



Universidade Federal de Pernambuco  
Centro de Informática  
Graduação em Ciência da Computação

## **Business Intelligence e Análise de Sentimentos no Contexto de Redes Sociais Online**

Leonardo José de Andrade Costa Santos

Proposta de Trabalho de Graduação

Orientador: Fernando da Fonseca de Souza

Recife  
Agosto de 2015

# Resumo

A utilização de redes sociais tem crescido de forma assustadora nos últimos anos, o que contribuiu para um aumento exponencial na produção de informações na Web. Este crescimento tem atraído o interesse de diversas organizações por consistir numa oportunidade de investigar o que seus consumidores falam sobre suas marcas e obter informações sobre estes clientes, visto que uma análise eficaz deste tipo de informação pode ajudar a guiar decisões corporativas. Este trabalho tem como objetivo a realização de um estudo de técnicas e ferramentas de extração de largas quantidades de dados não estruturados oriundos de redes sociais online. Será discutido como este tipo de informação pode ser recuperado e como ela pode ser útil em diferentes contextos. Aliado ao processo de extração de dados, o trabalho seguirá explorando algoritmos de análise e tratamento de dados. Em particular, técnicas de análise de sentimentos, que possuem o objetivo de avaliar sentimentos expressos em fragmentos de texto. Por fim, será mostrado um caso de estudo desenvolvido que ilustra como soluções de Business Intelligence e análise de sentimentos podem ser combinadas dentro do contexto de redes sociais para diversos tipos de análise que podem ser extremamente úteis para corporações.

**Palavras-chave:** extração de dados, big data, integração de dados, dados não estruturados, dados sociais, mineração de opinião.

# Abstract

The usage of social networks has grown substantially in the past years and this has contributed to an exponential increase of the production of information on Web. This growth has attracted the interest of several organizations since it is an opportunity to investigate what their customers talk about their brands and to obtain meaningful information about these clients. This is particularly relevant because an efficient analysis of this kind of information can help enormously to guide corporate decisions. This work intends to make a study of the available techniques and tools to extract large amounts of non-structured data from online social networks. It will discuss how this sort of information can be retrieved and how it can be useful in several different contexts. Along with this data extraction process, the work then follows with exploring data analysis algorithms. In particular, sentiment analysis techniques, which are useful for classifying sentiments expressed in text fragments. Finally, the work will show a developed case study that illustrates how Business Intelligence and sentiment analysis can be combined in the context of online social networks to perform a number of different types of analysis that can be valuable to corporations.

**Keywords:** data extraction, big data, data integration, non-structured data, social data, opinion mining.

# Sumário

1. Introdução .....	01
2. Objetivos .....	03
3. Estrutura do Trabalho .....	04
4. Cronograma .....	05
5. Possíveis Avaliadores .....	06
Assinaturas .....	07
Referências Bibliográficas .....	08

# Introdução

Na última década, observou-se um crescimento imenso da utilização de redes sociais no Brasil e no mundo. Este crescimento tem sido particularmente acelerado com a inclusão digital e popularização de smartphones e acesso à rede de baixo custo. Websites como Twitter<sup>1</sup>, Facebook<sup>2</sup>, LinkedIn<sup>3</sup>, Instagram<sup>4</sup>, Pinterest<sup>5</sup>, entre vários outros, tem atraído usuários, que se conectam com outros membros e se comunicam de diferentes formas sobre os mais variados temas, sentindo-se livres para expressar suas opiniões e potencialmente influenciar outros usuários (POLONI; TOMAÉL, 2014).

Essa ampla utilização de redes sociais traz consigo uma oportunidade para que empresas possam entender as necessidades de seus consumidores, descobrir o que tem sido falado sobre suas marcas, monitorar seus concorrentes, traçar perfis de seus clientes, entre outras estratégias relevantes para adquirir informações sobre o seu público (WEBER, 2009). Assim, encontrar, processar e armazenar dados relevantes no mercado de mídias sociais tem sido uma das maiores preocupações de organizações nos dias atuais, que tem investido cada vez mais em soluções com esta finalidade.

Para possibilitar diferentes análises da informação compartilhada nestas diferentes mídias sociais, é necessário realizar a coleta dos dados que são veiculados nestas plataformas. A fim de satisfazer essa necessidade de extrair este tipo de informação, atualmente, as principais redes sociais provêm interfaces de programação (*Application Programming Interfaces* – API), que permitem a captura total ou parcial de seus dados (FRANÇA et al., 2014). Este tipo de ferramenta de consumo viabiliza o uso de dados de redes sociais na alimentação de diversos tipos de aplicação.

Dispondo dos dados coletados, é preciso analisá-los para que se possa inferir informações relevantes. Devido à rápida produção de informação que se observa em redes sociais, uma análise manual deste tipo de dados torna-se inviável, fazendo-se necessária a busca por formas automáticas de classificar e interpretar textos publicados nestas mídias. Uma área de pesquisa neste contexto que tem ganhado atenção da academia é a de análise de sentimentos ou mineração de opinião, que se refere ao tratamento computacional de um texto com o intuito de identificar se este representa uma sentença positiva, negativa ou neutra sobre certo tópico (COSTA et al., 2012). Assim, conceitos de análise de sentimentos podem contribuir fortemente para uma compreensão mais detalhada de dados de redes sociais.

---

<sup>1</sup> <https://www.twitter.com/>

<sup>2</sup> <https://www.facebook.com/>

<sup>3</sup> <https://www.linkedin.com/>

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

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

Para entender de forma completa os dados analisados, também é preciso incorporar dados temporais e ferramentas de visualização apropriadas. Soluções de Business Intelligence (BI), que abrangem técnicas e ferramentas para transformação de dados puros em informação útil e relevante para análises de negócios, já conquistaram seu espaço entre grandes organizações por suprirem essa necessidade e continuam crescendo com um dos maiores índices no mercado de software (SALLAM et al., 2011). Este tipo de solução também pode ser usado para visualizar dados e explorar tendências em mídias sociais. Por meio da análise e geração de relatórios em tempo real, observações mais sofisticadas são possíveis, permitindo às organizações construir correlações que a observação humana por si só não seria capaz de replicar.

Dessa forma, é possível observar que a extração de dados de mídias sociais, combinada com a análise de sentimentos e soluções de Business Intelligence podem contribuir imensamente para o mundo corporativo para a análise de dados sociais e todas as suas vantagens associadas. Sendo assim, torna-se importante a realização de um estudo sobre redes sociais online, formas disponíveis de extrair dados deste tipo de plataforma e como as áreas previamente apresentadas podem ser combinadas em uma arquitetura que permita análises relevantes para organizações.

# Objetivos

O objetivo deste trabalho é fazer um estudo sobre redes sociais online e técnicas computacionais que podem ser utilizadas dentro desse contexto para viabilizar e otimizar análises úteis a diversos tipos de organizações. Especificamente, pretende-se explorar o processo de extração de dados das principais plataformas sociais atuais e como técnicas de análise de sentimentos e soluções de Business Intelligence podem ser usadas para permitir consulta e visualização de informações relevantes a partir destes dados.

# Estrutura do Trabalho

O trabalho será composto pelos seguintes capítulos:

- Introdução introduzirá o tema, mostrando a motivação para execução do trabalho, detalhando seus objetivos e justificando a utilização das tecnologias apresentadas no contexto de redes sociais online;
- Redes Sociais Online aborda com maior profundidade o conceito de redes sociais online e detalha cada uma das principais plataformas atuais, com foco em suas peculiaridades e como é possível extrair dados de cada uma delas;
- Análise de Sentimentos percorrerá sobre o estado da arte da área e como ela pode ser útil para classificação de conteúdo social. Segue destacando principais algoritmos propostos e testes realizados com conteúdo de redes sociais;
- Business Intelligence avaliará soluções de Business Intelligence disponíveis no mercado, com foco em sua utilização para consultas e análises inteligentes de dados sociais;
- Caso de Estudo apresentará uma implementação de uma arquitetura baseada em extração de dados sociais, análise de sentimentos e Business Intelligence para ilustrar sua eficácia, mostrando os resultados obtidos;
- Conclusão: É mostrado um breve resumo do estudo realizado e da arquitetura apresentada. Em seguida, são analisados os resultados da pesquisa e implementação realizados e desafios encontrados na área. Finalmente, serão apresentadas as limitações e sugestões de trabalhos futuros;
- Referências Bibliográficas: Lista as referências utilizadas na pesquisa;
- Anexos: Contém possíveis anexos complementares ao trabalho.

# Cronograma

As atividades pretendem ser desenvolvidas de acordo com o cronograma abaixo.

Atividade	Agosto				Setembro				Outubro				Novembro			
Revisão bibliográfica	■	■	■	■	■	■										
Levantamento e estudo das principais plataformas sociais				■	■	■										
Pesquisa sobre os processos de extração de dados nas plataformas levantadas					■	■	■	■								
Pesquisa e elaboração do capítulo de análise de sentimentos							■	■	■	■						
Pesquisa e elaboração do capítulo de Business Intelligence									■	■	■	■				
Desenvolvimento do caso de estudo que ilustre o funcionamento											■	■	■	■		
Elaboração do relatório final														■	■	
Preparação da defesa															■	
Defesa																■

## **Possíveis Avaliadores**

São possíveis avaliadores do trabalho a ser produzido conforme as especificações nessa proposta:

- Flavia de Almeida Barros (CIn/UFPE);
- Germano Crispim Vasconcelos (CIn/UFPE);
- Robson do Nascimento Fidalgo (CIn/UFPE).

# Assinaturas

---

Fernando da Fonseca de Souza  
Orientador

---

Leonardo José de Andrade Costa Santos  
Aluno

## Referências Bibliográficas

- [1] COSTA, P. R. S.; SOUZA, F. F.; TIMES, V. C.; BENEVUTO, F. “Towards Integrating Online Social Networks And Business Intelligence”, Proceedings of the International Conferences Web Based Communities and Social Media. Lisboa, p. 21-32, 2012.
- [2] POLONI, K. M.; TOMAÉL, K. I., “Coleta de Dados em Plataformas de Redes Sociais: Estudo de Aplicativos”, Anais do III Workshop de Pesquisa em Ciência da Informação – III WPCI ‘14. Londrina - PR, 2014.
- [3] SALLAM, R. L.; HOSTMANN, B.; SCHLEGEL, S.; TAPADINHAS, J.; PARENTAU, J., “Gartner Magic Quadrant for Business Intelligence Platforms – 2015”. [Online] Disponível em: <<http://www.gartner.com/technology/reprints.do?id=1-2ACL1P&ct=150220&st=sb>> [Acesso em 26 Jul 2015].
- [4] WEBER, L., “Marketing to the Social Web: How Digital Customer Communities Build Your Business”. 2. ed. Hoboken: John Wiley & Sons, out. 2009.
- [5] FRANÇA, T. C.; FARIA, F. F.; RANGEL, R. M.; FARIAS, C. M.; OLIVEIRA, J., “Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais”, XXIX Simpósio Brasileiro de Banco de Dados – SBBDD '14. Curitiba – PR, 2014.
- [6] MAIER, W.; RADOIU, D., “Unstructured Social Networks Data for Business Context Analysis”, Scientific Bulletin of the Petru Maior University of Tirgu Mures, vol. 9, n. 2, set. 2012.
- [7] EISENBERG, E. M.; JOHNSON, Z.; PIETERSON, W., “Leveraging Social Networks for Strategic Success”, International Journal of Business Communication 2015, vol. 52, p. 143–154, jul. 2015.