

# Melhorando a Gestão de Relacionamento Com Clientes Por Meio de Mídias Sociais: um estudo sobre o boca-a-boca eletrônico

Gustavo R. Almeida<sup>1</sup>, José S. D'Aquino Netto<sup>1</sup>, Jorge L. F. Silva Junior<sup>1</sup>  
Antonio F. L. Jacob Jr.<sup>2</sup>, Fábio M. F. Lobato<sup>1</sup>

<sup>1</sup>Instituto de Engenharia e Geociências –  
Universidade Federal do Oeste do Pará (UFOPA) – Santarém – BR – Brasil

<sup>2</sup>Departamento de Engenharia da Computação –  
Universidade Estadual do Maranhão (UEMA) – São Luís – MA – Brasil

{gr.almeida00 , jorgeluzfigueira, jnetto.ti}@gmail.com,  
jacobjr@engcomp.uema.br, fabio.lobato@ufopa.edu.br

**Identificação do Trabalho:** O artigo está baseado no trabalho de iniciação científica do discente do curso de Bacharelado de Ciência da Computação (Bolsista PIBIC da UFOPA, 2016-2017) e no trabalho de Iniciação em Desenvolvimento Tecnológico e Inovação (Bolsista PIBIT do CNPQ, 2018, processo número 163321/2018-3) do discente do curso de Bacharelado em Sistemas da Informação. Os três discentes que atuaram nos planos de trabalho são alunos da Universidade Federal do Oeste do Pará e foram supervisionados pelo Prof. Dr. Fábio Manoel França Lobato e co-orientado pelo Prof. M.Sc. Antonio F. L. Jacob Jr.

**Abstract.** *The 2.0 Web dramatically changed the way consumers make purchase decisions, requiring a new management methodology for social networks data, known as Social Customer Relationship Management. Moreover, the electronic Word-of-Mouth emerged with these new environments, allowing users the ability to view or express opinions on social networks. Consequently, important user generated content reside within these networks, in form of reviews and complaints that, upon analysis, can result in competitive advantages to companies. The results of this work confirms that the tool is a technological innovation. Therefore, the built tool will allow a better company-customer interaction, generating greater reliability in the decision-making process of companies in sales and customer services.*

**Resumo.** *A Web 2.0 modificou radicalmente a forma pela qual os consumidores tomam decisões de compra, necessitando de uma nova abordagem de gerenciamento de dados de redes sociais, conhecida por Social Customer Relationship Management. Além disso, nesses novos ambientes surgiu o boca-a-boca eletrônico, sendo a capacidade de disponibilizar opiniões sobre produtos e serviços por meio de redes sociais com qualquer outro usuário. Consequentemente, dentro dessas redes residem dados gerados por usuários em forma de comentários e reclamações que, sob análise, podem trazer em vantagens competitivas para empresas. Os resultados deste trabalho confirmam que a ferramenta é uma inovação tecnológica, que permitiu uma melhor interação empresa-consumidor, gerando maior confiabilidade nos processos decisórios de vendas e de atendimento aos clientes.*

## 1. Introdução

A *Web 2.0* está cada vez mais inserida nos ambientes sociais e econômicos [Kaplan and Haenlein 2010]. Com a capacidade de qualquer usuário criar conteúdo *online*, a interação nas redes sociais aumentou drasticamente, necessitando-se de novos processos para entender e trabalhar esses dados [Fan and Gordon 2014, Stephen and Lehmann 2016]. Como consequência, empresas ficaram sobrecarregadas com o volume de dados gerados por seus clientes, levando a necessidade de uma estratégia mais eficiente para gerência do relacionamento com os clientes, abordagem conhecida por *Social Customer Relationship Management (Social CRM)* [Lobato et al. 2017]. Este mecanismo pode ser definido como a combinação das práticas de gerenciamento de clientes já utilizadas em sistemas de *CRM* tradicionais, como o monitoramento e análise de mídias sociais [Kubina and Lendel 2015].

Outra consequência dos novos ambientes *online* foi a modificação do comportamento do consumidor, devido não só à rapidez e à facilidade com que as informações são encontradas, mas também a maior interação entre as pessoas [Lima 2014]. A partir destas interações formou-se o boca-a-boca eletrônico, também conhecido por *Electronic Word-of-Mouth (eWoM)*. O *eWoM* pode ser definido como a capacidade de uma pessoa de comunicar suas opiniões por meio das mídias sociais com amigos, seguidores e outros usuários [Rodrigues 2014]. Este fenômeno tem um aspecto viral, pois quanto mais as pessoas estão envolvidas no processo, maior é a capacidade dele se expandir entre outros usuários [Wang et al. 2016]. Sítios *web* de comércio eletrônico e *MarketPlaces* disponibilizam áreas para os usuários darem suas opiniões sobre produtos e serviços (*reviews*).

Neste contexto, [Kim et al. 2018] analisaram a influência dos *reviews* nos negócios. No estudo, os autores concluíram que há uma falta de presença social em sites que contêm opiniões/avaliações de consumidores, e um crescente ceticismo sobre a credibilidade das informações de boca-a-boca eletrônico, pois os consumidores passaram a duvidar da credibilidade das fontes [Kim et al. 2018]. Um fenômeno contrário acontece nas plataformas de reclamações por dois aspectos principais. Primeiro por que os consumidores confiam nas informações ali contidas. Por exemplo, [Eslami et al. 2018] investigaram quais *reviews* de consumidores influenciavam mais as decisões de compra, concluindo que as avaliações mais úteis tinham tamanho médio, baixa pontuação e uma argumentação negativa. Já o segundo motivo da importância das plataformas de reclamação é que por meio da análise dos relatos de problemas enfrentados pelos clientes, é possível corrigir eventuais falhas nos processos internos ou em produtos [Almeida et al. 2017].

Alguns trabalhos dedicaram-se a analisar plataformas de reclamação e de reputação a fim de trazer vantagens competitivas às corporações. No entanto, verificou-se que há uma lacuna, tanto no estado da arte quanto no estado da prática, no tocante à ferramentas dedicadas a aquisição e análise automática de dados gerados nestas plataformas. À luz de tais fatos, definiu-se como objetivo deste trabalho o desenvolvimento de uma ferramenta para aquisição e análise automática de dados advindos de sistemas de reputação, a fim de auxiliar na tomada de decisões pelos gestores. O trabalho foi desenvolvido sob a ótica do *Design Science Research (DSR)* [Peppers et al. 2007]. Como estudo de caso, a plataforma ReclameAqui foi adotada considerando sua relevância no cenário nacional. Para a análise dos dados, a modelagem de tópicos foi conduzida, assim como, dois métodos de visualização de dados. Como contribuições tangíveis deste

trabalho destacam-se: i) o desenvolvimento de uma ferramenta para extração de dados da plataforma ReclameAqui; ii) a organização de conjuntos de dados de reclamações extraídas a partir da plataforma em estudo; iii) o desenvolvimento de uma ferramenta de análise e visualização de informações dos dados de reclamações.

O restante do artigo está estruturado como segue. Uma breve fundamentação teórica sobre gestão de relacionamento com os clientes é apresentada na Seção 2. Na Seção 3 os estudos que aplicam extração de dados utilizando *web crawlers* e análises de dados das plataformas online são discutidos. Na Seção 4 é apresentado o método de pesquisa utilizado, a experimentação executada e a avaliação. Na Seção 5 o estudo de caso e os resultados obtidos são apresentados e discutidos, respectivamente. Por fim, na Seção 6 as considerações finais são apresentadas.

## 2. Fundamentação Teórica

A dinamicidade proporcionada pelas Tecnologias da Informação e Comunicação (TICs) mudou drasticamente a orientação dos negócios, saindo de um foco em vendas e marketing, para algo mais centrado no consumidor [Ranjit 2002]. Por exemplo, com a proliferação de bases de dados sobre transações comerciais, as empresas passaram a explorar os dados dos clientes a fim de segmentá-los, possibilitando atendimento e marketing personalizados. Esta segmentação permitiu que as empresas atendessem melhor as expectativas dos clientes [Ranjit 2002]. Enquanto no início dos anos 2000 a Gestão de Relacionamento com os Clientes, mais conhecida pelo seu acrônimo *CRM* (de *Customer Relationship Management*), era considerada um fenômeno emergente [Zablah et al. 2004], atualmente o *CRM* se estabeleceu como uma estratégia mercadológica imprescindível para a sustentabilidade dos negócios [Greve and Schlüschen 2018, Kasemsap 2018].

Em seu trabalho seminal, Payne e Frow resumizam a literatura de *CRM* com o intuito de obter uma definição mais concisa [Payne and Frow 2005]. Para os autores, o *CRM* é uma abordagem estratégica que se preocupa em criar valor para os negócios por meio do desenvolvimento de relacionamentos apropriados com os principais clientes e segmentos de clientes. Os autores ainda completam afirmando que o *CRM* une o potencial das estratégias de *marketing* de relacionamento e TICs para criar relacionamento rentáveis e de longo prazo com os clientes e outras partes interessadas, requerendo uma integração interfuncional de processos, pessoas, operações e recursos de marketing que é ativada por meio de informações, tecnologia e aplicações [Payne and Frow 2005].

O *CRM* traz à luz diversos outros conceitos que acabam por se confundir com a própria gestão de relacionamento com os clientes, como o caso do marketing um-a-um [Peppers and Rogers 2000]. Isso ocorre devido a sua capilaridade nos negócios. A fim de facilitar o entendimento da dimensão do *CRM*, [Payne and Frow 2005, Peppers and Rogers 2000, Khodakarami and Chan 2014], além de outros autores, categorizam o *CRM* em três tipos, a saber:

- **Operacional:** é o componente que auxilia nos processos diários relacionados ao cliente a fim de melhorar a eficiência no atendimento, incluindo a automatização de vendas, atendimento ao cliente e marketing direcionado. O *CRM* Operacional está ligado diretamente com os *Enterprises Resources Planing Systems* (ERPs) e sistemas de vendas.

- **Colaborativo:** este elemento pode ser visto como o centro de comunicação que fornece a conexão entre as empresas, clientes, fornecedores e parceiros de negócio. Diferentemente do Operacional, este não está ligado com tarefas diárias ou aspectos relacionados à venda direta. Relacionam-se ao CRM Colaborativo canais como *e-mail*, comunidades, conferências e centros de interação com relacionamento via web. Estas plataformas aumentam a produtividade do trabalho e permitem o estabelecimento de uma parceria com os clientes, colaboradores e fornecedores.
- **Analítico:** este componente trabalha sinergicamente com os CRM colaborativos e operacional. Para [Greenberg 2001] o CRM analítico é constituído de procedimentos que envolvem a captação, armazenamento, processamento e transmissão de dados acerca das interações entre a empresa e seus clientes. Por este ser o componente mais tangível e também ubíquo nos diversos setores da empresa, [Swift 2000] propõem uma sub-categorização de acordo com as quatro dimensões que podem ser vistas como o ciclo fechado do sistema de gestão de clientes. As sub-categorias propostas por [Swift 2000] são descritas a seguir.

A primeira dimensão consiste na *Identificação do Cliente*. Este módulo inclui a segmentação dos consumidores e análise de clientes-alvo. A segmentação consiste na subdivisão do conjunto dos clientes em segmentos, os quais agrupam os clientes com características semelhantes. Já a análise de clientes-alvo define os possíveis *leads* com base em suas características. Ou seja, clientes interessados em determinado produto/serviço considerando as suas características. Na segunda dimensão, de *atração do cliente*, com os clientes identificados, as empresas concentram esforços para atraí-los, de acordo com sua segmentação. Dentre os impulsionadores para a captura dos *leads* (potenciais clientes), destacam-se vantagens competitivas como preço, promoções, *marketing direto etc.*

A terceira subcategoria é a de *retenção do cliente*, sendo esta uma das dimensões mais trabalhadas no CRM analítico, onde a satisfação do cliente é a principal preocupação para se efetivar a retenção dos mesmos. A satisfação pode ser definida como a comparação das expectativas do cliente com as percepções resultados da experiência como usuário do produto/serviço. Os elementos dessa dimensão incluem o *marketing* personalizado, programas de fidelidade, gerenciamento de reclamações e estratégias para contenção de danos. Por fim, mas não menos importante, tem-se a dimensão de *desenvolvimento do cliente*, que visa aumentar o valor do cliente, em outras palavras, a sua lucratividade individual. Análises conhecidas como “valor da vida útil do cliente” (*Customer Lifetime Cycle* tem como objetivo aumentar o número de produtos e serviços associados ao cliente. O desenvolvimento utiliza-se das ferramentas apresentadas nas dimensões anteriores.

Com a popularização da Internet, novos canais de atendimento surgiram, tais como *e-mails*, lojas virtuais, e, mais recentemente, as mídias sociais. Conforme mencionado anteriormente, estas últimas trouxeram consigo uma revolução no CRM, fazendo surgir o conceito de Gestão de Relacionamento com Clientes por meio de Mídias Sociais, o *Social CRM*. Neste sentido é importante observar que as dimensões do CRM não modificaram, apenas foi adicionado um novo e poderoso canal de comunicação, as mídias sociais, as quais representam também uma importante fonte de dados a respeito dos consumidores e suas preferências [Lobato et al. 2017].

### 3. Trabalhos Correlatos

Os sistemas de *CRM* têm evoluindo notoriamente, tanto na aplicação de métodos de análise mais rebuscados como a aplicação de inteligência computacional [Rodrigues Chagas et al. 2018], quanto na utilização de diferentes fontes de dados para enriquecimento das análises, como no caso do *Social CRM*. Por exemplo, [Duan et al. 2008, Liu et al. 2017] que demonstraram os efeitos *eWoM* em uma compra e também mudanças de atitude frente à uma marca. É importante destacar que *eWoM* não ocorre exclusivamente em plataformas de propósito específico. De acordo com [Lobato et al. 2017], redes sociais de propósito geral, como Facebook ou Twitter, também são utilizadas e a maior parte dos trabalhos utilizam-nas pois fornecem *Application Programming Interfaces* para aquisição de dados.

Plataformas de *eWoM* em sua maioria não disponibilizam *API's* para acesso aos seus dados, portanto investigou-se a adoção de um processo para extração de dados de páginas *web* sem *API*, identificado como método de *Web Crawling*. [Mislove et al. 2007] foi um dos precursores a utilizar este método, extraindo dados de redes sociais populares como *Flickr*, *Youtube*, *LiveJournal* e *Orkut*, realizando um “*crawling*” dos dados públicos dos usuários. Neste mesmo contexto, [Olmedilla et al. 2016] propôs uma metodologia para coleta de grandes volumes de dados gerados por usuários de sítios *web* de *eWoM* utilizando uma *framework* de *web crawler* conhecida por *Scrapy*. Como estudo de caso, os autores extraíram dados do Ciao UK<sup>1</sup>. Os resultados dessa extração mostraram-se promissores e o método desenvolvido foi apto a coletar dados do sítio *web* em estudo.

De acordo com [Li et al. 2016], a mineração de dados está sendo utilizada abrangentemente no *E-Commerce*. A maioria destas pesquisas está relacionada à análise de sentimento nas opiniões dos consumidores, identificando a tendência à geração de comentários positivos e negativos [Cirqueira et al. 2017a, Cirqueira et al. 2017b, Cirqueira et al. 2018]. Porém, a análise de sentimento não é aplicada à sítios *web* específicos de reclamações como o ReclameAqui, uma vez que os usuários postam apenas opiniões negativas [Fernandes and Fernandes 2017]. Portanto, uma análise mais pontual baseada em correlação e frequência de palavras deve ser utilizada.

O *LDA* tem sido utilizado em uma variedade de aplicações, demonstrando seu potencial na detecção de tópicos e problemas de classificação de *Corpus* linguísticos. Em [Feuerriegel et al. 2016], os autores tentaram identificar os efeitos de tópicos escondidos em notícias do setor financeiro, como preços de ações, usando *LDA*. Em uma linha semelhante, o trabalho de [Lee et al. 2015] fez uma categorização de tópicos de notícias usando modelos de *LDA* e uma representação esparsa para aprimorar sistemas de recomendação de notícias. [Qian et al. 2015] aplicou uma abordagem *LDA* multi-modal supervisionada para auxiliar na classificação de eventos de um grande número de dados gerados por usuários, dispostos em mídias sociais.

### 4. Metodologia

Nesta seção a metodologia seguida é apresentada, de acordo com *Design Science Research* [Peppers et al. 2007]. Este é definido como um modelo de processos para desenvolvimento de projetos nas áreas de tecnologia (Sistemas de Informação e Computação). A

---

<sup>1</sup><http://www.ciao.uk/>

seguir estão apresentados as etapas do processo, divididos em seis fases, de acordo com a Figura 1.



**Figura 1. Metodologia de desenvolvimento *DSR* para a ferramenta desenvolvida.**

Convém destacar que o *DSR* é um processo orientado a ciclos e a Figura 1 apresenta o resultado final dos diversos ciclos conduzidos. Em resumo, o processo de construção da solução apresentada neste artigo passou por três ciclos. No primeiro, construiu-se uma ferramenta para extração dos dados. O artefato obtido foi de fácil demonstração e avaliação, onde o resultado era constituído apenas das bases com os dados obtidos. O segundo ciclo consistiu na concepção, projeto e desenvolvimento do módulo de análise. Por fim, a integração destes dois módulos e construção de um sistema que permitisse a visualização do conhecimento extraído de forma intuitiva foi realizada no terceiro ciclo.

A avaliação dos dois últimos ciclos está intimamente relacionada com a compreensão do domínio de aplicação, a qual ocorreu em duas fases. Primeiramente, conduziu-se uma revisão da literatura acerca do *eWoM* e sistemas de reputação. Como resultado,

identificou-se que *eWoM* consiste no compartilhamento de avaliações sobre produtos e serviços por meio de plataformas *on-line* [Barreto 2014]. Geralmente, estas avaliações são baseadas em texto e vêm acompanhadas de uma pontuação que visa refletir o nível de satisfação do cliente. Esta etapa também teve o suporte de especialistas do *Social CRM Research Center* (SCRC), com base principal da universidade de *Leipzig*, Alemanha. Os especialistas do SCRC que auxiliaram na tarefa possuem formação em Comunicação com ênfase em Marketing e Inovação em Redes Sociais; e Sistemas de Informação, com ênfase em Sistemas de Informação Sociais. Todos desenvolvem pesquisa aplicada e prestam consultoria para empresas da área.

#### 4.1. Estudo de caso: a plataforma ReclameAqui

A escolha da plataforma ReclameAqui como estudo de caso levou em consideração os seguintes aspectos: i) relevância da plataforma para o cenário de reclamações [Rocha 2010]; ii) viabilidade na obtenção dos dados; iii) possibilidade de geração de insights para as empresas. Sobre o primeiro aspecto, considerou-se que a quantidade de empresas (mais de 120 mil) e usuários (mais de 18 milhões de consumidores) cadastrados; Além disso, o seu posicionamento no *ranking* Alexa demonstra sua relevância. Por exemplo, em Janeiro de 2019 a plataforma apareceu entre os 1.000 sites mais influentes de acordo com esse *ranking*.

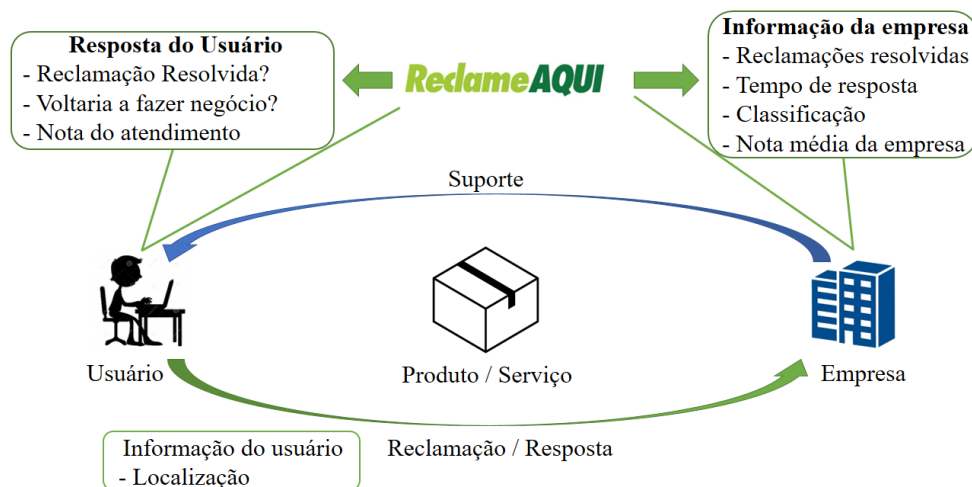
Em relação ao segundo aspecto, a plataforma não requer identificação do usuário (*login*) para acesso aos seus dados. Além disso, por se tratar de conteúdo gerado pelo usuário e por disponibilizar meta-dados como localização do cliente, dia-mês-ano e hora da reclamação, pontuação da reputação e a própria reclamação em si, os especialistas consultados visualizaram diversos cenários de aplicabilidade da solução desenvolvida, podendo auxiliar empresas na detecção de problemas recorrentes e na otimização de processos internos, por exemplo. Ademais, realizou-se também uma pesquisa exploratória de sistemas de reputação online, onde outras concorrentes como *eBit* e *BondFaro* demonstraram grande incidência de dados interessantes, porém o ReclameAqui demonstrou maior inserção de empresas em seu sistema, e uma quantidade de reclamações que propiciariam análises mais acuradas. A Figura 2 apresenta o funcionamento básico da plataforma.

O processo descrito na Figura 2 inicia com uma reclamação por parte do usuário. O segundo passo fica a cargo da empresa, a qual deve prover uma possível solução ou esclarecimento. Em seguida, o usuário pode replicar e avaliar o atendimento, por meio de um questionário. A Tabela 1 a seguir descreve os dados com relevância para extração da plataforma e como eles estão formatados.

#### 4.2. Extração de Dados

O ReclameAqui é um *sítio web* baseado em *Javascript*. Portanto, a abordagem utilizada por [Olmedilla et al. 2016] não funciona apropriadamente. Por esta razão, adaptamos o *framework* acima considerando as particularidades do ReclameAqui. O esquema de funcionamento é apresentado na Figura 3.

A Figura 3 apresenta alguns detalhes técnicos necessários para o correto funcionamento do processo de extração de dados, a qual inicia-se com a conversão dos páginas de *Javascript* para *HyperText Markup Language (HTML)* usando o navegador *PhantomJS*. O segundo passo (*crawling*) extrai os links de cada reclamação, sendo que todas as páginas



**Figura 2. Esquema de funcionamento da plataforma ReclameAqui.**

Nome	Formato	Descrição
Título	Corpo de Texto	Título da Reclamação
ID do Usuário	ID: 01234567	ID de identificação do usuário.
Reclamação	Corpo de texto	Texto da reclamação de um usuário.
Localização	Cidade-Estado	Cidade e Estado do usuário reclamante.
Data/Hora	DD/MM/AA às 00h00	Data e horário da reclamação.
Resposta da Empresa	Corpo de texto	Resposta da empresa à reclamação.
Resposta do Usuário	Corpo de texto	Possível resposta do usuário à empresa.
Declaração Final	Corpo de texto	Resposta final do usuário à reclamação.
Voltaria a Negociar?	Sim/Não	Resposta se voltaria a negociar com a empresa.
Reclamação Resolvida?	Sim/Não	Resposta do usuário se a reclamação foi resolvida.
Pontuação	0 - 10	Pontuação do usuário ao atendimento recebido.

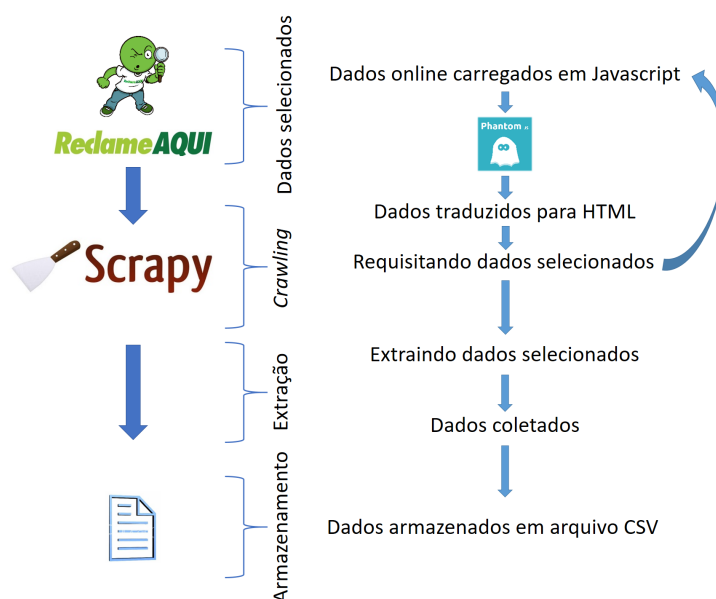
**Tabela 1. Descrição dos dados disponibilizados na plataforma.**

são convertidas para *HTML*, a fim de permitir que a informação seja extraída na terceira etapa. Os dados coletados são então organizados e armazenados em um arquivo no formato *Comma-Separated Values (CSV)*. Este formato foi escolhido considerando a implementação do método de análise, o qual é descrito a seguir. E por fim, conduziu-se a preparação dos dados, na qual os itens cujo formato eram “Corpo Textual” passaram pelas etapas básicas de pré-processamento aplicados à mineração de textos, a saber: remoção de *stop words* e redução de dimensionalidade tal como em [Cirqueira et al. 2018].

### 4.3. Avaliação

Primeiramente, foi conduzida uma análise baseada em modelagem de tópicos a fim de reduzir o tamanho descritivo e revelar a estrutura inter/intra estatística de documentos, por meio do uso de modelos geradores probabilísticos [Wallach 2006]. Dois algoritmos foram testados, o *LDA* e o *LSA*, considerando uma combinação dos parâmetros de quantidade de palavras por tópicos variando par a par de 5, 10 e 15. Por exemplo, foram executadas modelagens de cinco tópicos e cinco palavras (5x5) ou dez tópicos e quinze palavras (10x15). Ao todo, nove combinações foram testadas para cada algoritmo. A técnica escolhida para realizar a modelagem foi o *LDA*. Em seguida, dois autores anotaram os





**Figura 3. Diagrama de fluxo dos processos da extração de dados.**

tópicos, ou seja, atribuíram um sentido para as palavras encontradas naquele tópico. Um terceiro validara a anotação. A escolha da parametrização a ser adotada como padrão leva em conta a facilidade com a qual foi possível anotar o tópico e também o conjunto mínimo de palavras-tópico. Neste sentido, o cenário escolhido foi de cinco tópicos e cinco palavras por tópico. Vale ressaltar que cabe ao usuário a escolha da parametrização. Isso foi descrito na seção de resultados

Optou-se adicionalmente por utilizar outros métodos de visualização de dados. O Mapa de Calor, apresentando a quantidade de reclamações por região do território brasileiro. O método de nuvem de palavras foi utilizado para gerar uma visualização dos termos mais utilizados e a sua frequência nos títulos das reclamações. Todos os resultados obtidos foram validados pelos especialistas consultados por meio da apresentação das análises feitas para os estudos de caso selecionados.

## 5. Resultados Obtidos

Nesta seção são expostos os resultados adquiridos ao longo da pesquisa e implementação da ferramenta apresentada. Desse modo, tem-se a ferramenta desenvolvida, apresentando algumas características da mesma, os resultados experimentais do estudo de caso definido a seguir e os *insights* advindos dos artefatos produzidos nas análises.

### 5.1. Ciclos de Desenvolvimento e Arquitetura do Sistema

O processo de implementação do sistema proposto foi organizado em três fases, conforme descrito na seção de Metodologia. A primeira foi direcionada ao desenvolvimento de um *web crawler* para a aquisição dos dados da plataforma *ReclameAqui*. Na etapa seguinte foram implementados os algoritmos de análise, mais especificamente a modelagem de tópicos (*LDA*), e outros dois algoritmos responsáveis por gerar a análise de mapa de calor e nuvem de palavras. Na terceira etapa, projetou-se e implementou-se a interface *web* para interação do usuários com as funcionalidades e a realizou-se a integração dos módulos.

A Figura 4 apresenta o diagrama arquitetural do sistema e o seu esquema de funcionamento.

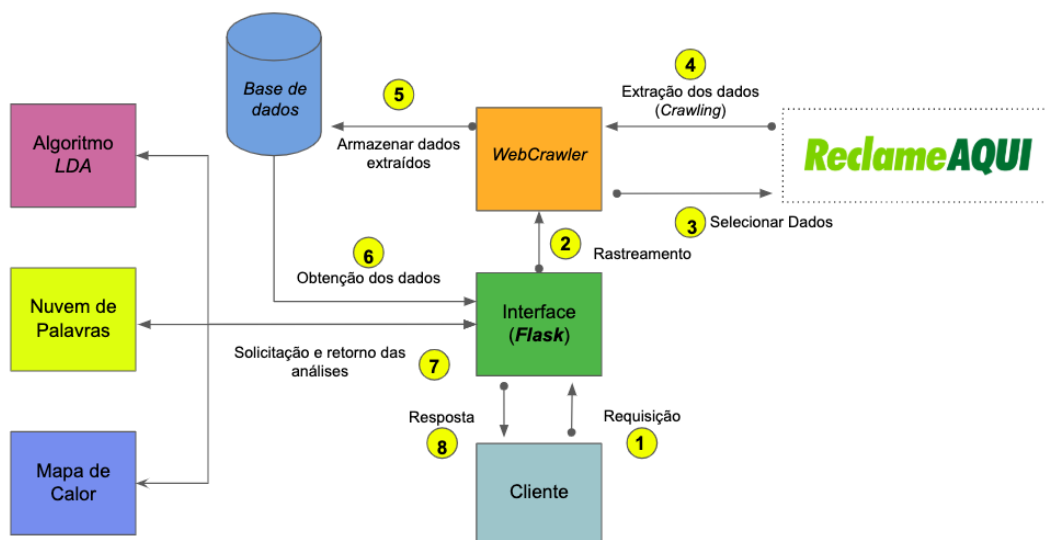


Figura 4. Arquitetura do Sistema.

As etapas do sistemas são discriminadas da seguinte forma. Na **Requisição** (1) o usuário realiza uma solicitação pela Interface Web, fornecendo o nome da empresa e quantidade de páginas a serem coletadas. Então, de forma automática o sistema passa para a etapa de **Rastreamento** (2), onde o *webcrawler* é iniciado, navega pela plataforma *ReclameAqui* com a finalidade de capturar os dados da empresa. Nas etapas de **Seleção e Extração dos dados** (3 e 4), as informações relevantes da empresa são selecionadas e extraídas, para posterior **Armazenamento** (5).

Na interação (6), descrita como **Obtenção dos Dados**, o usuário é redirecionado para uma página a qual exibe todas as informações coletadas. Então ele pode solicitar a **Execução das Análises** (7), podendo escolher entre as seguintes opções: modelagem de tópicos, nuvem de palavras e mapa de calor. O primeiro aciona o algoritmo *LDA* e ao final de sua execução o sistema retorna com uma página que exibe um conjunto de termos e divididos por tópicos. Quanto ao botão nuvem de palavras, é exibido uma página uma visualização semelhante à Figura 7, no qual o tamanho de cada palavra é proporcional ao seu nível de incidência. Por fim, a análise de mapa de calor indica a região de maior índice de reclamações. Por fim, a interação **Resposta** (8) tem como resultado a exibição dos dados extraídos e suas devidas análises.

Os processos de 2 à 5 foram implementados no primeiro ciclo de desenvolvimento, cujo artefato foi o coletor de dados. O segundo ciclo teve como artefato o conjunto de algoritmos de análise, os quais são utilizados no processo 7. O último ciclo, de integração, conectou todos os processos acima descritos, resultado nas etapas 1, 6 e 8. A interface construída é apresentada na subseção a seguir.

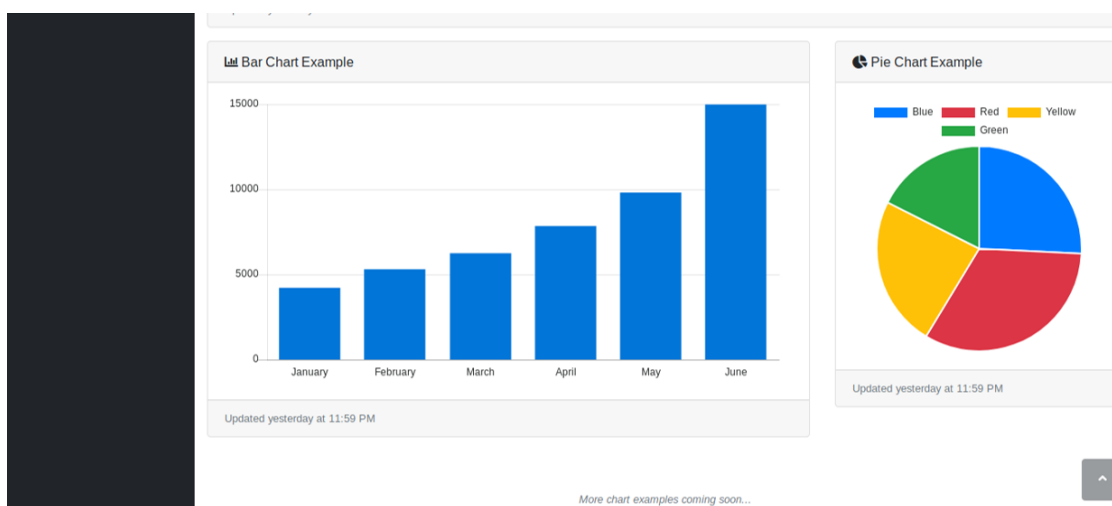
## 5.2. Interface de Utilização

Um dos grandes desafios das pesquisas que trabalham com dados é o monitoramento e a evolução dos resultados de acordo os objetivos traçados. Dessa forma, foi desenvolvida

uma Painel de Indicadores (*dashboard*), que visa mostrar as métricas e indicadores de forma visual, facilitando a compreensão das informações geradas. A aplicação foi estruturada por um *micro-framework* baseado em *Python* chamado *Flask*. Por conseguinte, todas as configurações foram padronizadas. As duas principais páginas são o *dashboard* e a tela de indicadores, respectivamente apresentadas nas Figuras 5 e 6.



**Figura 5. Visão Geral da Dashboard.**



**Figura 6. Visualização com mais indicadores.**

### 5.3. Estudo de Caso

Os dados extraídos da plataforma chegam a aproximadamente 180.000 reclamações, sendo que as reclamações estão datadas no período de seis meses para 15 empresas e de um período de um ano para 18 empresas. Para avaliar a metodologia proposta, duas empresas foram selecionadas como estudo de caso. A Tabela 2 sumariza as informações das empresas, referenciadas no texto como Empresas A e B, respectivamente. O nome das empresas serão preservados para garantir a confidencialidade dos dados.

Denominação	Sector do Mercado	Nº de Reclamações	Período Observado
Empresa A	Eletrônicos	5537	Um ano
Empresa B	Automotivo	1060	Seis meses

**Tabela 2. Resumo dos estudos de caso.**

Os resultados para as análises das empresas A e B são apresentadas a seguir. A Tabela 3 apresenta os tópicos extraídos. Convém destacar que o usuário pode configurar a sua parametrização (número de tópicos e número de palavras por tópicos).

<b>Empresa A</b>				
<b>Tópico identificado</b>	<b>Termos mais significativos</b>			
Atendimento / Assistência	Loja	Atendimento	Cabo	Shopping
Produto / Uso	Produto	MARCA A	Uso	Comprei
Garantia / Assistência	Aparelho	Assistência	Garantia	Reparo
Dados do aparelho	MARCA A	Contato	Conta	Senha
Atendimento / Defeito	Loja	Atendimento	Cabo	Shopping
<b>Empresa B</b>				
<b>Tópico identificado</b>	<b>Termos mais significativos</b>			
Concessionária / Carro	Carro	Concessionária	Veículo	Problemas
Concessionária / Peça	Veículo	Carro	Concessionária	Peça
Carro	Carro	Comprei	Concessionária	Concessionária
Oficina	Meio	Referente	Retorno	Oficina
Garantia / Revisão	Km	Garantia	Revisão	Carro

**Tabela 3. Modelagem de Tópicos resultante da utilização do LDA para as empresas A e B.**

Por meio dos dados dispostos na Tabela 3 é possível concluir que para a empresa A verifica-se a maior quantidade de reclamações relacionadas principalmente a: 1- Atendimento; 2- Assistência; 3- Produto/Uso (MARCA A); 4- Defeito/Garantia. Há uma presença maior de reclamações sobre o termo "atendimento", envolvendo o próprio atendimento ao cliente, a assistência técnica e o funcionamento do benefício de garantia. Para a empresa B as reclamações são focadas em: 1- Concessionária; 2- Peças de concessionárias; 3- Oficina/Revisão. Desta empresa é verificado a existência de problemas de logística nas oficinas e concessionárias, causando problemas em peças e no próprio conserto dos veículos. Diferente da primeira empresa, a questão do atendimento não foi identificada nos resultados. Estes dados provêm informações importantes de posicionamento e foco de esforços das empresas em pontos específicos de seus produtos e serviços prestados.

Partindo para as visualizações de dados, primeiramente temos as nuvens de palavras. As figuras mostram a frequência de termos nos títulos das reclamações, com o objetivo de verificar se o conteúdo mais citado nos títulos se assemelha ao das reclamações, além de verificar a forma como os usuários apresentam sua reclamação. Quanto mais frequente a palavra, maior é a sua escala na imagem. A Figura 7 apresenta um exemplo referente à Empresa A.

Por meio das visualizações apresentadas vemos que para a empresa A os termos mais citados nas reclamações são relacionados à: 1 - Garantia e Assistência; 2 - Bateria, tela e carregador; 3 - Atualização, defeito travamento; 4 - Cobrança. Vemos que mais



Como contribuições tangíveis deste trabalho destacam-se: i) o desenvolvimento de uma ferramenta para extração de dados da plataforma ReclameAqui; ii) a organização de conjuntos de dados de reclamações extraídas a partir da plataforma em estudo; iii) o desenvolvimento de uma ferramenta de análise e visualização de informações dos dados de reclamações.

Os resultados obtidos demonstraram que os dados do ReclameAqui podem ser utilizados para agregar valor aos negócios, auxiliando no aprimoramento dos atuais sistemas de *Social CRM*. O painel de indicadores implementado é intuitivo e foi validado por especialistas em *marketing e administração*. Adicionalmente, conjuntos de dados obtidos também possibilitam que outros pesquisadores da área extrapolem as análises realizadas. Como trabalhos futuros pretende-se finalizar a integração da interface de usuário com os demais módulos, implementar novas análises e também expandir a coleta de dados para outras plataformas como a *ebit* e o *consumidor.gov.br*.

## Referências

- Almeida, G. R. T. d., Lobato, F., and Cirqueira, D. (2017). Improving Social CRM through electronic word-of-mouth: a case study of ReclameAqui. In *XIV Workshop de Trabalhos de Iniciação Científica*.
- Barreto, A. M. (2014). The word-of-mouth phenomenon in the social media era. *International Journal of Market Research*, 56(5):631–654.
- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., and Santana, A. (2018). A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 746–749.
- Cirqueira, D., Jacob, A., Lobato, F., de Santana, A. L., and Pinheiro, M. (2017a). Performance Evaluation of Sentiment Analysis Methods for Brazilian Portuguese. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 245–251. Springer International Publishing, Cham.
- Cirqueira, D., Vinícius, L., Pinheiro, M., L., J. J. A. F., Lobato, F., and Santana, Á. (2017b). Opinion Label : A Gamified Crowdsourcing System for Sentiment Analysis Annotation. In *XVI Workshop de Ferramentas e Aplicações*.
- Duan, W., Gu, B., and Whinston, A. (2008). The dynamics of online word-of-mouth and product sales-an empirical investigation of the movie industry. *Journal of Retailing*, 84(2):233–242.
- Eslami, S. P., Ghasemaghaei, M., and Hassanein, K. (2018). Which online reviews do consumers find most helpful? a multi-method investigation. *Decision Support Systems*, 113:32 – 42.
- Fan, W. and Gordon, M. D. (2014). The Power of Social Media Analytics. *Commun. ACM*, 57(6):74–81.
- Fernandes, T. and Fernandes, F. (2017). *Social Media and Tourism: The Case of E-Complaints on TripAdvisor (An Extended Abstract)*, pages 825–829. Springer International Publishing, Cham.

- Feuerriegel, S., Ratku, A., and Neumann, D. (2016). Analysis of how underlying topics in financial news affect stock prices using latent dirichlet allocation. In *49th Hawaii International Conference on System Sciences, HICSS 2016, Koloa, HI, USA, January 5-8, 2016*, pages 1072–1081.
- Greenberg, P. (2001). *Crm at the Speed of Light: Capturing and Keeping Customers in Internet Real Time*. McGraw-Hill Professional.
- Greve, G. and Schlüschen, A. (2018). From Customer Relationship Management to Influencer Relationship Management. pages 80–91.
- Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1):59–68.
- Kasemsap, K. (2018). The Fundamentals of Customer Relationship Management. pages 44–60.
- Khodakarami, F. and Chan, Y. E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information & Management*, 51(1):27–42.
- Kim, S., Kandampully, J., and Bilgihan, A. (2018). The influence of ewom communications: An application of online social network framework. *Computers in Human Behavior*, 80:243 – 254.
- Kubina, M. and Lendel, V. (2015). Successful Application of Social CRM in The Company. *Procedia Economics and Finance*, 23(October 2014):1190–1194.
- Lee, Y. S., Lo, R., Chen, C. Y., Lin, P. C., and Wang, J. C. (2015). News topics categorization using latent dirichlet allocation and sparse representation classifier. In *2015 IEEE International Conference on Consumer Electronics - Taiwan*, pages 136–137.
- Li, J., Fong, S., Yan, Z., and Khoury, R. (2016). Hierarchical classification in text mining for sentiment analysis of online news. *Soft Comput.*, 20(9):3411–3420.
- Lima, V. (2014). O boca a boca virtual como aliado do consumidor na era das mídias sociais: Um estudo exploratório do site reclame aqui. Technical report, Universidade Federal Fluminense.
- Liu, J., Li, C., Ji, Y. G., North, M., and Yang, F. (2017). Like it or not: The Fortune 500’s Facebook strategies to generate users’ electronic word-of-mouth. *Computers in Human Behavior*, 73:605–613.
- Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2017). Social CRM: Biggest Challenges to Make it Work in the Real World. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 221–232. Springer International Publishing, Cham.
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC ’07*, pages 29–42, New York, NY, USA. ACM.

- Olmedilla, M., Martínez-Torres, M., and Toral, S. (2016). Harvesting Big Data in social science: A methodological approach for collecting online user-generated content. *Computer Standards & Interfaces*, 46:79–87.
- Payne, A. and Frow, P. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69(4):167–176.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77.
- Peppers, D. and Rogers, M. (2000). *CRM Series Marketing 1-to-1: um guia executivo para entender e implantar estratégias de Customer Relationship Management*. São Paulo.
- Qian, S., Zhang, T., Xu, C., and Hossain, M. S. (2015). Social event classification via boosted multimodal supervised latent dirichlet allocation. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(2):27:1–27:22.
- Ranjit, B. (2002). Customer relationship management: key components for IT success. *Industrial Management & Data Systems*, 102(2):89–97.
- Rocha, M. A. V. (2010). A rede social como instrumento de análise pós-consumo de produtos de moda-vestuário: um estudo exploratório. *Rio de Janeiro, RJ*, 55.
- Rodrigues, D. (2014). A influência do word of mouth online na decisão de compra do consumidor. Master's thesis, IADE - Creative University.
- Rodrigues Chagas, B. N., Nogueira Viana, J. A., Reinhold, O., Lobato, F., Jacob, A. F. L., and Alt, R. (2018). Current Applications of Machine Learning Techniques in CRM: A Literature Review and Practical Implications. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 452–458. IEEE.
- Stephen, A. T. and Lehmann, D. R. (2016). How word-of-mouth transmission encouragement affects consumers' transmission decisions, receiver selection, and diffusion speed. *International Journal of Research in Marketing*, 33(4):755–766.
- Swift, R. (2000). *Accelerating Customer Relationships: Using Crm and Relationship Technologies<sup>TM</sup>*. Prentice Hall Press, Upper Saddle River, NJ, USA, first edition.
- Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, pages 977–984, New York, NY, USA. ACM.
- Wang, T., Yeh, R. K.-J., Chen, C., and Tsydypov, Z. (2016). What drives electronic word-of-mouth on social networking sites? perspectives of social capital and self-determination. *Telematics and Informatics*, 33(4):1034 – 1047.
- Zablah, A. R., Bellenger, D. N., and Johnston, W. J. (2004). An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon. *Industrial Marketing Management*, 33(6):475–489.