

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Renato Prata Stacciarini Takenaka

**Avaliação de características de mensagens
encaminhadas via WhatsApp**

Uberlândia, Brasil

2019

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Renato Prata Stacciarini Takenaka

**Avaliação de características de mensagens encaminhadas
via WhatsApp**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Rodrigo Sanches Miani

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2019

Renato Prata Stacciarini Takenaka

Avaliação de características de mensagens encaminhadas via WhatsApp

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 10 de abril de 2019:

Prof. Dr. Rodrigo Sanches Miani
Orientador

Professor

Professor

Uberlândia, Brasil
2019

À minha mãe Renata, pessoa que cuidou e batalhou muito para eu chegar aonde estou e conquistar tudo que sempre sonhei.

Agradecimentos

Quero agradecer, em primeiro lugar, a Deus, pela força e coragem durante toda esta longa caminhada e que me abençoa todos os dias com todo o seu amor.

Aos meus pais Renata e Hiromo que sempre batalharam muito para me oferecer uma educação de qualidade, ao meu irmão Raphael e aos meus familiares que sempre me incentivaram para que tudo desse certo nesta etapa da minha vida, que não mediram esforços para que eu pudesse atingir o final da graduação, e também à minha namorada Natália por sempre acreditar em mim durante essa árdua caminhada.

A todos os professores que compartilharam seus conhecimentos e experiências durante o meu processo de formação acadêmica, profissional e pessoal. Em especial eu agradeço ao meu orientador Rodrigo por todo auxílio, incentivo, dedicação e paciência durante todo o período de pesquisa e escrita desta monografia.

A todos os meus amigos de longa data por todo apoio e incentivo. Aos amigos que fiz durante a graduação e a cada colega da turma 51 do curso de Ciência da Computação por fazerem parte desta caminhada, por todas as horas de estudo e risadas juntos.

A cada um dos meus amigos mais que especiais denominados Ferinhas por toda parceria e tempo juntos. Agradeço por todos os momentos de estudos e brincadeiras, de tristezas e alegrias, de dificuldades e vitórias, de desânimo e de ânimo em que atravessamos e nos fortaleceram.

Muito obrigado a todos que participaram e contribuíram de maneira direta ou indireta durante a conclusão desta importante etapa da minha vida.

“Nós nascemos, vivemos por um breve instante, e morremos. Sempre assim aconteceu durante imenso tempo. A tecnologia não muda muito isso - se é que muda alguma coisa.” - Jobs, Steve

Resumo

O WhatsApp tornou-se uma importante ferramenta no cotidiano da população brasileira. Dados recentes apontam que o aplicativo é utilizado diariamente por 89% dos usuários de smartphones no Brasil dentro de um universo de 1,5 bilhão de usuários ativos por mês. As mensagens trocadas com a auxílio de tal aplicativo são usadas com diferentes intuítos e em diferentes contextos. O presente trabalho busca analisar diferentes características de mensagens encaminhadas usando o aplicativo WhatsApp. Foram utilizadas as linguagens de programação Java e SQL para o desenvolvimento do *script* e exportação dos dados para o banco de dados, e a ferramenta TextBlob para a análise de polaridade (neutro, positivo e negativo) das mensagens encaminhadas. O desenvolvimento do projeto foi realizado à partir da coleta de mensagens em grupos públicos do WhatsApp analisando-se as características quantitativas comuns de suas mensagens. Os resultados obtidos demonstram que as mensagens com maior número de compartilhamentos dentro do período de realização do experimento se referem, quase que na totalidade, aos principais e mais relevantes acontecimentos no Brasil. Além disso, a análise de polaridade revelou uma neutralidade nas mensagens encaminhadas.

Palavras-chave: Análise de dados, WhatsApp, Mensagens Encaminhadas, Análise de sentimentos, TextBlob.

Lista de ilustrações

Figura 1 – Linha do tempo dos principais SRSs. Adaptado de Boyd e Ellison (2007)	17
Figura 2 – Usuários ativos das principais plataformas sociais. Extraído de KEMP (2019).	18
Figura 3 – Uso de Mídia Social em Abril de 2019. Extraído de KEMP (2019).	19
Figura 4 – Crescimento anual de Mídia Social de 2019. Extraído de KEMP (2019).	20
Figura 5 – Compartilhamento de Conexões Móveis. Extraído de KEMP (2019)	21
Figura 6 – Assinaturas de Dispositivos Móveis. Extraído de Ericsson (2019)	22
Figura 7 – Significado das marcas em escala do WhatsApp. Extraído de Rashdi (2015)	23
Figura 8 – Organização da SQL. Extraído de Parreira (2018)	26
Figura 9 – Escolha do grupo para exportação	33
Figura 10 – Abrir configuração da conversa	33
Figura 11 – Exportar Conversa	34
Figura 12 – Escolha de arquivo sem mídia	34
Figura 13 – Nuvem de Palavras Total. Extraído de Davies (2015)	41
Figura 14 – Nuvem de Palavras Reduzida. Extraído de Davies (2015)	41
Figura 15 – Análise sobre o dia da difusão de mensagens	45

Lista de tabelas

Tabela 1 – Informações dos grupos do WhatsApp	38
Tabela 2 – Sentenças relacionadas às palavras	39
Tabela 3 – Nuvem de Palavras	42
Tabela 4 – Análise sobre polaridade	44

Lista de códigos

Código 3.1 – Consulta de duplicatas	36
Código 3.2 – Consulta do registro completo	36
Código 4.1 – Remoção de <i>Stop words</i>	40
Código 4.2 – Análise de polaridade	43
Código A.1 – Classe Usuario	52
Código B.1 – <i>Script</i> para exportação dos dados	54
Código C.1 – Configuração do banco de dados	60
Código D.1 – Classe JPAUtil	61

Lista de abreviaturas e siglas

BD	Banco de Dados
DBA	Database Administrator
DDL	Data Definition Language
DML	Data Manipulation Language
DQL	Data Query Language
DTL	Data Transaction Language
DCL	Data Control Language
IBM	International Business Machines
JPA	Java Persistence API
JVM	Java virtual machine
ORM	Object-Relational Mapping
PC	Personal Computer
SRSs	Serviços de Redes Sociais
SQL	Structured Query Language
SSL	Secure Sockets Layer
URLs	Uniform Resource Locator

Sumário

1	INTRODUÇÃO	13
1.1	Objetivos	14
1.2	Organização do Trabalho	14
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Histórico, definição e outros pontos sobre Redes Sociais	16
2.2	Propagação de mensagens em redes sociais	19
2.3	Funcionamento do WhatsApp	22
2.4	Tecnologias Empregadas no Desenvolvimento do Trabalho	24
2.4.1	Java	24
2.4.2	Eclipse	24
2.4.3	SQL	25
2.4.4	Hibernate/JPA	26
2.4.5	MySQL	26
2.5	Trabalhos Relacionados	27
3	DESENVOLVIMENTO	30
3.1	Requisitos para o experimento	30
3.1.1	Como coletar mensagens do WhatsApp?	31
3.1.2	Qual critério será utilizado para verificar se uma mensagem foi encaminhada?	31
3.1.3	Por quanto tempo o experimento deverá ser executado?	32
3.2	Exportação dos dados	32
3.3	<i>Script para parsing</i>	35
3.3.1	Classe Usuario	35
3.3.2	Classe <i>Parsing</i>	35
3.4	Criação do banco de dados	35
3.5	Filtro para a busca dos resultados	36
4	RESULTADOS	37
4.1	Visão geral do experimento	37
4.2	Mensagens encaminhadas	38
4.3	Nuvem de palavras	39
4.4	Análise de polaridade	42
4.5	Considerações finais	45
5	CONCLUSÃO	47

REFERÊNCIAS	48
ANEXOS	51
ANEXO A – CLASSE USUARIO	52
ANEXO B – <i>SCRIPT</i> PARA EXPORTAÇÃO DOS DADOS	54
ANEXO C – CONFIGURAÇÃO DO BANCO DE DADOS	60
ANEXO D – CLASSE JPAUTIL	61

1 Introdução

O avanço dos aplicativos instantâneos de mensagens móveis, como o WhatsApp, permitem que usuários enviem mensagens de texto em tempo real para indivíduos ou grupos de amigos sem nenhum custo. Embora exista um vasto conjunto de pesquisas sobre práticas tradicionais de mensagens de texto, pouco se entende sobre como e por que as pessoas adotaram e se apropriaram desse tipo de aplicação (CHURCH; OLIVEIRA, 2013). O presente trabalho visa estudar o comportamento das mensagens encaminhadas (retransmitidas) via WhatsApp, um exemplo de praticidade ao dia a dia onde uma grande quantidade de mensagens são enviadas diariamente que podem representar um grande auxílio no modo de comunicação da sociedade, porém, também consegue ser um risco pela facilidade em que o uso pode servir para disseminar falsos rumores.

Em 2013, ocorreu um atentado à Maratona de Boston, onde duas bombas feitas com panelas de pressão explodiram durante a Maratona de Boston, o que causou a morte de três pessoas e feriu outras 264. Usando como base esse evento, foi realizado um dos primeiros exames sobre a retransmissão de mensagens de alerta em resposta ao ataque. Considerando o conteúdo, estilo e estrutura das mensagens disseminadas, foram examinados quais recursos e características de uma mensagem concisa enviada sob condições iminentes de ameaça provocam sua retransmissão entre as pessoas do público (SUTTON et al., 2015). Ou seja, a retransmissão de mensagens é o aspecto central da difusão de informações e em um contexto de desastre, a transmissão desses avisos oficiais também servem como indicador comportamental de difusão de mensagens, sugerindo que notícias específicas são percebidas pelo público como digna e valiosa para compartilhar com os outros.

Este trabalho tem o intuito de fornecer uma análise sobre as características das mensagens encaminhadas usando a ferramenta WhatsApp no Brasil. As características estudadas são: conteúdo, tamanho, frequência e polaridade da mensagem (neutra, positiva ou negativa). Para isso, foram realizadas coletas de mensagens em grupos públicos do aplicativo. Tais dados foram anonimizados e organizados em um banco de dados relacional o que permitiu a devida análise de conteúdo, tamanho e frequência. Já a análise da polaridade ou do sentimento da mensagem foi feita com o auxílio da ferramenta Textblob. (LORIA, 2018)

Com o intuito de analisar as características e finalidades nos compartilhamentos de mensagens, foram capturados momentos de conversas de alguns grupos públicos do WhatsApp que servem como grupo de controle para a observação do comportamento das notícias enviadas e da pessoa que as enviou, armazenando-as em bancos de dados. A partir

da análise geral feita no banco de dados das informações encaminhadas, pode ser observada também a dimensão de propagação dos diversos assuntos que foram compartilhados no início do ano de 2019. Os dados podem ser analisados de acordo com o tamanho dos grupos e da capacidade de abrangência em relação aos receptores e à volatilidade dos assuntos tratados, tais como política – principalmente pela época em que se deu a pesquisa, variações na economia, notícias de famosos, e, inclusive, a relevância e volume de mensagens simples que podem até serem consideradas spam por se tratarem de correntes.

A partir dos resultados obtidos, foi possível concluir que boa parte dos usuários tendem a encaminhar os principais e mais relevantes acontecimentos no Brasil. Há também uma maior quantidade de informações consideradas neutras, classificadas pela ferramenta TextBlob. Além disso, o turno de maior incidência de mensagens retransmitidas foi no período da manhã, o que provavelmente também indica que nesse período, há uma maior circulação de notícias nas redes sociais.

1.1 Objetivos

O objetivo geral deste trabalho consiste em analisar e extrair características quantitativas comuns das mensagens encaminhadas (tamanho, número de palavras, frequência de palavras), além de sua polaridade, para assim mensurar os impactos que essas informações provocam ao serem disseminadas. Para o estudo desse problema é considerado o próprio teor das informações compartilhadas.

O objetivo geral pode ser dividido nos seguintes objetivos específicos:

- Foi desenvolvido um algoritmo para realização do *parsing* do arquivo .txt exportado da conversa do WhatsApp em informações que possam ser inseridas no banco de dados;
- Foi criada uma base de dados real de mensagens encaminhadas. A base foi construída a partir de mensagens enviadas de grupos públicos e privados do WhatsApp, sendo realizado no período de um mês, ou seja, compreendido entre o mês de janeiro e fevereiro de 2019;
- Compartilhar a base de dados com a comunidade científica;
- Foram realizadas análises dos resultados obtidos através do estudo das características quantitativas comuns das mensagens encaminhadas.

1.2 Organização do Trabalho

O restante deste trabalho é organizado da seguinte maneira.

- Capítulo 2 fornece a fundamentação teórica necessária para o desenvolvimento e compreensão do estudo é apresentada, incluindo uma visão geral sobre redes sociais, principalmente o WhatsApp, e sua importância no mundo contemporâneo, além da descrição de conceitos e citação de trabalhos relacionados;
- Capítulo 3 expõe os métodos utilizados para o desenvolvimento do trabalho, descrevendo todos os requisitos necessários e todo o processo realizado para a execução da exportação da conversa do WhatsApp, realizando o *parsing* para o banco de dados com o objetivo de analisar as características quantidades comuns das conversas encaminhadas;
- Capítulo 4 detalha os resultados obtidos a partir do filtro e classificação sobre as mensagens encaminhadas;
- Capítulo 5 mostra as principais contribuições deste trabalho, além de sugestões de trabalhos futuros.

2 Fundamentação Teórica

Aplicativos de trocas de mensagens instantâneas permitem a união de pessoas com o interesses mútuos para a realização de várias atividades como o compartilhamento de informações, divulgação, educação e até o comércio.

Esta seção expõe assuntos relativos ao Histórico, definição e outros pontos sobre Redes Sociais (2.1), Propagação de mensagens em redes sociais (2.2), Funcionamento do WhatsApp (2.3), Tecnologias Empregadas (2.4), a qual contém subseções explicando um pouco sobre a linguagem Java (2.4.1), a IDE Eclipse (2.4.2), a linguagem SQL (2.4.3), o framework Hibernate e a interface JPA (2.4.4) e o gerenciador de banco de dados MySQL (2.4.5), e por fim é exemplificado alguns Trabalhos Relacionados (2.5).

2.1 Histórico, definição e outros pontos sobre Redes Sociais

Para [Boyd e Ellison \(2007\)](#), as redes sociais são serviços baseados na Web que permitem aos indivíduos construir um público ou semi-público, dentro de um sistema limitado, articular uma lista de usuários com quem eles compartilham uma conexão, visualizar e percorrer esta lista de conexões e aquelas feitas por outros dentro do sistema. A natureza e nomenclatura dessas conexões podem variar de site para site.

O que faz os serviços de redes sociais únicos (SRSs) não são que eles permitem que os indivíduos se relacionem com pessoas desconhecidas, mas sim que eles permitem os usuários articularem e tornarem visíveis as suas redes sociais. Isso pode resultar em conexões entre indivíduos que de outra forma não são feitas, mas que muitas vezes não é o objetivo. De acordo com [Consalvo et al. \(2004\)](#), em muitos dos grandes SRSs, os participantes não estão necessariamente procurando conhecer novas pessoas; em vez disso, eles estão se comunicando principalmente com pessoas que já fazem parte de sua rede social, facilitando assim a comunicação no dia a dia.

Como discutido em [Boyd e Ellison \(2007\)](#), a exibição pública de conexões é um componente crucial dos SRSs. A lista de amigos contém links do perfil de cada amigo, permitindo que outros usuários possam atravessar o gráfico de rede clicando através dessa lista. Na maioria dos sites, a lista de amigos é visível para quem está autorizado a visualizar o perfil, embora haja exceções. Além disso, fornecem um mecanismo para que os usuários possam deixar mensagens para seus amigos. Esta característica pode ser tanto positiva quanto negativa, dependendo de sua polaridade. Há também um recurso de mensagem privada semelhante ao *webmail*. Embora ambas as mensagens privadas e pública serem populares na maioria dos principais SRSs, elas não são universalmente disponíveis.

Os SRSs variam muito em suas características e base de usuários. Alguns têm de compartilhamento de fotos ou de compartilhamento de vídeo; outros apresentam um sistema integrado de *blogging* e tecnologia de mensagens instantâneas. Muitos serviços são também específicos de regiões geográficas ou grupos linguísticos, embora isso nem sempre determina a circunscrição do site. Orkut, por exemplo, foi lançado nos Estados Unidos com uma interface somente em inglês, mas os brasileiros de língua portuguesa tornaram-se rapidamente o grupo de usuários dominante (KOPYTOFF, 2014). A Figura 1 mostra a linha do tempo dos principais serviços de redes sociais de 1997 até 2006.

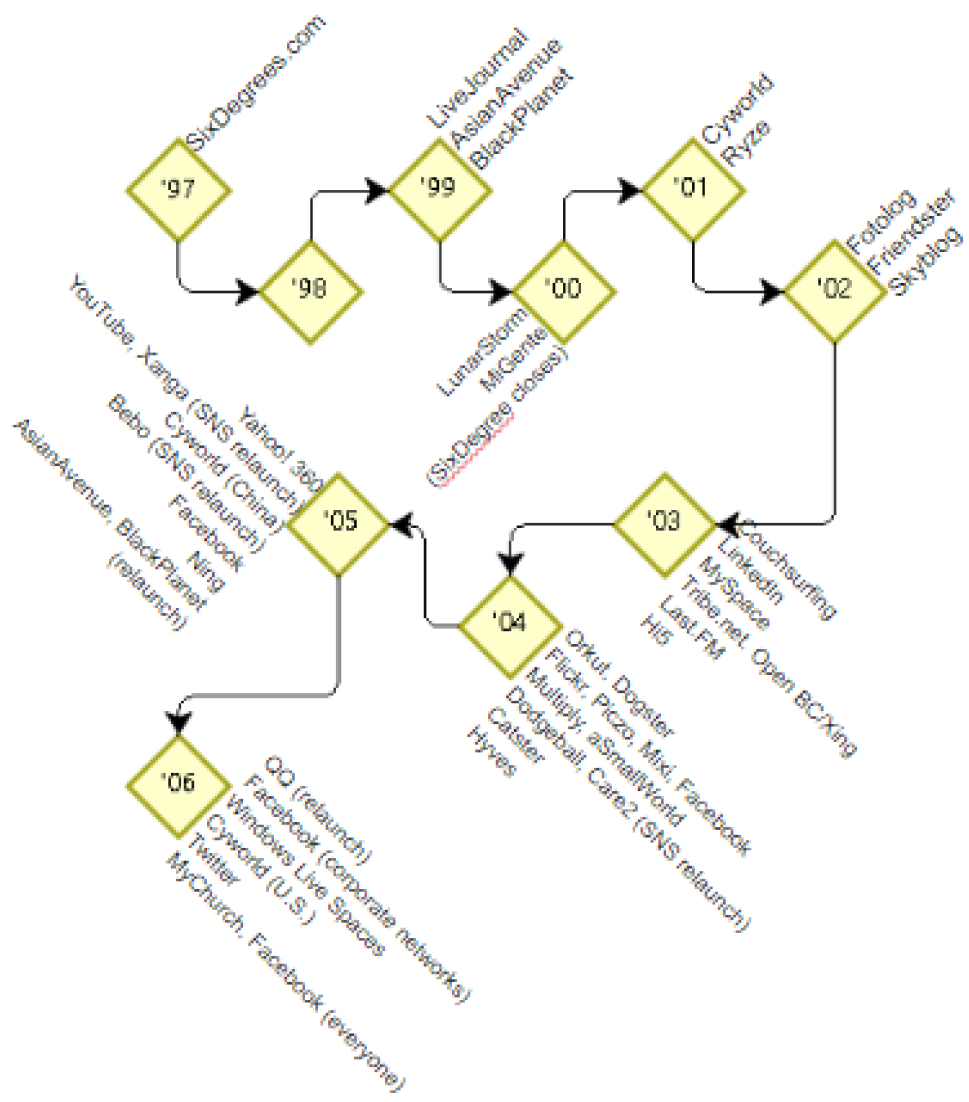


Figura 1 – Linha do tempo dos principais SRSs. Adaptado de Boyd e Ellison (2007)

Toda essa transformação e ascensão dos SRSs indica uma mudança na organização de comunidades on-line. Enquanto sites dedicados a comunidades de interesse ainda existem e prosperam, SRSs são organizados principalmente em torno de pessoas, e não de interesses. São estruturados como redes pessoais, com o indivíduo no centro da sua própria comunidade. Isto reflete com mais precisão estruturas sociais não-mediadas, em

que o mundo é composto de redes, não grupos (WELLMAN; BERKOWITZ, 1988). A introdução de características dos SRSs introduziu um novo quadro organizacional para comunidades on-line e, com isso, um novo contexto de pesquisa.

Passados anos de crescimento dos serviços de redes sociais, explorar os últimos números de público-alvo de publicidade de cada plataforma oferece uma perspectiva valiosa sobre o uso de mídia social. Portanto, é essencial observar os principais dados de desenvolvimentos e tendências nos últimos três meses. No geral, os serviços da Facebook Inc. continua a dominar o panorama social global, com o forte crescimento do sistema WhatsApp, como mostrado na Figura 2, pesquisa realizada por KEMP (2019) em abril de 2019,.

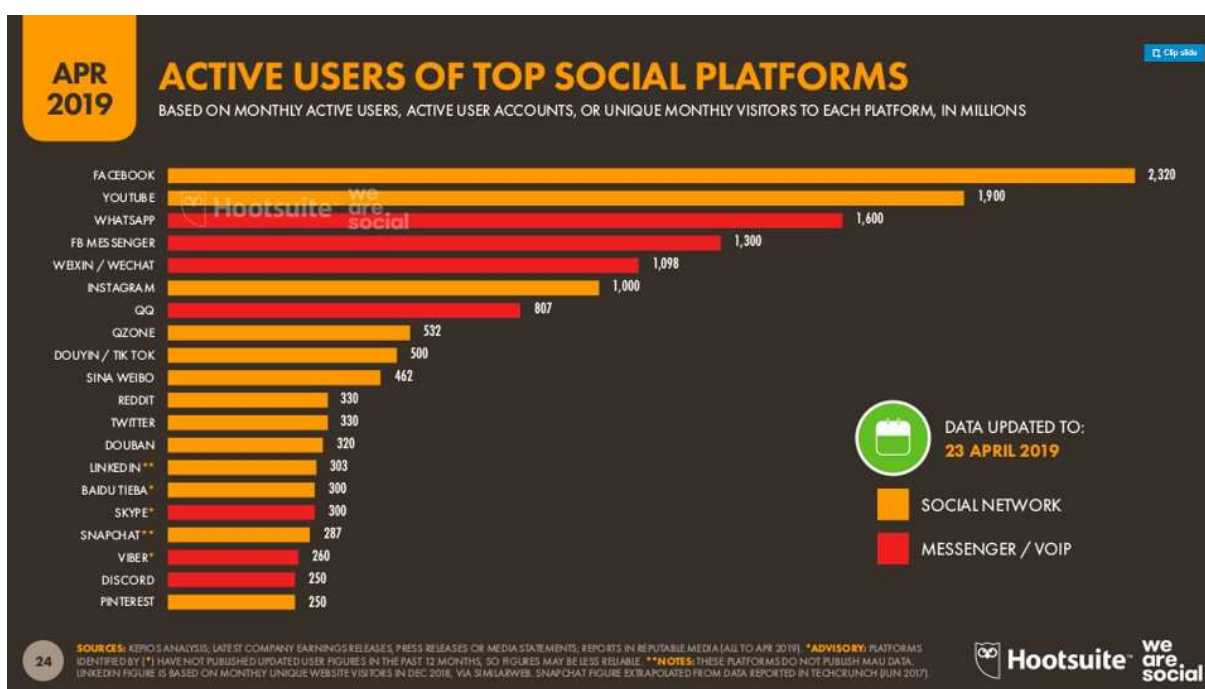


Figura 2 – Usuários ativos das principais plataformas sociais. Extraído de KEMP (2019).

De acordo com Siqueira (2013), as políticas de privacidade são meios nos quais os sites se protegem do mau uso de seus usuários, ou mesmo dos que agem com más intenções, é um contrato eletrônico no qual quem concorda se responsabiliza pelo que está escrito, isto é, a segurança com o usuário, muitas pessoas assinam sem ler imaginando que pode não haver algo sério, essa prática não é recomendada, pois não deve se assinar contratos ou concordar em dividir suas informações ou participar de algo sem antes conhecer todas as cláusulas. Para a segurança do próprio aplicativo ou página são aplicados outros tipos de estratégias e tecnologias, tais como SSL, que permite a troca de informações entre servidor e cliente com grande segurança.

Ou seja, os SRSs são ferramentas que unem pessoas e informações por meio de códigos de dados digitais, e são comumente utilizadas para que indivíduos se expressem

à sua maneira, garantindo sua liberdade dentro do que estiver na política de privacidade de cada endereço eletrônico. Empresas também fazem uso para disseminar seu trabalho, sendo a internet um ótimo lugar para divulgar notícias e direcionar propagandas (inclusive direcionadas com uso de *cookies*), entretanto nem tudo são maravilhas e às vezes a segurança desses *websites* podem ser quebradas e os mesmos invadidos, havendo perda de informações, sendo estas muito valiosas.

2.2 Propagação de mensagens em redes sociais

As plataformas de mídia social forneceram aos usuários um meio de comentar suas opiniões sobre qualquer assunto e torná-lo uma tendência. Um mero conjunto de páginas da Web que permitia que as pessoas criassem perfis, compartilhassem fotos e pensamentos, conectassem-se com amigos em comum, expressassem emoções por meio de emoticons - tudo isso se transformou em grandes oportunidades de engajamentos. E agora há uma era de marcas participando dessas interações para ganhar a confiança e a fidelidade de seus clientes por meio de plataformas de mídia social.

De acordo com [KEMP \(2019\)](#), a figura dos usuários ativos de mídia social atingiu 3,499 bilhões, o que representa 45% da população total do mundo. Isso indica um aumento de 6% em relação ao ano anterior. Mas não é isso, 3.429 bilhões de pessoas que são 45% do total da população acessam as mídias sociais através de dispositivos móveis. É um sinal claro para os principais serviços de mídias sociais otimizarem os aplicativos sociais de tempos em tempos com base no comportamento e na preferência dos usuários.



Figura 3 – Uso de Mídia Social em Abril de 2019. Extraído de [KEMP \(2019\)](#).

Através de todo o avanço tecnológico e pesquisas realizadas pela [KEMP \(2019\)](#), é possível perceber que na Figura 3, mediante os dados informados, uma média de quase 1 milhão de pessoas entrou na Internet pela primeira vez a cada dia no ano passado até os dias atuais, e ainda continuando com o forte crescimento.

Comparando os dados da Figura 3 e Figura 4, também é possível perceber que:

- Os usuários globais da Internet cresceram 8,6% nos últimos doze meses, com 350 milhões de novos usuários contribuindo para um total de 4,437 bilhões até o início de abril de 2019.
- Os números de usuários de mídias sociais também registraram um sólido crescimento, aumentando em mais de 200 milhões desde o ano passado, para chegar a quase 3,5 bilhões no momento da publicação. No entanto, mudanças recentes na metodologia de relatórios do Facebook significam que o crescimento real foi provavelmente ainda maior.
- Existem hoje mais de 5,1 bilhões de pessoas em todo o mundo usando um telefone celular - um aumento ano a ano de 2,6% - com os smartphones respondendo por mais de dois terços de todos os dispositivos em uso atualmente.
- Aproximadamente 98% dos usuários de mídia social do mundo - mais de 3,4 bilhões de pessoas - acessam plataformas sociais por meio de dispositivos móveis.

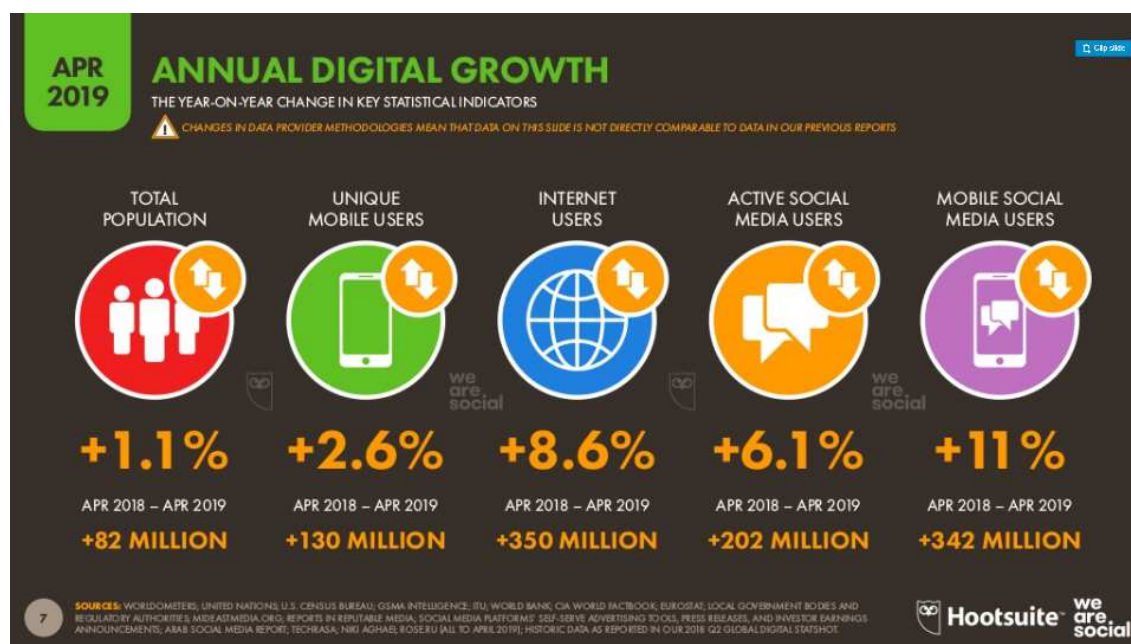


Figura 4 – Crescimento anual de Mídia Social de 2019. Extraído de [KEMP \(2019\)](#).

Através dos dados da Figura 4, observa-se que a propagação de mensagens aumentou muito com a criação dos aplicativos para aparelhos móveis, e o desenvolvimento da

tecnologia nos últimos tempos acelerou o compartilhamento de informações e como estas são compartilhadas de forma geral, com poucos cliques no mouse ou pressionar de dedos (como no WhatsApp por exemplo) pode-se enviar algo para quantas pessoas desejar, o que a princípio parece limitado se expande com a capacidade de retransmissão de dados, a capacidade de dissipação de mensagens entre usuários é ilimitada. Como dito em ??), o aplicativo já ultrapassou a marca de mais de 1 bilhão de mensagens enviadas em um único dia, ou seja, 41,666,667 mensagens por hora, 694,444 mensagens por minuto e 11,574 por segundo.

Conforme os dados mais recentes de [Ericsson \(2019\)](#), os smartphones são responsáveis por 68% das conexões móveis, com a participação de telefones celulares diminuindo abaixo de 30% pela primeira vez, demonstrado também na Figura 5.

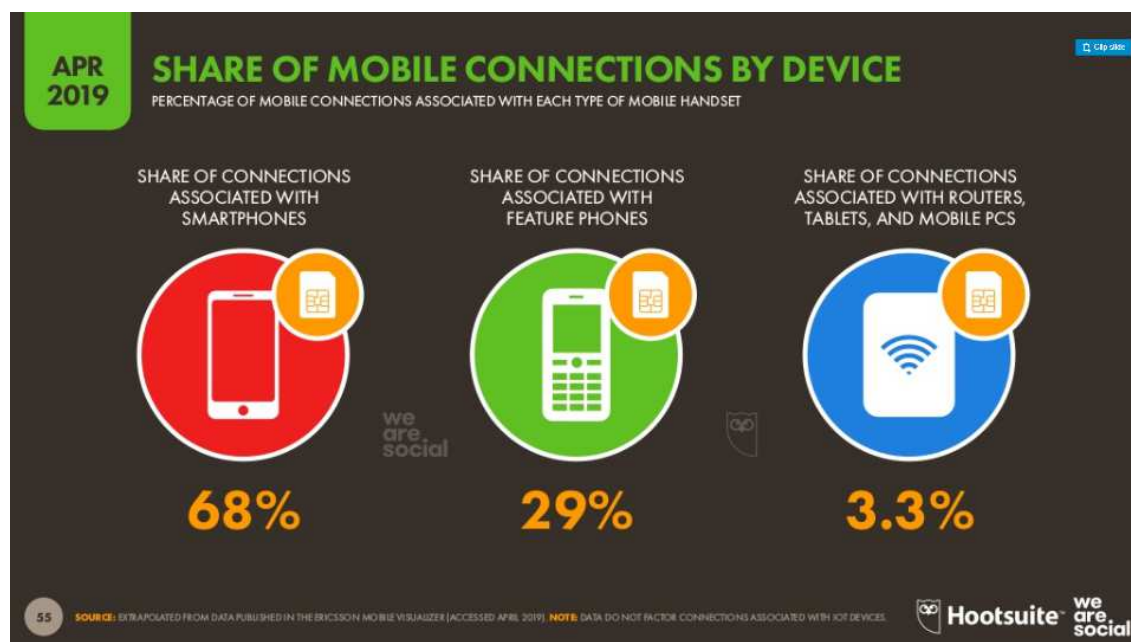


Figura 5 – Compartilhamento de Conexões Móveis. Extraído de [KEMP \(2019\)](#)

Ou seja, com esse grande número de mensagens enviadas por dia utilizando esse tipo de aplicação, a possibilidade de compartilhamento facilitada com o encaminhamento de mensagens, e os dados presentes na Figura 6, é possível que esse número já tenha aumentado e continue a crescer, o que pelo lado negativo também afeta o número de falsas informações compartilhadas.

Mobile subscriptions

Unit: Million

Source: Ericsson (Feb 2019)

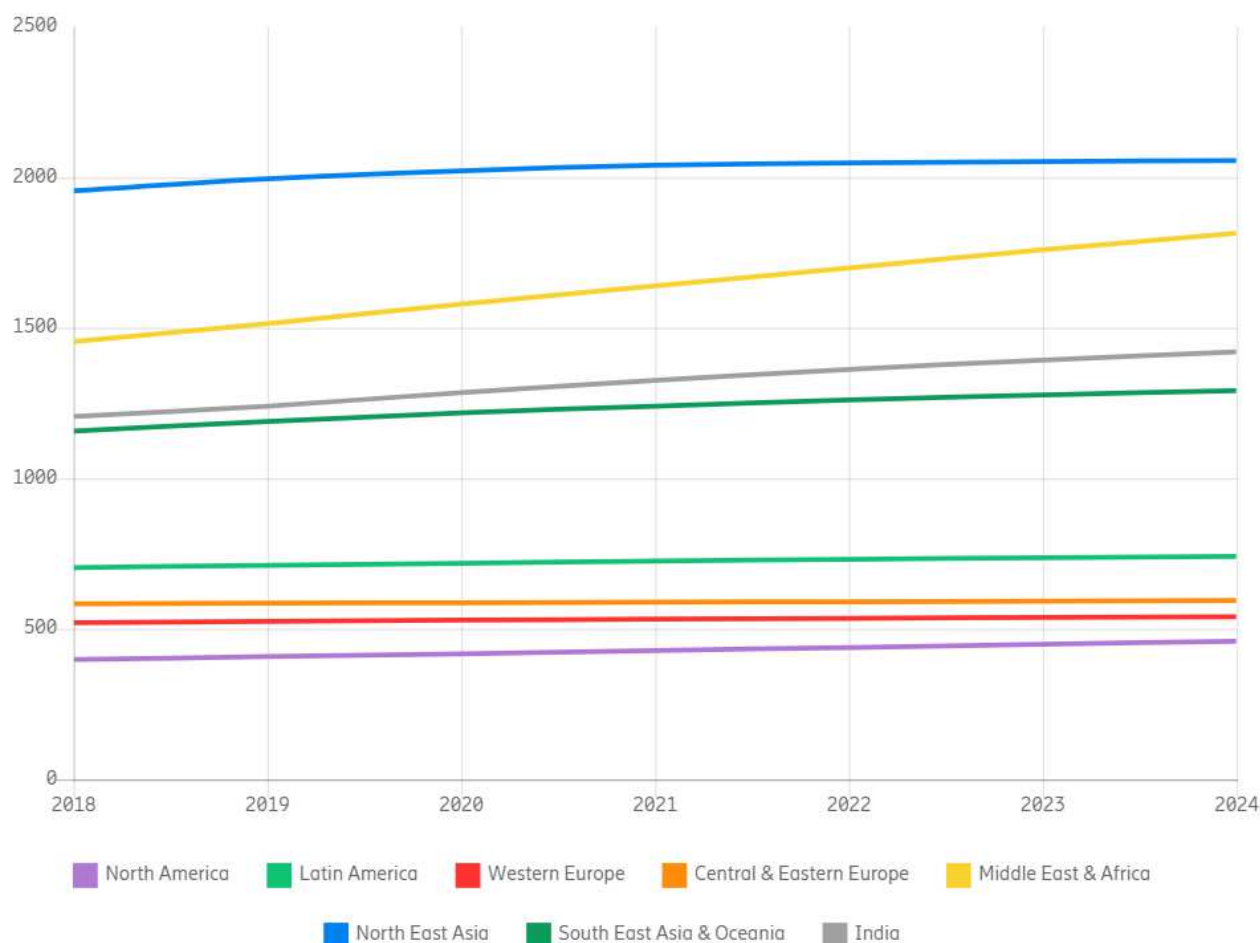


Figura 6 – Assinaturas de Dispositivos Móveis. Extraído de [Ericsson \(2019\)](#)

2.3 Funcionamento do WhatsApp

WhatsApp é um aplicativo para smartphones que opera em quase todos os tipos atuais de dispositivos e sistemas operacionais. Ele inclui uma variedade de funções, tais como mensagens de texto, imagens anexadas, arquivos de áudio, arquivos de vídeo e links para endereços da web. Ao longo dos últimos anos, o aplicativo tornou-se muito popular. Tecnicamente, WhatsApp pode ser visto como uma rede social que permite às pessoas acessarem uma grande quantidade de informações rapidamente. De acordo com [Bouhnik, Dshen e Gan \(2014\)](#), ele permite a comunicação com qualquer um que possua um smartphone, tem uma ligação ativa à Internet para registrar-se no aplicativo e executar ações no mesmo, ao realizar o registro do número de telefone consegue-se contato com uma interface simples e intuitiva com variadas opções, dentre elas criação de grupos, opção de pesquisa de palavras e pessoas, configurações, entre outras.

O primeiro contato principal com a interface é com as conversas realizadas com outros usuários (comunicação pessoa a pessoa). Também é possível acessar nos contatos telefônicos alguém que possua o aplicativo para realizarem conversa por chat, chamada simples ou vídeo chamada, para a conversa por chat também é permitido o envio de áudios que aumentam a dinâmica de uso. O compartilhamento de imagens, vídeos, GIFs, figurinhas e emojis também são algo que pode tornar informal e divertir a comunicação entre os usuários, com a principal função de conseguir reduzir mensagens longas para uma simples interpretação de imagem.

Conforme explicado em [Rashdi \(2015\)](#), outra característica do WhatsApp, é que ele permite que seus usuários possam ver quando seus amigos estão online, ou a última vez que acessou o aplicativo. Outro recurso interessante, é o de notificação, indicando quando uma mensagem é entregue e quando ela é lida. Uma marca de verificação cinza ao lado da mensagem significa que a mensagem foi enviada de telefone do remetente, mas ainda não foi recebida no telefone do destinatário; duas marcas de verificação cinza significa que a mensagem foi entregue ao telefone do destinatário, mas ainda não foi lida, e essas marcas de seleção fica azul, uma vez que o destinatário lê a mensagem. Um carimbo de tempo aparece ao lado destas marcas de seleção que indicam o tempo que a mensagem foi recebida como ilustrado na Figura 7:



Figura 7 – Significado das marcas em escala do WhatsApp. Extraído de [Rashdi \(2015\)](#)

Uma opção que surgiu através de sua tentativa de interação com o Snapchat é a permissão de postar fotos temporárias com duração de 24 horas a partir da postagem, e de enviar diretamente para pessoas ou grupos, inclusive. O processo de criação de grupos é bem intuitivo e permite privilégios de administrador a quem cria ou é delegada a função, esta torna possível a inclusão e exclusão de mais pessoas no grupo sem a necessidade de aprovação por parte dos demais membros. Com o tanto de estilos para mensagem propostos, há também a opção de silenciar o recebimento dessas mensagens individualmente (para um grupo ou pessoa específico), com duração de 8 horas, uma semana ou um ano, além da opção de apagar mensagens ou arquivá-las.

2.4 Tecnologias Empregadas no Desenvolvimento do Trabalho

2.4.1 Java

No início dos anos 90, um pequeno grupo de engenheiros da Sun Microsystems (fabricante de computadores) chamado de Green Team, acreditava que a próxima etapa na computação seria a união de dispositivos digitais e computadores. Liderada por James Gosling, a equipe trabalhou sem parar e criou a linguagem de programação que revolucionaria nosso mundo, o Java.

A Green Team demonstrou seu novo idioma com um controlador de entretenimento doméstico portátil e interativo, originalmente voltado para a indústria de televisão digital a cabo. Infelizmente, o conceito estava muito avançado para a equipe na época. Mas foi o ideal para a Internet, que estava apenas começando a decolar. Em 1995, a equipe anunciou que o navegador de *Internet Netscape Navigator* incorporaria a tecnologia Java (JOHARI, 2018).

Java é uma linguagem de programação que possui sua própria estrutura, regras de sintaxe e paradigma de programação, baseando-se na orientação a objetos. Estruturalmente, a linguagem Java possui classes, métodos, variáveis, atributos, etc. Utilizando a plataforma Java, o código-fonte irá formar arquivos .java para depois serem compilados. O compilador verifica o código nas regras de sintaxe da linguagem e depois grava bytecode em arquivos .class. Bytecode é um conjunto de instruções destinadas a executar em uma JVM, que em tempo de execução, lê e interpreta os arquivos .class e executa as instruções do programa.

2.4.2 Eclipse

O programa utilizado foi o Eclipse, nele foram escritos os códigos de programação necessários para captar as conversas dos grupos no período de tempo em que foi realizada a coleta de dados. A Eclipse Project foi originalmente criado pela IBM em novembro de 2001, se tornando uma organização sem fins lucrativos em janeiro de 2004, a fundação oferece serviços como IP management, Ecosystem Development, Development Process, IT Infrastructure, cada um com suas especialidades.

O primeiro serviço é o “IP Management”, este representa a capacidade dos usuários do programa para usar a tecnologia open source (código aberto), que significa sem a cobrança de licença de utilização e valores sobre propriedade intelectual ou direito autoral. “Ecosystem Development” engloba os produtos produzidos a partir da comunidade Open Source do Eclipse, que forma o nome de ecossistema e serve para promover tudo relacionado ao Eclipse, como treinar os provedores de serviço, revistas e livros, portais online, entre outros.

O *Development Process* se certifica que todos os projetos pela Eclipse sejam abertos, transparentes e sigam um comportamento meritocrático, organizando inclusive uma comunidade para que usuários se ajudem. *IT Infrastructure* é a parte que verifica os códigos, possuindo repositórios, fóruns sobre desenvolvimento, rastreador de bugs, entre outros, e foi designada para auxiliar com um serviço confiável e escalável, citado em [Eclipse \(2019\)](#).

2.4.3 SQL

Desenvolvido no início dos anos 70 na IBM por Raymond Boyce e Donald Chamberlin, o SQL foi lançado comercialmente pela Relational Software Inc. (agora conhecida como Oracle Corporation) em 1979. A versão SQL padrão atual é voluntária, compatível com o fornecedor e monitorada pela American Instituto Nacional de Padrões (ANSI).

Conforme [Techopedia \(2019\)](#), a *Structured Query Language* é uma linguagem de programação para gerenciamento de banco de dados relacional e manipulação de dados. O SQL é usado para consultar, inserir, atualizar e modificar dados. A maioria dos bancos de dados relacionais oferece suporte ao SQL, o que é um benefício adicional para os administradores de bancos de dados (DBAs), pois eles geralmente são necessários para oferecer suporte a bancos de dados em diversas plataformas diferentes.

A linguagem SQL é dividida em tipos de acordo com a funcionalidade dos comandos, demonstrado também na Figura 8:

- DDL - Data Definition Language - Linguagem de Definição de Dados. São os comandos que interagem com os objetos do banco.

São comandos DDL : CREATE, ALTER e DROP

- DML - Data Manipulation Language - Linguagem de Manipulação de Dados. São os comandos que interagem com os dados dentro das tabelas.

São comandos DML : INSERT, DELETE e UPDATE

- DQL - Data Query Language - Linguagem de Consulta de dados. São os comandos de consulta.

São comandos DQL : SELECT

- DTL - Data Transaction Language - Linguagem de Transação de Dados. São os comandos para controle de transação.

São comandos DTL : BEGIN TRANSACTION, COMMIT E ROLLBACK

- DCL - Data Control Language - Linguagem de Controle de Dados. São os comandos para controlar a parte de segurança do banco de dados.

São comandos DCL : GRANT, REVOKE E DENY.

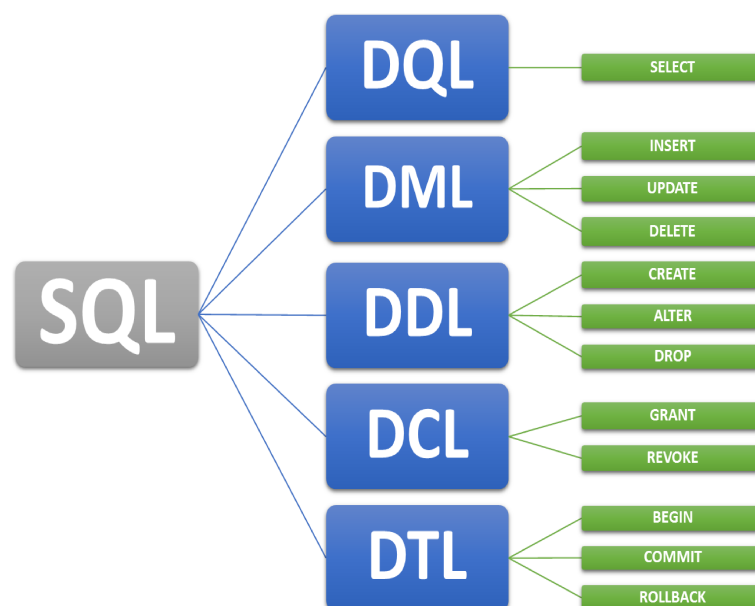


Figura 8 – Organização da SQL. Extraído de [Parreira \(2018\)](#)

2.4.4 Hibernate/JPA

Neste trabalho foi utilizado uma aplicação Java seguindo o paradigma de orientação a objetos, em que informações são representadas por classes e atributos, sendo necessário uma ferramenta para aumentar a produtividade na codificação de consultas em linguagem SQL. De acordo com [Caelum \(2011\)](#) Hibernate é um framework ou ferramenta de mapeamento objeto-relacional (ORM) utilizada para auxiliar e ajudar nesta tarefa de "transformar" os objetos em registros para o banco de dados e registros em objetos, ou seja, ele vai gerar o código SQL através de implementações utilizando a JPA que é uma interface que define um meio de mapeamento objeto-relacional para frameworks de persistência de dados.

2.4.5 MySQL

De acordo com [Date \(2004\)](#) “sistema de banco de dados pode ser considerado como uma sala de arquivos eletrônica”, o qual pode ser conceituado como um conjunto de hardware, software, dados e usuários, e nada mais é do que uma coleção de dados que se relacionam entre si com uso de alguma interface. E de acordo com [Silberschatz, Sundarshan e Korth \(2016\)](#) “uma coleção de dados inter-relacionados, representando informações sobre um domínio específico”, fazendo com que as informações armazenadas possuam um valor agregado para que possam ser utilizadas a qualquer momento quando necessárias, trazendo um leque de opções similares consigo.

O programa utilizado para o gerenciamento de banco de dados e para o armazenamento das informações necessárias para o desenvolvimento do trabalho foi o MySQL. Um sistema de gerenciamento de banco de dados (SGBD) possui os recursos necessários para a manipulação dos bancos de dados, são comumente utilizados por permitir uma interação facilitada com o usuário (RICARDO, 2006).

2.5 Trabalhos Relacionados

Esta seção tem como finalidade descrever trabalhos relacionados e que influenciaram na dissertação desta monografia. Os trabalhos apresentados por Shi et al. (2017), Bavelas (1948), Shaw (1954), Shimbil (1953), Abu-Shaqra (2014), Rivera (2015), Koohikamali e Kim (2015), Lee, Agrawal e Rao (2015) e Sutton et al. (2015) serão apresentados a seguir. Ao final, será exposto brevemente a relação destes trabalhos com a proposta deste trabalho.

Bavelas (1948) e Shaw (1954) descrevem que uma pessoa é central nas comunicações quando se liga a outras por estarem estrategicamente localizada entre elas, e que com essa posição pode influenciar o grupo distorcendo ou retendo informações. Pessoas com esse poder centralizado são detentoras de grande responsabilidade e, mesmo uma opinião pessoal sobre determinado assunto tem possibilidade de se transformar em notícia – caso de digital influencers e bloggers – mesmo sem intenção. Desde o início de civilizações pessoas seguem e são influenciadas por outras através da comunicação e outros gestos. Shimbil (1953) enfatiza a responsabilidade das pessoas que ocupam posições relevantes no espectro da comunicação. Já Bavelas (1948) argumenta que mensagens distribuídas pela pessoa mais centralizada na rede de comunicações tendem a se espalhar no menor tempo possível, sugestão lógica com base na proximidade do conceito sobre o ponto de centralidade.

Com isso, o peso da importância que as redes de transmissão de mensagens possui aumenta no caso de disseminação de notícias falsas ou modificadas. A brincadeira do “telefone sem fio” consegue descrever bem, em seu modo, como mesmo sem querer há alterações na conotação diferente que uma informação toma ao longo dos níveis em que é transmitida, quanto mais pessoas atinge, quando mais níveis avança se espalhando, já levou a muitas interpretações que podem nem ter fidelidade à mensagem original.

O tema abordado por Abu-Shaqra (2014), faz uma pequena síntese do livro de Rivera (2015), demonstrando que os SRSs possuem em seus pontos positivos e negativos, a condição em dar poder para o usuário ao mesmo tempo em que o expõe. O autor discute que esse é o preço que se paga por estar conectado à rede e exercer seu livre arbítrio. Contudo, ao mesmo tempo que essa exposição pode auxiliar no momento de detectar quem espalha notícias de cunho duvidoso, por outro lado pode ser atrapalhada

pela velocidade com que a notícia pode propagar, dependendo da pessoa centralizada na distribuição.

De acordo com [Shi et al. \(2017\)](#) o comportamento de cada indivíduo desempenha um papel fundamental na propagação de informações sobre mídia social. Foi realizado uma análise de um conjunto de dados do painel do Twitter, que consiste de 440 usuários válidos em seis anos, indicando que todos os principais componentes de um processo de comunicação social têm impactos significativas sobre o *retweeting* individual, dentre eles: fonte de informação (por exemplo, confiabilidade da fonte), a riqueza da informação, a relação entre a fonte e o receptor e os fatores contextuais (por exemplo, efeito de movimento). Entre eles, a força da ligação e da relevância social com o receptor são os fatores mais importantes, seguido pela riqueza da informação e dos fatores contextuais.

[Koohikamali e Kim \(2015\)](#) explica que os SRSs permitem que seus usuários criem conteúdo e compartilhem com os outros. Durante a crise social ocorrida em Baltimore em 2015, a propagação de informações falsas e verdadeiras poderiam ter impactos profundos sobre os usuários. A falta de estudos para comparar as diferenças entre os padrões de difusão de rumores e verdades durante a crise social seria a motivação do estudo realizado. Neste estudo, foi examinado o papel de credibilidade da informação, a ansiedade, o envolvimento pessoal, e os laços sociais sobre rumor e verdade que se espalharam durante os protestos. Foi proposto um modelo de pesquisa para examinar as diferenças entre disseminação de boatos e verdades. Para a realização de testes utilizando o modelo de pesquisa, foram usados dados do Twitter, que foram coletados durante os motins de Baltimore em 2015.

O trabalho de [Lee, Agrawal e Rao \(2015\)](#) contribuiu com a adaptação da inovação do modelo de difusão explorando a difusão de mensagens no sistema Twitter durante eventos extremos. Mostra que o Twitter, sendo este um dos principais sistemas de rede social, desempenha um papel significativo nos meios de comunicação, especialmente durante os eventos extremos. Foram examinados impactos de características de tweets sobre a difusão de dois tipos de mensagens em 2013: rumores relacionados e não relacionados à tragédia da maratona em Boston. Foi executado uma análise binomial negativa, revelando que as características do tweet, tais como tempo de reação, número de seguidores, e uso de hashtags, possuem um impacto significativo sobre a difusão de notícias sobre a tragédia. O número de seguidores apresentou uma correlação positiva, no entanto, a relação entre o tempo de reação e difusão de mensagem foi negativa. Os tweets que não incluíam hashtags foram mais difundidas que mensagens que continham hashtags.

A retransmissão da mensagens é um aspecto central da difusão da informação. Em um contexto de desastres, a transmissão de mensagens de aviso oficiais por membros do público também serve como um indicador comportamental de mensagem de relevância, o que sugere que as notícias particulares são (ou não são) percebidas pelo público para ser

notável e valioso o suficiente para compartilhar com outros. O estudo realizado por [Sutton et al. \(2015\)](#) fornece o primeiro exame de retransmissão de avisos oficiais em resposta a um ataque terrorista, o bombardeio à Maratona de Boston em 2013. Usando mensagens postadas de funcionários públicos de contas do Twitter que estavam ativas durante o período do bombardeio, foram examinadas as características das notícias que estão associadas com a sua retransmissão. Para saber quais características de uma mensagem concisa enviadas em condições de ameaça iminente, as quais causam sua retransmissão entre os membros do público, foram considerados o conteúdo da informação, estilo, estrutura, e também relações em rede de remetentes de mensagens. Empregando o modelo binomial negativo, a conclusão obtida foi que todos esses fatores influenciava conjuntamente na retransmissão de tweets oficiais durante a resposta do bombardeio de Boston.

Apesar dos trabalhos propostos conduzirem estudos semelhantes em relação a disseminação de informações ou rumores em condições extremas, ou de ameaça iminente, todos analisavam e comparavam como são realizadas as difusões dessas mensagens em sistemas de redes sociais, ou se estão sujeitas a maiores manipulações, o que, de maneira diferente, se assemelha com este estudo em relação aos objetivos buscados.

A principal diferença entre os trabalhos estudados com relação a esta monografia, é que neste estudo é analisado dados da exportação de uma conversa do sistema WhatsApp. Com relação a metodologia, o trabalho que mais se assemelha a essa monografia é o trabalho proposto por [Shi et al. \(2017\)](#), o qual também foi utilizado a ferramenta TextBlob para análise de polaridade das mensagens. Contudo, será utilizado outro conjunto de métricas e análises estatísticas, como também um *script* pessoal desenvolvido por mim. Assim, será analisado quais as características quantitativas comuns das mensagens encaminhadas pelo WhatsApp, diferente dos demais trabalhos estudados nessa secção.

3 Desenvolvimento

Neste capítulo são apresentados todos os passos que foram seguidos para o desenvolvimento deste trabalho, desde os métodos a serem seguidos, a escolha do programa a ser utilizado até a criação dos *scripts*, do experimento e por fim a análise dos resultados encontrados.

A Seção 3.1 detalha todo o processo e todos os requisitos necessários para a realização do desenvolvimento e experimento do trabalho.

A Seção 3.2 descreve como foi realizada a coleta das mensagens do WhatsApp para a realização do *parsing* e posteriormente compor a base de dados utilizada para o estudo.

A Seção 3.3 apresenta e explica toda a construção do algoritmo/*script* para a realização do *parsing* dos arquivos .txt exportados do WhatsApp em informações para serem inseridas no banco de dados.

A Seção 3.4 explica a criação do banco de dados através do framework Hibernate utilizando a Java Persistence API (JPA), que são usados para facilitar e auxiliar na tarefa de gerar todo o código SQL, ou seja, construção da tabela, linhas e colunas necessárias para a produção do banco de dados utilizado.

A Seção 3.5 expõe a consulta utilizada para a busca das mensagens iguais/semelhantes encaminhadas.

3.1 Requisitos para o experimento

Esta seção consiste em explicar e detalhar os critérios e requisitos necessários para a realização da análise dos impactos de fatores extraídos da difusão de mensagens do WhatsApp. Para isso, será necessário levar em consideração os principais pontos para o experimento:

- Como coletar mensagens do WhatsApp?
- Qual critério será utilizado para verificar se uma mensagem foi encaminhada?
- Por quanto tempo o experimento deverá ser executado?

Cada um desses itens será discutido nas subseções a seguir.

3.1.1 Como coletar mensagens do WhatsApp?

Publicado pelo próprio [WhatsApp \(2009\)](#), um dos aspectos mais marcantes e utilizado durante todo o trabalho é o surgimento da opção de exportação do histórico de uma conversa individual ou de grupo específicos, a partir do qual será criado um documento .txt, através de duas opções:

- Se você escolher anexar mídia, os arquivos de mídia mais recentes serão adicionados separadamente à mensagem.
- Quando você opta por adicionar arquivos de mídia, você poderá enviar apenas as 10,000 mensagens mais recentes. Sem arquivos de mídia, você poderá enviar 40,000 mensagens. Estes limites existem devido ao tamanho máximo permitido para mensagens de e-mail.

Foram selecionados cinco grupos/conversas públicos(as) e um grupo/conversa privado(a) de WhatsApp, com aproximadamente 150 a 200 membros cada, e a estratégia utilizada para a coleta dos dados para a exportação ao banco de dados e análise das características quantitativas das mensagens foi a opção de exportar o histórico da conversa sem arquivo de mídia.

3.1.2 Qual critério será utilizado para verificar se uma mensagem foi encaminhada?

A observação central é que nem todas as mensagens possuem a mesma probabilidade de ser transmitida por outros; assim é possível identificar os fatores que aumentam ou inibem a transmissão da mensagem, por meio de análises estatísticas. Cada mensagem tem seu estilo particular e características de conteúdo, cada um dos quais podem servir para aumentar ou suprimir a probabilidade de que um determinado destinatário da mensagem irá passar adiante. Da mesma forma, a probabilidade de que um usuário irá retransmitir a mensagem pode ser positivamente ou negativamente afetada por características do remetente ou pelo contexto em que a mensagem foi enviada, podendo haver fatores adicionais, aspectos imprevisíveis do ambiente de comunicação que também afetam a probabilidade de retransmissão ([SUTTON et al., 2015](#)).

Muitas pessoas também são influenciadas na retransmissão de rumores ou informações verdadeiras, porque elas tendem a procurar opiniões em tais circunstâncias. A qualidade do conteúdo pode ser explicada pela credibilidade, relevância, integridade e representação concisa. Assim como é informado em [Koohikamali e Kim \(2015\)](#), se o conteúdo da informação é credível, há menos dúvidas para interpretar corretamente a mensagem, ou seja, a credibilidade da informação influencia positivamente na disseminação de rumores

e informações verdadeiras. Condições incertas causam maior ansiedade entre os usuários e pessoas ansiosas são as mais propensas a buscar informações verdadeiras, influenciando positivamente na difusão das mesmas e de rumores. Expressões de envolvimento pessoal em rumores está positivamente relacionada com propagação de rumor, já que as pessoas transmitem mais rumores se elas sentem que estão envolvidas, ou seja, influencia negativamente para a difusão de informações verdadeiras. Laços sociais também influenciam positivamente na propagação de rumores e mensagens verdadeiras, pois são baseados em uma relação social existente da mensagem que é recebida de uma pessoa conhecida.

As mensagens com informações mais ricas possuem uma maior qualidade de informação e são mais úteis para o público, e, portanto, têm uma maior chance de ser compartilhada por usuários. Informações mais ricas implicam incertezas mais baixas e tendem a ser mais úteis para os usuários a tomar decisões. Além disso, mostram que a quantidade de informação é especialmente benéfica para os usuários se a informação pode ser obtida sem custos adicionais de pesquisa e informações.

Neste contexto, é possível notar que o número de pessoas expostas a qualquer mensagem é geralmente muito grande, e que a probabilidade de passagem de mensagens por qualquer indivíduo é geralmente muito pequena.

3.1.3 Por quanto tempo o experimento deverá ser executado?

Utilizando uma abordagem de estudo de caso, para lidar com nossas questões de pesquisa e quanto tempo o experimento deverá ser executado, realizamos análise de conteúdo e análises estatísticas sobre o conjunto de mensagens, verificando o seu conteúdo, estilo e características estruturais, e como isso afeta a retransmissão de mensagem pelo público por um período de três meses aproximadamente, no período que compreende os meses de janeiro e março de 2019.

3.2 Exportação dos dados

Nesta seção são explicados os passos para a realização da exportação das informações contidas nas conversas selecionadas do WhatsApp em um arquivo .txt.

1. Inicialmente, a coleta dos dados será feita através da escolha de um grupo, acessar tal grupo e clicar nas configurações do mesmo, conforme mostrado na Figura 9;



Figura 9 – Escolha do grupo para exportação

2. Após a escolha dos grupos, é preciso selecionar na própria configuração do WhatsApp a opção "Mais" para seguir com a exportação, demonstrado na Figura 10;

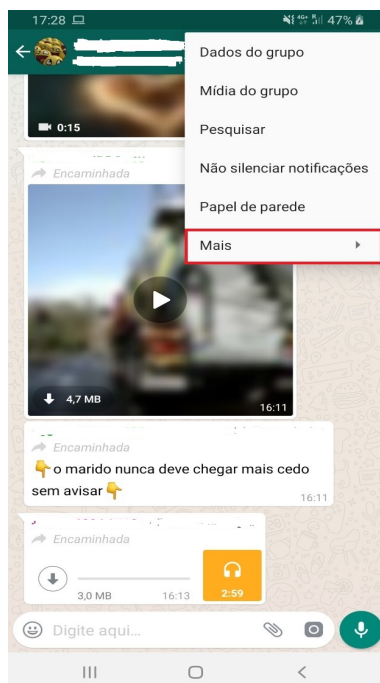


Figura 10 – Abrir configuração da conversa

3. Escolhendo a opção "Mais", será aberto novas opções para seleção. Para prosseguir com a exportação da conversa, é preciso selecionar a opção "Exportar conversa", exibido na Figura 11;



Figura 11 – Exportar Conversa

4. E por fim, foi determinado escolher a opção "Sem mídia" para melhor fluxo e desempenho durante o processo de *parsing* para o banco de dados, gerando assim, o arquivo .txt, exposto na Figura 12.

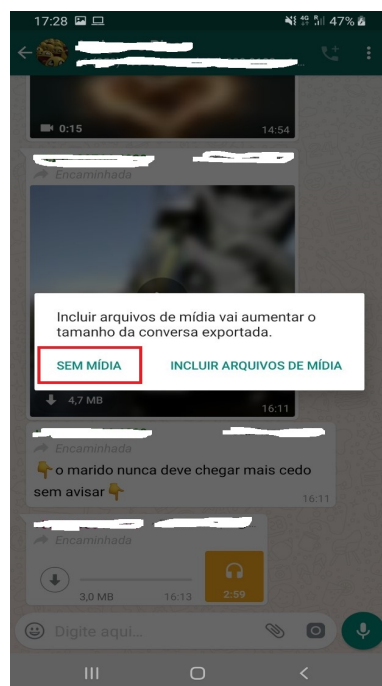


Figura 12 – Escolha de arquivo sem mídia

3.3 *Script para parsing*

Para todo o desenvolvimento do algoritmo/*script* foram utilizadas ferramentas necessárias para o funcionamento de todo o código. A linguagem Java (veja na Seção 2.4.1), a linguagem SQL (veja na Seção 2.4.3), a IDE Eclipse (veja na Seção 2.4.2), o framework Hibernate, utilizando JPA (veja na Seção 2.4.4) e o banco de dados MySQL (veja na Seção 2.4.5).

3.3.1 Classe Usuario

Inicialmente, foi criada a classe "Usuario", referenciando a tabela "usuario" no banco de dados, onde será armazenado todos os dados, informações e mensagens dos grupos selecionados do WhatsApp, separados pelas colunas "id", "data", "pessoa" e "mensagem", mostrado no *Script Código A.1*.

3.3.2 Classe Parsing

O *Script Código B.1* descreve todo o fluxo, desde a leitura dos dados/mensagens dos grupos do WhatsApp até a inserção no banco de dados.

Primeiramente, é realizado através do método "leArquivo()" a leitura da primeira conversa ou grupo do WhatsApp, nomeado de "Conversa1.txt". Foi definido um padrão de inserção e persistência no banco de dados antes da criação do *script*, segundo o qual todas as mensagens novas de usuário seguiria um modelo único que seria registrado, tanto de data e hora para o campo "data", da pessoa para o campo "pessoa" e da mensagem para o campo "mensagem". Todas as validações efetuadas durante o desenvolvimento do algoritmo seguiram o padrão mencionado anteriormente.

É executado um comando de percorrer todo o arquivo, validando linha por linha o formato da data e hora que foi enviado a mensagem para poder inserir na coluna "data" da tabela "usuario" do banco de dados corretamente. Para salvar os dados de todas as colunas no banco de dados separadamente, foi utilizado o método "salvarDados()". Neste, era definido uma regra que caso a linha possuísse a data, ou seja, era o início de uma nova mensagem e todos os campos eram divididos pelo método "split" do Java, separando por listas a mensagem e a pessoa que a enviou e caso fosse uma mensagem que não seguia o padrão de iniciar com alguma data, como no caso da pessoa sair ou entrar no grupo, foi aplicado um recurso chamado "regex", também do Java.

3.4 Criação do banco de dados

Foi criado um arquivo chamado "persistence.xml", que contém todas as configurações necessárias para a criação do banco de dados do trabalho. É possível identificar

também a implementação da configuração do framework Hibernate e da interface JPA, assim como mostrado no *Script Código C.1*.

O *Script Código D.1* mostra a classe chamada JPAUtil, a qual é instanciada para a realização de todas as ações de persistência ao BD. O EntityManager é o serviço central que provê APIs para criar consultas, buscando objetos, sincronizando objetos, e inserindo objetos no banco de dados. Também pode prover caching e pode administrar a interação entre uma entidade e serviços transacionais em um ambiente Java. (WENDELL, 2007)

3.5 Filtro para a busca dos resultados

Para a identificação das características comuns das mensagens iguais/semelhantes, ou encaminhadas, são utilizadas 2 consultas. O *Script Código 3.1* mostra a primeira consulta, simples e de forma resumida para verificar apenas se existem duplicatas na tabela. Se alguma linha for retornada, isso significa que foi identificado mensagens iguais ou encaminhadas.

```
1 SELECT pessoa, mensagem, COUNT(*)
2 FROM usuario
3 WHERE mensagem != '<Arquivo de midia oculto>'
4 GROUP BY pessoa, mensagem
5 HAVING COUNT(*) > 1;
```

Código 3.1 – Consulta de duplicatas

Para a busca do registro completo para cada linha duplicada, foi utilizado outra consulta, assim como no *Script Código 3.2*, a qual seleciona cada coluna na tabela de usuario e, em seguida, a une com a tabela de dados duplicados da consulta inicial. E caso retorne alguma linha, o conjunto de resultados irá incluir todos os IDs dos registros.

```
1 SELECT a.*
2 FROM usuario a
3 JOIN (SELECT pessoa, mensagem, COUNT(*)
4 FROM usuario
5 GROUP BY pessoa, mensagem
6 HAVING count(*) > 1 ) b
7 ON a.pessoa = b.pessoa
8 AND a.mensagem = b.mensagem
9 AND a.mensagem != '<Arquivo de midia oculto>'
10 ORDER BY a.mensagem;
```

Código 3.2 – Consulta do registro completo

4 Resultados

Este capítulo tem como objetivo analisar os dados obtidos através da consulta das mensagens compartilhadas. Para isso, diferentes comparações foram realizadas.

A Seção 4.1 apresenta uma visão geral sobre quais critérios foram utilizados para a realização do experimento. Tais como, em quanto tempo foi realizado, como foi realizado a coleta de dados, quantos e quais grupos foram usados e como os dados foram exportados.

A Seção 4.2 expõe algumas das sentenças/frases mais encaminhadas e também os principais assuntos considerados mais relevantes, ou seja, mais importantes para o compartilhamento das mensagens. Foram através dessas informações que foi criada a nuvem de palavras.

A Seção 4.3 apresenta a nuvem de palavras das mensagens compartilhadas mais relevantes e como ela foi gerada, incluindo a quantidade de mensagens, o horário e data de envio, como também o conteúdo da mensagem.

A Seção 4.4 descreve a análise da polaridade das sentenças utilizadas, identificando se a opinião que foi expressada em um determinado texto, é positiva, negativa ou neutra.

A Seção 4.5 apresenta os resultados e responde se os objetivos da pesquisa foram alcançados.

4.1 Visão geral do experimento

Seguindo todo o processo exposto no Capítulo 3, foram selecionados cinco grupos/-conversas públicos(as) e um grupo/conversa privado(a) de WhatsApp, com aproximadamente 150 a 200 membros cada, para a coleta de informações e realização do experimento. Para este estudo, as mensagens foram coletadas entre o período de janeiro e fevereiro de 2019, ou seja, 31 dias, para posteriormente, através do *script* desenvolvido, realizar a exportação desses dados para o banco, e serem feitos todas as consultas e análises necessárias para obtenção dos resultados.

Para a coleta das informações da Tabela 1, observa-se que em cada conversa selecionada durante o experimento, foram obtidos inicialmente os dados quantitativos em relação a cada grupo, No Grupo 1, é possível perceber que é o que contém o maior número de mensagens, 64.386, sendo 1.591 mensagens encaminhadas, no Grupo 2, o total de mensagens é de 15.191 e 222 encaminhadas, no Grupo 3, 37.983 e 1.691 encaminhadas, no grupo 4, 6.287 e 208 encaminhadas, no Grupo 5, 8.016 e 398 encaminhadas e no Grupo 6, o único privado, 585 mensagens e 8 encaminhadas. Também foram obtidas as relações de

mensagens encaminhadas por dia, sendo que no Grupo 1 foram encaminhadas aproximadamente 53 mensagens por dia, no Grupo 2, 7, no Grupo 3, 56, no Grupo 4, 7, no Grupo 5, 13 e no Grupo 6, 0,3. E em cada conversa também é possível verificar o tamanho do seu arquivo .txt, em bytes, de acordo com a quantidade de informações contidas.

Tabela 1 – Informações dos grupos do WhatsApp

	Nº de Mensagens	Bytes de Conversa (.txt)	Nº de Mensagens Encaminhadas	Nº de Mensagens Encaminhadas por Dia
GRUPO 1	42.142	3.004.096	8.786	283
GRUPO 2	5.579	524.014	1.990	64
GRUPO 3	36.976	2.395.174	10.243	330
GRUPO 4	5.669	367.644	1.491	48
GRUPO 5	4.426	486.912	1.146	37
GRUPO 6	576	38.656	20	0,6

4.2 Mensagens encaminhadas

Depois de toda a exportação das informações das conversas selecionadas do WhatsApp, foi realizada uma consulta para obter todas as mensagens encaminhadas da base de dados. De acordo com a Seção 4.1 nota-se que há uma grande quantidade de mensagens enviadas durante todo o período do experimento. Para uma melhor representação das mensagens encaminhadas, um filtro para apresentar apenas os assuntos considerados mais importantes e as respectivas frases associadas a eles foi desenvolvido. A Tabela 2 apresenta os principais resultados de tal filtro.

É possível notar que existem diversas mensagens encaminhadas relacionadas ao mesmo assunto. Isso indica a existência de diferentes usuários encaminhando mensagens sobre um mesmo assunto. Um exemplo é o assunto sobre o rompimento da barragem em Brumadinho. Diferentes mensagens com praticamente o mesmo conteúdo inundaram os grupos do WhatsApp. O mesmo aconteceu com os outros assuntos listados na Tabela 2. Além disso, é notável a alta incidência de compartilhamento das mensagens com cunho político ou acontecimentos que afetam a economia.

Tabela 2 – Sentenças relacionadas às palavras

Assuntos	Frases por palavra		
Caio Junqueira	Ator da Globo Caio Junqueira morre após sofrer acidente	Após uma semana da cirurgia ator Caio Junqueira morre	Caio Junqueira morre após cirurgia
Brumadinho	Acidente em Brumadinho	Tragédia da Vale em Brumadinho	Número de vítimas em Brumadinho
SISU	Piores Faculdades para tentar o SISU	Data de início do SISU	Médias de Medicina para o SISU
Barragem	Município de Minas Gerais sofre com rompimento de barragem	Acidente com barragem da Vale	Barragem rompida tem várias vítimas
Big Brother Brasil	Participantes do BBB	Favoritos a vencer o Big Brother Brasil 2019	Novo Big Brother Brasil 2019
Exame Nacional do Ensino Médio	Divulgação das notas do ENEM 2019	Tentativas de Fraude no ENEM	Data de início do ENEM 2019
Carnaval	Dias de feriado do Carnaval 2019	Melhores lugares para ficar no carnaval brasileiro de 2019	Feriados do carnaval 2019
IPVA	Pagamentos IPVA 2019	Data de vencimento do IPVA pela placa do veículo	Datas início para pagamento do IPVA em 2019
Jair Bolsonaro	Família Bolsonaro e seu legado político	O que muda com o governo Bolsonaro	Bolsonaro não é homofóbico
Emprego	Piora na expectativa de empregos	Novos empregos podem ser realidade em 2019	Melhora no desemprego em 2019
Dólar	Comparativo entre as últimas cotações de dólar	Valor do dólar com o novo governo	Dólar cresce ou cai com Jair Bolsonaro?

4.3 Nuvem de palavras

Nuvem de palavras (word cloud) é um gráfico digital que mostra o grau de frequência das palavras em um texto. Quanto mais a palavra é utilizada, mais chamativa é a representação dessa palavra no gráfico. As palavras aparecem indicando o que é mais relevante e o que é menos relevante no contexto.

De acordo com as sentenças obtidas, foram geradas duas nuvens de palavras. A primeira contém todas as mensagens encaminhadas e a segunda contém apenas algumas das que mais foram disseminadas. Para a produção de cada nuvem, foi utilizado um *script* para a remoção de *stop words*, ou palavras que são removidas antes ou após o processamento de um texto, no qual os motores de busca ignoram em seus resultados, ou seja, são consideradas irrelevantes para o conjunto de resultados a ser exibido pela busca. Exemplos de *stop word* na língua portuguesa incluem as preposições e os artigos visto que tais palavras não acrescentam conteúdo semântico a uma sentença.

O *Script Código 4.1* exibe um exemplo de filtro para remoção de *stop word*.

```
1 # Carregar biblioteca
2 from nltk.corpus import stopwords
3
4 # Download do conjunto de stop words
5 import nltk
6 nltk.download('stopwords')
7
8 # Criação dos tokens da sentença
9 tokenized_words = ['Caio Junqueira', 'morre', 'em', 'acidente']
10
11 # Carregar as stop words
12 stop_words = stopwords.words('portuguese')
13
14 # Remover as stop words
15 [word for word in tokenized_words if word not in stop_words]
16
17 # Mostrar stop words
18 stop_words[:5]
```

Código 4.1 – Remoção de *Stop words*

Na Figura 13, é mostrada a nuvem de palavras que agrupa todas as mensagens encaminhadas consideradas mais relevantes com a remoção das *stop words*, com o total de noventa e sete frases envolvendo todos os assuntos que fazem parte do experimento.

Ator”, “Brumadinho”, “Sistema de Seleção Unificada (SISU)”, “Barragem”, “Big Brother Brasil (BBB)”, “Exame Nacional do Ensino Médio (ENEM)”, “Carnaval”, “IPVA”, “Jair Bolsonaro”, “Emprego”, “Dólar dos Estados Unidos – Moeda”, como representado na Tabela 3. Pelo fato de ter sido uma época movimentada com acontecimentos marcantes para o país, há uma alta incidência de retransmissão em mensagens que possuam essas palavras, principalmente as que envolvem política ou coisas relacionadas ou que afetam essa área.

Tabela 3 – Nuvem de Palavras

Assunto	Abreviação	Nº de Abreviações Possíveis	Nº de palavras
Caio Junqueira	Caio	1	2
Brumadinho	-	0	1
SISU	-	0	1
Barragem	-	0	1
Big Brother Brasil	BBB	1	3
Exame Nacional do Ensino Médio	ENEM	1	5
Carnaval	-	0	1
IPVA	-	0	1
Jair Bolsonaro	Jair/Bolsonaro	2	2
Emprego	-	0	1
Dólar	-	0	1

A Tabela 3 é uma visão geral simplificada da relação entre as palavras, suas abreviações, número de palavras e abreviações possíveis, isso para situar sobre os assuntos que serão tratados os resultados sobre o conteúdo de encaminhamento e retransmissão das mensagens.

4.4 Análise de polaridade

De acordo com Gupta (2018), a Análise de Sentimento é uma mineração contextual de texto que identifica e extrai informações subjetivas no material de origem. Normalmente utilizada para entender o sentimento social de uma marca, produto ou serviço enquanto monitora as conversas on-line. Ou seja, é uma ferramenta de classificação de texto que analisa uma mensagem recebida e informa se sua polaridade é positiva, negativa ou neutra.

Para analisar a polaridade de todas as mensagens encaminhadas do experimento, foi utilizado o TextBlob, que é uma biblioteca Python (2 e 3) para processar dados textuais. Ela fornece uma API simples para abstrair-se em tarefas comuns de processamento de linguagem natural (NLP), como marcação de parte da fala, extração de frase de substantivo, análise de sentimento, classificação, tradução e muito mais. (LORIA, 2018) Para

uma melhor utilização da biblioteca, e usando o modelo padrão do TextBlob, todas as sentenças foram traduzidas para o inglês.

Para a realização da análise de polaridade é utilizado a implementação padrão *Pattern Analyser* (com base na biblioteca de padrões). De acordo com [Drupal \(2018\)](#), *Pattern* possui ferramentas para mineração de dados (API do Google, Twitter e Wikipedia, um rastreador da Web, um analisador HTML DOM), processamento de linguagem natural, aprendizado de máquina e análise de rede. É um pacote de ferramentas de processamento de linguagem natural (NLP), identificando os tipos de palavras e suas inflexões.

O *Script Código 4.2* expõe um exemplo de como realizar a análise da polaridade das sentenças.

```
1 # Instalacao do TextBlob
2 pip install -U textblob
3
4 # Download do modelo padrao do TextBlob
5 python -m textblob.download_corpora
6
7 # Importacao.
8 from textblob import TextBlob
9
10 # Criacao do TextBlob
11 blob = TextBlob("Caio Junqueira dies in accident")
12
13 # Analise de polaridade
14 blob.sentiment.polarity
```

Código 4.2 – Análise de polaridade

A Tabela 4 mostra todas as vezes que os assuntos apareceram repetidamente, ou seja, foram encaminhados, em um total de 9.493 mensagens, além da polaridade de cada frase disseminada e suas porcentagens.

Tabela 4 – Análise sobre polaridade

Assunto	Vezes que aparecem	Polaridades			Porcentagens		
		+	Neutra	-	+	Neutras	-
Caio Junqueira	271	0	206	65	0%	76%	24%
Brumadinho	1.038	228	685	125	22%	66%	12%
SISU	327	0	173	154	0%	53%	47%
Barragem	1.088	0	925	163	0%	85%	15%
Big Brother Brasil	137	36	101	0	26%	74%	0%
Exame Nacional do Ensino Médio	252	0	166	86	0%	66%	34%
Carnaval	567	272	295	0	48%	52%	0%
IPVA	272	0	272	0	0%	100%	0%
Jair Bolsonaro	3.778	793	2834	151	21%	75%	4%
Emprego	1.388	472	916	0	34%	66%	0%
Dólar	442	274	168	0	62%	38%	0%

De modo geral, a maioria das incidências de neutralidade surgiram por mensagens sem posicionamento sobre os assuntos das palavras abordadas, apenas de cunho expositivo. Os posicionamentos e a análise de polaridade foram analisadas com base na Análise de Sentimentos, ou Mineração de Opinião, onde a organização das informações disponíveis no mundo podem ser explicadas como opiniões e fatos (LIU et al., 2010).

Essa análise quantitativa sobre o número de mensagens em cada espectro da polaridade, positiva ou negativa, reflete o posicionamento da opinião que o usuário exerceu sobre a mensagem e retransmissões da mesma. Mensagens positivas, como o exemplo do caso de brumadinho, partilham da notícia reforçando comentários como “pessoas salvas”. Comentários negativos reforçam “quantidade de mortos”, outro exemplo é o caso das mensagens sobre Jair Bolsonaro que as positivas refletem seus apoiadores compartilhando notícias que enalteçam seu governo e expectativas, as negativas em grande parte por seus críticos.

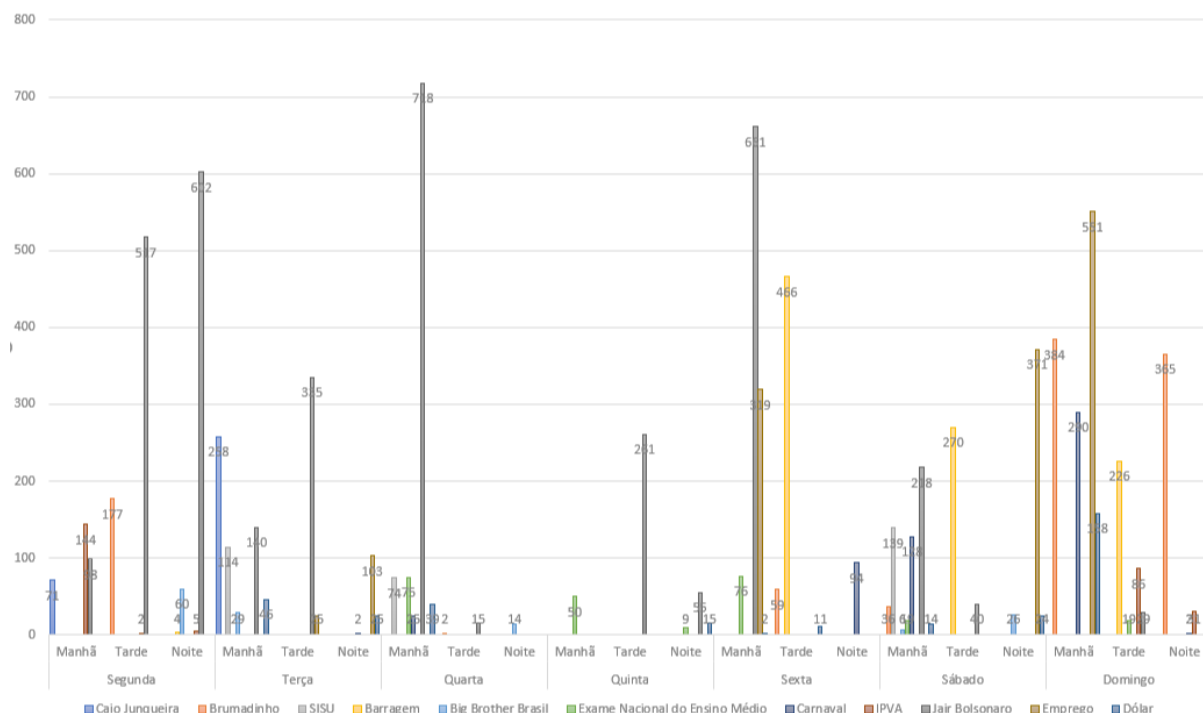


Figura 15 – Análise sobre o dia da difusão de mensagens

A Figura 15 é construída a partir das datas (segunda a domingo) e turnos (manhã, tarde e noite) das mensagens encaminhadas durante todo o período da realização do experimento. O turno da manhã se situa entre as 00:01 horas até as 12:00 horas, o turno da tarde das 12:01 horas até as 18:00 horas, e o turno da noite das 18:01 horas até as 24:00 horas. A partir dos valores obtidos, observa-se que no período da manhã, há mais mensagens compartilhadas de modo geral. Notícias como as da “Barragem” e “Brumadinho” por exemplo, influenciaram os dias da semana em que foram mais compartilhadas pela data do acidente que se iniciou em uma sexta-feira na data de 25 de janeiro de 2019, tendo seus dias subsequentes movimentado com várias mensagens compartilhadas e retransmitidas.

4.5 Considerações finais

Neste capítulo, uma visão geral sobre as mensagens mais encaminhadas em conversas no WhatsApp foi apresentada, incluindo assuntos relacionados a acontecimentos ocorridos no período do experimento, principalmente que envolvem política ou coisas relacionadas ou que afetam essa área. Também foi realizada uma análise sobre a quantidade de informações disseminadas dos dados que foram coletados e como esses dados podem ser retransmitidos.

Após a filtragem de todas as mensagens encaminhadas por aquelas mais relevantes ou fortemente disseminadas, foi possível escolher o TextBlob como a biblioteca de análise de polaridade para ser utilizado neste trabalho. Em seguida, ela foi aplicada sobre as

principais mensagens encaminhadas e foi possível avaliar os sentimentos das mensagens encaminhadas pelo WhatsApp.

A partir dos resultados obtidos, foi possível concluir que boa parte dos usuários tendem a encaminhar as principais notícias mais mencionadas nos noticiários de jornais, programas de televisão e de rádio, redes sociais, sites de notícias, etc. Também é possível perceber que há uma maior quantidade de informações consideradas neutras do que positivas e negativas, classificadas pela análise de polaridade do TextBlob. Além disso, através dos resultados obtidos com a análise de difusão de mensagens, o turno de maior incidência de mensagens encaminhadas foi no período da manhã, o que provavelmente também indica que nesse período, há uma maior circulação de notícias nas redes sociais.

5 Conclusão

Neste trabalho, foi apresentado um breve histórico sobre as redes sociais e sua importância em diferentes aspectos. Foram analisadas as características de notícias disseminadas, seguindo o modelo de trabalho de [Bavelas \(1948\)](#), [Shaw \(1954\)](#), [Shimbel \(1953\)](#), [Abu-Shaqra \(2014\)](#), [Rivera \(2015\)](#), [Koochikamali e Kim \(2015\)](#), [Lee, Agrawal e Rao \(2015\)](#) e [Sutton et al. \(2015\)](#), a fim de verificar uma correlação de mensagens encaminhadas via WhatsApp. Nos dedicamos ao estudo deste aplicativo por ser ele uma rede social com grande e atual potencial de crescimento, além da interessante abordagem a ser explorada referente à avaliação das características das mensagens nessa plataforma diariamente movimentada.

Quando comparado com o trabalho de [Shi et al. \(2017\)](#) foi observado que neste estudo os principais componentes envolvidos em um processo de comunicação social também obteve impactos no comportamento de compartilhamento de notícias. Os principais acontecimentos que estão mais relevantes no Brasil dentro do período de realização do experimento são praticamente aquelas que estão sendo mais disseminadas. Ou seja, notícias da atualidade e riqueza de informações são grandes impactadores de difusão de mensagens.

Dessa forma, o objetivo deste trabalho foi realizar análises estatísticas baseadas em grupos selecionados de WhatsApp. Para isso, foram realizadas as etapas da tarefa de coleta de dados das conversas, produção do *script* para *parsing* dos dados para o banco, construção de uma base de dados rotulada, classificação dos textos e validação dos resultados. Em suma, foi possível concluir que boa parte dos usuários tende a encaminhar as principais notícias mais mencionadas nos noticiários de jornais, programas de televisão e de rádio, redes sociais, sites de notícias, etc.

Em um trabalho futuro, espera-se investigar um maior número de grupos/conversas, contemplando um maior número mensagens, bem como utilizar novas ferramentas e verificar a ocorrência de outras características e fatores que impactam no compartilhamento de mensagens. Também seria interessante desenvolver uma nova forma de exportação de dados, com a inclusão de arquivos de mídia, baseada na linguagem de programação e *script* utilizados no desenvolvimento, visto que quanto maior o número de informações e dados obtidos, maior a variedade de características pode que pode ser obtida.

Referências

- ABU-SHAQRA, B. The dialectics of empowerment and exploitation of using social media. *Global Media Journal*, Mahmoud Eid, v. 7, n. 2, p. 75, 2014. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 27 e 47.
- BAVELAS, A. A mathematical model for group structures. *Applied anthropology*, JSTOR, v. 7, n. 3, p. 16–30, 1948. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 27 e 47.
- BOUHNİK, D.; DESHEN, M.; GAN, R. Whatsapp goes to school: Mobile instant messaging between teachers and students. *Journal of Information Technology Education: Research*, v. 13, n. 1, p. 217–231, 2014. Acesso em: 2019-07-10. Citado na página 22.
- BOYD, D. M.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, Wiley Online Library, v. 13, n. 1, p. 210–230, 2007. Acesso em: 2019-07-10. Citado 3 vezes nas páginas 7, 16 e 17.
- CAELUM. *Uma introdução prática ao JPA com Hibernate*. 2011. Disponível em: <<https://www.caelum.com.br/apostila-java-web/uma-introducao-pratica-ao-jpa-com-hibernate/>>. Acesso em: 2019-07-10. Citado na página 26.
- CHURCH, K.; OLIVEIRA, R. D. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional sms. In: ACM. *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*. [S.l.], 2013. p. 352–361. Acesso em: 2019-07-10. Citado na página 13.
- CONSALVO, M. et al. *Internet Research Annual: Selected Papers from the Association of Internet Researchers Conferences 2000-2002*. [S.l.]: Peter Lang, 2004. v. 19. Acesso em: 2019-07-10. Citado na página 16.
- DATE, C. J. *Introdução a sistemas de bancos de dados*. [S.l.]: Elsevier Brasil, 2004. Acesso em: 2019-07-10. Citado na página 26.
- DAVIES, J. *Word Cloud*. 2015. Disponível em: <https://www.jasondavies.com/wordcloud/?fbclid=IwAR3kuziqxJ0zDHvfTDLidN-rSz2uLYSknnYit49ifcfxz16Q9oRq_e5amKQ#%2F%2Fwww.jasondavies.com%2Fwordcloud%2Fabout%2F>. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 7 e 41.
- DRUPAL. *Pattern*. 2018. Disponível em: <<https://www.clips.uantwerpen.be/pattern>>. Acesso em: 2019-07-10. Citado na página 43.
- ECLIPSE. *Eclipse Foundation*. 2019. Disponível em: <<https://www.eclipse.org/org/>>. Acesso em: 2019-07-10. Citado na página 25.
- ERICSSON. *Mobile subscriptions*. 2019. Disponível em: <<https://www.ericsson.com/en/mobility-report/mobility-visualizer?f=1&ft=1&r=2,3,4,5,6,7,8,9&t=8&s=1,2,3&u=1&y=2018,2024&c=1>>. Acesso em: 2019-07-10. Citado 3 vezes nas páginas 7, 21 e 22.

- GUPTA, S. *Sentiment Analysis: Concept, Analysis and Applications*. 2018. Disponível em: <<https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>>. Acesso em: 2019-07-10. Citado na página 42.
- JOHARI, A. *What Is Java? A Beginner's Guide to Java and Its Evolution*. 2018. Disponível em: <<https://www.edureka.co/blog/what-is-java/>>. Acesso em: 2019-07-10. Citado na página 24.
- KEMP, S. *The state of digital in april 2019: All the numbers you need to know*. 2019. Disponível em: <<https://wearesocial.com/uk/blog/2019/04/the-state-of-digital-in-april-2019-all-the-numbers-you-need-to-know>>. Acesso em: 2019-07-10. Citado 5 vezes nas páginas 7, 18, 19, 20 e 21.
- KOOHIKAMALI, M.; KIM, D. J. Rumor and truth spreading patterns on social network sites during social crisis: Big data analytics approach. In: SPRINGER. *Workshop on E-Business*. [S.l.], 2015. p. 166–170. Acesso em: 2019-07-10. Citado 4 vezes nas páginas 27, 28, 31 e 47.
- KOPYTOFF, V. *Google's Orkut puzzles experts / Internet watchers ponder reason for social network site*. 2014. Disponível em: <<https://www.sfgate.com/business/article/Google-s-Orkut-puzzles-experts-Internet-2633049.php>>. Acesso em: 2019-07-10. Citado na página 17.
- LEE, J.; AGRAWAL, M.; RAO, H. R. Message diffusion through social network service: The case of rumor and non-rumor related tweets during boston bombing 2013. *Information Systems Frontiers*, Springer, v. 17, n. 5, p. 997–1005, 2015. Acesso em: 2019-07-10. Citado 3 vezes nas páginas 27, 28 e 47.
- LIU, B. et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, v. 2, n. 2010, p. 627–666, 2010. Acesso em: 2019-07-10. Citado na página 44.
- LORIA, S. *TextBlob: Simplified Text Processing*. 2018. Disponível em: <<https://textblob.readthedocs.io/en/dev/>>. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 13 e 42.
- PARREIRA, G. *Guia Completo de SQL*. 2018. Disponível em: <<https://www.devmedia.com.br/guia/guia-completo-de-sql/38314>>. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 7 e 26.
- RASHDI, F. A. *Forms and functions of emojis in WhatsApp interaction among Omanis*. Tese (Doutorado) — Georgetown University, 2015. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 7 e 23.
- RICARDO. *Conceitos Fundamentais de Banco de Dados*. 2006. Disponível em: <<https://www.devmedia.com.br/conceitos-fundamentais-de-banco-de-dados/1649>>. Acesso em: 2019-07-10. Citado na página 27.
- RIVERA, J. de. Reverse engineering social media: Software, culture and political economy in new media capitalism. *Teknokultura*, Grupo de Investigación Cibersomosaguas. Universidad Complutense de Madrid, v. 12, n. 1, p. 211–221, 2015. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 27 e 47.

- SHAW, M. E. Group structure and the behavior of individuals in small groups. *The Journal of psychology*, Taylor & Francis, v. 38, n. 1, p. 139–149, 1954. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 27 e 47.
- SHI, J. et al. Understanding and predicting individual retweeting behavior: Receiver perspectives. *Applied Soft Computing*, Elsevier, v. 60, p. 844–857, 2017. Acesso em: 2019-07-10. Citado 4 vezes nas páginas 27, 28, 29 e 47.
- SHIMBEL, A. Structural parameters of communication networks. *The bulletin of mathematical biophysics*, Springer, v. 15, n. 4, p. 501–507, 1953. Acesso em: 2019-07-10. Citado 2 vezes nas páginas 27 e 47.
- SILBERSCHATZ, A.; SUNDARSHAN, S.; KORTH, H. F. *Sistema de banco de dados*. [S.l.]: Elsevier Brasil, 2016. Acesso em: 2019-07-10. Citado na página 26.
- SIQUEIRA, A. *Por que é importante e como montar uma política de privacidade para seu site*. 2013. Disponível em: <https://resultadosdigitais.com.br/blog/como-montar-uma-politica-de-privacidade/?fbclid=IwAR0LLMrBRK9gyFHnwcoTnl60Njz8_XPFSdbvGnu5IT0BDXWtyizrEdT4Ftg>. Acesso em: 2019-07-10. Citado na página 18.
- SUTTON, J. et al. What it takes to get passed on: message content, style, and structure as predictors of retransmission in the boston marathon bombing response. *PLoS one*, Public Library of Science, v. 10, n. 8, p. e0134452, 2015. Acesso em: 2019-07-10. Citado 5 vezes nas páginas 13, 27, 29, 31 e 47.
- TECHOPEDIA. *Structured Query Language (SQL)*. 2019. Disponível em: <<https://www.techopedia.com/definition/1245/structured-query-language-sql>>. Acesso em: 2019-07-10. Citado na página 25.
- WELLMAN, B.; BERKOWITZ, S. D. *Social structures: A network approach*. [S.l.]: CUP Archive, 1988. v. 2. Acesso em: 2019-07-10. Citado na página 18.
- WENDELL. *Introdução ao EntityManager*. 2007. Disponível em: <<https://www.devmedia.com.br/introducao-ao-entitymanager/5206>>. Acesso em: 2019-07-10. Citado na página 36.
- WHATSAPP. *Salvando seu histórico de conversas*. 2009. Disponível em: <https://faq.whatsapp.com/pt_br/android/23756533/?category=5245251>. Acesso em: 2019-07-10. Citado na página 31.

Anexos

ANEXO A – Classe Usuário

```
1 package br.com.renato.models;
2
3 import javax.persistence.Column;
4 import javax.persistence.Entity;
5 import javax.persistence.GeneratedValue;
6 import javax.persistence.GenerationType;
7 import javax.persistence.Id;
8
9 @Entity
10 public class Usuario {
11
12     @Id
13     @GeneratedValue(strategy=GenerationType.IDENTITY)
14     private Integer id;
15     private String data;
16     private String pessoa;
17
18     @Column(columnDefinition="varchar(10000)")
19     private String mensagem;
20
21
22     public Integer getId() {
23         return id;
24     }
25     public void setId(Integer id) {
26         this.id = id;
27     }
28     public String getData() {
29         return data;
30     }
31     public void setData(String data) {
32         this.data = data;
33     }
34     public String getPessoa() {
35         return pessoa;
36     }
37     public void setPessoa(String pessoa) {
38         this.pessoa = pessoa;
```

```
39     }
40     public String getMensagem() {
41         return mensagem;
42     }
43     public void setMensagem(String mensagem) {
44         this.mensagem = mensagem;
45     }
46
47 }
```

Código A.1 – Classe Usuario

ANEXO B – *Script* para exportação dos dados

```
1 import java.io.BufferedReader;
2 import java.io.FileInputStream;
3 import java.io.FileNotFoundException;
4 import java.io.IOException;
5 import java.io.InputStreamReader;
6 import java.util.regex.Matcher;
7 import java.util.regex.Pattern;
8
9 import javax.persistence.EntityManager;
10
11 import br.com.renato.models.Usuario;
12 import br.com.renato.util.JPAUtil;
13
14 public class Parsing {
15
16     private static final int TAMANHO_LISTA = 2;
17
18     public static void main(String[] args) throws IOException {
19
20         try {
21
22             Usuario usuario = null;
23             EntityManager em = new JPAUtil().getEntityManager();
24             String terceiroTermo = "";
25             String sextotermo = "";
26
27             //Le o arquivo
28             BufferedReader br = leArquivo();
29
30             String linha = br.readLine();
31
32             //Percorrer todo o arquivo
33             while (linha != null) {
34
35                 //Linha em branco
36
```

```
37         if (validarLinha(linha)) {
38
39             int i;
40             String dataFormato = "";
41
42             //Percorre os 10 caracteres
43             for (i = 0; i < linha.length(); i++) {
44                 if (i == 10) {
45                     dataFormato = linha.substring(0, i);
46                     break;
47                 }
48             }
49
50             //Verifica se a frase contem mais de 6 caracteres
51             if (linha.length() >= 6) {
52                 terceiroTermo = linha.substring(2, 3);
53                 sextotermo = linha.substring(5, 6);
54             } else {
55                 //terceiroTermo = linha.substring(2, 3);
56                 terceiroTermo = "";
57                 sextotermo = "";
58             }
59
60             //Verifica se o terceiro e o sexto termo sao "/"
61             if (terceiroTermo.equals("/") && sextotermo.
62                 equals("/")) {
63
64                 String[] dt = dataFormato.split("/");
65
66                 String dtDia = dt[0];
67                 String dtMes = dt[1];
68                 String dtAno = dt[2];
69
70                 //Verifica se o formato da data esta correto
71                 if(formatoCorreto(dtDia, dtMes, dtAno)) {
72
73                     usuario = new Usuario();
74                     salvarDados(usuario, linha);
75                 } else {
76                     usuario.setMensagem(usuario.getMensagem()
77                         + linha);
```



```
77         }
78
79     } else {
80         if (usuario != null) {
81             usuario.setMensagem(usuario.getMensagem()
82                 + linha);
83             System.out.println(usuario.getMensagem());
84             ;
85         }
86     }
87
88     linha = br.readLine();
89
90     } else {
91         linha = br.readLine();
92     }
93
94     em.getTransaction().begin();
95     em.persist(usuario);
96     em.getTransaction().commit();
97
98     //Fecha as conexoes
99     br.close();
100    em.close();
101
102    } catch (Exception e) {
103        System.out.println(e.getMessage());
104    }
105 }
106
107 private static BufferedReader leArquivo() throws
108     FileNotFoundException {
109     FileInputStream fis = new FileInputStream("Conversa1.txt"
110         );
111     InputStreamReader isr = new InputStreamReader(fis);
112     BufferedReader br = new BufferedReader(isr);
113     return br;
114 }
115
116 private static boolean validarLinha(String linha) {
```

```
115     String linhaStr = linha.trim();
116     if (linhaStr.isEmpty()) {
117         return false;
118     }
119     return true;
120 }
121
122 private static boolean formatoCorreto(String dtDia, String
123     dtMes, String dtAno) {
124     return dtDia.length() == 2 && dtMes.length() == 2 &&
125         dtAno.length() == 4;
126 }
127
128 private static void salvarDados(Usuario usuario, String linha
129     ) {
130
131     int inicioPalavraProcurada = 0;
132     String frase = "", frasePessoa = "", fraseMensagem = "";
133     Boolean flag = false;
134     String[] list = new String[TAMANHO_LISTA];
135     list = linha.split(" - ");
136     frase = list[1].toString();
137
138     usuario.setData(list[0]);
139     System.out.println("Data: " + list[0]);
140
141     if (list[1].contains(": ")) {
142         usuario.setPessoa(list[1].split(": ")[0]);
143         System.out.println("Pessoa: " + list[1].split(": ")
144             [0]);
145         usuario.setMensagem(list[1].split(": ")[1]);
146         System.out.println("Mensagem: " + list[1].split(": ")
147             [1]);
148     } else {
149         //Se nao contem ": " (A pessoa saiu ou entrou no grupo)
150     } else {
151         //Busca a palavra "saiu"
152         Pattern p = Pattern.compile("saiu");
153         Matcher m = p.matcher(frase);
154         while (m.find()) {
155             inicioPalavraProcurada = m.start();
156             System.out.println(m.start() + " " + m.group() + "
157                 "+ m.end());
```

```
151         flag = true;
152     }
153     if (flag) {
154         setarPessoaEMensagem(usuario,
155             inicioPalavraProcurada, frase, frasePessoa,
156             fraseMensagem);
157     } else {
158         //Busca a palavra "entrou"
159         Pattern p2 = Pattern.compile("entrou");
160         Matcher m2 = p2.matcher(frase);
161         while (m2.find()) {
162             inicioPalavraProcurada = m2.start();
163             System.out.println(m2.start() + " " + m2.
164                 group() + " " + m2.end());
165             flag = false;
166         }
167         setarPessoaEMensagem(usuario,
168             inicioPalavraProcurada, frase, frasePessoa,
169             fraseMensagem);
170     }
171 }
172 }
173 }
174
175 private static void setarPessoaEMensagem(Usuario usuario, int
176     inicioPalavraProcurada, String frase,
177     String frasePessoa, String fraseMensagem) {
178     int i;
179     if (inicioPalavraProcurada != 0) {
180         for (i = 2; i < inicioPalavraProcurada - 2; i++) {
181             frasePessoa += frase.charAt(i);
182             usuario.setPessoa(frasePessoa);
183             System.out.println(frasePessoa);
184         }
185         for (i = inicioPalavraProcurada; i < frase.length();
186             i++) {
187             fraseMensagem += frase.charAt(i);
188             usuario.setMensagem(fraseMensagem);
189             System.out.println(fraseMensagem);
190         }
191     }
192 }
```

186

187 }

Código B.1 – *Script* para exportação dos dados

ANEXO C – Configuração do banco de dados

```

1 <persistence xmlns="http://java.sun.com/xml/ns/persistence"
2   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xsi:schemaLocation="http://java.sun.com/xml/ns/persistence
4     http://java.sun.com/xml/ns/persistence/persistence_2_0.xsd"
5     version="2.0">
6   <!-- unidade de persistencia com o nome whatsapp -->
7   <persistence-unit name="whatsapp">
8     <!-- Implementa o do JPA, no nosso caso Hibernate -->
9     <!-- -->
10    <provider>org.hibernate.jpa.HibernatePersistenceProvider
11    </provider>
12    <!-- Aqui s o listadas todas as entidades -->
13    <class>br.com.renato.models.Conta</class>
14    <properties>
15      <!-- Propriedades JDBC -->
16      <property name="javax.persistence.jdbc.driver" value="
17        com.mysql.jdbc.Driver"/>
18      <property name="javax.persistence.jdbc.url" value="
19        jdbc:mysql://localhost/whatsapp"/>
20      <property name="javax.persistence.jdbc.user" value="
21        root"/>
22      <property name="javax.persistence.jdbc.password"
23        value="root"/>
24      <!-- Configura es especificas do Hibernate -->
25      <property name="hibernate.dialect" value="org.
26        hibernate.dialect.MySQL5InnoDBDialect"/>
27      <property name="hibernate.hbm2ddl.auto" value="update
28        "/>
29      <property name="hibernate.show_sql" value="true"/>
30      <property name="hibernate.format_sql" value="true"/>
31    </properties>
32  </persistence-unit>
33 </persistence>

```

Código C.1 – Configuração do banco de dados

ANEXO D – Classe JPAUtil

```
1 package br.com.renato.util;
2
3 import javax.persistence.EntityManager;
4 import javax.persistence.EntityManagerFactory;
5 import javax.persistence.Persistence;
6
7 public class JPAUtil {
8
9     private static EntityManagerFactory emf = Persistence.
        createEntityManagerFactory("whatsapp");
10
11     public EntityManager getEntityManager() {
12         return emf.createEntityManager();
13     }
14
15 }
```

Código D.1 – Classe JPAUtil