



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
Faculdade de Ciências Exatas e Tecnologia
Curso de Ciência da Computação

GABRIEL WILLIS PINHEIRO GUERRA
TERSO WILLIS PINHEIRO GUERRA

APLICAÇÃO DE ANÁLISE DE REPUTAÇÃO NAS MÍDIAS SOCIAIS

São Paulo
Dezembro de 2018

**GABRIEL WILLIS PINHEIRO GUERRA
TERSO WILLIS PINHEIRO GUERRA**

APLICAÇÃO DE ANÁLISE DE REPUTAÇÃO NAS MÍDIAS SOCIAIS

Trabalho de Conclusão de Curso apresentado ao curso de Ciência da Computação da Pontifícia Universidade Católica de São Paulo – PUCSP, como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Maurício Nacib Pontuschka.

São Paulo
Dezembro de 2018

RESUMO

Pessoas e empresas estão procurando ter acesso à própria reputação e de terceiros. Porém, as técnicas existentes não são tão eficientes ou acessíveis para qualquer um.

Há técnicas não digitais e bastante antigas, como levantamento de questionários, assim como há ferramentas tecnológicas com alto custo de utilização. Ambas são acessíveis apenas para grandes empresas, excluindo a maioria da população e as empresas menores que também almejam essas informações.

Neste projeto, é descrita a arquitetura de um software que, de forma automatizada, coleta comentários em mídias sociais e gera gráficos ilustrando métricas de reputação. Além dos diagramas descritores da arquitetura, também há especificações do programa desenvolvido, que executa as principais funcionalidades do sistema proposto.

A solução proposta acessa a API da mídia social para coletar os comentários relacionados à entidade definida pelo usuário, aplica algoritmos de processamento de linguagem natural para classificar o conteúdo recebido e gera gráficos que ilustram o resultado das métricas de reputação calculadas.

A metodologia de desenvolvimento utilizada é o ICONIX. Diagramas foram elaborados para ajudar na compreensão da análise, desenvolvimento e funcionamento do protótipo.

A monografia ainda aborda, de forma introdutória e convidativa, os temas mídia social, consumismo e estratégias empresariais e seus reflexos no comportamento da sociedade contemporânea.

Palavras-chave: mídias sociais, análise de comentários, gráficos, processamento de linguagem natural, empresas, software

ABSTRACT

People and businesses are seeking to gain access to their own reputation and to others. Notwithstanding, the existing techniques are not as efficient or accessible to anyone.

There are non-digital and very old techniques, such as survey questionnaires, as there are technological tools with high cost of use. Both are accessible only to large companies, excluding the majority of the population and the smaller companies that also crave this information.

In this project, the architecture of software that, in an automated way, collects comments in media and generates graphs that illustrate reputation metrics is described. In addition to the architectural descriptive diagrams, there are also specifications of the developed program, which performs the main functionalities of the proposed system.

The proposed solution accesses the Social Media API to collect comments related to the user-defined entity, applies natural language processing algorithms to classify the content received, and generates graphs that illustrate the result of the calculated reputation metrics.

The development methodology used is ICONIX. Diagrams were developed to assist in the understanding of the analysis, development and operation of the prototype.

The term paper also addresses, in an introductory and inviting way, social media topics, consumerism and corporate strategies and their effects on the behavior of contemporary society.

Keywords: social media, comment analysis, graphics, natural language processing, business, software.

Agradecimentos

Agradecemos com amor esse trabalho aos nossos pais, Ana Carla Pinheiro Freitas e Willis Santiago Guerra Filho, que nos inspiram e investem no nosso desenvolvimento acadêmico e pessoal. Este é o fechamento de toda trajetória de estudos e desenvolvimento intelectual dos 4 aos 21 anos. Eles sempre colocaram a educação, o estudo e o conhecimento como prioridade e são 2 acadêmicos renomados e nosso maiores exemplos.

Agradecemos também à Sofia Marshallowitz Apuzzo pela ajuda na revisão deste TCC, pelo exemplo de conquista nos estudos acadêmicos e por todo apoio.

À Deoclides Augusto Gomes Neto, pela cooperação na idealização desse projeto, pela mentoria e investimento na nossa carreira profissional, que trouxe a aplicação dos conhecimentos teóricos da faculdade e tantos conhecimentos práticos do mundo de negócios.

Ao Orientador Maurício Nacib Pontuschka, por ensinar não só à diagramar, desenvolver e apresentar um bom TCC, mas também pelos ensinamentos sobre a vida profissional e por nos inspirar sendo sempre uma pessoa positiva e bem humorada.

Aos Professores que tiveram um papel essencial no ensinamento das nossas bases de conhecimento matemático e computacional e que nos cativaram de uma forma especial:

Cristiana Abud da Silva Fusco, Lisbete Madsen Barbosa, Edith Ranzini, Ítalo Santiago Vega, Satoshi Nagayama, Eduardo Savino Gomes, Carlos Eduardo de Barros Paes.

À Letícia da Cruz Silva, pelo apoio e motivação para chegarmos até aqui.

LISTA DE FIGURAS

FIGURA 1 - EXEMPLO CÉLEBRE DA PRIMEIRAS INTERAÇÃO DA NETFLIX NA MÍDIA A CAUSAR GRANDE COMOÇÃO

FIGURA 2 - INTERAÇÃO EM MÍDIA SOCIAL ENTRE DIVERSAS EMPRESAS E USUÁRIOS, POR SEQUÊNCIA DE MENSAGENS, QUE CAUSOU GRANDE REPERCUSSÃO.

FIGURA 3 - À DIREITA, PARTE ILUSTRATIVA DO VÍDEO DE PROPAGANDA DA EMPRESA MCDONALD'S, REPLETA DE MEMES. À ESQUERDA, MEME DO PICA-PAU NAS CATARATAS DO NIÁGARA

FIGURA 4 - MAPA MENTAL DO CONTEXTO EM QUE O TRABALHO ESTÁ INSERIDO

FIGURA 5 - MAPA MENTAL DAS POSSIBILIDADES DE UTILIZAÇÃO DA FERRAMENTA DESENVOLVIDA

FIGURA 6 - MAPA MENTAL DO CONTEÚDO DE PROCESSAMENTO DE LINGUAGEM NATURAL

FIGURA 7 - PROPAGANDA DA EMPRESA BURGER KING CONTRA DISCURSOS DE ÓDIO NAS REDES

FIGURA 8 - MAPA MENTAL DO CONTEÚDO DE MÍDIAS SOCIAIS

FIGURA 9 - MAPA MENTAL DO CONCEITO DE REPUTAÇÃO

FIGURA 10 - PROCESSO DE DIAGRAMAÇÃO DA METODOLOGIA ICONIX

FIGURA 11 - WORDCLOUD DAS PALAVRAS QUE COMPÕEM ESTA MONOGRAFIA ATÉ O SUBCAPÍTULO 3.1

FIGURA 12 - DIAGRAMA DOS CASOS DE USO DO PROJETO

FIGURA 13 - MODELO DE DOMÍNIO DO PROJETO

FIGURA 14 - TIPOS DE OBJETOS DE DIAGRAMAS DE ROBUSTEZ

FIGURA 15 - DIAGRAMA DE ROBUSTEZ DE COLETA DE COMENTÁRIOS

FIGURA 16 - DIAGRAMA DE ROBUSTEZ DA INICIALIZAÇÃO DA ANÁLISE DE REPUTAÇÃO

FIGURA 17 - DIAGRAMA DE ROBUSTEZ DA GERAÇÃO DE GRÁFICO DE ANÁLISE

FIGURA 18 - DIAGRAMA DE ROBUSTEZ DE VALIDAÇÃO DA ANÁLISE

FIGURA 19 - DIAGRAMA DE ROBUSTEZ DE EDIÇÃO DA COLETA DE ANÁLISE

FIGURA 20 - DIAGRAMA DE SEQUÊNCIA DE MENSAGENS DA COLETA DE COMENTÁRIOS

FIGURA 21 - DIAGRAMA DE SEQUÊNCIA DE MENSAGENS DE INICIAÇÃO DA ANÁLISE DE REPUTAÇÃO

FIGURA 22 - DIAGRAMA DE SEQUÊNCIA DE MENSAGENS DA GERAÇÃO DE GRÁFICO DE ANÁLISE

FIGURA 23 - DIAGRAMA DE SEQUÊNCIA DE MENSAGENS DA VALIDAÇÃO DE ANÁLISE

FIGURA 24 - DIAGRAMA DE SEQUÊNCIA DE MENSAGENS DA EDIÇÃO DA COLETA DE ANÁLISE

FIGURA 25 - DIAGRAMA DE CLASSES CANDIDATAS

FIGURA 26 - ILUSTRAÇÃO DE PARTE DOS DADOS ESTRUTURADOS NO ARQUIVO JSON REFERENTES À COLETA DA PUCSP

FIGURA 27 - WORDCLOUD DAS REFERÊNCIAS À PUCSP NO TWITTER

FIGURA 28 - GRÁFICO DE SETORES DAS REFERÊNCIAS À PUCSP NO TWITTER

FIGURA 29 - VALORES EXIBIDOS PELA INTERFACE DA APLICAÇÃO, REFERENTES À PUCSP NO TWITTER

FIGURA 30 - WORDCLOUD DAS REFERÊNCIAS À UNINOVE NO TWITTER

FIGURA 31 - GRÁFICO DE SETORES DAS REFERÊNCIAS À UNINOVE NO TWITTER

FIGURA 32 - VALORES EXIBIDOS PELA INTERFACE DA APLICAÇÃO, REFERENTES À UNINOVE NO TWITTER

FIGURA 33 - WORDCLOUD DAS REFERÊNCIAS À BOLSONARO NO TWITTER

FIGURA 34 - GRÁFICO DE SETORES DAS REFERÊNCIAS À BOLSONARO NO TWITTER

FIGURA 35 - VALORES EXIBIDOS PELA INTERFACE DA APLICAÇÃO, REFERENTES À BOLSONARO NO TWITTER

FIGURA 36 - WORDCLOUD DAS REFERÊNCIAS À HADDAD NO TWITTER

FIGURA 37 - GRÁFICO DE SETORES DAS REFERÊNCIAS À HADDAD NO TWITTER

FIGURA 38 - VALORES EXIBIDOS PELA INTERFACE DA APLICAÇÃO, REFERENTES À HADDAD NO TWITTER

LISTA DE TABELAS

TABELA 1 - SESSÃO DAS NOTÍCIAS MAIS REPLICADAS NO TWITTER

TABELA 2 - SESSÃO DAS NOTÍCIAS MAIS REPLICADAS NO FACEBOOK

SUMÁRIO

1. INTRODUÇÃO	12
1.1. Motivação	12
1.2. Objetivo	16
1.2.1. Objetivo Geral:	19
1.2.2. Objetivos Específicos:	19
1.3. Delimitações	19
1.4. Contribuição	21
1.5. Método de Trabalho	22
1.6. Organização do Texto	23
2. REVISÃO BIBLIOGRÁFICA	25
2.1. Fundamentação Teórica	25
2.2. Trabalhos Relacionados	35
2.2.1. Mineração de Opiniões Aplicada a Mídias Sociais	35
2.2.2. Visibilidade de Notícias no Twitter e no Facebook	38
3. DESENVOLVIMENTO DO PROTÓTIPO	42
3.1. Descrição do sistema proposto	43
3.2. Análise de Caso de Uso	46
3.2.1. Diagrama dos Casos de Uso	47
3.2.2. Descrição dos Casos de Uso	48
3.2.2.1. Caso de Uso: Iniciar análise de reputação	48
3.2.2.2. Caso de Uso: Gerar gráficos de análise	49
3.2.2.3. Caso de Uso: Validar análise	49
3.2.2.4. Caso de Uso: Editar coleta da análise	50
3.3. Modelo de Domínio	51
3.4. Diagrama de Robustez	52
3.4.1. Coleta de comentários	53
3.4.2. Iniciar análise de reputação	54

3.4.3. Gerar gráfico de análise	55
3.4.4. Validar análise	57
3.4.5. Editar coleta de análise	58
3.5. Diagrama de sequência de mensagens	59
3.5.1. Coleta de comentários	59
3.5.2. Iniciar análise de reputação	63
3.5.3. Gerar gráfico de análise	66
3.5.4. Validar análise	69
3.5.5. Editar coleta de análise	72
3.6. Diagrama de classes candidatas	75
4. DECISÕES ARQUITETURAIS	77
4.1. Decisões de Implementação	77
4.2. Codificação	78
4.3. Amostra de execução da aplicação	82
5. CONCLUSÃO E TRABALHOS FUTUROS	89
5.1. Conclusão	89
5.2. Trabalhos Futuros	90
6. REFERÊNCIAS BIBLIOGRÁFICAS	91

1. INTRODUÇÃO

Descrição dos motivos que levaram à escolha de desenvolver esse projeto, por meio da constatação que se teve da importância do conhecimento da reputação de empresas e seus concorrentes nas mídias sociais. Ele contém a motivação para escolher tratar do tema, o objetivo do presente projeto, suas delimitações funcionais, as contribuições propostas para o meio computacional e o método de trabalho utilizado para o seu desenvolvimento. Assim, há uma introdução ao tema e ao direcionamento do projeto.

1.1. Motivação

As empresas devem atentar às mídias sociais. Eis o local onde as pessoas expressam suas ideias espontaneamente e onde ocorre a interação dessas ideias capaz de favorecer ou arruinar a reputação de alguém. Como disse Daniel Motta em seu consagrado livro *Liderança Essencial*, as redes sociais são a nova “Força de Porta” a que os líderes de uma corporação devem atentar para que sua empresa não seja desbancada repentinamente, sem que ao menos saibam como (Motta, 2017).

A relação entre empresas diversas e mídias sociais se mostra cada vez mais acentuada, embora passe muitas vezes despercebida pelos olhos desatentos. Já existem instituições que investem na obtenção de uma posição relevante no espaço das mídias sociais. Com esse investimento, conseguem aumentar sua visibilidade e ganhar maior proximidade com seus clientes.

É conhecido que a Netflix foi pioneira em ganhar destaque nas mídias sociais. Ela, como várias outras, passou a prestar os serviços que envolvem comunicação com o cliente (como suporte básico, reclamações e recomendações) dentro do ambiente das mídias sociais desde seu início. Esse serviço que, ainda hoje, é comumente realizado em sessões precárias do site da própria empresa, através da seção de FAQs¹ ou via e-mail ou ligação ao setor de atendimentos, ganha com o

¹ Do inglês “frequently asked questions”, é a coleção das perguntas mais frequentes feitas pelos clientes sobre problemas com os serviços prestados pela empresa.

advento das mídias sociais uma comunicação muito mais próxima com o cliente (diálogo dentro do próprio meio de socialização dele), acontecendo de forma mais natural a ele. Porém, o pioneirismo da Netflix vai além quando seus empregados, representantes da marca nas mídias, passaram a responder de forma descontraída, informal, criativa e cômica as postagens de usuários que a citam. Fazendo piada das perguntas! Eles deixam de lado a formalidade pré-estabelecida da relação cliente-empresa e acabam por se tornar celebridade. Isso se deve à ação incitar um vínculo com esse seu potencial cliente, que se comunicaram diretamente, e com cada outro possível cliente que visualiza a interação na mídia, visto que esses diálogos se propagam na rede, sendo visualizados e repassados por milhares de outros usuários, consagrando assim uma imagem positiva e de destaque da marca nesse ambiente. Seguem duas imagens retiradas das mídias Twitter e Facebook, respectivamente, que ilustram bem esse fenômeno.

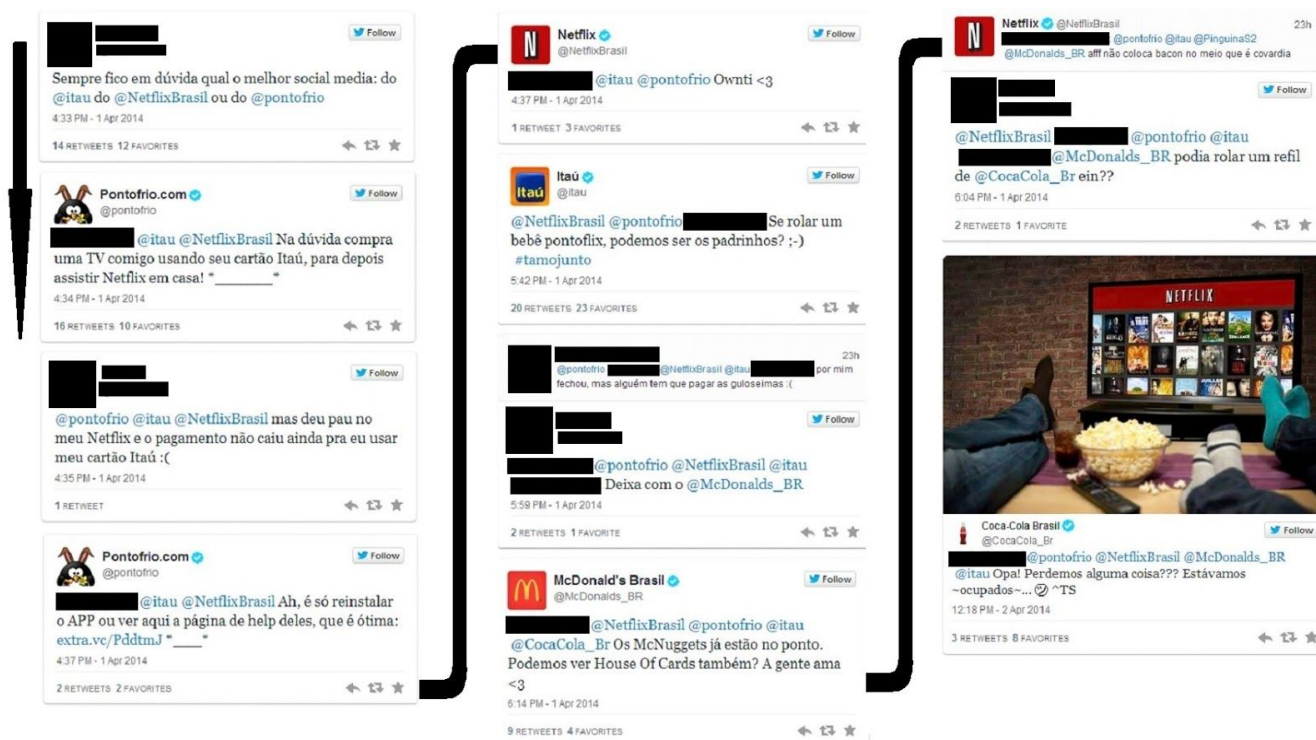
Figura 1 - exemplo célebre da primeiras interação da Netflix na mídia a causar grande comoção



Fonte: Insoonía². Edição de censura no nome e foto de perfil do usuário da mídia social realizada pelos membros do projeto.

² Disponível em: <<https://www.insoonía.com/top13-as-melhores-respostas-que-a-netflix-ja-deu-nas-redes-sociais/>>. Acesso em novembro de 2018.

Figura 2 - interação em mídia social entre diversas empresas e usuários, por sequência de mensagens, que causou grande repercussão.



Fonte: Criatives³. Edição de censura no nome e foto de perfil de usuários da mídia social, assim como adição de setas que indicam o fluxo do diálogo e união das imagens de diálogo previamente separadas, todas realizadas pelos membros do projeto.

Outra estratégia bastante utilizada por grandes companhias é o engajamento na cultura das mídias sociais para auxiliar a criação de propagandas. Utiliza-se o conhecimento dos memes⁴, da linguagem da internet em geral (o conhecido “internetês”) e dos assuntos mais difusos nas mídias sociais no momento para posicionar a marca como participante desse fenômeno, conquistando assim a atenção e afeição do cliente usuário das mídias sociais. Vale ressaltar que a estratégia funciona para propagandas difundidas tanto dentro do meio das mídias sociais quanto fora, em programas de televisão e outdoors, por exemplo, pois os usuários das redes, que se identificam com esse conteúdo, permeiam todo tipo de ambiente.

³ Disponível em: <<http://www.criatives.com.br/2014/04/netflix-itaú-ponto-frio-e-outras-marcas-conversam-pelo-twitter/>>. Acesso em novembro de 2018.

⁴ Conceito humorado na forma de imagem, vídeo ou GIF, vastamente difundido através da Internet, tornando-se parte fundamental da cultura digital.

Figura 3 - à direita, parte ilustrativa do vídeo de propaganda da empresa McDonald's, repleta de memes. À esquerda, Meme do Pica-Pau nas Cataratas do Iguaçu



Fonte: Youtube⁵ e Estadão⁶. Edição na junção de uma parte do vídeo da primeira fonte com a imagem da segunda fonte, postas lado a lado e separadas por um traço negro, realizada pelos membros do projeto.

Todavia, a utilização mais interessante que qualquer entidade pode fazer das mídias sociais é a análise da opinião das pessoas e da sua reputação dentro desse meio. Compreender a opinião real das pessoas sempre foi um desafio para a maioria das organizações, e as mídias sociais se mostram como uma possibilidade para a solução dessa dificuldade.

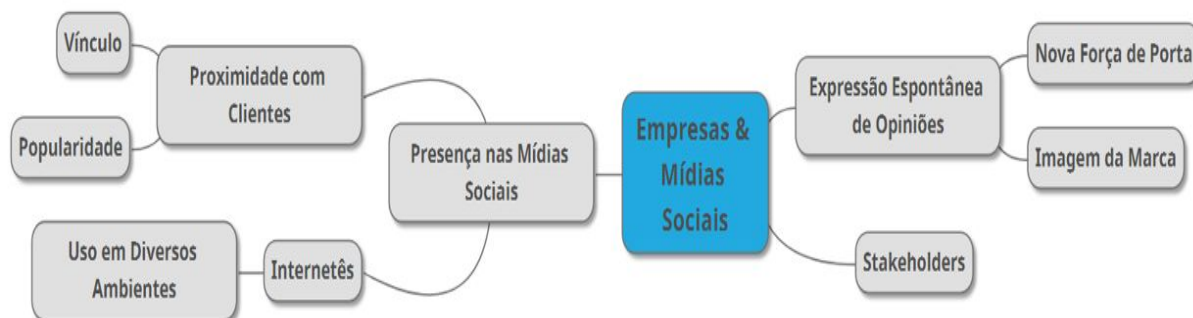
Esse conhecimento é preciosíssimo, especialmente para os negócios, pois parte essencial dos *Stakeholders*⁷ de todo negócio reside nas redes. Porém, é difícil ter uma visão ampla da reputação do seu empreendimento apenas analisando a rede como um usuário comum. É preciso ter algum mecanismo capaz de proporcionar uma análise ampla, capaz de lidar com toda complexidade e volume de informação contido no sistema de rede social, processando-o e gerando conhecimento, e é exatamente esse o tipo de sistema a ser desenvolvido neste trabalho.

⁵ Disponível em: <<https://www.youtube.com/watch?v=CMFhfr42QLg>>. Acesso em novembro de 2018.

⁶ Disponível em: <<https://emails.estadao.com.br/noticias/comportamento,pica-pau-recria-cena-classica-do-desenho-nas-cataratas-do-iguacu,70002016269>>. Acesso em novembro de 2018.

⁷ Os elementos essenciais ao planejamento estratégico de negócios (FREEMAN, 1984).

Figura 4 - mapa mental do contexto em que o trabalho está inserido



Fonte: Própria (2018).

A Figura 4 é um mapa mental que sintetiza os conceitos abordados neste subcapítulo, assim como as relações entre os conceitos. Esse mapa facilita o entendimento do conteúdo que será aprofundadas à diante.

1.2. Objetivo

A proposta é de um sistema capaz de analisar automaticamente um grande volume de informação contida na mídia social. Para efetuar essa análise, pretende-se que o *software* diferencie as sentenças que os usuários das mídias sociais publicam que estejam relacionadas ao assunto em questão, informando se cada comentário exprime uma opinião positiva, negativa ou neutra e quantificando a intensidade desse discurso (se é fervoroso ou brando). Então, deve fazer a contabilização e comparação dos dados de pesquisas e, por fim, deve gerar gráficos ilustrando o resultado da análise para o usuário.

Os coeficientes gerados por essa análise são muito significativos. Eles são capazes de mostrar uma realidade bastante concreta e muitas vezes surpreendente sobre a opinião do público. É capaz de evitar que um líder tome decisões baseadas em noções equivocadas da sua marca ou produto, o que igualmente pode prevenir a falha de um projeto pela inesperada reação negativa de seus clientes, que poderia ter sido prevista com o uso da ferramenta. Também podem ser utilizados para

proporcionar *insights* de estratégias e direções a serem exploradas baseadas nesse tipo de *feedback*⁸.

O objeto de análise do sistema pode ser a empresa usuária do aplicativo. Para a organização, as vantagens de consultar a própria reputação nas mídias sociais incluem entender os motivos que levam à execração ou à adoração da marca, tomando medidas nessa direção para então prosperar. Outras práticas benéficas advêm da comparação da reputação prévia e posterior a um acontecimento conhecido, como uma mudança ou inovação interna, medindo assim a aderência do público à mudança. Ainda, é possível, ao detectar qualquer alteração abrupta da reputação, evidenciar um evento que merece ser estudado, como nos exemplos a seguir.

- Um ataque coletivo à imagem da marca nas mídias, possibilitando a instituição a reagir às ofensas em tempo real, sem perder o *timing*⁹ para reparar os danos. Também, pode oferecer um esclarecimento do porquê de os negócios terem piorado repentinamente.
- Ou alguma contribuição externa para a reputação da companhia, proporcionando oportunidade de aprender uma maneira efetiva de promover a marca, de patrocinar esse possível aliado e de compreender o porquê dos negócios terem melhorado repentinamente.

O alvo a ser analisado pelo sistema também podem ser empresas concorrentes. Essa análise é muito interessante no sentido de conhecer a participação do concorrente nas redes e a relação disso com a sua posição na competição, além de aprender com seus erros e conquistas ao praticar o mesmo estudo que se faz com a própria instituição com a ferramenta.

Produtos também podem ser o objeto da análise, o que afetaria de forma mais palpável e imediata o negócio, pois entender que uma de suas mercadorias não agrada como esperado, enquanto outras que recebem menos investimento são

⁸ Termo utilizado na área de Administração de Empresas para um parecer sobre uma pessoa ou grupo de pessoas, revelando pontos positivos e negativos de suas ações tendo em vista a melhoria do mesmo.

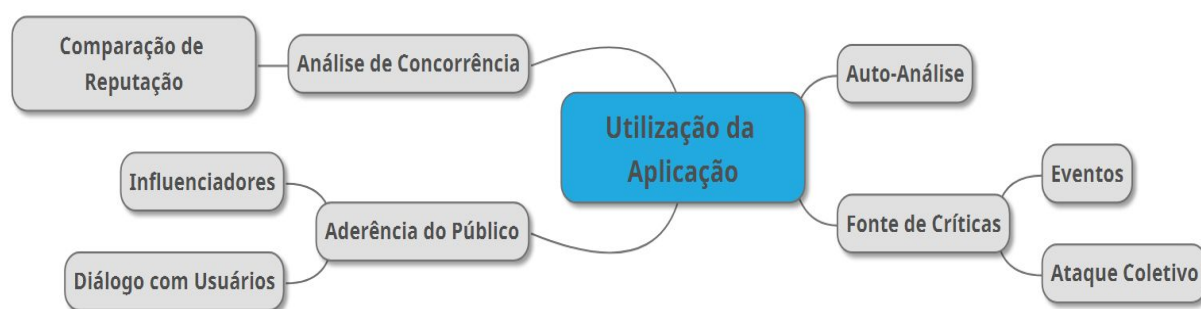
⁹ Ter sensibilidade para perceber o momento propício de realizar ou de perceber a ocorrência de algo.

mais estimadas, pode sugerir uma completa reestruturação de projetos e prioridades de negócio. Pesquisar também a adesão dos internautas por produtos que são investimentos futuros em potencial pode evidenciar uma possibilidade a ser explorada.

Uma última possível aplicação do *software*, desta vez não sendo voltada a dar suporte às empresas, se refere às possibilidades que o *software* oferece para o usuário comum, bem como possíveis implicações sociais. Um usuário comum de mídia social pode utilizar a aplicação¹⁰ para obter acesso às poucas referências a ele na rede que não estão disponíveis através da navegação comum. Qualquer usuário minimamente popular poderia usá-la para gerir sua imagem, o que é fundamental, principalmente se esse é um microempreendedor no ambiente da mídia social. Há pessoas que usam sua influência na rede para oferecer à empresas um serviço remunerado de propaganda no seu perfil da mídia social, que provavelmente será visualizado por uma grande quantidade de usuários, sendo a gestão de imagem uma prática essencial para o exercício da prestação informal de serviço.

O mais importante é que a aplicação torna essa tecnologia acessível por conta da sua gratuidade e interface amigável. Ela torna um conhecimento mais democrático para qualquer possível interessado na reputação das entidades influentes do país.

Figura 5 - mapa mental das possibilidades de utilização da ferramenta desenvolvida



Fonte: Própria (2018)

¹⁰ Termo usado como sinônimo de software e aplicativo, não de aplicabilidade.

O mapa mental da Figura 5 oferece um resumo visualmente amigável das funcionalidades da aplicação descritas anteriormente. Se trata da utilidade que a ferramenta pretende ter, os serviços que pretende oferecer aos usuários.

1.2.1. Objetivo Geral:

Construir um *software* que torne acessível as técnicas computacionais de análise de reputação nas mídias sociais para trazer transparência à reputação das entidades influentes do país ao público geral e para agregar à gestão de marcas empresariais.

1.2.2. Objetivos Específicos:

- Projetar aplicação que consulte as mídias sociais, extraindo automaticamente comentários sobre a instituição ou mercadoria em questão.
- Habilitar essa aplicação a interpretar automaticamente o conteúdo coletado, que contém linguagem natural humana.
- Disponibilizar indicadores que determinem a reputação dessa instituição ou mercadoria analisados.

1.3. Delimitações

A solução aqui proposta possui as seguintes delimitações:

- Apenas postagens públicas nas mídias sociais selecionadas para implementação serão consideradas para coleta.
- São considerados apenas os textos de postagens citando o nome da marca, na forma que foi explicitada na aplicação, ou postagens no portal da marca indicado pelo usuário, se houver.

- Os comentários coletados se restringem aos disponibilizados pela API¹¹ das Mídias Sociais selecionadas.
- A aplicação depende de acesso à internet para executar a coleta de dados nas mídias sociais.
- O processamento de linguagem natural tem seu desempenho restrito ao *framework* LeIA (descrito no capítulo 4), que não garante interpretação perfeita para todas sentenças. Há uma porcentagem de precisão variável, porém suficiente para garantir o exercício da função.
- A arquitetura elaborada suporta múltiplas Mídias Sociais, porém apenas o Twitter foi escolhido para o protótipo da aplicação.

O Twitter foi a mídia social escolhida por proporcionar a API mais acessível e abrangente dentre as mídias mais utilizadas (de acordo com o estudo de diversas APIs conduzido pelos projetistas deste trabalho), além de suas características como rede social serem particularmente interessantes para as necessidades da aplicação. Seus usuários costumam produzir diversos comentários espontâneos e sinceros diariamente, sobre os mais diversos aspectos da vida. Muitos usam essa rede como um “diário online público”, assim como um local para discutir e se informar sobre os acontecimentos mais atuais. Essa expressão descompromissada de opiniões e impressão cotidiana que reproduzem na rede é o que abastece substancialmente uma pesquisa sobre empresas e produtos.

¹¹ Sigla em inglês para “Interface de Programação de Aplicações”, ou seja, a interface que um *software* pode oferecer para que outros sistemas acessem parte dos seus dados, seguindo um conjunto de rotinas e padrões pré-estabelecidos pelo *software*

1.4. Contribuição

A aplicação resultante do projeto apresenta potencial para agregar à área de análise de dados. Ela não avança os estudos da área de inteligência artificial responsável por capacitar máquinas a interpretar sentenças com linguagem informal. Também não cria um novo método de buscas nas redes, mas maneja de maneira eficiente ambas as tecnologias, usando as melhores ferramentas *open source*¹² encontradas para disponibilizar um serviço de qualidade que torna acessível a análise de reputação.

Embora existam outros programas e bibliotecas que proporcionem um serviço parecido com o desse projeto, são menos acessíveis nos quesitos do custo financeiro e simplicidade de ser manuseados e instalados. Quando não, não são tão completos e performáticos. Eis os resultados do estudo efetuado pelo grupo sobre as ferramentas mais populares com propostas semelhantes à aplicação do projeto.

Os exemplos das ferramentas mais robustas encontradas foram a NetBase e Stilingue. Ambas abrangem todas as funcionalidades da presente aplicação, bem como tantas outras. No entanto, são pagas e os valores são acessíveis apenas para empresas de maior porte. No momento em que foi analisada para esse trabalho, foi constatado que a NetBase cobra uma mensalidade fixa de custo elevado para disponibilizar acesso total às suas funcionalidades, enquanto o Stilingue oferece pagamento equivalente à utilização da ferramenta, ultrapassando o preço estipulado pela NetBase caso o usuário utilize a grande parte de suas funções constantemente.

Exemplos de outros programas *open source* com funcionalidades parecidas com o deste projeto, encontrados numa busca superficial pela rede, assim como um usuário comum buscaria, foram “Zoho Social”, “Kaput”, “Steamcrab”, “Meaningcloud” e “SocialHarvest”. No entanto, todos, menos o último citado, não contemplam todas funcionalidades propostas pela aplicação ou simplesmente são menos

¹² Código aberto. diz respeito ao código-fonte de um software que está disponível gratuitamente e pode ser utilizado e adaptado legalmente.

performáticos. Já a ferramenta “SocialHarvest” se parece com a proposta neste trabalho, porém tem uma maiores limitações para sua instalação.

A ferramenta desenvolvida no projeto é *open source* e de fácil instalação e manuseio, trazendo uma democratização dessa tecnologia. Ela se dedica em abordar do melhor jeito encontrado todas funcionalidades necessárias para entregar informação correta, relevante e de fácil interpretação para o usuário.

1.5. Método de Trabalho

Segue a descrição da metodologia adotada para a produção do trabalho aqui posto:

1. **Concepção do Tema:** Conforme as diretrizes mais usuais deste trabalho e os interesses do grupo, a tipo pesquisa foi definido como Pesquisa-Ação. Foi identificado, através dos próprios conhecimentos e experiências dos participantes, uma problemática da sociedade a ser explorada e solucionada através de tecnologia. Assim, foi idealizada a intervenção que dá nome ao trabalho.
2. **Levantamento da Bibliografia:** Após estabelecidas as diretrizes iniciais do trabalho, foram levantados diversos recursos bibliográficos (artigos, livros, outros trabalhos de conclusão de curso e projetos semelhantes) acessíveis ao grupo para verificar a viabilidade do projeto, garantir a fundamentação sólida dos conceitos relacionados ao tema e proporcionar referências para uma escrita e projeção de ferramenta bem embasadas.
3. **Leitura da Bibliografia:** Os integrantes estudaram os princípios de Processamento de Linguagem Natural, Mídias Sociais, Reputação, Análise de Dados, Liderança nos Negócios e Marketing, contidos na literatura obtida, além de técnicas, ferramentas e *frameworks* relacionados para agregar ao projeto de aplicação e à própria formação acadêmica.

4. **Concepção do Texto:** Os conceitos aprendidos foram relacionados e estruturados em um texto coeso e articulado perante o tema, tendo em vista o formato convencionado do documento aqui posto (explicado em Organização Textual, próximo subcapítulo). Os textos contêm todo embasamento teórico do trabalho e determinação das especificações funcionais da aplicação.
5. **Análise da Aplicação:** Início do estudo da solução do problema através da elaboração de diagramas e textos descritivos associados a eles.
6. **Desenvolvimento do Protótipo:** A implementação em si do *software*, contendo todas as fases de desenvolvimento, validação e documentação.
7. **Conclusão:** Terminada a etapa anterior, são constatados os resultados obtidos no projeto e considerados os possíveis acréscimos e planejamentos para a continuação dele.

1.6. Organização do Texto

Este trabalho está organizado nos seguintes capítulos:

- Capítulo 1: descrição dos motivos que levaram à escolha de desenvolver esse projeto, por meio da análise da importância de ter conhecimento da reputação de sua marca e de seus concorrentes nas mídias sociais. Além disso, esse capítulo contempla os objetivos, delimitação e contribuição esperada do projeto.

- Capítulo 2: descrição da base conceitual que sustenta o desenvolvimento deste trabalho, por meio da análise de artigos, livros envolvendo os temas de reputação nas mídias sociais e processamento de linguagem natural e análise de trabalhos relacionados.

- Capítulo 3: descrição por escrito da aplicação ideal a ser desenvolvida, assim como a apresentação do método utilizado para efetuar a fase de análise deste produto e os diagramas gerados no processo.

- Capítulo 4: descrição da modelagem efetuada, baseada no método ICONIX, a ser descrito também. Todo desenvolvimento arquitetural desse capítulo se baseia na análise posta no capítulo passado.

- Capítulo 5: descrição dos resultados alcançados, de trabalhos futuros a serem feitos para agregar ao produto gerado e conclusão do trabalho como um todo.

2. REVISÃO BIBLIOGRÁFICA

Este capítulo tem como finalidade a descrição da base conceitual do trabalho, abordando seu fundamentos teóricos e analisando trabalhos relacionados. Na fundamentação teórica, temas como processamento de linguagem natural e reputação em mídias sociais são descritos, aplicando-os ao escopo do trabalho. Em trabalhos relacionados, há um resumo de outros trabalhos com temas paralelos e a descrição de como serviram de referência.

2.1. Fundamentação Teórica

Neste capítulo, as bases conceituais que guiaram esse projeto são apresentadas. Dentre elas, processamento de linguagem natural, mídias sociais, consumo contemporâneo e reputação são os principais tópicos abordados.

De acordo com Steven Birds, em processamento de linguagem natural, “linguagem natural” significa a linguagem utilizada no cotidiano de comunidades humanas, como Inglês e Português (Birds, 2009). Isso diverge de linguagens artificiais, como linguagens de programação e notações matemáticas. A característica da linguagem natural de estar em constante mutação torna difícil definir regras explícitas de interpretação. Assim, em termos gerais, Processamento de Linguagem Natural significa qualquer manipulação computacional da linguagem natural. Isso abrange desde simplesmente contar a frequência que certas palavras são usadas, com o intuito de detectar diferentes estilos de escrita, até algum nível de entendimento das expressões humanas para promover análises úteis (Birds, 2009).

O processamento de linguagem natural é uma aplicação da área de conhecimento da computação que estuda os problemas da geração e compreensão automática da linguagem humana natural. Uma vertente dele são os sistemas de

geração de linguagem natural, que convertem dados computacionais em linguagem compreensível ao ser humano. Os aplicativos de GPS são um exemplo de aplicação que utiliza essa tecnologia, pois para produzir comandos de voz que guiarão seu cliente, precisam transformar dados da rota em linguagem natural.

Já sistemas de compreensão de linguagem natural representam linguagem humana em dados estruturados e categorizados para o processamento computacional, como é o caso do software desenvolvido neste trabalho. Outro exemplo de aplicações desse tipo são as aplicações de propagandas do Google¹³, que processam os conteúdos acessados e as palavras que os usuários acabam por disponibilizar para o sistema durante a utilização da plataforma para identificar os produtos que seus usuários poderiam se interessar para determinar quais anúncios cada usuário irá receber.

Um último exemplo de aplicação é esta que utiliza geração de linguagem natural e processamento de linguagem natural, os *Chatbots*. São programas de computador que tentam simular um ser humano em uma conversação com uma pessoa. Eles são utilizados em diversos sites prestadores de serviço atualmente, principalmente na seção de suporte e dúvidas do cliente, onde precisam interpretar a linguagem natural que esse cliente está utilizando para entender a ajuda que ele precisa, consultar seu sistema para encontrar dados relevantes para a situação dele e transformar os dados em um texto coeso que será enviado para o cliente.

A concordância permite a análise de palavras em contextos. O autor afirma que, ao listar as palavras relevantes que se relacionam a um termo, é possível extrair o sentido desse termo dentro do contexto inserido (Birds, 2009). Assim, um mesmo termo pode gerar diferentes interpretações para diferentes textos, podendo ter conotações positivas ou negativas.

Como definido por Birds, o “*token*” é um nome técnico para uma sequência de caracteres tratada como um grupo (Birds, 2009). A contagem de tokens de um texto é a contagem de ocorrências dessas sequências. Assim, o processamento de

¹³ Referência: <<https://blog.mateada.com/links-patrocinados/tudo-sobre-o-google-ads>>. Acesso em novembro de 2018.

linguagem natural tem como alguns dos seus principais desafios resumidos em “tokenização” e desambiguação do sentido das palavras (Birds, 2009).

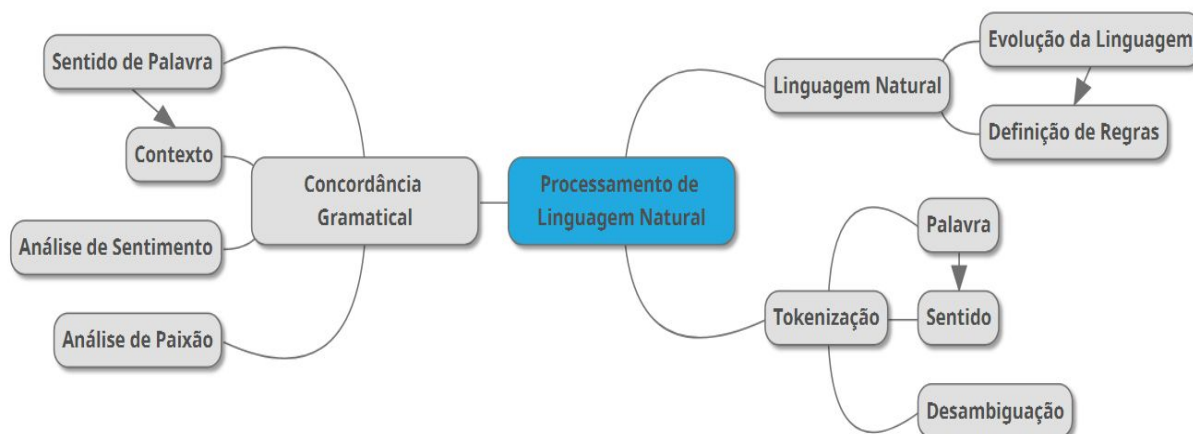
A “tokenização” é o processo de, a partir de frases de um texto, secciona-lo em menores partes. Assim, tanto as palavras menos importantes, que dão menos sentido ao texto, quanto as terminações das palavras são removidas. O texto que sobra é separado então em pares ordenados, que relacionam cada palavra relevante com uma indicação do sentido dela na frase. Por fim, essa relação é estruturada em n-gramas.

A desambiguação do sentido das palavras é o processo de, a partir de uma ambiguidade no sentido de palavras em uma frase, conseguir identificar a relação das palavras e caracterizá-las, através da “tokenização”, para então definir qual é o sentido das palavras no contexto inserido.

O processo de aprendizado de máquina abordado no processamento de linguagem natural segue os seguintes passos:

1. Receber texto e requisição de classificação.
2. Representar os dados usando atributos numéricos.
3. Usar o algoritmo específico para a classificação selecionada, retornando o modelo requisitado.

Figura 6 - mapa mental do conteúdo de processamento de linguagem natural



Fonte: Própria (2018).

O mapa mental da Figura 6 oferece um resumo visualmente amigável dos conceitos envolvidos no processamento de linguagem natural.

Os conceitos abordados definem a base conceitual para o processamento de linguagem natural. A seguir, será apresentado uma reflexão sobre as mídias sociais, sua influência no mundo, na vida das pessoas e principalmente nos negócios.

Para se ter uma análise profunda sobre as mídias sociais, é preciso começar pelo seu componente mais central; a humanidade contemporânea. Para Motta, a principal característica dela provém da relativização das estruturas básicas que suportavam a sociedade até então, como família, nação e espiritualidade (Motta, 2017). A adoção, estabelecimento e não discussão desses valores fundamentais pelas gerações passadas manteve uma mentalidade comum entre as pessoas, pouco se questionando sobre outro sentido da vida. Então, com a crítica e relativização de todos valores pré-estabelecidos nos tempos modernos, tornamos seres inconscientemente frustrados em busca do que signifique sua vida. Porém, essa busca de um significado transformou-se em uma jornada egocêntrica, com objetivo de satisfazer o ego e o anseio de pertencimento a uma tribo que vá nos definir e diferenciar da massa. A consequência disso se apresenta no “consumo aspiracional” (sede de consumir algo pouco acessível e muito estimado pelo seu grupo) de bens e serviços, junto a sua ostentação em locais públicos, principalmente nas redes sociais (Motta, 2017).

O autor define que as mídias sociais são esse espaço de relacionamento interpessoal limitado pelo interesse de estar conectado, onde o indivíduo concretiza as suas aspirações (e muitas vezes carências) pessoais através do coletivo, se expressando, se promovendo e interagindo com semelhantes e estranhos nas redes sociais, com o intuito de afirmar-se alguém dentro do grupo social pertencido ao ganhar visibilidade e atenção de outros (Motta, 2017).

Entender as forças que ditam o consumo atualmente é fundamental para as empresas prosperarem. O consumo tem sido cada vez mais influenciado por dois aspectos da natureza do humano contemporâneo e, conseqüentemente, pelas redes sociais também.

Por um lado, o consumo é tribal, regido pelo anseio de pertencimento ao grupo dos indivíduos hedonistas de redes sociais. O consumidor vive um momento hedonista, no sentido de cada relação e atividade da sua vida deve atender às necessidades e expectativas pessoais (Motta, 2017). Todavia, essas necessidades e expectativas pessoais são ditadas pela tribo que o ser se vê pertencente. Logo, o consumo é ditado pelos prazeres que são mais estimados pelo nicho social do indivíduo, o que gera o consumo aspiracional citado anteriormente.

Por outro lado, com a disseminação do conhecimento e um certo nível de consciência coletiva da sociedade globalizada e interligada, há no indivíduo uma intenção de ser mais engajado e racional para ser mais inteligente e assim ser bem-sucedido na vida e garantidor de seus privilégios e prazeres (Motta, 2017). Tal comportamento ativa um senso crítico constante que vai fazê-lo contrapor com o que se tem estabelecido, gerando um rompimento com grupos de ideias opostas que encontra, o que o exclui de algumas tribos, fazendo se verem como opostos e afirmando-o nas tribos que têm confluência de ideias. É importante para uma instituição não ser vista como aliada aos conceitos das tribos opostas à dos seus consumidores e, também, ser vista como integrante da tribo dos seus consumidores, mostrando-se engajada nos costumes, valores e ideias que as identificam. Ainda, é preciso manter cautela e atenção aos ativistas de redes sociais; pessoas com prestígio e influência que conseguem representar grupos de pessoas e colocar a opinião dessas à favor ou contra algo ou alguém a partir da expressão das próprias opiniões.

Figura 7 - Propaganda da empresa Burger King contra discursos de ódio nas redes



Fonte: Twitter¹⁴. Edição na união das imagens de duas partes do vídeo, postas lado a lado e separadas por um traço negro, realizada pelos membros do projeto.

A Figura 7 contém duas partes de um vídeo de propaganda da empresa Burger King que utiliza a estratégia citada anteriormente, de demonstrar engajamento nos valores e ideias que identificam-se uma parcela interessante dos seus clientes. A empresa pode estar tentando ganhar aderência desses clientes por diversos motivos. À esquerda na Figura 7, é ilustrado um comentário com opinião intensa e negativa em relação ao estabelecimento. Diz “Odeio burger King mas odeio mesmo só o cheiro enjoa”. Colocar um comentário do tipo na própria propaganda pode arriscar a qualidade da sua reputação, mas fazem isso para fortalecer seu argumento e o impacto da propaganda. Deixam claro que, mesmo sendo depreciativa, é uma opinião válida. Porém, o conjunto de postagens ilustradas à direita na Figura 7 não são consideráveis como opinião, mas sim expressão de preconceito, como distinguem no título do vídeo e nas imagens restantes. A propaganda ensina a diferenciação de opinião negativa e preconceito em um vídeo simples e se posicionam nessa questão ao mesmo tempo.

As mídias sociais mostram-se, então, um componente fundamental da vida moderna. O fenômeno analisado não se restringe às novas gerações, embora seja

¹⁴ Disponível em: <<https://twitter.com/BurgerKingBR/status/1060622993104273409>>. Acesso em novembro de 2018.

mais natural a elas questionarem o status quo (Motta, 2017). Até os idosos permeiam também as redes e indagam como deverão viver suas próximas décadas. É válido considerar que o público das mídias sociais abrange a maioria da sociedade, e é essa relevância e influência que possui que é ilustrada pelos dados a seguir.

Dados de uso de redes sociais por volume de informação consumido em alguns veículos:

- 700 bilhões de minutos são gastos por mês no Facebook;
- 3 mil fotos são publicadas por minuto no Flickr;
- 190 milhões de mensagens são enviadas por dia no Twitter;
- 2,9 bilhões de horas são dedicadas por ano aos vídeos publicados no YouTube.

(Motta, 2017).

Estima-se que 20% da população mundial seja usuária de redes sociais. Desses, 58% utilizam primordialmente o Facebook, sendo que 46% buscam interagir com empresas para resolver seus problemas com produtos e serviços (É interessante observar que 22% dos consumidores esperam uma resposta no mesmo dia e 29% esperam uma resposta em até 2 horas! E no Twitter, o nível de impaciência é ainda maior com 30% dos consumidores que aguardam uma resposta em até 30 minutos para suas reclamações) (Motta, 2017).

Adolescentes são ainda mais ativos nas redes sociais. 75% dos adolescentes norte-americanos têm perfil em alguma rede social, 68% enviam mensagens de texto todos os dias, 41% consideram-se viciados em telefone celular, 20% consideram-se viciados em redes sociais e 43% gostariam (mas não conseguem) de se desconectar de vez em quando (Motta, 2017).

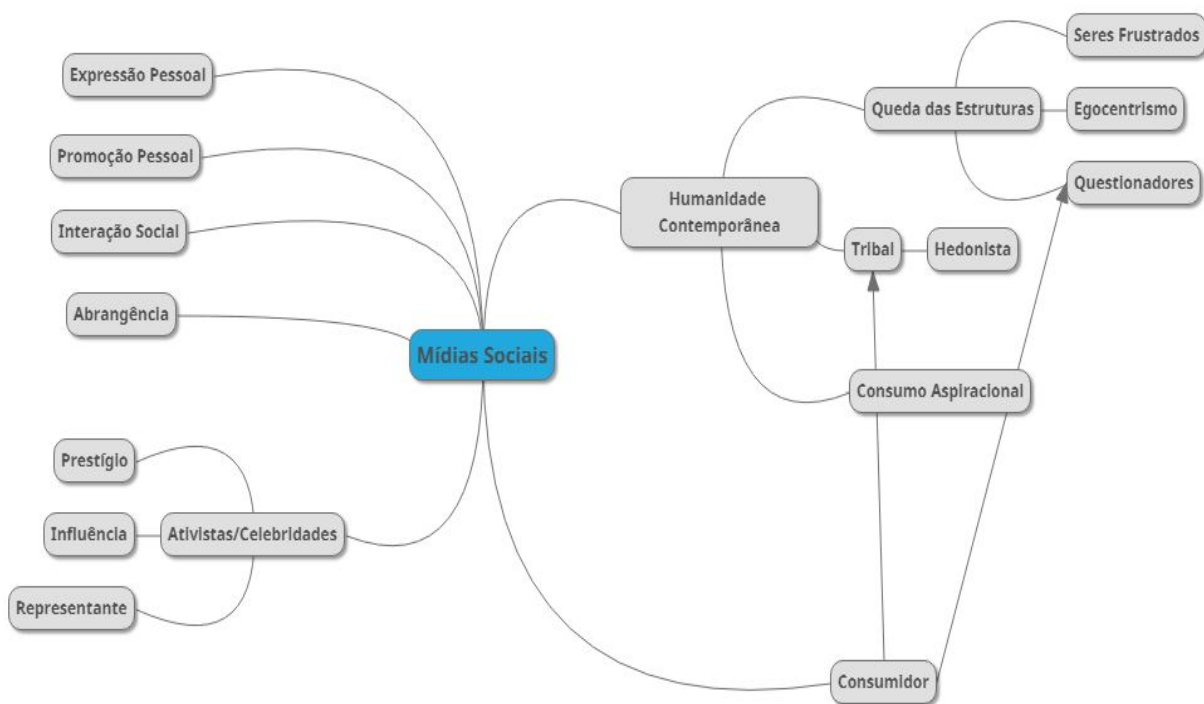
O impacto desse ambiente conectado sobre as organizações: Consumidores socialmente digitais não prestam mais atenção em mensagens corporativas do que são influenciados pela opinião de seus amigos. Um consumidor socialmente digital compartilha suas opiniões sobre uma empresa com aproximadamente 42 pessoas, 88% das quais serão significativamente influenciadas. Insta observar que apenas

40% das empresas dispõem-se a interagir com seus clientes nas redes sociais, embora consumidores engajados nas redes sociais tendem a gastar entre 20% e 40% acima da média de consumo com determinado produto (Motta, 2017).

A conclusão diante os dados é simples: as empresas precisam participar das redes sociais. Não investir em um perfil nas redes para se relacionar diretamente com seu cliente e não aproveitar as informações que as pessoas disponibilizam espontaneamente é perder as grandes possibilidades de mercado que os avanços tecnológicos deste século proporcionam. É perda de dinheiro.

Em suma, é por esse ambiente das mídias sociais ser extremamente essencial e pessoal para a vivência da humanidade como um todo que é o canal ideal para se relacionar com elas. Através dele, se obtém o poder de acessar o íntimo da vida das pessoas para fazer sua marca participar dela e conseguir informações essenciais para um negócio. Ainda, com o produto tecnológico desenvolvida no projeto, é possível superar a limitação das relações consensuais nas mídias sociais (variável definidora das mídias, previamente citada) para se ter acesso a todas informações que dizem respeito ao objeto pesquisado e processá-las automaticamente.

Figura 8 - mapa mental do conteúdo de mídias sociais



Fonte: Própria (2018).

O mapa mental da Figura 6 oferece um resumo visualmente amigável dos conceitos que envolvem o tema mídias sociais.

Tamanho dimensão das mídias sociais se mostra inegável. As possibilidades de sua aplicação nos negócios são vastas. A seguir, serão descritos os motivos da análise de reputação ter sido elegida como a nossa aplicação das tecnologias, citadas anteriormente, no contexto das mídias sociais.

Para o uso do termo “reputação” e estudo desse conceito, é fundamental o entendimento do seu significado e dos outros conceitos que o compõem. Para lasbeck, ela é a imagem consolidada de algo, baseada em argumentos lógicos e convicções, enquanto essa imagem é uma construção mental gerada pelo conjunto de impressões, dos mais diversos tipos, daquele que se depara com o discurso (lasbeck, 2007)..

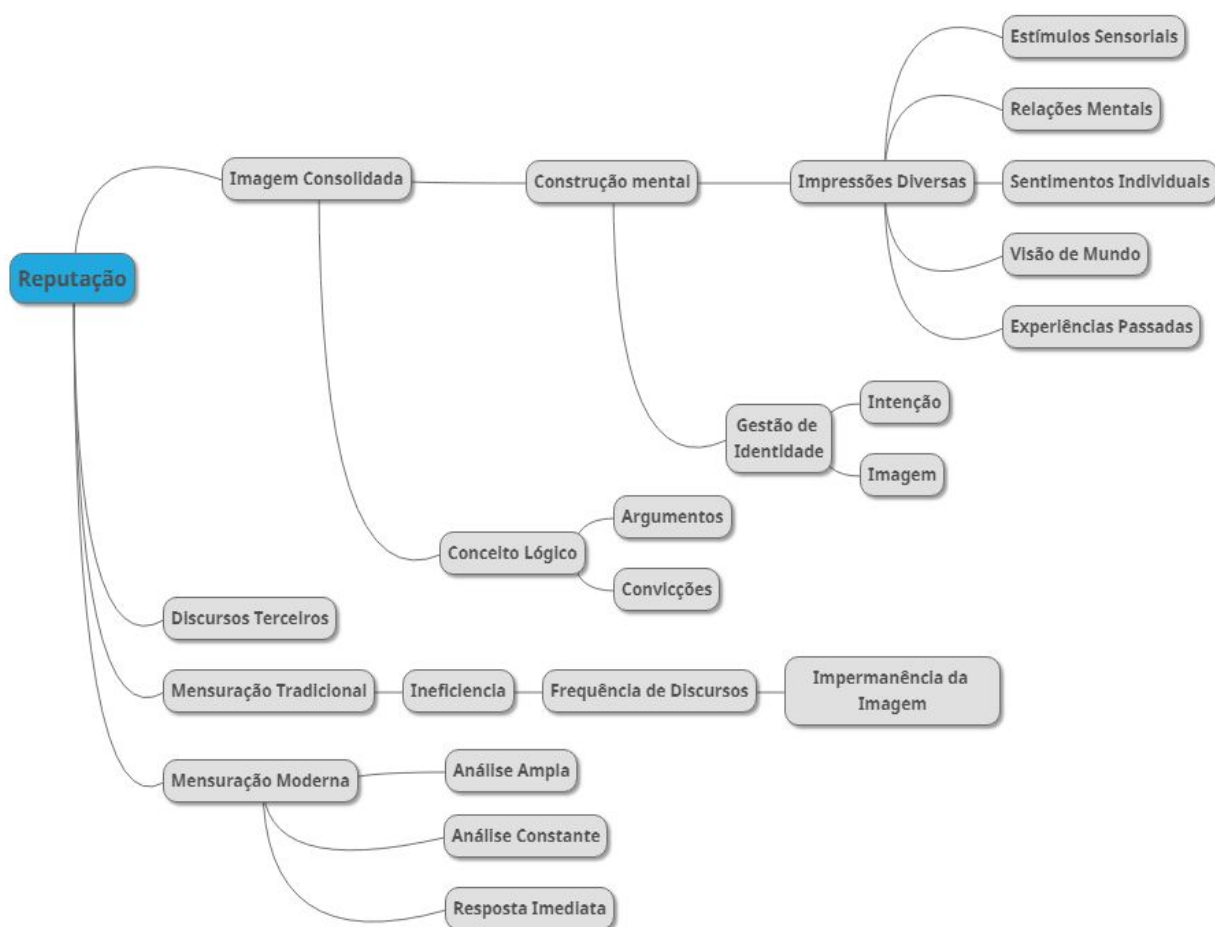
A natureza das impressões são vastas, baseando-se em todos estímulos sensoriais envolvidos no contexto e nas relações mentais envolvidas na situação,

passando sempre pelos sentimentos do indivíduo, pela sua visão de mundo e até alguma experiência pessoal sua com a empresa em questão (lasbeck, 2007). Assim, a atividade de manipular as impressões a serem passadas para os clientes é extremamente complexa. Tal complexidade se deve ao fato de que a maioria dos discursos que interagem com o sujeito perante a marca não provém exclusivamente da empresa, mas de indivíduos terceiros que ele se relacione ocasionalmente. Porém, essa tarefa não é impossível, uma vez que a imagem que outros passam para o cliente vai ser, em algum ponto, fortemente baseadas na imagem que a própria empresa passa, sendo o discurso dela um dos mais relevantes (lasbeck, 2007).

A gestão da identidade, ramo que o produto do projeto aqui proposto pretende oferecer suporte, consiste, então, em tornar compatíveis as intenções do discurso com a imagem que pode gerar, com o intuito de consolidar uma reputação positiva e particularmente interessante ao público-alvo. Essa gestão inclui manejar as sensações e ideias que o discurso estimula e a situação que ele está inserido, (lasbeck, 2007).

O autor ainda evidencia que maneiras tradicionais de medir a reputação perderam sua eficiência nos tempos modernos. Com o avanço das tecnologias de informação, ocorreu o aumento da frequência que as pessoas recebem discursos sobre as empresas, tornando-se mais impermanente e imprevisível sua imagem da instituição e mais difícil de entender a reputação empresarial que agora oscila a todo momento, principalmente ao tentar utilizar-se de técnicas antigas de medição de reputação, como levantamento de questionário (a abordagem de pessoas em público com questionários sobre a sua opinião em relação à marca). Mostra-se urgente a implementação de métodos de análise ampla, constante e de resposta imediata para o um mapeamento da imagem organizacional que vá subsidiar uma reputação sustentável (lasbeck, 2007), como o da tecnologia aqui proposta.

Figura 9 - mapa mental do conceito de reputação



Fonte: Própria (2018)

2.2. Trabalhos Relacionados

A seguir, a descrição dos dois trabalhos escolhidos como referências práticas para o projeto aqui posto. O nome de cada subcapítulo equivale ao do projeto descrito. Eles foram selecionados por agregarem como base do desenvolvimento do sistema e serão descritos brevemente.

2.2.1. Mineração de Opiniões Aplicada a Mídias Sociais

A dissertação de mestrado que dá nome a esse subcapítulo (SOUZA, 2012) se detém a desenvolver uma ferramenta que coleta comentários dos brasileiros contidos na mídia social Twitter e processa seu conteúdo automaticamente, dispondo ao usuário dados de opinião relacionados à amostra sem passarem por um tratamento de dados que tornaria mais fácil sua interpretação. A informação gerada pode ser manipulado para auxiliar em atividades diversas na área da Inteligência Competitiva, área essa que o próprio trabalho define como o processo de monitorar o ambiente competitivo de negócio.

Na época em que foi projetado, ainda não era tão disseminado o estudo de processamento de linguagem natural. Não haviam tantos projetos na área para suportá-lo e não foi possível encontrar frameworks com a função de interpretar textos em português (parametrizando-os) que fosse independente de domínio (com os dados disponíveis ao público). Foi feito, então, o desenvolvimento do próprio, baseado em conteúdo estrangeiro. Hoje, há diversos *frameworks open source* que executam essa função, como o que usamos.

As partes de desenvolvimento do sistema foram destrinchadas pelo autor em 5 etapas, descritas a seguir.

A primeira foi o desenvolvimento de um *framework* que ordena a execução das diferentes funções do processo. É um controlador de todos processos são executados no sistema.

A segunda é um léxico de opiniões, parte fundamental para a análise de sentimentos, que potencializa a capacidade de identificação de expressões. Foram coletadas mais de 10.000 expressões da língua portuguesa, assim como um dicionário de *emojicons* e *hashtags* (metadados de mídias sociais que podem determinar o contexto de uma frase ou mensagem inteira ou ainda alterá-lo por completo, sugerindo ironia, por exemplo), com seus respectivos valores e métricas. Para sua construção, foram extraídas e analisadas frases de resenhas de filmes junto à inserção de sentido, feita manualmente.

A terceira parte foi um conjunto de heurísticas para normatização dos textos, devido à natureza da escrita no Twitter. A heurística é a associação de palavras ou

frases com grafias diferentes porém mesmo sentido. Boa parte dos casos se originam de erros ortográficos, abreviações e outras variações linguísticas comuns no contexto das redes sociais. Para sua criação, foi desenvolvida uma variação para a língua portuguesa do algoritmo Metaphone (funcional para a língua inglesa) que compara palavras consideradas semelhantes por terem as mesmas consoantes nas mesmas posições. Entretanto, teve seu uso limitado pelo fato do algoritmo ter alta taxa de erro.

A quarta foi a parte de Mineração de Opinião, que é a abordagem utilizada para desenvolver a Análise de Sentimentos. Sobre a modelagem para identificação de diversas ambiguidades no escopo da negação, que tem forte ligação com o contexto sintático e discursivo, foram testados então dois tipos de escopos mais simples. O termo de negação não necessariamente deve aparecer adjacente à locução negada, porém mais usualmente o termo da negação se apresenta dessa forma, vizinho à locução que nega. Como a modelagem da negação teve pouco impacto nos testes executados, foi decidido se ater à modelagem baseada em vizinhança por acreditar-se que essa modela melhor o comportamento “padrão”.

E a quinta, e última, é o Reconhecimento de Entidades Nomeadas, que se resume em reconhecer que entidade, conceito ou ideia se refere a frase analisada, relacionando o sentido de suas palavras, para então aliar essa entidade às outras e gerar um valor, através de dados numéricos. É a execução em si da interpretação de textos pela máquina, uma vez que todos subsídios para a interpretação de cada palavra e de expressões específicas já foram desenvolvidos e estão disponíveis.

Tanto este trabalho quanto o nosso propõem a extração de dados de mídias sociais e sua interpretação automática (através do processamento de linguagem natural) para dar suporte às atividades empresariais. Entretanto, boa parte dos esforços do autor foram destinados a viabilizar o uso dessa tecnologia para a língua portuguesa, não sendo aproveitados neste trabalho por atualmente existirem conteúdos de uso livre que prestam esse serviço. Mesmo assim, o restante do seu desenvolvimento, que ainda é maioria do projeto, entra em confluência com o desenvolvimento do nosso trabalho.

2.2.2 Visibilidade de Notícias no Twitter e no Facebook

O artigo (ZAGO, 2013) identifica e compara as notícias com maior repercussão nas mídias sociais Twitter e Facebook com o intuito de realizar um estudo social dos seus usuários em cinco países. Foram escolhidos os 8 jornais de maior circulação desses países, sendo eles Brasil, Alemanha, Espanha, Estados Unidos e Reino Unido.

Durante as duas primeiras semanas de outubro de 2012, foram coletadas as postagens¹⁵ no perfil oficial dos próprios jornais no Facebook e Twitter e as postagens de usuários do Twitter contendo o *link*¹⁶ de reportagens, considerando apenas o conteúdo das 25 mensagens mais replicadas em cada rede social, para cada jornal.

A execução da coleta de dados no Twitter e no Facebook ocorreu através das suas próprias APIs. A restrição, mencionada anteriormente, de não serem utilizadas postagens de terceiros que contém *link* de notícia no Facebook se deu às limitações impostas pela API dessa mídia social.

Os critérios para determinar a visibilidade das notícias, e sua consequente relevância em cada país, foram o volume de “retuïtes”, “compartilhadas” e “curtidas” (os termos se referem às interações que os usuários das mídias em questão podem ter com postagens, demonstrando interesse do usuário e replicando o conteúdo com sua rede social). Foi associado um entre os seguintes tipos de tema de matéria para viabilizar a contagem e comparação dos conteúdos dessas: política, economia, notícias locais ou nacionais, entretenimento, opinião e esportes. Os resultados numéricos foram ilustrados em tabelas de porcentagem de replicação de países por seus 3 principais tópicos e gráficos de proporção relativa dos assuntos mais replicados, como mostram as Tabelas 1 e 2.

¹⁵ Conteúdo divulgado em uma mídia social por um de seus usuário

¹⁶ Termo usual para hiperligação, ou seja, uma palavra, texto ou imagem que quando é clicada pelo usuário, o encaminha para outra página na internet, que pode conter outros textos ou imagens.

Tabela 1 - Sessão das notícias mais replicadas no Twitter

País	Tópico 1	Tópico 2	Tópico 3
Alemanha	Política (28,63%)	Economia (14,09%)	Notícias locais (8,63%) e Nacionais (8,63%)
Brasil	Política (29,05%)	Notícias locais (15,24%)	Entretenimento (14,76%)
Espanha	Política (20,4%)	Notícias locais (13,43%)	Economia (11,94%)
Estados Unidos	Opinião (12,93%)	Entretenimento (12,43%)	Política (9,45%)
Reino Unido	Esportes (17,7%)	Opinião (12,92%)	Nacional (10,52%)

Fonte:(ZAGO, 2013).

Tabela 2 - Sessão das notícias mais replicadas no Facebook

País	Tópico 1	Tópico 2	Tópico 3
Alemanha	Entretenimento (21,18%)	Cultura (15,27%)	Política (14,28%)
Brasil	Política (20,45%)	Notícias locais (18,75%)	Cultura (17,04%)
Espanha	Notícias locais (13,79%)	Cultura (13,10%)	Política (11,72%)
Estados Unidos	Política (22,22%)	Notícias locais (16,62%)	Esportes (12,12%)
Reino Unido	Opinião (15,07%)	Nacional (12,56%)	Política (9,09%)

Fonte:(ZAGO, 2013).

A equipe ainda contrapõe os números obtidos com a análise da situação socioeconômica e política de cada país durante o período em questão, pois constata que a última é uma variável que altera de forma nítida a consciência coletiva e os assuntos mais pertinentes.

Os dados apontaram que em países como Alemanha, Brasil e Espanha, “Política” foi o assunto mais replicado no Twitter. “Entretenimento” e “Cultura” aparecem dentre os assuntos mais curtidos no Facebook em parte significativa dos jornais analisados. As eleições representam 11% dos tópicos dentre os 25 *tweets*

mais replicados em cada jornal, emergindo como o assunto mais repercutido no Twitter no período, devido ao fato que três dos cinco países analisados atravessarem período eleitoral (eleições municipais e regionais no Brasil e na Espanha e eleição nacional nos Estados Unidos), com um predomínio de notícias sobre Economia na Espanha e Política no Brasil, Estados Unidos e Espanha. Foram percebidos padrões diferentes de temas mais relevantes para cada país, porém uma equivalência na cobertura de imprensa entre todos países para “escândalos”, o que contribuiu substancialmente para o estudo social do artigo. Os resultados ainda ofereceram uma confirmação do senso comum generalizado perante o perfil dos usuários de cada rede, onde o Facebook é utilizado majoritariamente para acessar conteúdos ligados ao entretenimento e curiosidades, já o Twitter para a cobertura e de eventos políticos, econômicos e sociais.

Ambos, o artigo e o trabalho aqui expostos, visam entender opiniões e mentalidades da sociedade ao coletar as expressões espontâneas das pessoas nas mídias sociais e então analisá-las. Chegam a usar a mesma tecnologia para acessar o conteúdo das redes (acesso à API), mas se diferem na maneira de processar esse conteúdo, com volumetria¹⁷, eficiência e acurácia distintas devido às técnicas utilizadas. Os próprios autores fizeram manualmente a interpretação e categorização de sua amostra para conseguirem captar os assuntos discutidos sem utilizarem qualquer tipo de inteligência artificial, o que restringe o volume de dados e demanda o demorado processo de análise humana mas garante maior precisão na classificação do conteúdo. Analisar os dados dos compartilhamento das notícias mais relevantes nas mídias é uma técnica bastante inteligente para se medir, com uso restrito de tecnologia, os assuntos mais relevantes em cada país e rede social, porém muito limitada em comparação às possibilidades que a tecnologia do sistema desenvolvido neste projeto propõe.

Entretanto, ambos os trabalhos concluem seus objetivos por meio de um estudo de métricas¹⁸, contagem e comparação amostral inteligente. O desenvolvimento da ferramenta proposta neste projeto precisa se basear em

¹⁷ No sentido de quantidade de elementos distintos.

¹⁸ Sentido de medições.

conhecimentos sólidos e referências de como associar, diferenciar e dispor os números com fim de se obter conclusões verídicas. Ainda, se diferem na finalidade da abordagem (um sendo aplicação para marcas e empresas e o outro, pesquisa social), o que pode gerar estranheza, à primeira vista, na associação dos dois trabalhos, mas com a correta interpretação da composição de cada um (como foi posto por meio deste), é evidente sua relação.

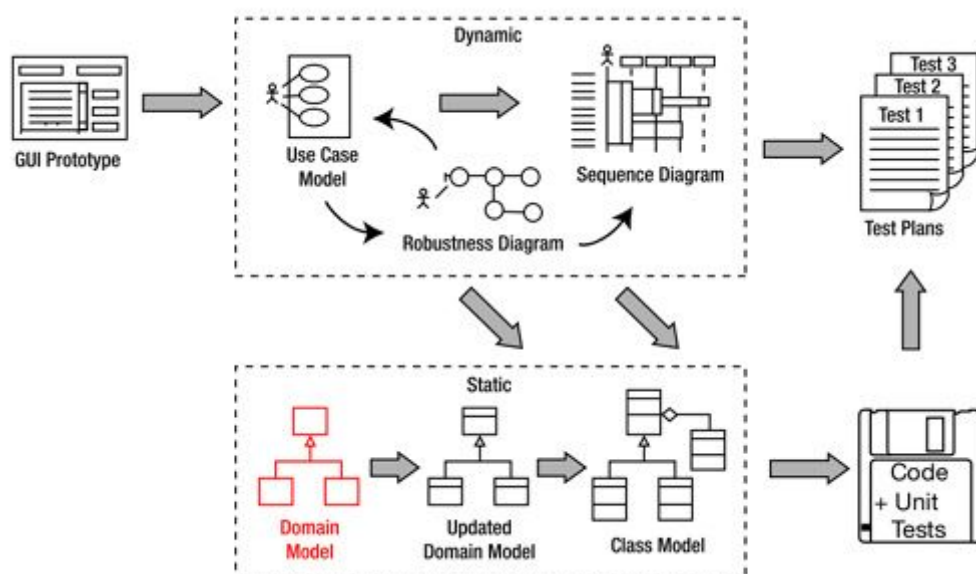
Por fim, o trabalho relacionado anterior (SOUZA, 2012), que desenvolve um programa parecido com o que propomos, se atenta a descrever as tecnologias usadas e a maneira de captar os dados e processá-los automaticamente, mas carece em maneiras de manipular seus dados e evidenciar informações valiosas para os interesses do usuário, o que é, por sua vez, o conteúdo majoritário do artigo (ZAGO, 2013) aqui posto. Unidos, oferecem a instrução necessária para a implementação do projeto.

3. DESENVOLVIMENTO DO PROTÓTIPO

No capítulo de desenvolvimento do protótipo, o processo de modelagem dos diagramas será descrito, com a finalidade de elaborar os componentes do *software*. Após a elaboração dos modelos, o código do programa pode ser produzido, baseando-se nos modelos e diagramas.

A metodologia de desenvolvimento de *software* utilizada foi o ICONIX, que indica o seguinte fluxo de diagramação, ilustrado pela Figura 10:

Figura 10 - processo de diagramação da metodologia ICONIX



Fonte: TheRegister¹⁹.

- Diagrama de Casos de Uso (Use Case Model): Levantamento das funcionalidades do sistema e dos agentes externos ao sistema que comunicam com ele. É um diagrama que descreve a estrutura do *software* no momento dinâmico
- Modelo de Domínio (Domain Model): Elaboração dos componentes de domínio (entidades) do sistema e de suas relações. É um diagrama que descreve a estrutura do *software* no momento estático.

¹⁹ Disponível em: <https://www.theregister.co.uk/Print/2007/11/13/domain_modelling_excerpt/>. Acesso em novembro de 2018.

- Diagrama de Robustez (Robustness Diagram): Elaboração de um diagrama com os componentes do sistema para cada funcionalidade da aplicação. É um diagrama que descreve a estrutura do *software* para o momento dinâmico.
- Diagrama de Sequência de Mensagens (Sequence Diagram): Elaboração de um diagrama que retrata a comunicação dos componentes do sistema, em ordem cronológica, para cada funcionalidade da aplicação. É um diagrama que descreve o comportamento do *software* no momento dinâmico.
- Diagrama de Classes (Class Model): Elaboração de todos componentes do sistema, descrevendo relações e características. É um diagrama que descreve a estrutura do *software* no momento estático.

A ordem de elaboração dos diagramas é a seguinte: diagrama de casos de uso, modelo de domínio, diagrama de robustez, diagrama de sequência de mensagens e então diagrama de classes. Seguir a ordem sugerida pela metodologia é importante pelo fato da ordem proporcionar um aprofundamento da compreensão das características do *software* de forma gradual. A própria natureza de cada diagrama sugere tal ordem proposta, pois a elaboração do próximo diagrama é o aprofundamento do diagrama anterior.

Vale ressaltar que, frequentemente, o desenvolvimento de um diagrama evidencia a necessidade de modificação de outro diagrama prévio, e é fundamental que ocorra esse aprimoramento constante do conjunto de diagramas como um todo para usufruir plenamente da metodologia.

As características de cada diagrama são aprofundadas nos sub capítulos posteriores à descrição do sistema proposto.

3.1. Descrição do sistema proposto

A função principal da aplicação consiste em três etapas; coleta de comentários nas mídias sociais, classificação dos comentários em relação a sentimento, paixão e repercussão e, por fim, a geração gráficos a partir das classificações de comentários.

A coleta de comentários nas mídias sociais é efetuada através do serviço de integração oferecido pelos mesmos, conhecido como API. Com esse serviço, é possível coletar tanto as postagens de usuários no portal da marca quanto fora deste portal, ao citarem-na de alguma maneira. Essa estratégia de coleta foi escolhida para contemplar tanto os comentários feitos diretamente para a marca, quanto os comentários públicos relacionados à marca sem ser direcionados à mesma.

Após a etapa da coleta, cada um dos comentários é então classificado utilizando as técnicas de processamento de linguagem natural. O texto de cada comentário é então processado, calculando o contexto em que o nome da marca é inserido no comentário e retornando classificação perante o sentimento e paixão. Dependendo das palavras no texto, que estão relacionadas ao nome da marca, esse comentário pode ser classificado com sentimento positivo, negativo ou neutro. Outra classificação é de paixão, indicando se a emoção expressada sobre a marca é forte, fraca ou neutra. A união dessas duas métricas resulta em uma mensuração consistente da reputação de uma marca, indicando se há uma boa ou má reputação e a intensidade dessa reputação perante os usuários das mídias sociais.

Outra classificação aplicada ao comentário coletado, sem envolver processamento de linguagem natural, é o coeficiente de repercussão. Essa classificação resulta em um número, calculado pela soma do número de compartilhamentos da postagem, de comentários na postagem e o número de *feedbacks* (como as “curtidas” nas principais redes sociais) da postagem. Essa métrica de repercussão do comentário qualifica a relevância dessa postagem na mídia social, de acordo com a quantidade de pessoas que esse comentário alcançou e pode ter influenciado.

Após a etapa de classificação, há uma somatória dos valores classificados dos comentários da marca, gerando gráficos para uma ampla visualização. Para a classificação de sentimento, há um gráfico de setor (mais conhecido como Gráfico de Pizza, ilustrado na Figura 28) demonstrando a porcentagem de comentários positivos, negativos e neutros em relação à marca, além de ser informado pela interface os valores absolutos. O mesmo é feito para a classificação de paixão, demonstrando a porcentagem de comentários fortemente e fracamente emocionado,

dia. Essa abordagem possibilita identificar a relação de oscilações na reputação da marca e acontecimentos que podem ter causado tal alteração. Assim, é possível entender o motivo para a mudança e conhecer mais seus clientes e fatores determinantes para sua imagem.

Além dos gráficos, o usuário também pode acessar a base de dados a fim de analisar os comentários coletados, aplicando filtros e ordenações. É possível avaliar a autenticidade dos resultados calculados pelo processamento de linguagem natural (pois a interpretação textual sempre está sujeita a falha) e alterar as classificações errôneas de sentimento e emoção, resultando em atualização dos gráficos.

É possível também manter uma coleta de comentários nas mídias sociais em monitoramento. Assim, a coleta é executada periodicamente para atualizar a base de dados ao longo do tempo. O resultado dessa função é manter o usuário com as informações de reputação mais recentes relacionadas ao objeto de análise específico. É oferecida a possibilidade de acompanhar o estado atual da reputação, proporcionando a informação no momento de um ataque ou promoção da imagem nas mídias sociais.

Enfim, a execução da análise de reputação, com todas as funcionalidades descritas à cima, pode ser criada para múltiplas marcas, o que permite analisar a reputação não apenas de uma marca, mas também de seus concorrentes, comparando seus resultados e gerando informação sobre a situação competitiva.

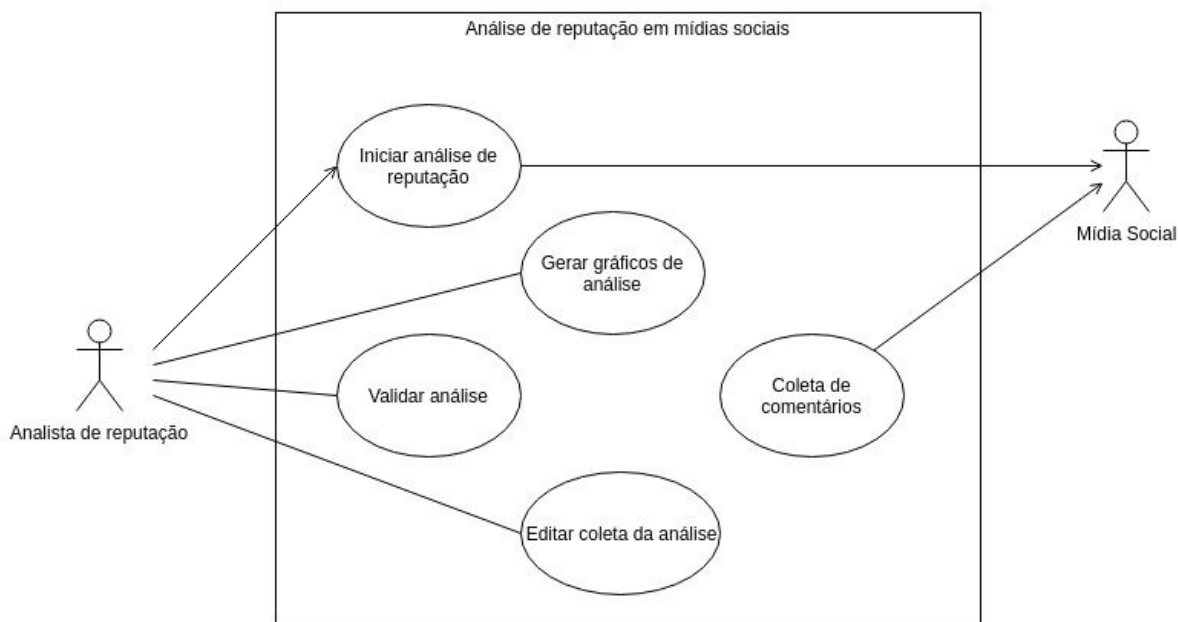
3.2. Análise de Caso de Uso

A análise de caso de uso é a união do diagrama dos casos de uso e da descrição dos casos de uso. O diagrama dos casos de uso ilustra a interação entre os atores, que são os agentes externos ao sistema, com os casos de uso, que são as funcionalidades do sistema. A descrição de casos detalha, em formato de texto, o funcionamento dos casos de uso.

3.2.1. Diagrama dos Casos de Uso

A Figura 12 demonstra o diagrama de casos de uso da aplicação. Esse diagrama representa como o usuário da aplicação, representado como “Analista de reputação”, interage com as funcionalidades.

Figura 12 - diagrama dos casos de uso do projeto



Fonte: Própria (2018).

Adiante, uma descrição detalhada do diagrama:

Há 2 atores: “Analista de reputação” e “Mídia Social”.

O ator “Analista de reputação” interage com 4 dos casos de uso: “Iniciar análise de reputação”, “Gerar gráficos de análise”, “Validar análise” e “Editar coleta da análise”. Em sua relação com “Iniciar análise de reputação”, o ator é o agente ativo.

O ator “Mídia Social” interage com 2 dos casos de uso: “Mídia Social” e “Coleta de comentários”, sendo sempre o agente passivo.

A funcionalidade “Coleta de comentários” não é acessível ao analista. Apenas essa funcionalidade e a “Iniciar análise de reputação” interagem com o ator “Mídia Social”.

3.2.2. Descrição dos Casos de Uso

A seguir, a especificação de cada caso de uso ilustrado no diagrama.

3.2.2.1. Caso de Uso: Iniciar análise de reputação

1 - Descrição do Caso de Uso

O Analista cria uma análise para a entidade alvo desejada e a coleta é iniciada.

2 - Pré-Condições

Não há pré-condições.

3 - Fluxo básico de eventos

1. A aplicação disponibiliza um menu.
2. O analista seleciona a opção “iniciar análise de reputação”.
3. A aplicação disponibiliza um formulário de cadastro de análise.
4. O analista preenche os seguintes campos:
 - Nome da análise.
 - Lista de nomes relacionados à entidade alvo por mídia social.
 - Lista de perfis público de posse da entidade alvo por mídia social, se houver.
 - Período limite de coleta, se houver.
5. A aplicação valida os campos, retorna a mensagem “iniciação de análise de reputação bem sucedida” e disponibiliza o menu inicial novamente.
6. O caso de uso é terminado.

4 - Fluxo alternativo de eventos

- No passo 5, caso os campos não sejam validados, uma mensagem de erro de validação do campo específico é retornada e o caso de uso volta para o passo 4.

5 - Pós-condições

- Análise criada.
- Coleta da análise em andamento.

3.2.2.2. Caso de Uso: Gerar gráficos de análise

1 - Descrição do Caso de Uso

O analista seleciona uma ou mais análises para gerar gráficos do que foi coletado, podendo aplicar filtros.

2 - Pré-Condições

- A(s) análise(s) alvo deve(m) ter sido previamente iniciada(s).

3 - Fluxo básico de eventos

1. A aplicação disponibiliza um menu.
2. O analista seleciona a opção “gerar gráficos”.
3. A aplicação disponibiliza a lista de análises e uma lista de comandos para aplicar filtros.
4. O analista seleciona uma ou mais análises (para comparação) e pode selecionar filtros.
5. A aplicação gera imagens de gráficos a partir das análises e filtros selecionados, retorna uma mensagem indicando o local dos arquivos e disponibiliza o menu inicial novamente.
6. O caso de uso termina.

4 - Fluxo alternativo de eventos

Não há fluxos alternativos.

5 - Pós-condições

- A visualização da reputação através de gráficos é emitida, para os dados coletados e selecionados.

3.2.2.3. Caso de Uso: Validar análise

1 - Descrição do Caso de Uso

O analista visualiza os dados coletados de análise, com a possibilidade de alterar valores de reputação qualificados por processamento de linguagem natural.

2 - Pré-Condições

- A análise alvo deve ter sido previamente iniciada.

3 - Fluxo básico de eventos

1. A aplicação disponibiliza um menu.
2. O analista seleciona a opção “validar análise”.

3. A aplicação disponibiliza a lista de análises.
4. O analista seleciona uma análise.
5. A aplicação disponibiliza uma tabela com todos dados coletados para a análise, podendo aplicar filtros.
6. O analista seleciona um registro, altera um valor qualificado por processamento de linguagem natural e seleciona a opção “salvar alterações”.
7. A aplicação edita o registro armazenado e disponibiliza a tabela atualizada.
8. O analista seleciona a opção “finalizar validação de análise”.
9. A aplicação disponibiliza o menu inicial novamente.
10. O caso de uso termina.

4 - Fluxo alternativo de eventos

- Após o passo 5, o caso de uso pode prosseguir para o passo 8.
- Após o passo 7, o caso de uso pode prosseguir para o passo 6.

5 - Pós-condições

- Todos os registros de uma análise foram visualizados.
- Um ou mais registros de uma análise podem ter seus valores de reputação alterados pelo analista.

3.2.2.4. Caso de Uso: Editar coleta da análise

1 - Descrição do Caso de Uso

O analista edita a coleta de uma análise, como parar coleta, prosseguir coleta, prosseguir monitoramento ou parar monitoramento.

2 - Pré-Condições

- A análise alvo deve ter sido previamente iniciada.

3 - Fluxo básico de eventos

1. A aplicação disponibiliza um menu.
2. O analista seleciona a opção “editar coleta de análise”.
3. A aplicação disponibiliza a lista de análises.
4. O analista seleciona uma análise e um comando entre os seguintes:
 - Parar coleta.
 - Prosseguir coleta.
 - Parar monitoramento.
 - Prosseguir monitoramento.

5. A aplicação valida o comando para a análise, retorna mensagem “edição bem sucedida” e disponibiliza o menu inicial novamente.

6. O caso de uso é terminado.

4 - Fluxo alternativo de eventos

- No passo 5, caso o comando não seja adequado para a coleta da análise selecionada, o caso de uso retorna uma mensagem de erro relacionada e volta para o passo 4.

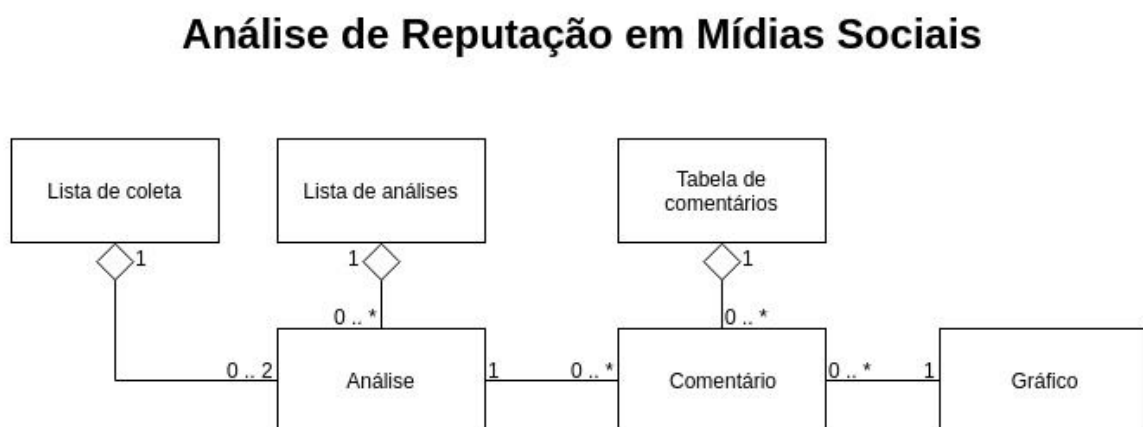
5 - Pós-condições

- A análise muda seu estado de coleta.

3.3. Modelo de Domínio

O modelo de domínio representa os conceitos principais do projeto, no que diz respeito ao domínio da aplicação. Estes conceitos são identificados a partir dos artefatos que detalham os casos de uso. Além da representação no diagrama dos conceitos como objetos, é descrito a relação entre eles em termo de multiplicidade e tipo de relação.

Figura 13 - modelo de domínio do projeto



Fonte: Própria (2018).

A Figura 13 demonstra o diagrama de modelo de domínio da aplicação. Esse diagrama representa que há 3 objetos principais, no que diz respeito ao domínio,

sendo eles “Análise”, “Comentário” e “Gráfico”, além de outros 3 que agrupam esses objetos.

A seguir, uma descrição detalhada do diagrama:

O modelo de domínio possui 6 objetos: “Lista de coleta”, “Lista de análises”, , “Tabela de comentários”, “Comentário” e “Gráfico”.

O objeto “Lista de coleta” pode se relacionar com nenhum até dois objetos “Análise”, sendo essa relação do tipo composição. O objeto “Lista de análises” pode se relacionar com qualquer número de objetos “Análise”, sendo essa relação do tipo composição.

Os objetos “Análise”, “Tabela de comentários” e “Gráfico” podem se relacionar com qualquer número de objetos “Comentário”, sendo a relação com “Tabela de comentários” do tipo composição.

3.4. Diagrama de Robustez

O diagrama de robustez representa todos objetos e atores de uma funcionalidade, com suas respectivas relações. Esses objetos são classificados como:

- Interface, que pode se relacionar com atores e é gerenciado por objetos do tipo controle,
- entidade, que diz respeito ao domínio da aplicação e é gerenciado por objetos do tipo controle,
- controle, que gerencia os objetos do tipo interface e entidade.

Suas representações são ilustradas pela Figura 14:

Figura 14 - tipos de objetos de diagramas de robustez



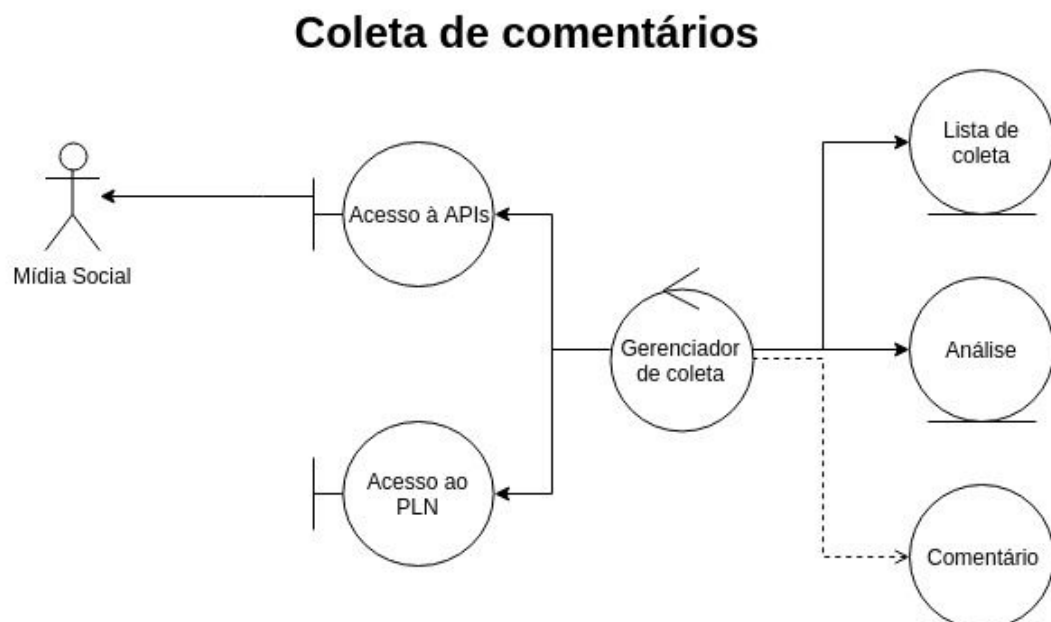
Fonte: Própria (2018).

Os subcapítulos subsequentes contêm as descrições de todos diagramas de robustez do projeto, um a um.

3.4.1. Coleta de comentários

A Figura 15 demonstra o diagrama de robustez referente ao caso de uso “Coleta de comentários”.

Figura 15 - diagrama de robustez de coleta de comentários



Fonte: Própria (2018).

Esse diagrama representa como o objeto de controle “Gerenciador de coleta” utiliza o objeto de interface “Acesso à APIs” para acessar o ator “Mídia Social”, o

objeto de interface “Acesso ao PLN” para acessar a classificação por processamento de linguagem natural e outros objetos de entidade para criar comentários. A seguir, uma descrição detalhada do diagrama.

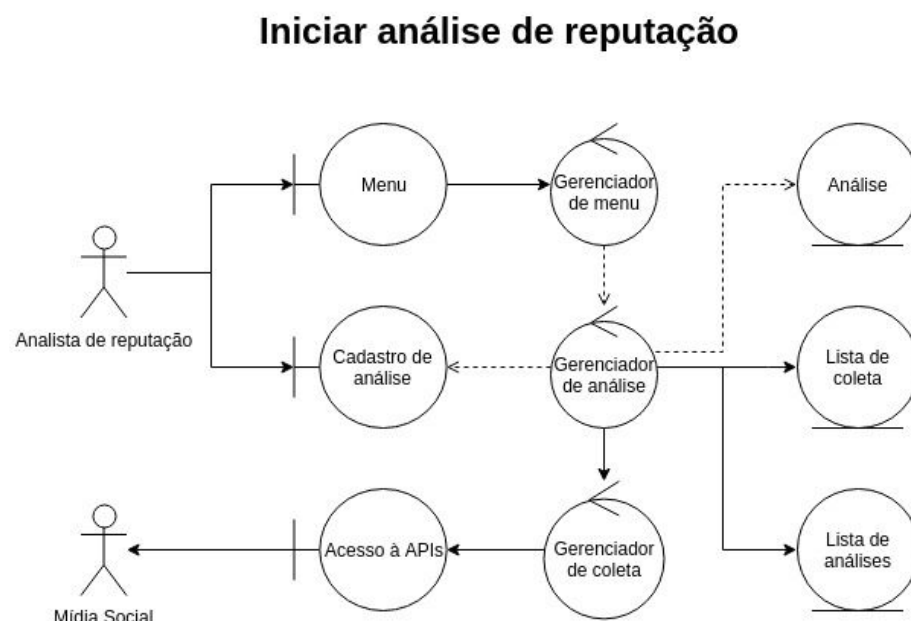
O diagrama possui apenas o ator “Mídia Social”, além de seis objetos: “Gerenciador de coleta”, “Acesso à APIs”, “Acesso ao PLN”, “Lista de coleta”, “Análise” e “Comentário”.

O objeto de controle “Gerenciador de coleta” se relaciona com o objeto de interface “Acesso à APIs”, que por sua vez se relaciona com o ator “Mídia Social”. O “Gerenciador de coleta” também se relaciona com o objeto de interface “Acesso ao PLN”. Por fim, o objeto “Gerenciador de coleta” se relaciona com os objetos entidade “Lista de coleta” e “Análise”, além de criar o objeto de interface “Comentário”.

3.4.2. Iniciar análise de reputação

A Figura 16 demonstra o diagrama de robustez referente ao caso de uso “Iniciar análise de reputação”.

Figura 16 - diagrama de robustez da inicialização da análise de reputação



Fonte: Própria (2018).

Esse diagrama representa como o usuário cria um objeto “Análise”, necessitando acessar o ator “Mídia Social” e os objetos “Lista de coleta” e “Lista de Análise”. A seguir, uma descrição detalhada do diagrama.

O diagrama possui dois atores, “Analista de reputação” e “Mídia Social”, além de nove objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de análise”, “Cadastro de análise”, “Gerenciador de coleta”, “Acesso à APIs”, “Lista de análises”, “Lista de coleta” e “Análise”.

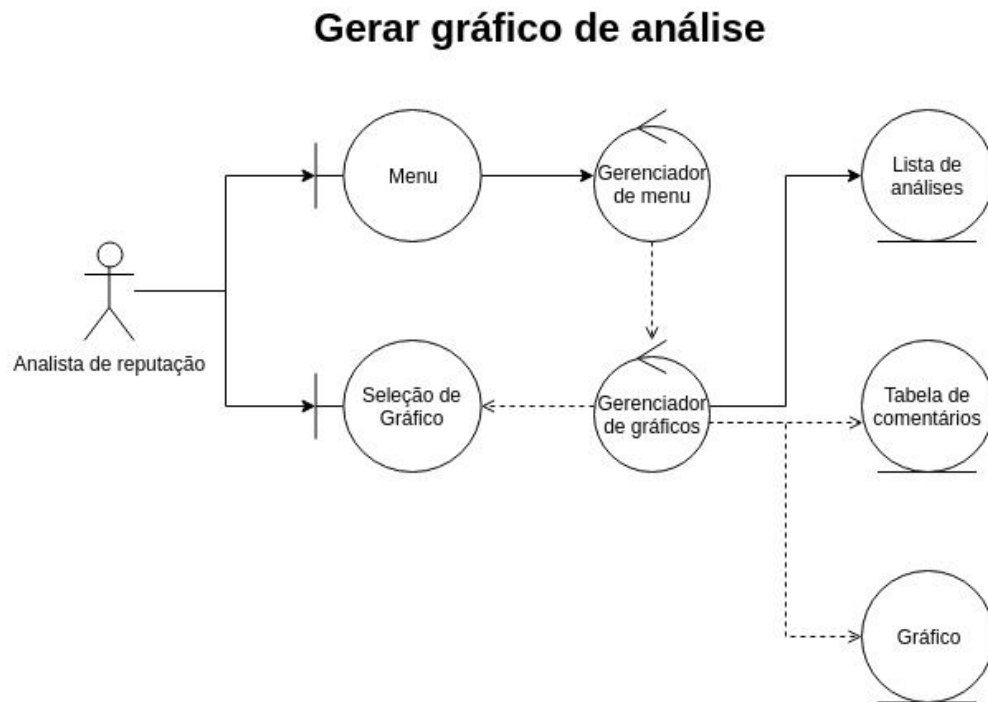
O ator “Analista de reputação” se relaciona com o objeto de interface “Menu”, que se relaciona com o objeto de controle “Gerenciador de menu”. O objeto “Gerenciador de menu” cria o objeto de controle “Gerenciador de análise” que cria o objeto de interface “Cadastro de análise”. O ator “Analista de reputação” também se relaciona com o objeto “Cadastro de análise”.

O objeto de controle “Gerenciador de análise” se relaciona com os objetos de entidade “Lista de coleta” e “Lista de análise”, criando o objeto de entidade “Análise”. Além disso, o objeto “Gerenciador de análise” se relaciona com o objeto de controle “Gerenciador de coleta”, que se relaciona com o objeto de interface “Acesso à APIs”, que por fim se relaciona com o ator “Mídia Social”.

3.4.3. Gerar gráfico de análise

A Figura 17 demonstra o diagrama de robustez referente ao caso de uso “Gerar gráfico de análise”.

Figura 17 - diagrama de robustez da geração de gráfico de análise



Fonte: Própria (2018).

A Figura 17 demonstra o diagrama de robustez referente ao caso de uso “Gerar gráfico de análise”. Esse diagrama representa a criação do objeto “Gráfico”, requisitado pelo usuário, ao criar o objeto “Tabela de comentários” e acessar o objeto “Lista de análises”. Adiante, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de sete objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de gráficos”, “Seleção de gráficos”, “Lista de análises”, “Tabela de comentários” e “Gráfico”.

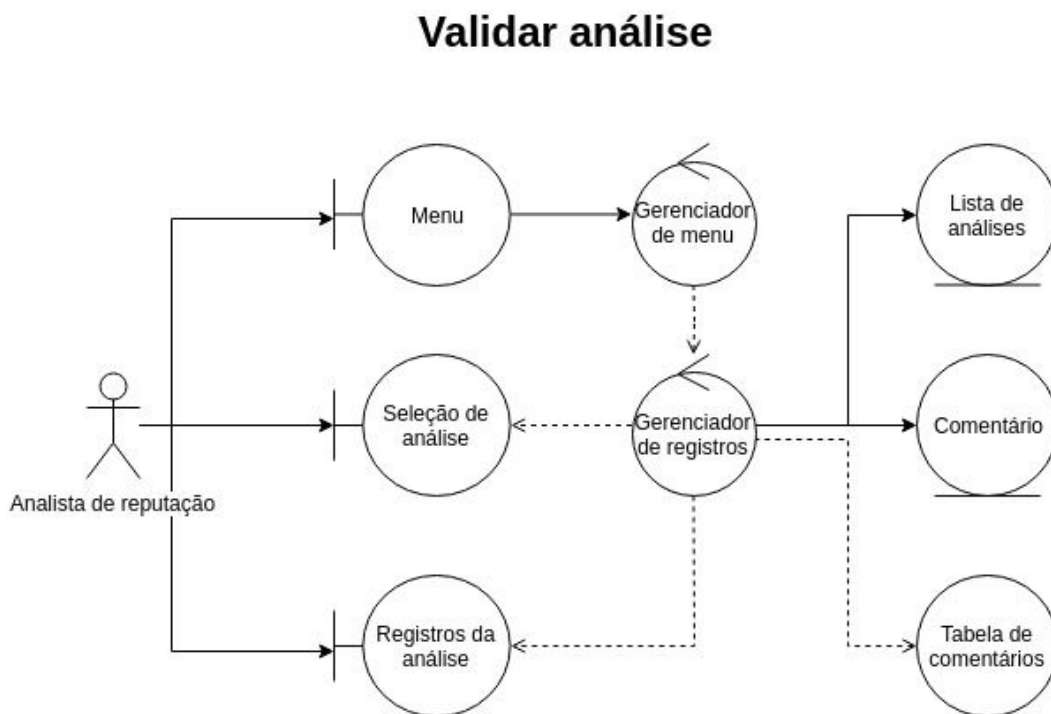
O ator “Analista de reputação” se relaciona com o objeto de interface “Menu”, que se relaciona com o objeto de controle “Gerenciador de menu”. O objeto “Gerenciador de menu” cria o objeto de controle “Gerenciador de gráficos”, que cria o objeto de interface “Seleção de Gráfico”. O ator “Analista de reputação” também se relaciona com o objeto “Seleção de Gráfico”.

O objeto de controle “Gerenciador de gráficos” se relaciona com os objetos de entidade “Lista de análises” e “Tabela de comentários”, criando o objeto de entidade “Gráfico”.

3.4.4. Validar análise

A Figura 18 demonstra o diagrama de robustez referente ao caso de uso “Validar análise”.

Figura 18 - diagrama de robustez de validação da análise



Fonte: Própria (2018).

Esse diagrama representa como o objeto “Comentários”, visualizado pelo objeto criado “Tabela de comentários”, pode ser validado e alterado pelo usuário. A seguir, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de oito objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de registros”, “Seleção de análise”, “Registros de análise”, “Lista de análises”, “Comentário” e “Tabela de comentários”.

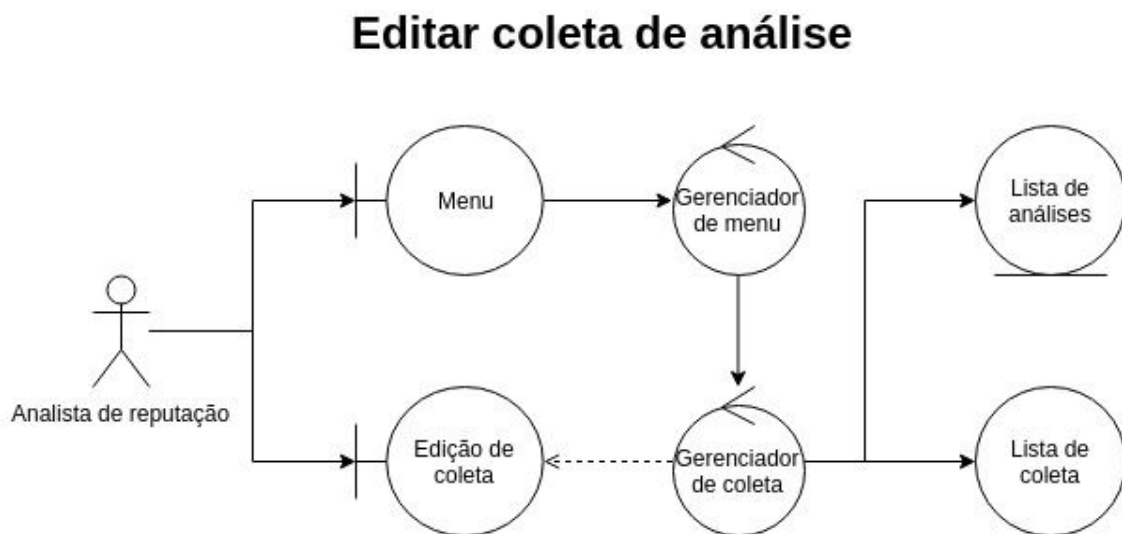
O ator “Analista de reputação” se relaciona com o objeto de interface “Menu”, que se relaciona com o objeto de controle “Gerenciador de menu”. O objeto “Gerenciador de menu” cria o objeto de controle “Gerenciador de registros”, que cria os objetos de interface “Seleção de análise” e “Registros de análise”. O ator também se relaciona com os objetos “Seleção de análise” e “Registros de análise”.

O objeto de controle “Gerenciador de registros” se relaciona com os objetos de entidade “Lista de análises” e “Comentário”, criando o objeto de entidade “Tabela de comentários”.

3.4.5. Editar coleta de análise

A Figura 19 demonstra o diagrama de robustez referente ao caso de uso “Editar coleta de análise”.

Figura 19 - diagrama de robustez de edição da coleta de análise



Fonte: Própria (2018).

Esse diagrama representa como o usuário pode editar o objeto “Lista de coleta” referente à “Lista de análises”. A diante, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de seis objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de coleta”, “Edição de coleta”, “Lista de análises” e “Lista de coleta”.

O ator “Analista de reputação” se relaciona com o objeto de interface “Menu”, que se relaciona com o objeto de controle “Gerenciador de menu”. O objeto “Gerenciador de menu” se relaciona com o objeto de controle “Gerenciador de coleta”, que cria o objeto de interface “Edição de coleta”. O ator “Analista de reputação” também se relaciona com o objeto “Edição de coleta”.

O objeto de controle “Gerenciador de coleta” se relaciona com os objetos de entidade “Lista de análises” e “Lista de coleta”.

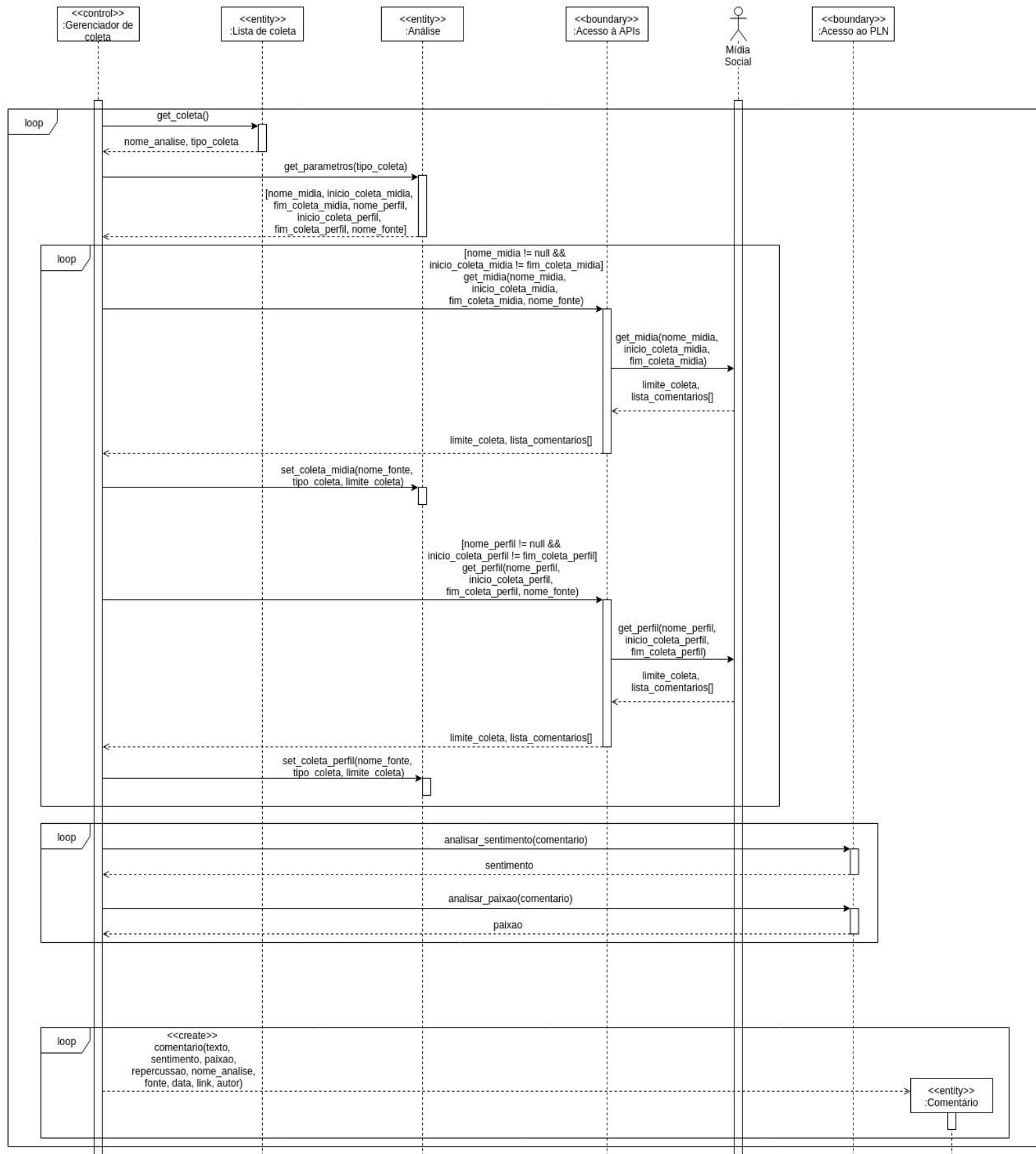
3.5. Diagrama de sequência de mensagens

O diagrama de sequência de mensagens enfatiza a ordem cronológica de mensagens trocadas entre os objetos participantes, e relações com os atores, no momento da execução.

3.5.1. Coleta de comentários

A Figura 20 demonstra o diagrama de sequência de mensagens referente ao caso de uso “Coleta de comentários”.

Figura 20 - diagrama de sequência de mensagens da coleta de comentários



Fonte: Própria (2018).

Esse diagrama representa como o objeto “Gerenciador de Coleta” se permanece em um laço infinito para processar as coletas do objeto “Lista de Coleta”. Então, os parâmetros da coleta são obtidos do objeto “Análise” e a coleta é executada trocando mensagens com o ator “Mídia Social”. Por fim, a classificação da reputação é feita para a lista de comentários retornados pelo ator “Mídia Social”, criando objetos “Comentário”. A seguir, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Mídia Social”, além de ser seis objetos: “Gerenciador de coleta”, “Acesso à APIs”, “Acesso ao PLN”, “Lista de coleta”, “Análise” e “Comentário”. Todos objetos, exceto “Comentário”, já foram previamente criados.

Toda execução se encontra em loop (laço) infinito, com o objeto de controle “Gerenciador de coleta” invocando o método “get_coleta()” do objeto de entidade “Lista de coleta”, para receber os parâmetros da análise que deve ser coletada:

- “nome_analise”, que é o identificador do objeto entidade “Análise” alvo da coleta,
- “tipo_coleta”, que é o tipo de coleta para ser executada, que pode ser retroativa ou progressiva (monitoramento).

Em seguida, o objeto “Gerenciador de coleta” invoca o método “get_parametros(tipo_coleta)” do objeto de entidade “Análise” referente ao identificador “nome_analise”. Esse método retorna os seguintes parâmetros de coleta:

- “nome_fonte”, que é o nome da Mídia Social referente aos dados a seguir,
- “nome_midia”, que é o termo que identifica o que deve ser coletado na mídia social,
- “inicio_coleta_midia”, que é o identificador de onde deve iniciar a coleta por termo,
- “fim_coleta_midia”, que é o identificador do limite da coleta por termo,
- “nome_perfil”, que é o nome que identifica o perfil que deve ser coletado na mídia,

- “inicio_coleta_perfil”, que é o identificador de onde deve iniciar a coleta por perfil,
- “fim_coleta_perfil”, que é o identificador do limite da coleta por perfil,

Então, o objeto “Gerenciador de coleta” entra em loop, iterando sobre a lista de parâmetros de coleta.

O objeto “Gerenciador de coleta” invoca o método “get_midia(nome_midia, inicio_coleta_midia, fim_coleta_midia, nome_fonte)” do objeto de interface “Acesso à APIs”, para fazer a coleta dos comentários relacionados ao termo da análise para a mídia social identificada. Porém, é necessário atender à condição “[nome_midia != null && inicio_coleta_midia != fim_coleta_midia]”, que assegura que haja um termo há ser buscado e que a coleta ainda não foi terminada.

O objeto “Acesso à APIs” invoca o método “get_midia(nome_midia, inicio_coleta_midia, fim_coleta_midia)” do ator “Mídia Social” identificado por “nome_fonte”, retornando para o objeto “Acesso à APIs” e então para o objeto “Gerenciador de coleta”:

- “lista_comentarios”, que é a lista dos dados de cada comentário retornado referentes aos parâmetros de busca,
- “limite_coleta”, que é o identificador do limite que a coleta pode alcançar.

O objeto “Gerenciador de coleta” invoca o método “set_coleta_midia(nome_fonte, tipo_coleta, limite_coleta)” do objeto “Lista de coleta”, para atualizar o atributo “inicio_coleta_midia” com o valor de “limite_coleta”, praticando a paginação na coleta da API.

Em seguida, o objeto “Gerenciador de coleta” executa o mesmo processamento anterior para a coleta por perfil, ao invés de ser por termo, e finaliza-se o loop.

Após a coleta dos comentários, o objeto “Gerenciador de coleta” entra em loop ao iterar a lista de comentários “lista_comentarios” e invoca os métodos “analisar_sentimento(comentario)” e “analisar_sentimento(comentario)” do objeto de

interface “Acesso ao PLN”, passando por parâmetro o texto do comentário e retornando respectivamente os valores “sentimento” e “paixão”, finalizado o loop.

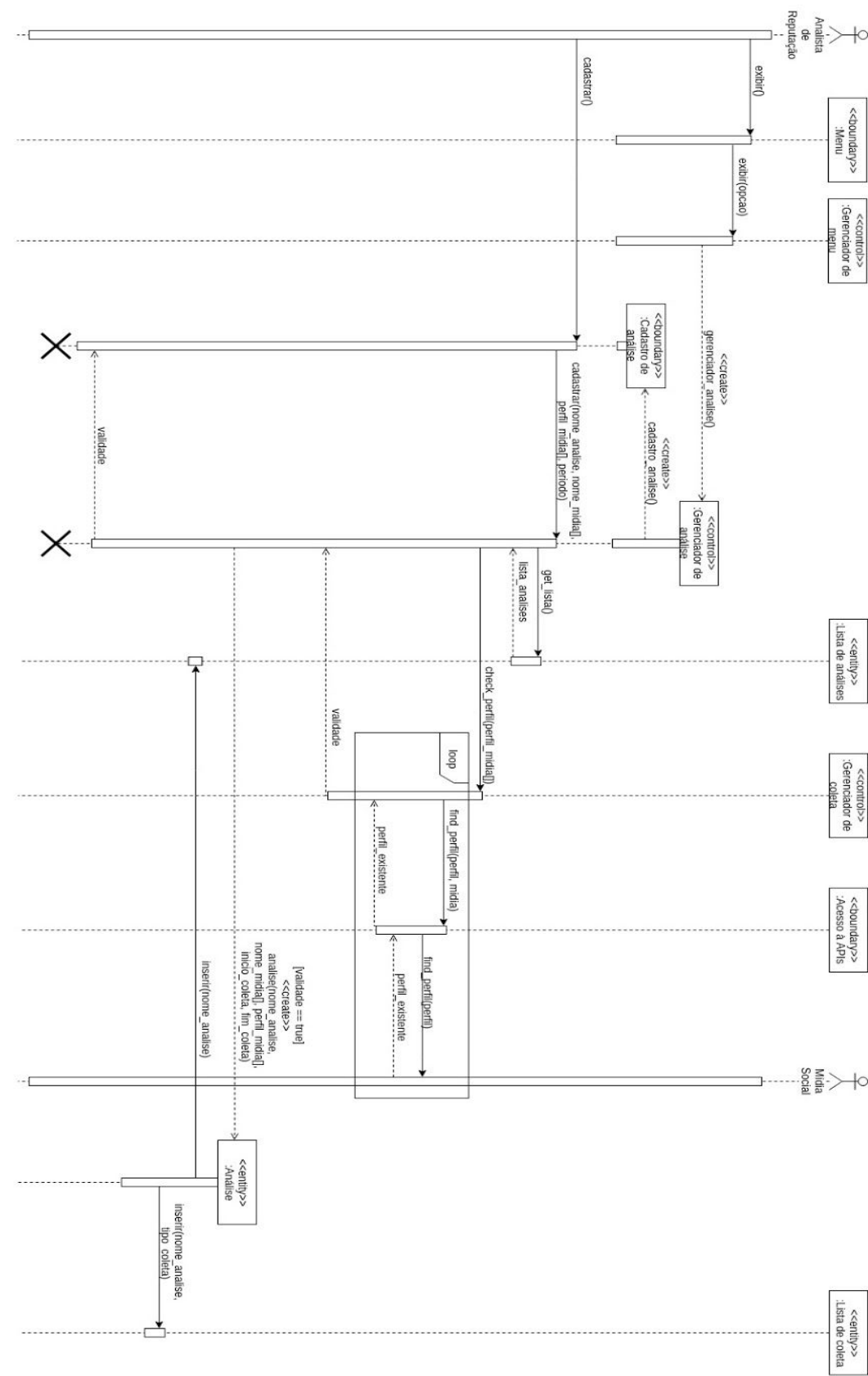
O objeto “Acesso ao PLN” entra em loop ao iterar a lista de comentários “lista_comentarios” e invoca os métodos “analisar_sentimento(comentario)” e “analisar_sentimento(comentario)” do ator “XXXXX”, retornando respectivamente os valores “sentimento” e “paixao”. Assim, o loop é finalizado e é retornado para o objeto “Gerenciador de coleta” o valor “lista_comentarios” atualizado com os indicadores de sentimento e paixão.

Por fim, o objeto “Gerenciador de coleta” entra em loop, iterando em “lista_comentarios”, para criar um objeto entidade “Comentário” utilizando o método “comentario(texto, sentimento, paixao, repercussao, nome_analise, fonte, data, link, autor)”.

3.5.2. Iniciar análise de reputação

A Figura 21 demonstra o diagrama de sequência de mensagens referente ao caso de uso “Iniciar análise de reputação”.

Figura 21 - diagrama de sequência de mensagens de iniciação da análise de reputação



Fonte: Própria (2018).

Esse diagrama representa como a requisição feita pelo usuário para criação de um objeto “Análise” requer validação no objeto “Lista de Análises” e existência de perfis no ator “Mídia Social”, para também ser referenciada nos objetos “Listas de análise” e “Lista de coleta”. A seguir, uma descrição detalhada do diagrama.

O diagrama possui dois atores, “Analista de reputação” e “Mídia Social”, além de nove objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de análise”, “Cadastro de análise”, “Gerenciador de coleta”, “Acesso à APIs”, “Lista de análises”, “Lista de coleta” e “Análise”. Apenas os objetos “Gerenciador de análise”, “Cadastro de análise” e “Análise” não foram previamente criados.

Inicialmente, o ator “Analista de reputação” interage com o objeto de interface “Menu” para selecionar uma funcionalidade. O objeto “Menu” invoca o método “exibir(opcao)” do objeto de controle “Gerenciador de menu”, que cria o objeto de controle referente à opção selecionada, sendo o objeto “Gerenciador de análise”. O objeto “Gerenciador de análise” cria o objeto de interface “Cadastro de análise”.

O ator “Analista de reputação” interage com o objeto de interface “Cadastro de análise” para preencher os campos de cadastro. Então, o objeto “Cadastro de análise” invoca o método “cadastrar(nome_analise, nome_midia[], perfil_midia[], periodo)” do objeto “Gerenciador de análise”, com parâmetros sendo:

- “nome_analise” o identificador único do objeto “Análise” à ser criado,
- “nome_midia” o termo que identifica o que deve ser coletado para cada mídia social,
- “perfil_midia” o endereço do perfil que deve ser coletado para cada mídia social,
- “periodo” o período entre a data limite de coleta e a data atual.

O objeto “Gerenciador de análise” invoca o método “get_lista()” do objeto de entidade “Lista de análises” para retornar o valor “lista_analises” e verificar se “nome_analise” é um valor único.

O objeto “Gerenciador de análise” invoca o método “check_perfil(perfil_midia[])” do objeto de controle “Gerenciador de coleta”, para

verificar se a lista de “perfil_midia” é válida. Então, o objeto “Gerenciador de coleta” entra em loop, iterando na lista “perfil_midia”, obtendo os valores:

- “perfil”, que é o endereço do perfil que deve ser coletado,
- “midia”, que é o identificador da mídia social cujo perfil pertence.

O objeto “Gerenciador de coleta” invoca o método “find_perfil(perfil, midia)” do objeto de interface “Acesso à APIs”, que invoca o método “find_perfil(perfil)” do ator “Mídia Social” e retorna o valor booleano “perfil_existente” para o objeto “Acesso à APIs”, que retorna o mesmo valor para o objeto “Gerenciador de coleta”, finalizando o loop. O objeto “Gerenciador de coleta” retorna para o objeto “Gerenciador de análise” o valor “validade”, referente à existência de todos perfis nas mídias sociais.

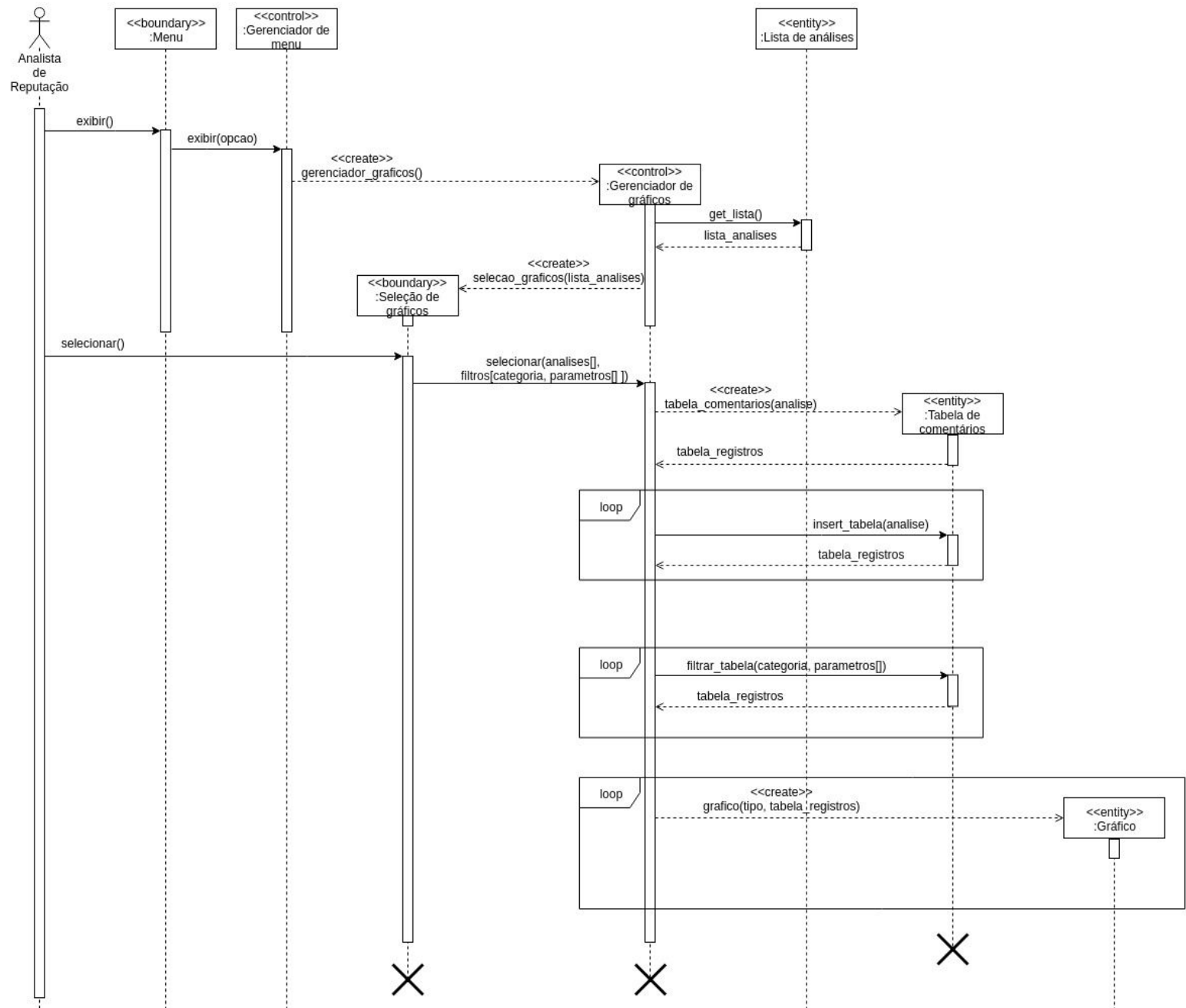
O objeto “Gerenciador de análise” cria o objeto de entidade “Análise” com o método “analise(nome_analise, nome_midia[], perfil_midia[], inicio_coleta, fim_coleta)”, caso a condição “[validade == true]” seja atendida. No processo de criação, o objeto “Análise” invoca o método “inserir(nome_analise)” do objeto “Lista de análises” e também invoca o método “inserir(nome_analise, tipo_coleta)” do objeto de interface “Lista de coleta”, inserindo o identificador da análise na lista de análises e a sua coleta retroativa na lista de coleta.

Por fim, o objeto “Gerenciador de análise” retorna o valor “validade” para o objeto “Cadastro de análise” e os dois objetos são terminados.

3.5.3. Gerar gráfico de análise

A Figura 22 demonstra o diagrama de sequência de mensagens referente ao caso de uso “Gerar gráfico de análise”.

Figura 22 - diagrama de sequência de mensagens da geração de gráfico de análise



Fonte: Própria (2018).

Nesse diagrama é representado como um gráfico requer a criação do objeto “Tabela de comentários” relacionado à objetos “Análise” selecionados, além de possíveis aplicar filtros nessa tabela, para gerar os gráficos disponíveis para esses registros. A seguir, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de sete objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de gráficos”, “Seleção de

gráficos”, “Lista de análises”, “Tabela de comentários” e “Gráfico”. Apenas os objetos “Tabela de comentários”, “Gerenciador de análise” e “Seleção de gráficos” não foram previamente criados.

Inicialmente, o ator “Analista de reputação” interage com o objeto de interface “Menu” para selecionar uma funcionalidade. O objeto “Menu” invoca o método “exibir(opcao)” do objeto de controle “Gerenciador de menu”, que cria o objeto de controle referente à opção selecionada, sendo o objeto “Gerenciador de gráficos”.

O objeto “Gerenciador de análise” invoca o método “get_lista()” do objeto de entidade “Lista de análises”, retornando o valor “lista_analises”, com a lista dos identificadores das análises. Então, o objeto “Gerenciador de análise” cria o objeto de interface “Seleção de gráficos” através do método “selecao_graficos(lista_analises)”.

O ator “Analista de reputação” interage com o objeto de interface “Seleção de gráficos”, que invoca o método “selecionar(analises[], filtros[])” do objeto “Gerenciador de análise”, com parâmetros sendo:

- “analises”, que é uma lista do identificadores das análises,
- “filtros”, que é uma lista contendo o identificador da categoria do filtro à ser aplicado nas análises e a lista de parâmetros necessários para esse filtro.

O objeto “Gerenciador de análise” cria o objeto de entidade “Tabela de comentários” invocando o método “tabela_comentarios(analise)” com o primeiro identificador de análise sendo o parâmetro e sendo retornado a lista de comentários da análise. Então, o objeto “Gerenciador de análise” entra em loop, iterando na lista “analises” para invocar o método “insert_tabela(analise)” do objeto entidade “Tabela de comentários” utilizando como parâmetro o valor recebido pela iteração.

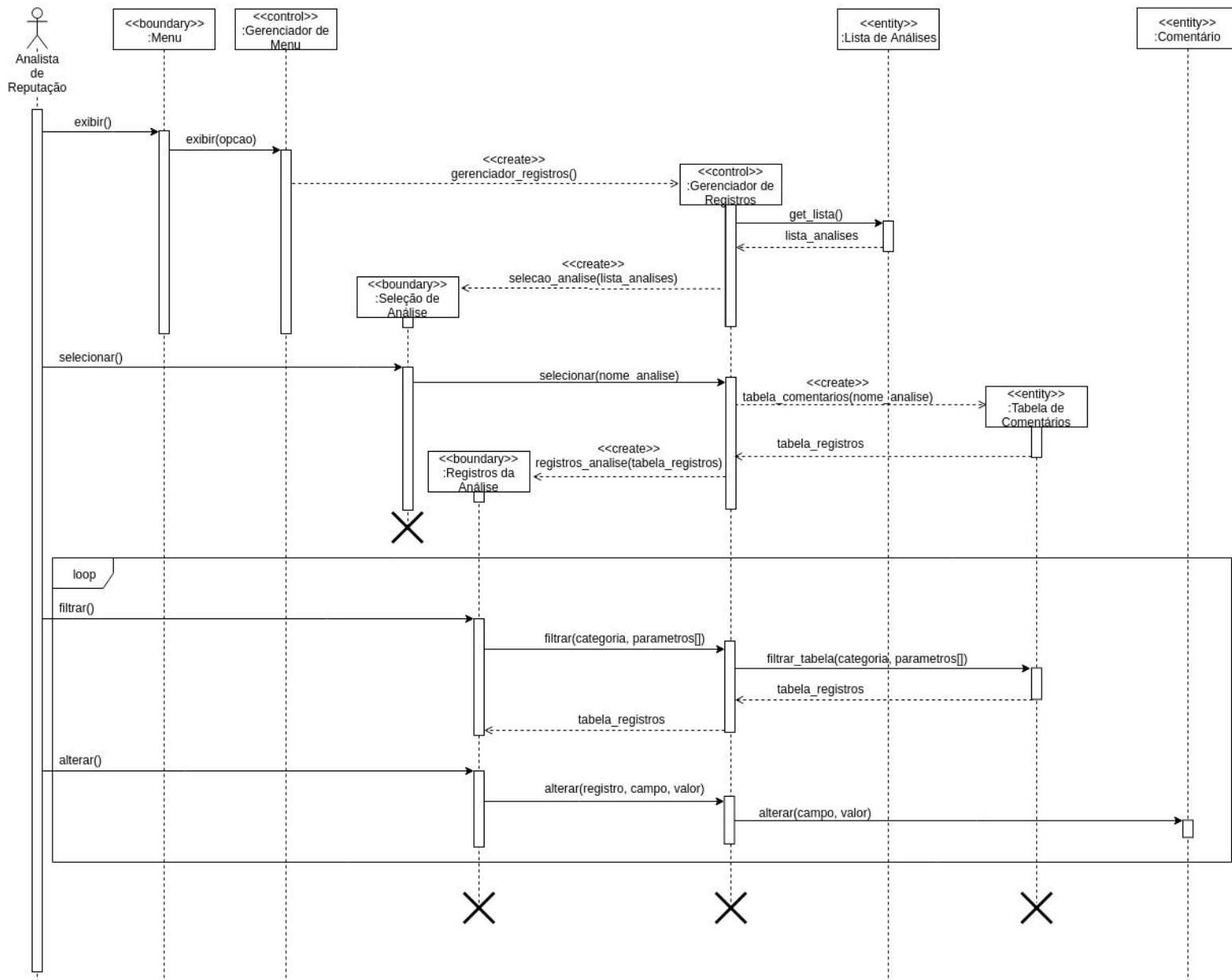
Após finalizar o loop anterior, o objeto “Gerenciador de análise” entra em loop novamente, iterando na lista “filtros” para invocar o método “filtrar_tabela(categoria, parametros[])” do objeto “Tabela de comentários” utilizando como parâmetro os valores recebidos pela iteração e recebendo o valor “tabela_registros” com a tabela contendo os comentários das análises após aplicação de filtros.

Por fim, o objeto “Gerenciador de análise” entra em loop iterando na lista de tipos de gráficos disponíveis, criando um objeto de entidade “Gráfico” para gerar o arquivo do tipo selecionado com os dados da tabela de registros final. Após o fim desse loop, os objetos “Tabela de comentários”, “Gerenciador de análise” e “Seleção de gráficos” são terminados.

3.5.4. Validar análise

A Figura 23 demonstra o diagrama de sequência de mensagens referente ao caso de uso “Validar análise”.

Figura 23 - diagrama de sequência de mensagens da validação de análise



Fonte: Própria (2018)

Esse diagrama representa como o objeto “Tabela de comentários” referente a um objeto “Análise” é disponibilizada para visualização, filtro e alteração do objeto “Comentários”. A seguir, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de oito objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de registros”, “Seleção de análise”, “Registros de análise”, “Lista de análises”, “Comentário” e “Tabela de

comentários”. Apenas os objetos “Menu”, “Gerenciador de menu”, “Lista de análises” e “Comentário” foram previamente criados.

Inicialmente, o ator “Analista de reputação” interage com o objeto de interface “Menu” para selecionar uma funcionalidade. O objeto “Menu” invoca o método “exibir(opcao)” do objeto de controle “Gerenciador de menu”, que cria o objeto de controle referente à opção selecionada, sendo o objeto “Gerenciador de registros”.

O objeto “Gerenciador de registros” invoca o método “get_lista()” do objeto de entidade “Lista de análises”, retornando o valor “lista_analises”, com a lista dos identificadores das análises. Então, o objeto “Gerenciador de análise” cria o objeto de interface “Seleção de análise” através do método “selecao_analise(lista_analises)”.

O ator “Analista de reputação” interage com o objeto de interface, que invoca o método “selecionar(nome_analise)” do objeto “Gerenciador de análise”. Então, o objeto “Gerenciador de análise” cria o objeto de entidade “Tabela de comentários” para os comentários relacionados ao identificador da análise selecionada, retornando o valor “tabela_comentarios”. Em seguida, o objeto “Gerenciador de análise” cria o objeto de interface “Registros de análise” invocando o método “registros_analise(tabela_registros)”.

Então, a o objeto de interface “Registros de análise” entra em loop para executar o processo de visualização, filtro e alteração dos registros da tabela até o ator “Analista de Reputação” terminar.

O ator “Analista de Reputação” interage com o objeto “Registros de análise” para invocar o método “filtrar(categoria, parametros[])” do objeto “Gerenciador de análise”, com os parâmetros sendo:

- “categoria”: o identificador do filtro para ser aplicado,
- “parametros”: a lista de parâmetros para o filtro.

O objeto “Registros de análise” invoca o método “filtrar_tabela(categoria, parametros[])” do objeto “Tabela de comentários”, que retorna o valor

“tabela_registros” da tabela filtrada, retornando também o valor do objeto “Gerenciador de análise” para o objeto “Registros de análise”.

O ator “Analista de Reputação” interage com o objeto “Registros de análise” para invocar o método “alterar(registro, campo, valor)” do objeto “Gerenciador de análise”, com os parâmetros sendo:

- “registro”: o identificador do comentário descrito na tabela,
- “campo”: o atributo do comentário selecionado para alteração,
- “valor”: o novo valor do atributo do comentário que foi selecionado.

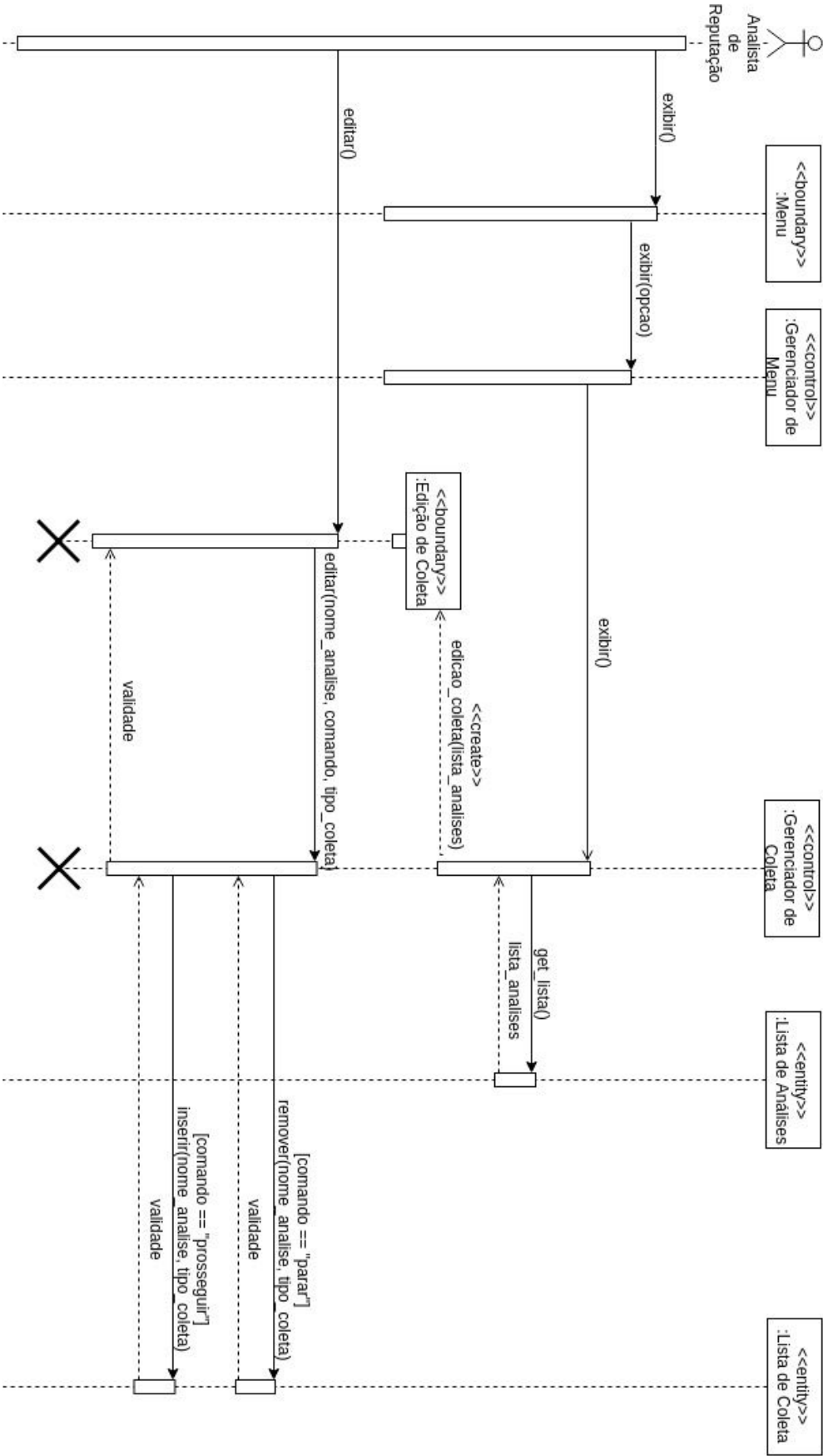
O objeto “Gerenciador de análise” invoca o método “alterar(campo, valor)” do objeto de entidade “Comentário” relacionado ao identificador “registro”, alterando o atributo do mesmo.

O loop é finalizado quando o ator “Analista de Reputação” selecionar essa opção e os objetos “Menu”, “Gerenciador de menu” e “Lista de análises” são terminados.

3.5.5. Editar coleta de análise

A Figura 24 demonstra o diagrama de sequência de mensagens referente ao caso de uso “Editar coleta de análise”.

Figura 24 - diagrama de sequência de mensagens da edição da coleta de análise



Fonte: Própria (2018).

Esse diagrama representa como a inserção ou remoção de uma coleta do tipo “padrão” ou “monitoramento”, referente a um objeto “Análise”, pode executada após validação do objeto “Lista de coleta”. A seguir, uma descrição detalhada do diagrama.

O diagrama possui apenas o ator “Analista de reputação”, além de seis objetos: “Menu”, “Gerenciador de menu”, “Gerenciador de coleta”, “Edição de coleta”, “Lista de análises” e “Lista de coleta”. Apenas o objeto “Edição de coleta” não foi previamente criado.

Inicialmente, o ator “Analista de reputação” interage com o objeto de interface “Menu” para selecionar uma funcionalidade. O objeto “Menu” invoca o método “exibir(opcao)” do objeto de controle “Gerenciador de menu”, que invoca o método “exibir()” do objeto de controle referente à opção selecionada, sendo o objeto “Gerenciador de coleta”.

O objeto “Gerenciador de coleta” invoca o método “get_lista()” do objeto de entidade “Lista de análises”, retornando o valor “lista_analises”, com a lista dos identificadores das análises. Então, o objeto “Gerenciador de coleta” cria o objeto de interface “Edição de coleta” através do método “edicao_coleta(lista_analises)”.

O ator interage com o objeto “Edição de coleta”, que invoca o método “editar(analise, comando, tipo_coleta)” do objeto “Gerenciador de coleta”, com os parâmetros sendo:

- “nome_analise”: identificador do objeto entidade “Análise” alvo da coleta,
- “comando”: identificador do método aplicado para a coleta, podendo ser “parar” ou “prosseguir,
- “tipo_coleta”: identificador do tipo de coleta, podendo ser “monitoramento” ou “padrão”.

O objeto “Gerenciador de coleta” invoca o método “remover(nome_analise, tipo_coleta)” no caso da condição “[comando == “parar”]” ser atendida, removendo o tipo de coleta relacionado à análise selecionada no caso do comando ter o valor

“parar” e retornando valor “validade”. O valor “validade” indica se é válido aplicar método requisitado ou não.

O objeto “Gerenciador de coleta” invoca o método “inserir(nome_analise, tipo_coleta)” no caso da condição “[comando == “prosseguir”]” ser atendida, removendo o tipo de coleta relacionado à análise selecionada no caso do comando ter o valor “prosseguir” e retornando valor “validade”.

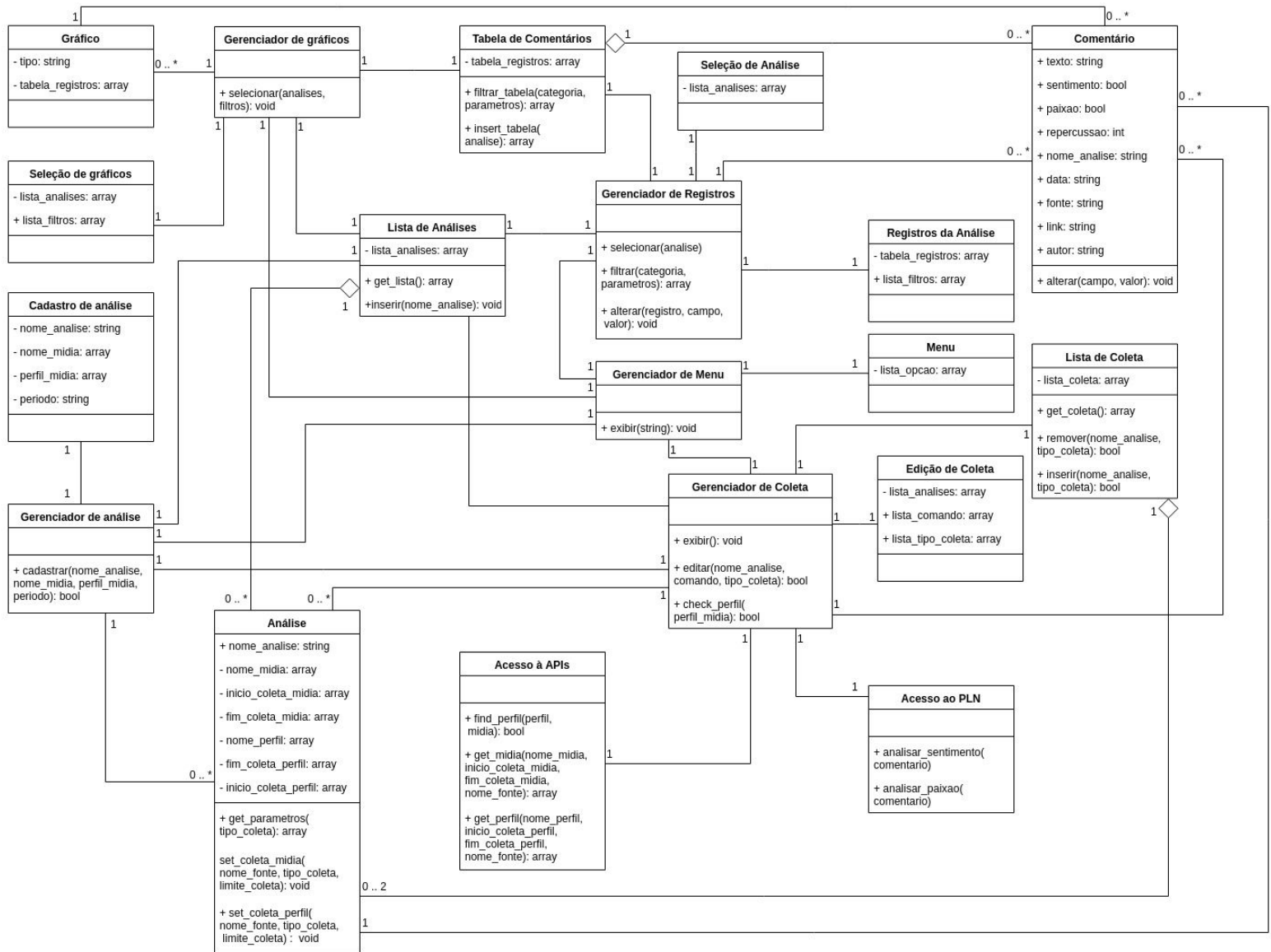
O objeto “Gerenciador de coleta” retorna o valor “validade” para o objeto “Edição de coleta” e os objetos “Gerenciador de coleta” e “Edição de coleta” são terminados.

3.6. Diagrama de classes candidatas

O diagrama de classes representa a estrutura e a relação dos objetos descritos nos outros diagramas, em momento estático. Os campos da classe são descritos no bloco central, abaixo do nome da classe. Esses campos possuem indicação de ser público (+) ou privado (-), além de nome e tipo de dado. Os métodos são descritos no bloco de baixo, com indicação de nome, parâmetros, tipo de dado retornado e tipo de acesso público (+) ou privado (-). A multiplicidade e o tipo da relação também são identificados.

Esse diagrama é denominado como de classes candidatas pois é uma opção que não necessariamente será estritamente seguida na implementação, porém é um resultado inspirado nos outros diagramas desenvolvidos.

Figura 25 - diagrama de classes candidatas



Fonte: Própria (2018).

A Figura 25 demonstra o diagrama de classes candidatas da aplicação. Esse diagrama representa todas as classes do sistema, compondo métodos e atributos referentes aos diagramas de sequência de mensagens e as relações, com seus tipos e multiplicidades, proveniente da mesma referência e do modelo de domínio, resultando em uma ampla visão da estrutura do sistema.

4. DECISÕES ARQUITETURAIS

O desenvolvimento do protótipo tem como objetivo a coleta de múltiplos comentários sobre uma entidade (seja pessoa, empresa ou produto) em mídias sociais e a entrega de uma análise de reputação dessa entidade, em forma de gráficos e lista de dados dos comentários, de acordo com a descrição no subcapítulo 3.1.

Os componentes fundamentais desse sistema são:

- Acesso à API do Twitter: um objeto que faz requisições para o servidor da API do Twitter, por meio da utilização da biblioteca “tweepy”. É enviado o termo que representa a entidade alvo, informado pelo usuário, e é retornado uma lista de objetos com dados dos comentários requisitados. Esse componente também estrutura a lista de objetos de “tweets” coletados para possuir apenas os campos necessários, gerando um arquivo no formato “json” para exportar essa estrutura.
- Acesso ao Processamento de Linguagem Natural (PLN): um objeto que acessa a biblioteca “LeIA” para enviar um texto de um comentário e receber a classificação de sentimento correspondente. Esse componente também lê o arquivo de texto com a lista de “stopwords” em português para remover as palavras irrelevantes de um texto, sendo utilizado para gerar o gráfico “wordcloud”.
- Gerador de Gráficos: um objeto que recebe os parâmetros de geração dos gráficos, sendo texto para o gráfico “wordcloud” e quantidade dos coeficientes de sentimento para o gráfico de setores (gráfico de pizza). A imagem do gráfico requisitado é gerada ao acessar as bibliotecas “matplotlib” para o gráfico de setores e “wordcloud” para o gráfico de “wordcloud”.

4.1. Decisões de Implementação

Para a programação do sistema, optou-se pela utilização de:

- A linguagem de programação Python para implementar os códigos

- A IDE PyCharm para auxiliar na programação
- A biblioteca de Python LeIA (Fonte: <<https://github.com/rafjaa/LeIA>>) para análise de sentimento da língua portuguesa
- A biblioteca de Python matplotlib (Fonte: <<https://matplotlib.org/>>) para gerar gráfico de setores (de pizza)
- A biblioteca de Python wordcloud (Fonte: <https://github.com/amueller/word_cloud>) para gerar gráficos wordcloud
- A biblioteca de Python tweepy (Fonte: <<https://www.tweepy.org/>>) para acessar a API do Twitter
- As bibliotecas de Python embutidas: json, io e re, para utilizar arquivos no formato json e aplicar regras de expressão regular.
- A lista de termos irrelevantes (stopwords) da língua portuguesa (Fonte: <<https://github.com/danielkansaon/BraSNAM2018-Dataset-Analise-de-sentimentos-em-tweets-em-portugues-brasileiro/blob/master/StopWords-pt>>) para filtrar os tokens de um texto.

A linguagem de programação Python e a IDE PyCharm foram escolhidas pela familiaridade que os integrantes do projeto têm com as mesmas. Outro fator de decisão da escolha da linguagem de programação é a vasta quantidade de bibliotecas disponíveis, com código aberto e ampla prestação de suporte da comunidade.

As bibliotecas de python LeIA, matplotlib, wordcloud, tweepy, json, io e re foram escolhidas por apresentarem maior popularidade nas pesquisas de bibliotecas de Python para sua funcionalidade, além de demonstrarem eficiência e praticidade durante a implementação e os testes.

4.2. Codificação

As principais funcionalidades implementadas no protótipo da aplicação são descritas a seguir:

- “AcessoTwitter.quick_query(string query, int limit)”

Acessa a biblioteca “tweepy” para coletar a lista dos últimos “tweets” citando o termo enviado, limitados aos 1500 comentários dos 10 últimos dias.

A biblioteca “tweepy” limita a coleta para comentários criados em até 10 dias retroativos, para o total de 1500 coletas por requisição e para 100 comentários por requisição. A API do Twitter limita a quantidades de requisições para 180 requisições a cada 15 minutos, 30 requisições a cada minuto e 10 requisições a cada segundo.

- “AcessoTwitter.struct_tweet_list(list tweet_list)”

Recebe a lista de objetos “tweets” coletados e estrutura para uma lista de objetos de comentários, contendo apenas os campos:

- “id”: identificador único do comentário para a API
- “text”: texto do comentário
- “author”: identificador do usuário autor do comentário
- “date”: data e hora de criação do comentário
- “link”: URL do endereço do comentário
- “share_count”: quantidade de compartilhamentos (retweets) do comentário
- “likes_count”: quantidade de curtidas (*feedback*) do comentário
- “repercussion_count”: coeficiente de repercussão do comentário, sendo a soma da quantidade de compartilhamentos com a de curtidas.

Não foi possível inserir a quantidade de comentários no coeficiente de repercussão, pois a API do Twitter não disponibiliza esse dado.

- “AcessoTwitter.save_tweets(list tweet_list, string query)”

Gera um arquivo no formato “json” com a lista dos comentários estruturados, ordenados por data de forma descendente e indentados para melhor visualização.

Figura 26 - ilustração de parte dos dados estruturados no arquivo json referentes à coleta da PUCSP

```
{
  "likes_count": 0,
  "repercussion count": 0,
  "id": 1065194690402807815,
  "share_count": 0,
  "link": "twitter.com/jlgoldfarb/status/1065194690402807815",
  "date": "21/11/2018 10:46:19",
  "author": "jlgoldfarb",
  "sentiment": true,
  "text": "Hoje tem lançamento da @educ_pucsp Tds convidados Viva a #poesianapuc https://t.co/NVcDwPe0zr"
},
{
  "likes_count": 1,
  "repercussion count": 1,
  "id": 1065193648671940608,
  "share_count": 0,
  "link": "twitter.com/jlgoldfarb/status/1065193648671940608",
  "date": "21/11/2018 10:42:10",
  "author": "jlgoldfarb",
  "sentiment": null,
  "text": "I'm at PUC-SP Campus Marquês de Paranaguá! https://t.co/2tB1ZPDZth"
},
{

```

Fonte: Própria (2018).

- “AcessoPLN.sentiment_analysis(string text)”

Acessa a biblioteca “LeIA” para executar a análise de sentimento do texto enviado.

É retornado da biblioteca um coeficiente do sentimento predominante do texto chamado “compound”, valorado de -1 para extremamente negativo até +1 para extremamente positivo. Como é instruído pela documentação da biblioteca, o coeficiente de sentimento é positivo para um “compound” maior que 0.05, negativo para um “compound” menor que -0.05 e neutro para um “compound” entre 0.05 e -0.05.

- “AcessoPLN.treat_stopwords(string text)”

Consulta a lista de “stopwords” em português para remover as palavras irrelevantes do texto, como “do” e “que”, e retornando os “tokens”. É utilizado na geração do gráfico “wordcloud”.

- “Grafico.pizza(list label_list, list size_list, list color_list, string start_name)”

Acessa a biblioteca “matplotlib” para gerar um arquivo no formato “png” do gráfico de setores (gráfico de pizza) com os rótulos, coeficientes e cores enviados de cada seção.

- “Grafico.wordcloud(string text, string start_name)”

Acessa a biblioteca “wordcloud” para gerar um arquivo no formato “png” do gráfico de “wordcloud” para um texto enviado.

- “Gerenciador.quick_execution(string query)”

Invoca, com o termo enviado, a coleta, estruturação e classificação de “tweets”, a contagem e impressão dos coeficientes e a geração das imagens de gráficos e arquivo no formato “json” contendo os dados.

- “Gerenciador.import_json(string json_file)”

Invoca a leitura de um arquivo “json” com o resultado de uma coleta validada. Com essa funcionalidade, o usuário pode manualmente acessar o arquivo estruturado no formato “json” e validar os coeficientes de análise de sentimento, além do filtro manual dos comentários, para ser processado na aplicação.

- “Menu.select_option()”

Executa a interface de comunicação com o usuário da aplicação, para receber a seleção do método a ser executado, entre “Gerenciador.quick_execution()” e “Gerenciador.import_json()”, e seu respectivo parâmetro para execução.

É importante ressaltar que há uma diferença entre as classes candidatas elaboradas no diagramas de classes e as classes contidas no programa final. Isso se deve ao fato da diagramação servir para inspirar a estrutura e o comportamento do programa, mas o produto final geralmente modifica partes da estrutura descrita.

Essa modificação ocorre pois a diagramação carrega funcionalidades muito específicas, sendo facilmente acoplável por outra classe que execute papel

semelhante, principalmente no caso de projetos menores, como acontece com a única classe “gerenciador” no programa deste trabalho que carrega as funções de três classes candidatas gerenciadoras (gerenciador de coleta, gerenciador de gráfico e gerenciador de análise).

O programa desenvolvido é um protótipo, uma prova de conceito, que implementa as funcionalidades e componentes principais do sistema arquitetado para comprovar sua eficiência e demonstrar os serviços mais fundamentais idealizados. Todavia, a monografia possui bases conceituais e de modelagem suficientes para amparar toda implementação dos trabalhos futuros da equipe, assim como ampara o apoio de terceiros ao projeto, pois o programa é *open source*, podendo contar com o apoio da comunidade para seu desenvolvimento completo.

Por fim, segue o endereço dos códigos da aplicação no site GitHub, onde estarão disponíveis publicamente para consulta, utilização, acompanhamento do projeto e colaboração no desenvolvimento dos trabalhos futuros <<https://github.com/tersowillis/reputacao>>.

4.3. Amostra de execução da aplicação

Nesta seção, o resultado de quatro execuções é descrito e ilustrado com a imagem dos gráficos e mensagens geradas pela aplicação.

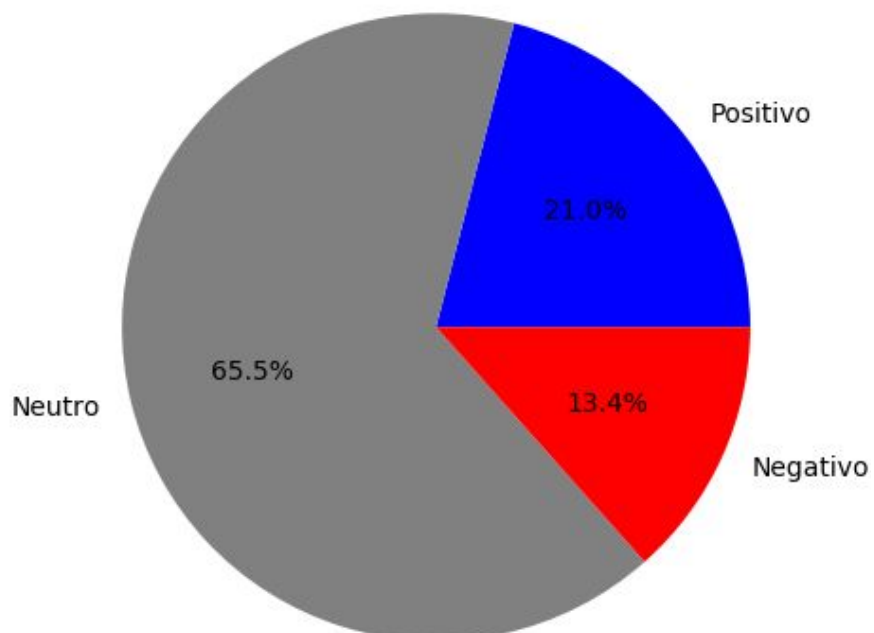
Ao executar a aplicação utilizando o termo “PUCSP”, os gráficos e dados resultantes podem ser vistos a seguir, referentes aos comentários no Twitter publicados entre o dia 21/11/2018 às 15:43:48 e o dia 11/11/2018 às 05:27:52.

Figura 27 - wordcloud das referências à PUCSP no twitter



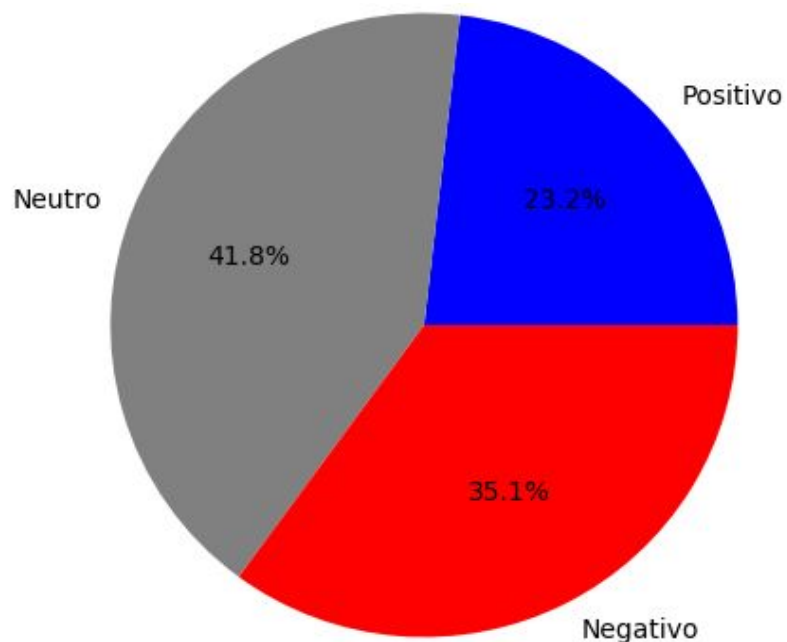
Fonte: Própria (2018).

Figura 28 - gráfico de setores das referências à PUCSP no twitter



Fonte: Própria (2018).

Figura 31 - gráfico de setores das referências à UNINOVE no twitter



Fonte: Própria (2018).

Figura 32 - valores exibidos pela interface da aplicação, referentes à UNINOVE no twitter

```
A soma dos coeficiente de repercussão é: 556
A soma das métricas de sentimento para os 328 comentários é:
Total de comentários positivos: 76
Total de comentários negativos: 115
Total de comentários neutros: 137
Fim da coleta da análise UNINOVE
```

Fonte: Própria (2018).

Ao executar a aplicação utilizando o termo “Bolsonaro”, os gráficos e dados resultantes podem ser vistos a seguir, referentes aos comentários no Twitter publicados entre o dia 21/11/2018 às 15:51:44 e o dia 21/11/2018 às 15:40:03.

Figura 35 - valores exibidos pela interface da aplicação, referentes à Bolsonaro no twitter

```
A soma dos coeficiente de repercussão é: 2212944
A soma das métricas de sentimento para os 1500 comentários é:
Total de comentários positivos: 378
Total de comentários negativos: 597
Total de comentários neutros: 525
Fim da coleta da análise Bolsonaro
```

Fonte: Própria (2018).

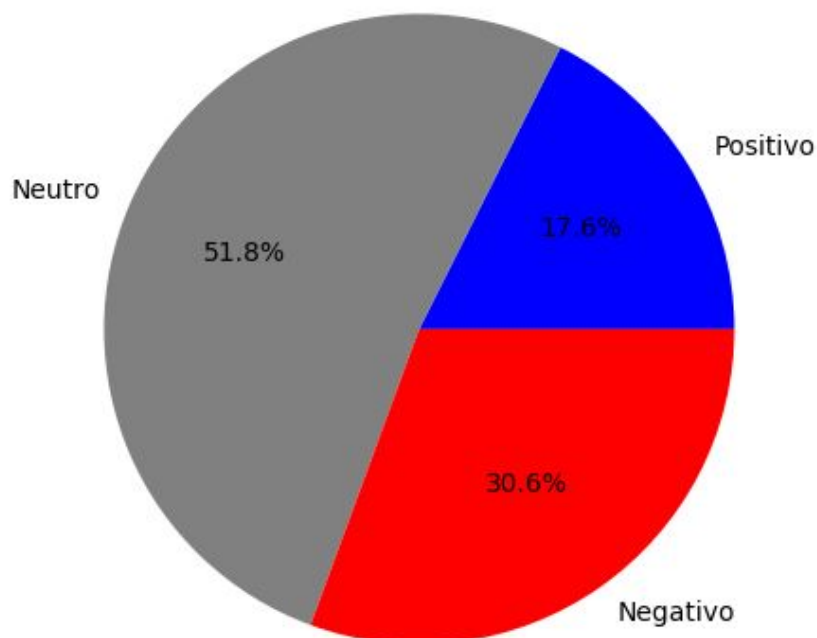
Ao executar a aplicação utilizando o termo “Haddad”, os gráficos e dados resultantes podem ser vistos a seguir, referentes aos comentários no Twitter publicados entre o dia 21/11/2018 às 15:54:59 e o dia 21/11/2018 às 15:18:55.

Figura 36 - wordcloud das referências à Haddad no twitter



Fonte: Própria (2018).

Figura 37 - gráfico de setores das referências à Haddad no twitter



Fonte: Própria (2018).

Figura 38 - valores exibidos pela interface da aplicação, referentes à Haddad no twitter

```
A soma dos coeficiente de repercussão é: 360724
A soma das métricas de sentimento para os 1500 comentários é:
Total de comentários positivos: 264
Total de comentários negativos: 459
Total de comentários neutros: 777
Fim da coleta da análise Haddad
```

Fonte: Própria (2018).

5. CONCLUSÃO E TRABALHOS FUTUROS

5.1. Conclusão

Este trabalho iniciou-se com a observação do ambiente plural e intrigante das mídias sociais. A forma como estão inseridas na vida da população mundial sugere a alta relevância que têm, e a utilização de ferramentas tecnológicas que extraiam valor do seu conteúdo é praticamente inevitável.

A solução tecnológica abordada é uma aplicação que, através das mídias sociais, extrai a reputação de uma entidade nesse ambiente e gera gráficos que informam o resultado.

Os ganhos proporcionados por esta análise de reputação partem da obtenção do conhecimento de fatos relacionados a um sujeito, baseados nas opiniões mais difundidas no ambiente das mídias sociais. Com isso, é possível utilizar opiniões públicas para tomar melhores decisões estratégicas.

Outro aspecto importante da aplicação é a acessibilidade que proporciona, por estar disponível gratuitamente e possuir todas funcionalidades até a visualização dos gráficos. Outras aplicações que contém todas as funcionalidades do sistema desenvolvido são menos acessíveis pelo seu alto custo de utilização.

A metodologia de desenvolvimento utilizada foi o ICONIX, para orientar o processo de elaboração da arquitetura da aplicação. O resultado é a descrição e ilustração da arquitetura em forma de diagramas, que instruem na criação dos códigos da aplicação e no seu entendimento.

Para o protótipo da aplicação desenvolvido, foi utilizado a linguagem de programação “Python”, a biblioteca “tweepy” para acesso à API do Twitter, a biblioteca “LelA” para aplicar análise de sentimento utilizando processamento de linguagem natural e as bibliotecas “matplotlib” e “wordcloud” para gerar os gráficos.

Dessa forma, o protótipo coleta comentários públicos do Twitter citando o termo que identifica a entidade alvo, classifica o sentimento (entre positivo, negativo e neutro) e o coeficiente de repercussão dos comentários, exporta esses dados de forma estruturada para um arquivo do formato “json”, gera um gráfico de setores para as porcentagens dos coeficientes de sentimento dos comentários coletados, gerar um gráfico *wordcloud* dos termos mais citados e informa os valores absolutos dos coeficientes de sentimento e repercussão.

Não foi possível desenvolver para esse protótipo, devido à complexidade, a classificação de paixão, a geração de gráfico de linha do tempo e os testes de confiabilidade da análise de sentimento. Também não foi possível expandir a coleta de comentários para monitoramento, para o perfil da entidade e para outras mídias sociais. Sua competitividade com outros produtos do tipo também pode melhorar com uma interface gráfica amigável.

Apesar disso, com pesquisa desenvolvida, modelagem elaborada com êxito e geração de um protótipo atendendo às principais funcionalidades, considera-se que o objetivo do trabalho foi alcançado. Mesmo assim, o desenvolvimento será continuado de forma colaborativa no portal de código fonte abertos chamado “GitHub”, através do *link* disponível no capítulo 4.

5.2. Trabalhos Futuros

- Implementar a classificação de paixão,
- Implementar a geração de gráfico de linha do tempo,
- Realizar testes de confiabilidade da análise de sentimento,
- Expandir a coleta de comentários para coletar em perfil da entidade,
- Expandir a coleta de comentários para manter execução de múltiplas análises ao longo do tempo, na forma de monitoramento.
- Ampliar a coleta para também acessar outras mídias sociais,
- Melhorar interface com o usuário, tornando-a mais amigável.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BIRDS, Steven; KLEIN, Ewan; LOPER, Edward. **Natural Language Processing with Python**. 1ª edição. O'Reilly Media, Inc. Sebastopol, USA. 2009.

FREEMAN, R. Edward; REED, David L. (). ***Stockholders and Stakeholders: A new perspective on Corporate Governance***. 3 ed. [S.l.]: California Management Review, Spring83, 1984. p. 88-106F

GIL, Antonio C. **Como Elaborar Projetos de Pesquisa**. 4ª edição. São Paulo: Editora Atlas S.A., 2002.

IASBECK, Luiz C. A. **Imagem e reputação na gestão da identidade organizacional**. 1ª edição. São Paulo: Organicom, 2007.

MOTTA, Daniel A. **A Liderança Essencial**. 2ª edição. São Paulo: BMI Editora e Publicações, 2017.

SOUZA, Marlo V. S. **Mineração de Opiniões Aplicada a Mídias Sociais**. PUCRS. Porto Alegre, 2012.

ZAGO, Gabriela S .; BASTOS, Marco T. **Visibilidade de Notícias no Twitter e no Facebook: Análise comparativa das notícias mais repercutidas na Europa e nas Américas**. SBPjor, Brasília, 2013.

