensagens de e-mail, etc.

Uma vez que as informações extraídas de redes sociais não estão estruturadas como uma descrição de processos, pois trata-se de linguagem informal, é necessário que se faça o pré-processamento destas antes de gerar-se o arquivo final que será utilizado na arquitetura dos autores.

4.5 Módulo de Processamento

Com base na descrição de como deve ser o texto de entrada da ferramenta de Silva and Thom (2019) e também os conceitos de BPM, são propostas, no presente trabalho, definições (D) para formatar a informação adquirida anteriormente nos módulos de extração de informação da plataforma *Slack*. As definições foram divididas em dois níveis, como mostrado na Tabela 4.2 e Tabela 4.3. O primeiro nível diz respeito à inclusão ou exclusão de termos em uma instâncias de fala, isto é, uma instância de fala é uma mensagem enviada em algum canal da plataforma *Slack*. O segundo nível trata de processar os dados com técnicas de NLP, com o intuito de aproximar o texto de um padrão de entrada, em um arquivo .txt, conforme especificações do trabalho de Silva and Thom (2019).

4.5.1 Primeiro Nível de Processamento

No primeiro nível (N1) de processamento, é executada a exclusão de sentenças e pontuação. Para isso, foram desenvolvidas três definições:

Definição 1 - Nível 1 trata de excluir termos de uma instância de fala, como perguntas, uma vez que, para ser suportado pela ferramenta de Silva and Thom (2019), o texto descritivo final não deve conter perguntas. Além disso, exclui-se também instâncias que iniciem com o caractere “@” porque este é reservado da plataforma para indicar ações

como “ingresso na sala”, por exemplo, que indica que um novo membro da equipe foi adicionado no canal de comunicação.

Definição 2 - Nível 1 trata também de exclusão, porém é da instância inteira de fala, quando esta possui menos de 10 caracteres de escrita. Isso acontece porque supõe-se que instâncias muito pequenas não possuem informações de relevância para processos de negócio. Escolheu-se arbitrariamente o valor de 10 caracteres, o que pode ser adaptado posteriormente, caso necessário.

Definição 3 - Nível 1 trata de incluir pontuação nas sentenças, uma vez que faz-se necessário separá-las, mantendo assim a sequencialidade do texto descritivo final. A pontuação é necessária porque denota o fim de uma sentença, ou seja, delimita a quantidade de texto que é analisado em uma só vez pelas ferramentas de processamento de texto.

Um resumo das definições do primeiro nível é mostrado na Tabela 4.2.

Tabela 4.2: Definições do primeiro nível de processamento

<i>Definição</i>	<i>Descrição</i>
D1N1	Excluir falas que possuem os caracteres “?” e “@” (perguntas e informações reservadas da plataforma slack)
D2N1	Uma instância de fala com menos de 10 caracteres é excluída
D3N1	Se não estiver pontuada, inserir ponto final a cada diálogo extraído

Fonte: Os Autores

4.5.2 Segundo nível de processamento

No segundo nível (N2) de processamento é feita análise de sentenças, com foco na sintaxe, para que seja possível a realização de modificações estruturais voltadas ao domínio de processos de negócio. Foram então desenvolvidas cinco definições adicionais, conforme tabela 4.3.

Definição 1 - Nível 2 Os eventos, atividades e tarefas representam acontecimentos ou ações, ou seja, estão fortemente relacionados com componentes verbais em uma sentença. Assim, a *Definição 1* trata de excluir instâncias de fala que não possuem verbo,

Tabela 4.3: Definições do segundo nível de processamento

<i>Definição</i>	<i>Descrição</i>
D1N2	Uma instância de fala que não possui verbo é excluída
D2N2	Se sujeito for “I”, apenas trocar por “the”+ “role”. Se o sujeito for um nome, trocar pelo devido cargo
D3N2	Configurar a estrutura para sujeito+verbo+objeto
D4N2	<subject>+<verb>+ <object>+ <conjunction>+ <verb>+ <object> A client <subject>calls <verb>the help desk <object>and <conjunction >makes <verb>a request <object>. Em caso de conjunção: transformar a primeira parte até a conjunção em <subject>+<verb>+ <object>, manter a conjunção e modificar a pessoa do verbo+objeto
D5N2	Se tiver sintaxe errada, se possível, corrigir antes de inserir no texto

Fonte: Os Autores

uma vez que estas sentenças não contêm elementos relevantes para descoberta e análise de processos de negócio.

Definição 2 - Nível 2 De acordo com Dumas et al. (2018), Frequentemente, em um modelo de processo, não nos referimos explicitamente a um recurso de cada vez, como por exemplo um funcionário "John Smith", mas nos referimos a um grupo de recursos que são intercambiáveis no sentido de que qualquer membro do grupo pode executar um determinado conjunto de atividades. Esses grupos são chamados de classes de recursos e podem ser uma organização inteira, uma unidade organizacional ou um papel. Assim, para o ciclo de vida do BPM é mais importante a obtenção das informações acerca do papel ou cargo de uma pessoa ou grupo de pessoas. Por exemplo, não é interessante fazer a relação *Marcelo verifica a disponibilidade de produto* e sim *O balconista verifica a disponibilidade de produto*, pois é mais genérico, serve para qualquer pessoa que ocupar aquele cargo ou desempenhar aquele papel. Por isso, esta Definição trata de fazer a troca do pronome pessoal pelo cargo, ou ainda, caso seja um nome pessoal que pertence a alguém dentro do canal, faz-se a mudança, para garantir que o texto continue genérico. Essa informação pode ser obtida por meio do método *users.info*, fornecido pela API da plataforma, passando *token* e *id* de usuário como parâmetro. O método retorna informa-

ções de um determinado usuário, em um JSON, dentre as quais encontra-se o campo *title*, que se refere ao cargo ou papel que este usuário desempenha no grupo de trabalho e que foi previamente cadastrado nas informações de conta de usuário.

Definição 3 - Nível 2 Nos trabalhos de Friedrich (2010) e Silva and Thom (2019), um dos estilos de marcação, ou criação de rótulos para processos de negócio, é o *verb-object*, ou seja, um verbo na forma imperativa sucedido do objeto de negócio ao qual este verbo se refere. Assim, além desses dois componentes da sentença, deseja-se também saber o sujeito que realiza a ação caracterizada pelo verbo. Por isso, definiu-se a estrutura sujeito+verbo+objeto para manter as informações importantes do processo e ainda preservar a sequencialidade do texto descritivo.

Definição 4 - Nível 2 Deseja-se manter o formato da definição anterior e ainda manter-se as conjunções, como *and*, *but* e *or*, por exemplo, pois elas denotam outro elemento de processos, o *gateway*.

Definição 5 - Nível 2 Pretende garantir uma das especificações do texto de entrada de Silva and Thom (2019), que diz respeito à sintaxe das sentenças do texto descritivo, que deve ser correta. Por isso, é importante a tentativa de correção das palavras, caso necessário, utilizando algum dicionário do idioma inglês obtido na *web*, antes de sua inclusão no arquivo final.

Na tabela 4.4 são mostrados exemplos da aplicação das definições em sentenças criadas para simular o tipo de comunicação em grupo de trabalho de uma plataforma social.

Tabela 4.4: Exemplos de aplicação das definições

<i>Comunicação</i>	<i>Transformação</i>
First test message	Não é uma sentença válida, pois não possui um verbo
Support Officer: I updated the data base	The Support Officer <subject> must update <verb> the data base <object>
Manager: I approved the report	The manager <subject> must approve <verb> the report <object>
Manager: Lia, send a email to Faurgs and request the contract.	The Secretary <subject> must send <verb> a email to Faurgs<object> and <conjunction> must request <verb> the contract <object>

Klein and Manning (2003a), Klein and Manning (2003b) e Socher et al. (2013) discutem métodos de marcação e processamento de linguagem natural. Mais do que isso, seus trabalhos formam a base da ferramenta *Stanford Parser*. Por meio da *API stanford-nlp-tagger*, faz-se acesso aos métodos de processamento do *Stanford Parser*. O método utilizado pela *API* para marcação das estruturas gramaticais está descrito na Figura 4.6.

Figura 4.6: Método de marcação para Stanford Parser

```

1
2 public function stanford_postag($sentence)
3 {
4     $pos = new \StanfordTagger\POSTagger();
5
6     $pos->setModel('../stanford-postagger-full-
7     2017-06-09/models/english-bidirectional-distsim.tagger');
8
9     $pos->setJarArchive('../stanford-postagger-
10    full-2017-06-09/stanford-postagger.jar');
11
12    $pos->tag('Sentença');
13 }

```

Fonte: Os Autores

É criada uma instância do parser (linha 4), em seguida define-se o modelo de linguagem a ser utilizado (escolhendo-se inglês e passando o caminho do modelo nas linhas 6 e 7), informa-se o caminho onde estão instalados os pacotes do parser (linhas 9 e 10) e, por fim, utiliza-se o método de marcação “tag(‘sentença’)” na linha 12, com a sentença que se deseja marcar, passada como parâmetro. De acordo com Rusu et al. (2007), uma sentença é representada pelo analisador como uma árvore que tem três filhos: uma frase nominal, uma frase verbal e o ponto final. O sujeito da sentença é encontrado realizando uma pesquisa de largura e selecionando o primeiro descendente da frase nominal que é um substantivo (marcados como NN, NNP, NNPS ou NNS). Já o verbo é marcado diretamente pela ferramenta (VB, VBD, VBG, VBN, VBP ou VBZ), extraído da frase verbal. Por fim, o objeto é a parte do predicado da sentença, geralmente adjetivos (marcados como JJ, JJR ou JJS). Em Taylor, Marcus and Santorini (2003), são expostas as formas de marcação de sentenças conforme suas “partes de fala”, seus significados e como as palavras estão relacionadas gramaticalmente. Em especial, deseja-se destacar os substantivos, verbos e adjetivos, que estão diretamente relacionados com os elementos a serem processados para garantir-se o formato <sujeito>+<verbo>+<objeto> no arquivo final. As marcações e seus significados podem ser vistos na tabela 4.5.

Tabela 4.5: Lista de marcação do Stanford POS Tagger

<i>Marcação</i>	<i>Significado</i>	<i>Marcação</i>	<i>Significado</i>
CC	Coordinating conjunction	PP\$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential there	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	to
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP\$	Possessive wh-pronoun
PP	Personal pronoun	WRB	Wh-adverb

Fonte: Taylor, Marcus and Santorini (2003)

4.6 Resumo do Capítulo

Neste capítulo discutiu-se a extração de informações relacionadas à processos de negócio a partir da plataforma *Slack* e métodos para processar os textos obtidos, com o intuito de aproximá-los de um texto descritivo de processo. Abordou-se a necessidade

de criar-se um *token* para acessar a API da plataforma *Slack* e também os métodos disponíveis por essa API para extrair dados dos canais, *groups.list* e *channel.list*, e para a obtenção das conversas desses canais, *conversations.history* e *channels.history*.

Discutiu-se também a arquitetura desenvolvida por Silva and Thom (2019), bem como as especificações para o arquivo de entrada desta arquitetura. Com base nessas especificações, criou-se definições para pré-processar a informação obtida da *Slack*, que foram divididas em dois níveis: o primeiro, com três definições, tratando de exclusão de sentenças e pontuação destas, já o segundo nível, com cinco definições, executando análise das sentenças e reestruturação destas para o formato <sujeito> + <verbo> + <objeto>.

5 DEMONSTRAÇÃO E VERIFICAÇÃO DA APLICAÇÃO DESENVOLVIDA

Este capítulo apresenta uma demonstração da ferramenta desenvolvida no presente trabalho, bem como os passos para a obtenção das informações de um determinado canal dentro da plataforma *Slack*. Além disso, aborda-se também uma metodologia de teste e verificação da ferramenta, a fim de demonstrar que ela está de acordo com as funcionalidades esperadas.

A Figura 5.1 mostra um trecho de conversa criado para a simulação de comunicação dentro de um canal da plataforma *Slack*. Existem dois participantes nesse diálogo, cadastrados na plataforma como *Manager* e o outro como *Marketing Director*. Este diálogo será utilizado na demonstração de utilização da ferramenta e na geração do arquivo final contendo um texto mais próximo de uma descrição de processo.

Figura 5.1: Conversas simuladas na plataforma Slack

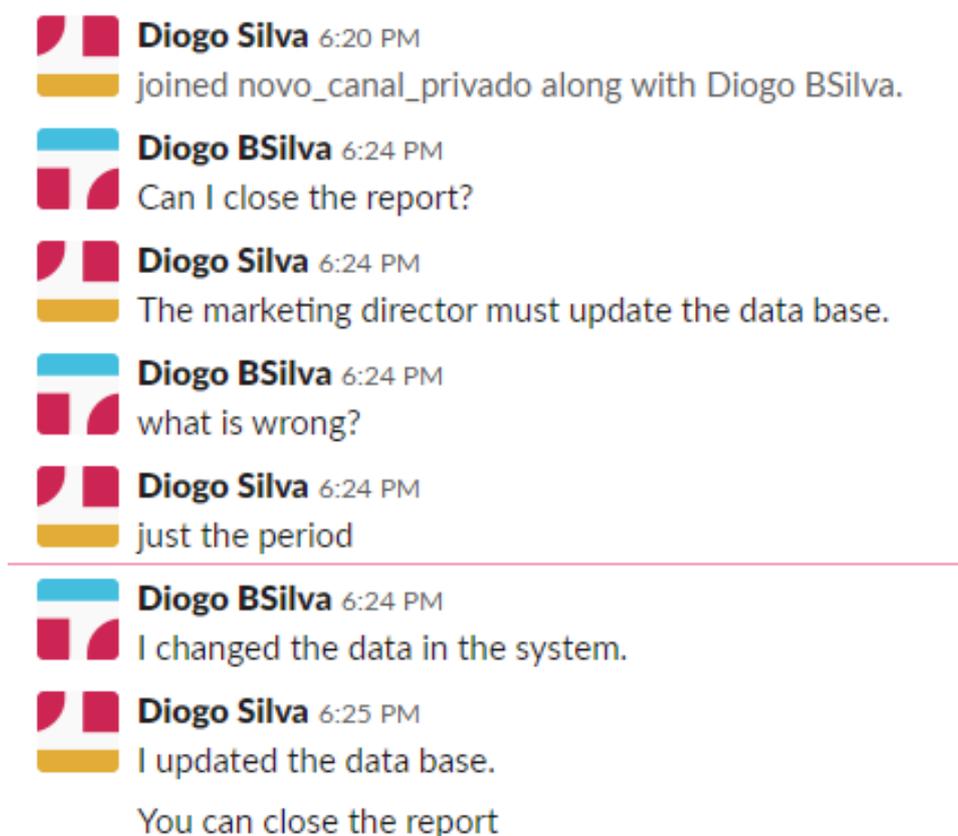


Figura 5.2: Tela de Cadastro

Início **Cadastrar**

Name

Access Token

Cadastrar

Fonte: Os Autores

5.1 Utilização da ferramenta desenvolvida

Para a utilização da ferramenta, deve-se cadastrar o *token* que foi gerado previamente. Uma tela de cadastro é exibida, onde deve-se fornecer um nome de conta, preferencialmente o nome da organização que se deseja obter as informações da plataforma *slack*. Esta tela é mostrada na Figura 5.2. Podem ser cadastrados quantos *tokens* forem necessários.

Após o cadastro do *token*, este será exibido na tela inicial, como uma possível escolha para a extração de informações, conforme mostra a Figura 5.3. Nesta tela é possível editar ou excluir o *token* da tabela de cadastro e escolher qual dos *workspaces*, geralmente é o nome da organização, ou o nome genérico para um time de trabalho, que será utilizado para a obtenção dos *logs* de conversas. Assim, ao clicar em “extrair” em algum dos *workspaces*, é exibida uma tela onde são disponibilizados todos os canais que aquele determinado *workspace* possui. Deve-se então, na tela da Figura 5.4, selecionar o canal que se tem interesse, clicando sobre o botão “extrair do canal”.

Figura 5.3: Tela Inicial

Início [Cadastrar](#)

Nome	Channel	Flag	Opções
teste_tcc_diogo	xoxp-468116830608-468682947540-468702433252-aeaafc1a145d6d8d1c28f77f29724575	Ativado	Editar Excluir Extrair
studycases	xoxp-449791295479-448641432947-448749972370-7151df7004b15ad2a9a43531b0964fd8	Desativado	Editar Excluir Extrair

Fonte: Os Autores

Figura 5.4: Tela de Escolha do Canal de Comunicação

Início [Cadastrar](#)

teste_tcc_diogo

terceiro [Extrair do canal](#)

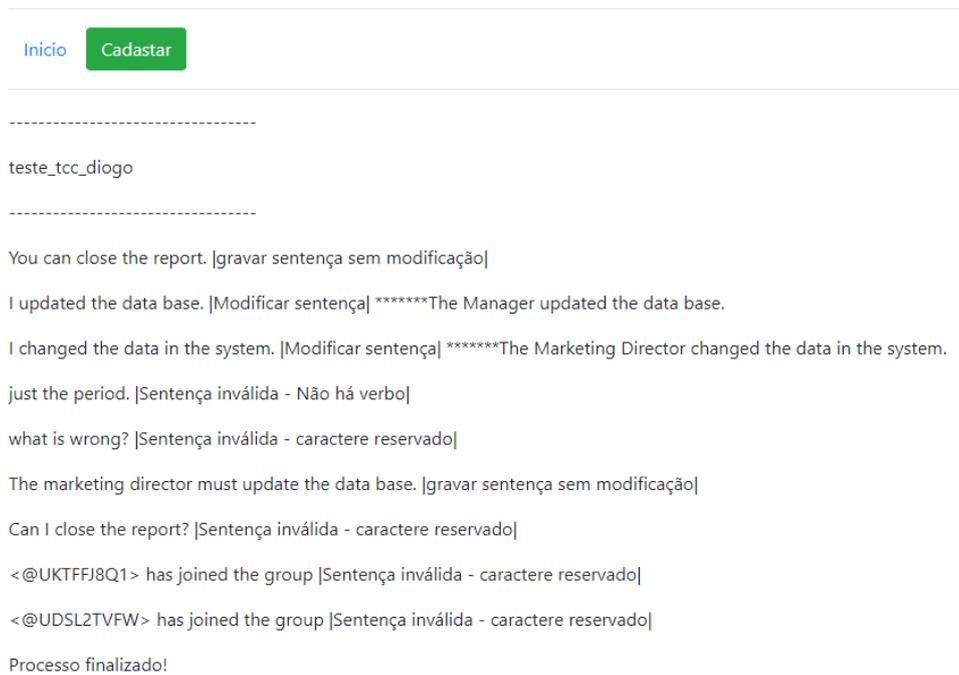
testes_tcc [Extrair do canal](#)

segundochannel [Extrair do canal](#)

Fonte: Os Autores

Já a Figura 5.5 representa a última etapa para extração e processamento das informações da plataforma *Slack*. Após o cadastramento do *token* de acesso, a escolha da conta e a escolha do canal que se deseja obter informações, ocorre todo o processamento em segundo plano. Isso requer alguns minutos e depende da quantidade de dados gerados no canal, ou seja, quanto mais conversas, mais tempo é necessário para o processamento. Por exemplo, uma conversa com 7 frases pode levar cerca de 20 segundos.

Figura 5.5: Tela de extração e processamento das informações



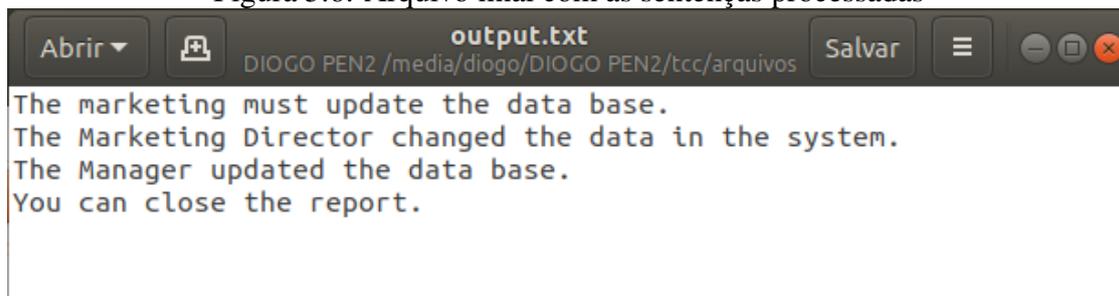
Fonte: Os Autores

Na imagem 5.5 é possível verificar ao lado de cada sentença um *status* desta, que identifica se é uma sentença válida, se não possui perguntas ou caracteres reservados e se deve ser gravada com ou sem modificação, condição que relaciona se a sentença se apresenta na estrutura sujeito + verbo + objeto ou não. Ao final é salvo um arquivo em formato .txt com as sentenças processadas de modo a melhor se aproximarem de uma descrição de processo. Este arquivo contém apenas as sentenças válidas, após serem processadas para o formato que se deseja, como mostrado na Figura 5.6.

5.2 Teste e verificação da abordagem

Segundo Pressman and Maxim (2016), uma estratégia para planejamento sistemático de teste, execução e controle começa considerando pequenos elementos do software

Figura 5.6: Arquivo final com as sentenças processadas



The screenshot shows a text editor window titled 'output.txt' with a dark theme. The window title bar includes 'Abrir', 'DIOGO PEN2 /media/diogo/DIOGO PEN2/tcc/arquivos', 'Salvar', and window control buttons. The text content is as follows:

```
The marketing must update the data base.
The Marketing Director changed the data in the system.
The Manager updated the data base.
You can close the report.
```

Fonte: Os Autores

e se encaminha para fora no sentido de abranger o software como um todo. Desta forma, classifica o teste em quatro etapas: *teste de unidade*, que foca na verificação na menor unidade de projeto do software, como os componentes ou módulos de *software*; *teste de integração*, cujo objetivo é construir uma estrutura de programa determinada pelo projeto a partir de componentes testados em unidade; *teste de validação*, que demonstra os requisitos de rastreabilidade do software, ou seja, como surgiram determinados requisitos ou a quais outros requisitos estão relacionados, por exemplo e; *teste de sistema*, que aprova o software quando ele é incorporado em um sistema maior.

De acordo com Pressman and Maxim (2016), o teste de unidade normalmente é considerado como um auxiliar para a etapa de codificação. Assim, testou-se progressivamente, durante a codificação, cada método novo da aplicação criada no presente trabalho, a medida que eram desenvolvidos. Desta forma, pretendeu-se cobrir os erros nos trechos de código considerados críticos, como na validade de um *token* ou na recuperação de informações na *API Slack*, por exemplo.

Após os testes de cada novo método, ou unidade do software, progrediu-se para testar se esses componentes se comportam de acordo com o esperado e funcionam em conjunto, ou seja, teste de integração. Para isso, criou-se casos de teste, como mostrado na Tabela 5.1, para testar o funcionamento integrado das unidades e a adequação da aplicação aos requisitos que se esperava implementar. Deseja-se, em trabalhos futuros, a realização de testes de usabilidade, ou seja, testes com outros usuários, a fim de proporcionar melhorias na ferramenta no contexto de interação com as interfaces.

De acordo com Ricca and Tonella (2001), no contexto de aplicações *web*, pode-se aplicar técnicas de teste estrutural, ou caixa branca, para avaliar o comportamento interno do componente de software. Os autores propõem critérios para este tipo de teste:

Teste de página todas as páginas do site são visitadas pelo menos uma vez em algum caso de teste;

teste de hiperlink todos os hiperlinks de todas as páginas do site são percorridos

Tabela 5.1: Casos de teste da integração das unidades

<i>Caso de teste</i>	<i>Requisito</i>	<i>Realiza</i>
C1	A aplicação faz corretamente o cadastro de uma conta?	Sim
C2	A aplicação faz corretamente a edição de uma conta?	Sim
C3	A aplicação faz corretamente a exclusão de uma conta?	Sim
C4	A aplicação reconhece a validade de um token?	Sim
C5	A aplicação consegue se comunicar com a API Slack?	Sim
C6	A aplicação consegue obter os canais e seus IDs internos?	Sim
C7	A aplicação reconhece canais públicos?	Sim
C8	A aplicação reconhece canais privados?	Sim
C9	A aplicação funciona com canais sem nenhuma mensagem?	Sim
C10	A aplicação captura as informações de um canal que possui mensagens?	Sim
C11	A aplicação reconhece as estruturas gramaticais de uma sentença?	Sim
C12	A aplicação modifica as sentenças para a estrutura sujeito + verbo + objeto?	Sim
C13	A aplicação gera um arquivo de saída corretamente?	Sim

Fonte: Os Autores

pelo menos uma vez;

teste de todos os usos todos os caminhos de navegação serão exercidos considerando todas as possibilidades de uso de variáveis, testando a dependência de dados;

teste de todos os caminhos todo caminho no site é percorrido em algum caso de teste pelo menos uma vez.

No contexto de teste de caixa branca, segundo a abordagem de Ricca and Tonella (2001), exercitou-se as páginas, ou telas, os usos e todos os caminhos possíveis dentro da aplicação. Não usou-se o critério de *hiperlinks*, pois na aplicação desenvolvida no presente trabalho essa relação não existe, ou seja, devido a inexistência de hiperlinks, não há a necessidade de testes neste sentido. A tabela 5.2 mostra os caminhos exercitados, bem como as páginas abordadas em cada caminho.

Os testes feitos na ferramenta desenvolvida estão detalhados no Apêndice A, cujos resultados são listados na ordem que os casos de teste foram criados, como mostrado na Tabela 5.1. Já o código fonte desenvolvido pode ser acessado em <https://github.com/dbsdih/BPM_Slack>.

Tabela 5.2: Teste caixa-branca de acordo com Ricca and Tonella (2001)

<i>Caminhos exercitados</i>	<i>Telas exercitadas</i>	<i>êxito</i>
Cadastrar > cadastrar > Registro incluído com sucesso!	Cadastro	Sim
Editar > Editar > Registro incluído com sucesso!	Edição	Sim
Excluir > Registro deletado com sucesso!	Exclusão	Sim
Cadastrar > Início	Cadastro, Início	Sim
Extrair > Token não existe!	Extrair	Sim
Extrair > Canal 1 público > Arquivo gerado com sucesso!	Extrair, Lista de canais, tela de conclusão de tarefa	Sim
Extrair > Canal 2 público > Arquivo gerado com sucesso!	Extrair, Lista de canais, tela de conclusão de tarefa	Sim
Extrair > Canal 1 privado > Arquivo gerado com sucesso!	Extrair, Lista de canais, tela de conclusão de tarefa	Sim
Extrair > Canal 2 privado > Arquivo gerado com sucesso!	Extrair, Lista de canais, tela de conclusão de tarefa	Sim

Fonte: Os Autores

5.3 Resumo do Capítulo

Neste capítulo abordou-se a utilização da ferramenta e demonstrou-se os passos para a extração de informações da plataforma *Slack*. Observou-se a criação do *token* de acesso da *API Slack*, a criação de um registro na ferramenta desenvolvida e a exibição dos canais existentes neste registro, para seleção e extração de informações.

Além disso, discutiu-se uma metodologia de teste e verificação da ferramenta, com base na abordagem de Pressman and Maxim (2016), executando-se assim testes de unidade e integração, por exemplo, por meio da criação de treze casos de testes. Utilizou-se também uma técnica de teste estrutural, ou caixa-branca, com base na abordagem de Ricca and Tonella (2001), cuja finalidade é exercitar as páginas, os caminhos e usos da ferramenta e assim, avaliar o comportamento interno dos componentes do software desenvolvido.

6 CONCLUSÕES

Com o desenvolvimento de novas redes sociais, os meios de comunicação nas organizações tendem a ficar também mais plurais. Assim, supõe-se que informações acerca de processos de negócios estejam presentes também nestas novas plataformas, porém de forma não tão disponível para acesso e análise por meio de ferramentas computacionais. Além disso, sabe-se que as etapas de identificação e descoberta de processos representam um custo relativamente alto no ciclo do BPM e assim, métodos automáticos e/ou parcialmente automáticos de realizá-las são altamente desejáveis. Desta forma, o presente trabalho apresentou uma abordagem para extrair as informações referentes a processos a partir da plataforma *slack* e processá-las a fim de criar uma estrutura que possa ser utilizada por outras ferramentas, com poucas adaptações.

Informações acerca de processos de negócio podem estar nos mais variados locais dentro de uma organização. Ter acesso a essas informações e sobretudo utilizá-las para mapear os processos de uma organização dá a possibilidade de diminuir custos operacionais e melhorar a utilização de recursos.

Neste trabalho, deu-se prioridade para as etapas de *identificação e descoberta de processos*, por meio dos módulos de extração de dados e NLP, para então fornecer um arquivo de entrada para a arquitetura desenvolvida em Silva and Thom (2019).

O trabalho foi dividido em duas partes: a primeira dedicada a extrair dados da plataforma *Slack*, ou seja, todas as conversas de um determinado canal de comunicação, onde os times trocam atividades e informações relevantes aos processos de trabalho ao qual estão relacionados. Já a segunda parte foi dedicada a criar meios de melhor estruturar os dados obtidos da primeira etapa, por meio de NLP, para posteriormente ser possível gerar um arquivo com texto sequencial que se aproxima de uma descrição de processo. A etapa de extração foi desenvolvida de forma a integrar uma interface *web* à *API* disponibilizada pela própria plataforma. Utilizou-se métodos específicos da *API* para a obtenção de dados sobre a conta, como *token* de acesso, grupo e canal de trabalho (privacidade e id interno dos canais) e assim coletar as conversas realizadas na plataforma, os participantes do grupo, seus cargos e demais informações que tenham relações relevantes com o estudo dos processos realizados.

Já a etapa de processamento, abrange principalmente estruturação de sentenças e criação de definições de adequação das conversas para o tipo de texto que se deseja obter como saída da ferramenta. O processamento das conversas é de extrema importância, uma

vez que a plataforma desenvolvida por Silva and Thom (2019) espera um texto descritivo, diferente dos textos originados de conversas em ambientes virtuais sociais, como o caso da *Slack*, *facebook* ou *whatsApp*, por exemplo.

Para isso, dividiu-se a etapa em duas fases de processamento, uma voltada à inclusão/exclusão de sentenças e pontuação e uma segunda fase voltada para análise de sintaxe, com foco nas estruturas gramaticais apresentadas. Foram utilizadas para isso as ferramentas de NLP do Stanford Core NLP, mais especificamente o Stanford Parser.

Assim, com o auxílio das ferramentas de processamento, foi possível desenvolver as definições de estruturação dos textos, de acordo com o padrão <sujeito> + <verbo> + <objeto>, de forma sequencial e sem sentenças de questionamentos, conforme os modelos discutidos no capítulo 4. Além disso, criou-se um mecanismo de exclusão de sentenças que não possuem verbo, ou ainda aquelas cujo número de caracteres é pequeno, pois supõe-se que estas não contêm elementos relevantes relacionados a processos.

6.1 Contribuições e limitações da ferramenta desenvolvida

No presente trabalho foi apresentada uma abordagem para extrair informações relacionadas a elementos de processos de negócio, tendo como fonte a plataforma *Slack*. Apresentou-se também modos de manipulação dessas informações, por meio de NLP, para adequá-las a uma estrutura de descrição de processos. Assim, criou-se um módulo de software que estende a arquitetura de Silva and Thom (2019) para suportar informações inicialmente em linguagem natural e provindas de redes sociais, possibilitando a utilização deste tipo de redes como fonte para a análise e descoberta de informações de processos, a fim de, futuramente, modelá-los de forma automática e com alto grau de correção. Entretanto, não fez-se testes com o arquivo de saída da ferramenta desenvolvida no presente trabalho em função da arquitetura dos autores, ou seja, a ferramenta limita-se a extrair as informações de processos e reestruturar as sentenças, conforme as definições do arquivo de entrada da arquitetura de Silva and Thom (2019). Deseja-se, em trabalhos futuros, testar esse arquivo de forma a analisar a capacidade de descoberta de elementos de processos da arquitetura dos autores, utilizando as trocas de mensagens provenientes da plataforma *Slack* como fonte.

Além disso, os diálogos simulados neste trabalho não são de conversas reais de times de uma organizações, em parte devido ao acesso de informações dessas organizações ser restrito. Entretanto, deseja-se a utilização de conversas reais na plataforma, com o intuito

de testar a eficiência da ferramenta e também melhorar os métodos de transformação de sentenças.

6.2 Trabalhos Futuros

Com os resultados obtidos por meio deste trabalho é possível conduzir outros estudos que relacionam NLP e processos de negócio, bem como na utilização das sentenças extraídas de redes sociais e transformadas pela ferramenta para mapeamento e análise dos processos de uma organização, por meio de BPMN, de forma automática ou parcialmente automática.

Durante o desenvolvimento do trabalho, foram verificados aspectos interessantes a serem melhorados. As seguintes possibilidades foram identificadas para trabalhos futuros:

- Expandir o módulo de extração para abranger outras plataformas sociais, como *facebook*, *Trello* e *whatsApp*, por exemplo.
- Implementar uma técnica elaborada para correção da ortografia das sentenças.
- Realizar testes de usabilidade, ou seja, testes com outros usuários, a fim de proporcionar melhorias na ferramenta no contexto de interação com as interfaces.
- Utilizar conversas reais de times de trabalho de organizações, com o intuito de testar a eficiência da ferramenta, bem como melhorar os métodos de transformação de sentenças.
- Testar similaridade dos modelos de processos gerados a partir do texto de saída da ferramenta e modelos gerados manualmente ou por outros métodos e instrumentos.
- Testar o arquivo de saída da ferramenta desenvolvida no presente trabalho na arquitetura desenvolvida por Silva and Thom (2019), de forma a analisar a capacidade de descoberta de elementos de processos da arquitetura dos autores, utilizando as trocas de mensagens provenientes da plataforma *Slack* como fonte.

REFERÊNCIAS

- AA, H. Van der et al. Challenges and opportunities of applying natural language processing in business process management. In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. p. 2791–2801.
- AAGESEN, G.; KROGSTIE, J. Bpmn 2.0 for modeling business processes. **Handbook on Business Process Management 1: Introduction, Methods, and Information Systems**, p. 219–250, 04 2015.
- Aalst, van der, W. **Process mining : discovery, conformance and enhancement of business processes**. Germany: Springer, 2011. ISBN 978-3-642-19344-6.
- ANDERSON, K. Getting acquainted with social networks and apps: picking up the slack in communication and collaboration. **Library Hi Tech News**, v. 33, n. 9, p. 6–9, 2016.
- API Methods. 2019. Available from Internet: <<https://api.slack.com/methods>>.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit**. Beijing: O'Reilly, 2009. ISBN 978-0-596-51649-9.
- BORDIGNON, A. Cláudia de A. et al. Natural language processing in business process identification and modeling: A systematic literature review. In: . Caxias do Sul, RS, Brasil: 14th Simpósio Brasileiro de Sistemas de Informação - SBSI 2018, 2018.
- DUMAS, M. et al. **Fundamentals of Business Process Management**. Berlin: Springer-Verlag GmbH Germany, part of Springer Nature, 2018.
- FRIEDRICH, F. Automated generation of business process models from natural language input. In: . Berlin, Germany: School Of Business And Economics Of The Humboldt-universitat Zu Berlin, 2010.
- FRIEDRICH, F.; MENDLING, J.; PUHLMANN, F. Process model generation from natural language text. In: MOURATIDIS, H.; ROLLAND, C. (Ed.). **Advanced Information Systems Engineering**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 482–496.
- GEIGER, M.; NEUGEBAUER, P.; VORNDRAN, A. Automatic standard compliance assessment of bpmn 2.0 process models. In: . [S.l.: s.n.], 2017.
- INDURKHYA, N.; DAMERAU, F. J. **Handbook of Natural Language Processing**. 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2010.
- International Organization for Standardization. Iso/iec 19510:2013 – information technology - object management group business process model and notation. In: . [S.l.: s.n.], 2013.
- KLEIN, D.; MANNING, C. D. Accurate unlexicalized parsing. In: **Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (ACL '03), p. 423–430. Available from Internet: <<https://doi.org/10.3115/1075096.1075150>>.

- KLEIN, D.; MANNING, C. D. Fast exact inference with a factored model for natural language parsing. In: BECKER, S.; THRUN, S.; OBERMAYER, K. (Ed.). **Advances in Neural Information Processing Systems 15**. MIT Press, 2003. p. 3–10. Available from Internet: <<http://papers.nips.cc/paper/2325-fast-exact-inference-with-a-factored-model-for-natural-language-parsing.pdf>>.
- LEONARD, P.; HUYSMAN, M.; STEINFELD, C. Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. **Journal of Computer-Mediated Communication**, v. 19, p. 1–19, 2013.
- LEOPOLD, H. **Natural Language in Business Process Models: Theoretical Foundations, Techniques, and Applications**. Switzerland: Springer International Publishing, 2013. ISBN 978-3-319-04174-2.
- MANNING, C. et al. The stanford corenlp natural language processing toolkit. In: . [S.l.: s.n.], 2014.
- PRESSMAN, R.; MAXIM, B. **Engenharia De Software: Uma Abordagem Profissional**. Porto Alegre, Brasil: MCGRAW HILL - ARTMED, 2016.
- RICCA, F.; TONELLA, P. Analysis and testing of web applications. In: . Toronto, Ontario, Canada: [s.n.], 2001. p. 25– 34.
- ROBERTSON, B. W.; KEE, K. F. Social media at work: The roles of job satisfaction, employment status, and facebook use with co-workers. **Computers in Human Behavior**, v. 70, p. 191 – 196, 2017.
- RUSU, D. et al. **Triplet Extraction from Sentences**. Ljubljana, Slovenia: Proceedings of the 10th International Multiconference - Information Society-IS, 2007.
- SCHUR, P. **PHP wrapper for the Stanford Natural Language Processing library**. 2019. Available from Internet: <<https://github.com/patrickschur/stanford-nlp-tagger>>.
- SILVA, S. T.; THOM, L. H. An architecture for enabling business process-oriented text generation. In: . Porto Alegre, RS, Brasil: Universidade Federal do Rio Grande do Sul, 2019.
- SILVA, T. S. et al. Empirical analysis of sentence templates and ambiguity issues for business process descriptions. In: **On the Move to Meaningful Internet Systems. OTM 2018 Conferences - Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, October 22-26, 2018, Proceedings, Part I**. Valletta, Malta: [s.n.], 2018. p. 279–297.
- SLACK. **Imagine o que você e sua equipe poderão fazer juntos**. 2019. Available from Internet: <<https://slack.com/intl/pt-br/>>.
- SOCHER, R. et al. Parsing with compositional vector grammars. In: **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Sofia, Bulgaria: Association for Computational Linguistics, 2013. p. 455–465.
- SOMMERVILLE, I. et al. **Engenharia de software**. São Paulo, Bra: Pearson Prentice Hall, 2011.

TAYLOR, A.; MARCUS, M.; SANTORINI, B. **The Penn Treebank: An Overview.** 2003.

The Object Management Group. **About The Business Process Model And Notation Specification Version 2.0.** 2011. Available from Internet: <<https://www.omg.org/spec/BPMN/2.0/>>.

The PHP Group. **O que é PHP?** 2019. Available from Internet: <https://php.net/manual/pt_BR/intro-what-is.php>.

APÊNDICE A — TESTES DA FERRAMENTA DESENVOLVIDA

Casos de teste:

C1- A aplicação faz corretamente o cadastro de uma conta?

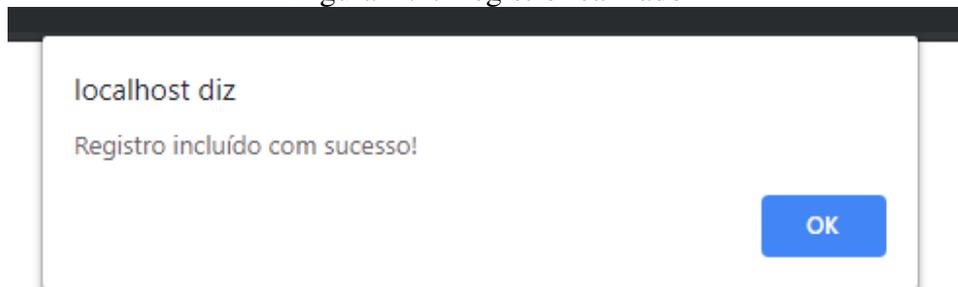
A Figura A.1 mostra que só é possível cadastrar um novo *token* preenchendo corretamente os campos de nome e *token* na tela de cadastro. Caso tenham sido preenchidos os campos, o registro é feito com sucesso (como mostra a Figura A.2). Já a Figura A.3 mostra a tela inicial com alguns *tokens* cadastrados.

Figura A.1: Preenchimento dos campos obrigatórios

A captura de tela mostra a interface de cadastro. No topo, há dois botões: 'Início' (azul) e 'Cadastrar' (verde). Abaixo, há dois campos de entrada: 'Name' e 'Access Token'. O campo 'Access Token' está em foco e possui uma mensagem de erro em uma caixa flutuante que diz 'Preencha este campo.' com um ícone de alerta. Abaixo dos campos, há um botão verde 'Cadastrar'.

Fonte: Os Autores

Figura A.2: Registro realizado



Fonte: Os Autores

Figura A.3: Tela Inicial com os registros criados

A captura de tela mostra a tela inicial com os registros criados. No topo, há dois botões: 'Início' (azul) e 'Cadastrar' (verde). Abaixo, há uma tabela com as seguintes colunas: Nome, Channel, Flag e Opções.

Nome	Channel	Flag	Opções
teste_tcc_diogo	xxxp-468116830608-468682947540-468702433252-aaaaf1a145d6d8d1c28f77f29724575	Ativado	Editar Excluir Extrair
token_errado	sdfsfdsfds	Desativado	Editar Excluir Extrair
studycases	xxxp-449791295479-448641432947-448749972370-7151df7004b15ad2a9a43531b0964fd8	Desativado	Editar Excluir Extrair

Fonte: Os Autores

C2- A aplicação faz corretamente a edição de uma conta?

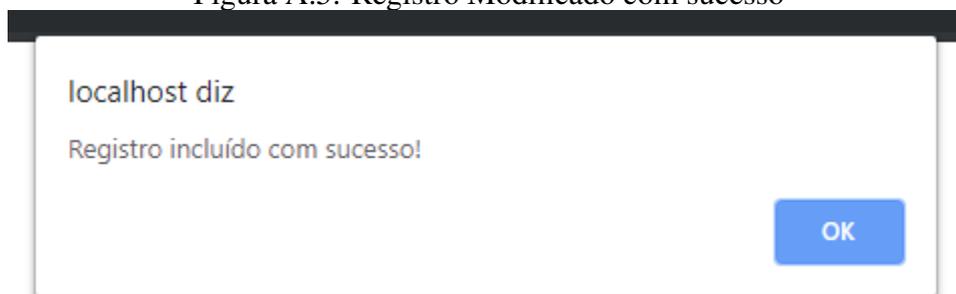
A Figura A.4 mostra a tela de edição. Caso tenham sido preenchidos os campos, o registro é editado com sucesso (como mostra a Figura A.5). Já a Figura A.6 mostra

a tela inicial com os *tokens* que existem cadastrados. Na entrada *token_errado* existe um novo valor, atribuído após a edição.

Figura A.4: Preenchimento dos campos obrigatórios para edição

Fonte: Os Autores

Figura A.5: Registro Modificado com sucesso



Fonte: Os Autores

Figura A.6: Tela Inicial com registros realizados

Nome	Channel	Flag	Opções
teste_tcc_diogo	xoxp-468116830608-468682947540-468702433252-aeaaf1a145d6d8d1c28f77f29724575	Ativado	Editar Excluir Extrair
token_errado	token_errado_depois_de_editar_no_cadastro	Ativado	Editar Excluir Extrair
studycases	xoxp-449791295479-448641432947-448749972370-7151df7004b15ad2a9a43531b0964fd8	Desativado	Editar Excluir Extrair

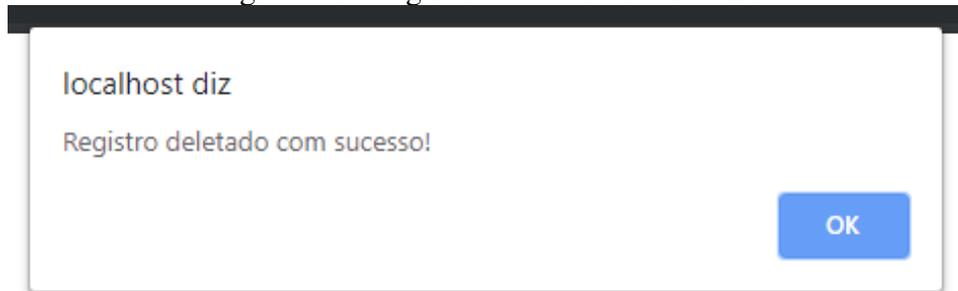
Fonte: Os Autores

C3- A aplicação faz corretamente a exclusão de uma conta?

A Figura A.7 é exibida após o clique sobre o botão *Excluir*, em um dos canais cadastrados anteriormente. Neste exemplo, excluiu-se o registro *token_errado*, apresentado

anteriormente na Figura A.6 e que já não pode ser visto na Figura A.8.

Figura A.7: Registro deletado com sucesso



Fonte: Os Autores

Figura A.8: Tela Inicial sem o registro deletado

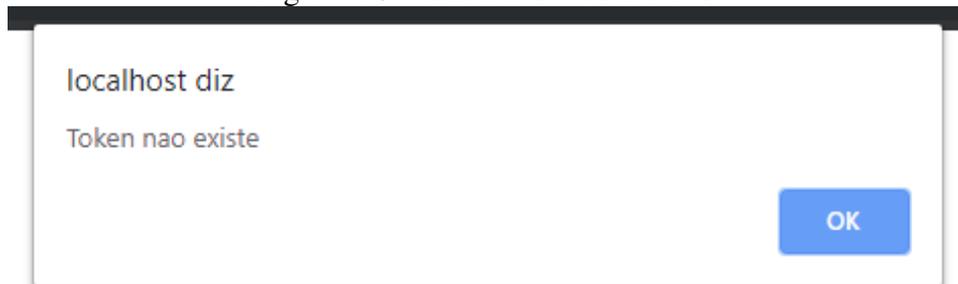
Nome	Channel	Flag	Opções
teste_tcc_diogo	xxxx-468116830608-468682947540-468702433252-aeaafc1a145d6d8d1c28f77f29724575	Ativado	Editar Excluir Extrair
studycases	xxxx-449791295479-448641432947-448749972370-7151df7004b15ad2a9a43531b0964fd8	Desativado	Editar Excluir Extrair

Fonte: Os Autores

C4- A aplicação reconhece a validade de um token?

Em caso de token não reconhecido, a ferramenta exibe uma tela de aviso, informando que o *token* não é válido, como mostrado na Figura A.9.

Figura A.9: Token não reconhecido



Fonte: Os Autores

C5- A aplicação consegue se comunicar com a API Slack?

C6- A aplicação consegue obter os canais e seus IDs internos?

A Figura A.10 mostra que a ferramenta consegue obter os *IDs* internos dos canais, pois lista todos os canais existentes na conta selecionada. Assim, mostra-se também que a ferramenta consegue se comunicar com a API e extrair informações por meio dela.

Figura A.10: Comunicação com a API Slack e obtenção de informação



Fonte: Os Autores

C7- A aplicação reconhece canais públicos?

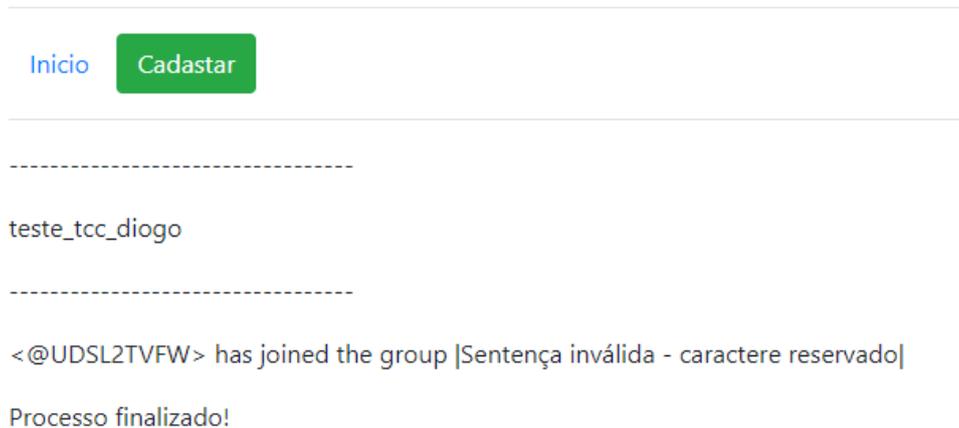
C8- A aplicação reconhece canais privados?

A Figura A.10 mostra os canais que existem na conta, onde “terceiro” é um canal público e os demais são privados. Logo, se todos os canais são listados, a ferramenta é capaz de reconhecer ambos os tipos de canais, ou seja, ela reconhece canais públicos e canais privados.

C9- A aplicação funciona com canais sem nenhuma mensagem?

A Figura A.11 mostra que a ferramenta reconhece que a única sentença presente no canal é uma ação, ou seja, um usuário entrou no chat. Assim, não há mensagens a serem processadas e o processamento é finalizado.

Figura A.11: Processamento de canal sem mensagens



Fonte: Os Autores

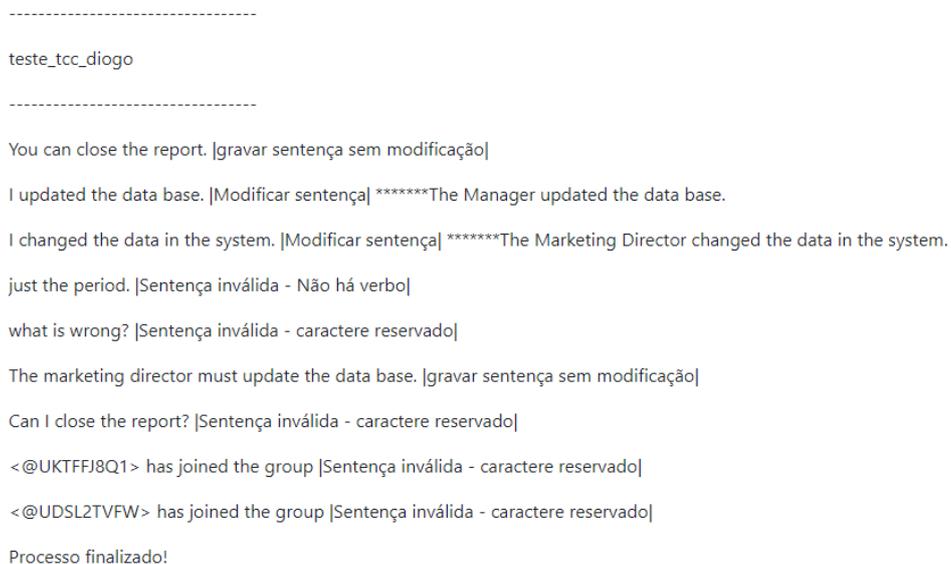
C10- A aplicação captura as informações de um canal que possui mensagens?

C11- A aplicação reconhece as estruturas gramaticais de uma sentença?

C12 - A aplicação modifica as sentenças para a estrutura sujeito + verbo + objeto?

A Figura A.12 mostra que a ferramenta consegue extrair todas as mensagens presentes em um canal e exibi-las na tela. Além disso, a ferramenta reconhece a estrutura gramatical das sentenças e também modifica as estruturas, quando necessário, para o formato sujeito + verbo + objeto.

Figura A.12: teste em canal com mensagens

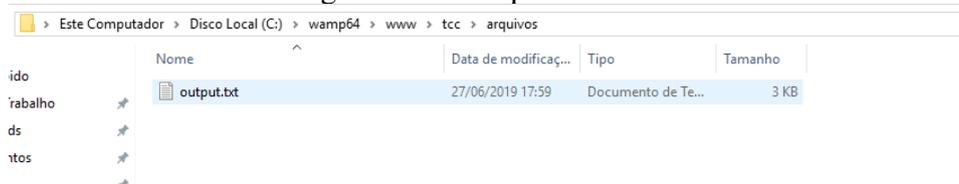


Fonte: Os Autores

C13- A aplicação gera um arquivo de saída corretamente?

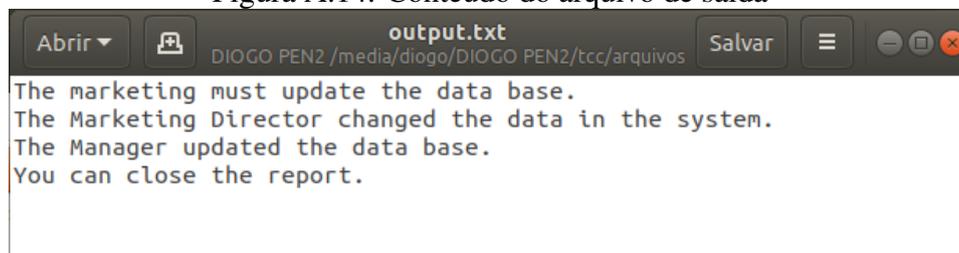
A Figura A.13 mostra o arquivo de saída criado a partir da extração e processamento das mensagens dentro de um canal. Já a Figura A.14 mostra o conteúdo deste arquivo, com as sentenças finais do exemplo apresentado.

Figura A.13: Arquivo de saída



Fonte: Os Autores

Figura A.14: Conteúdo do arquivo de saída



Fonte: Os Autores