

Brand Feeling: um sistema para analisar o sentimento dos usuários em relação a uma marca

Aluno: Luiz César Coppi Junior

Orientadora: Andreza Sartori

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos funcionais e não funcionais
- Especificação e Implementação
- Operacionalidade da Implementação
- Dificuldades e Limitações
- Resultados e Discussões
- Conclusões e Extensões

Introdução

- Redes sociais e a geração de dados
- Mineração de dados
- Análise de sentimentos
- Fornecer informações sobre uma marca para os usuários

Objetivo Geral

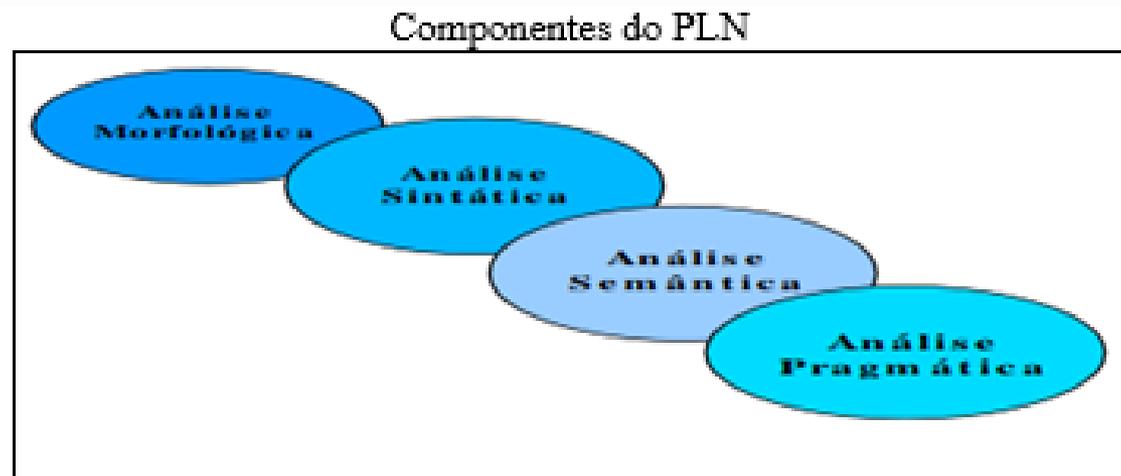
Desenvolver um sistema para analisar automaticamente o sentimento dos usuários em relação a uma marca.

Objetivos Específicos

- a) realizar uma pesquisa de opinião no sobre uma marca
- b) efetuar um filtro sobre os dados vindos do Twitter
- c) aplicar um algoritmo de análise de sentimentos visando medir a opinião dos usuários em relação a marca
- d) disponibilizar uma interface para que o usuário possa realizar a pesquisa

Fundamentação Teórica

- Processamento de Linguagem Natural (PLN)



Fonte: Müller (2003).

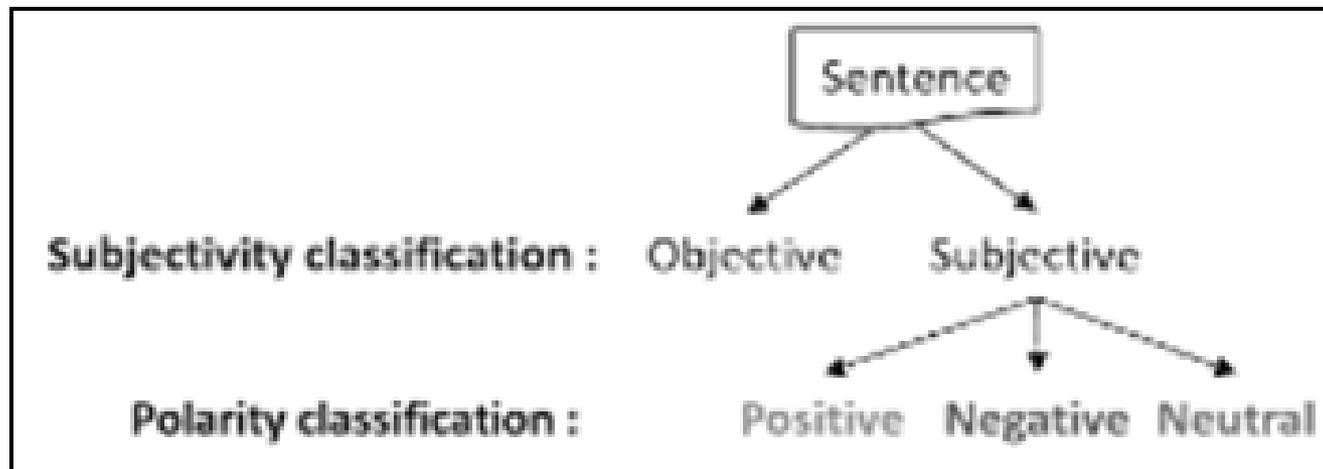
Fundamentação Teórica

- Filtragem de dados
 - Lematização
 - Tokenização
 - StopWords

```
private void populaStopList() {  
    String[] stopListArray =  
        {"a", "an", "and", "any", "are", "as",  
         "at", "be", "by", "for", "from", "has",  
         "he", "her", "i", "in", "is", "it",  
         "its", "me", "my", "of", "on", "so",  
         "that", "the", "to", "was", "were",  
         "will", "with", "you"};  
  
    for (String stopWord : stopListArray) {  
        this.adicionaTermoIgnorar(stopWord);  
    }  
}
```

Fundamentação Teórica

- Análise de Sentimentos
 - Objetividade
 - Polaridade



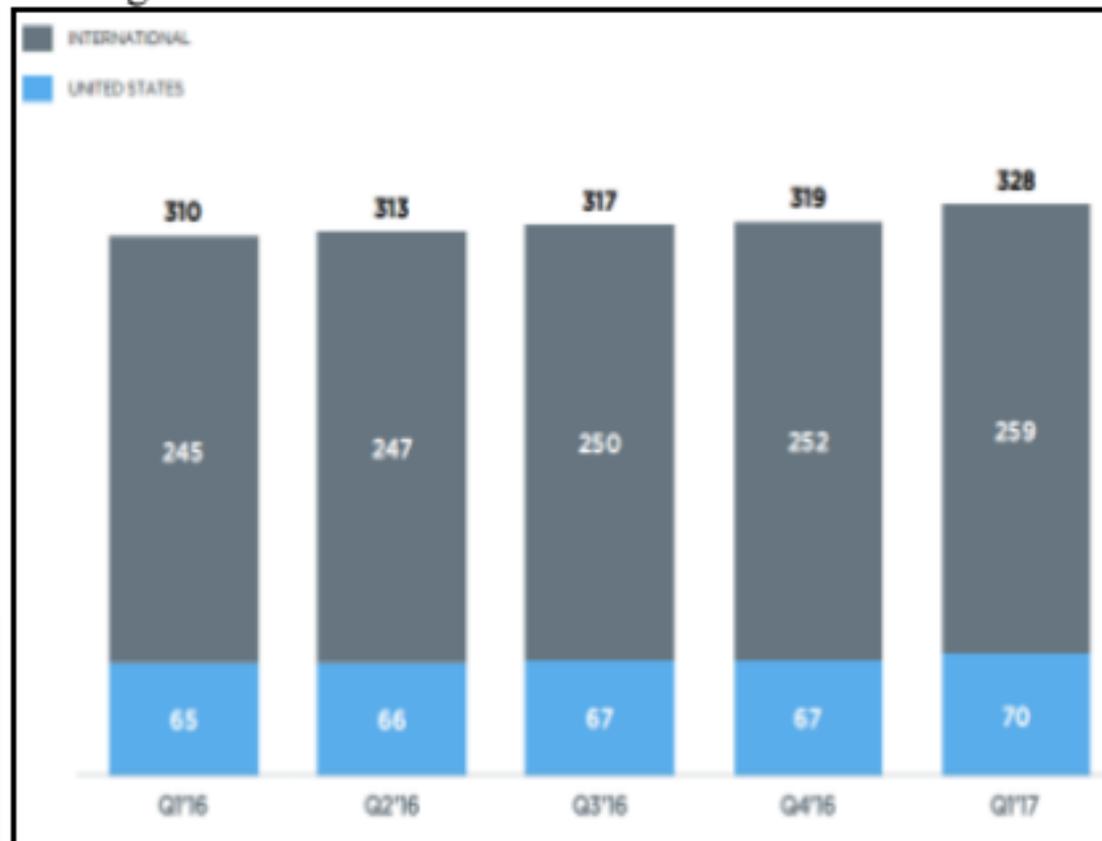
Fonte: Fersini et al (2016).

Fundamentação Teórica

- Mineração de Dados
 - Identificar padrões
 - Passo-a-passo
 - Extração da web

Fundamentação Teórica

- Rede Social



Fonte: Hutchinson (2017).

Trabalhos Correlatos

Título: Sentiment Strength Detection for the Social Web

características / trabalhos relacionados	Thelwall, Buckley e Paltoglou (2012)
Objetivo do trabalho	Verificar o desempenho do algoritmo SentiStrength
Geração de dados	Extraídos de redes sociais
Utiliza algoritmo de aprendizado de máquina	×
Algoritmo de análise de sentimentos utilizado	<i>SentiStrength 2.0</i>
Mostra resultados aos usuários	×

Trabalhos Correlatos

Título: Protótipo para mineração de opinião em redes sociais: estudo de casos selecionados usando o twitter

características / trabalhos relacionados	Santos (2010)
Objetivo do trabalho	Analisar opinião na web
Geração de dados	Utiliza um <i>crawler</i>
Utiliza algoritmo de aprendizado de máquina	<i>Support Vector Machine</i> (SVM)
Algoritmo de análise de sentimentos utilizado	Algoritmo próprio
Mostra resultados aos usuários	

Trabalhos Correlatos

Título: Protótipo para mineração de opinião em redes sociais: estudo de casos selecionados usando o twitter



Trabalhos Correlatos

Título: Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media

características / trabalhos relacionados	Paltoglou e Thelwall (2012)
Objetivo do trabalho	Validar uma abordagem baseada num dicionário de expressões
Geração de dados	Extraídos do Twitter, MySpace e Digg
Utiliza algoritmo de aprendizado de máquina	Classificador não supervisionado criado pelos autores
Algoritmo de análise de sentimentos utilizado	Classificador baseado em um dicionário de expressões pré-definidas
Mostra resultados aos usuários	✘

Requisitos

Requisitos funcionais	Casos de uso
RF01: exibir os dados das melhores marcas já classificadas	UC05
RF02: permitir ao usuário visualizar os dados de uma marca ou realizar uma pesquisa	UC01, UC02, UC04, UC06
RF03: permitir ao usuário selecionar duas marcas e fazer uma comparação entre elas	UC03
RF04: efetuar a pesquisa de opiniões em tempo real sobre uma marca no Twitter	UC01, UC02, UC04
RF05: filtrar os dados vindos do Twitter visando remover ruídos e otimizar sua qualidade	UC02

Requisitos

Requisitos funcionais	Casos de uso
RF06: utilizar um algoritmo de análise de sentimentos sobre os dados gerados na pesquisa para medir a opinião dos usuários	UC02
RF07: salvar as pesquisas num banco de dados para evitar buscas duplicadas num curto período de tempo	UC02
RF08: efetuar as pesquisas com base na categoria selecionada pelo usuário	UC07

Requisitos

Requisitos não funcionais

RNF01: apresentar uma interface responsiva para ser acessado tanto de computadores quanto de dispositivos móveis

RNF02: utilizar a técnica de votação em seus servidores, onde caso um erro aconteça em algum dos servidores este seja identificado e a informação correta retornada

RNF03: possuir seus servidores programados em Java e a interface em PHP e AngularJs para utilizar os recursos disponíveis nestas linguagens

RNF04: efetuar a análise de sentimentos utilizando como base a emoção das palavras definidas no SentiWordNet

RNF05: utilizar as técnicas de remoção de caracteres HTML, tokenização e de StopWords na filtragem de dados

Diagrama de Casos de Uso

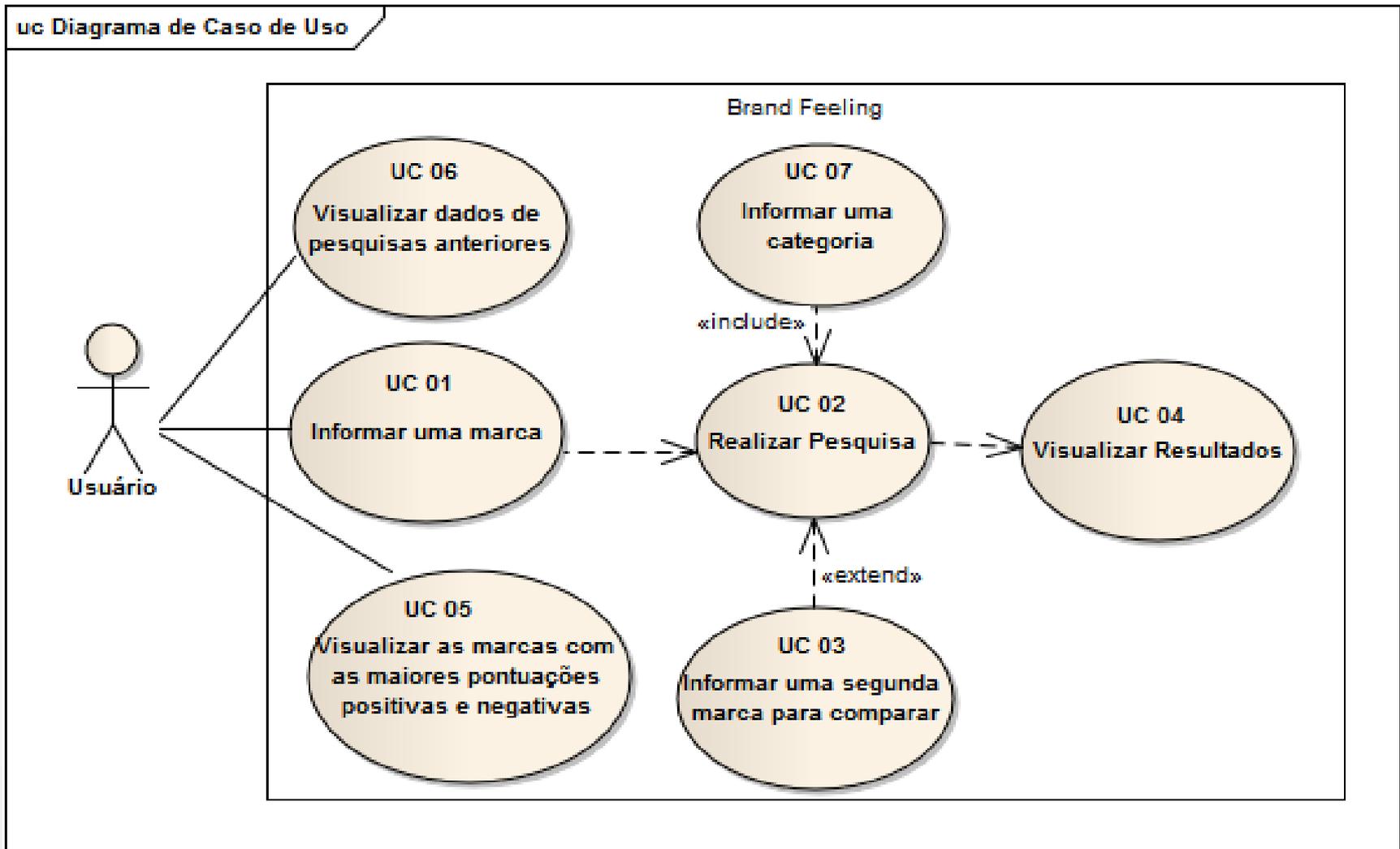
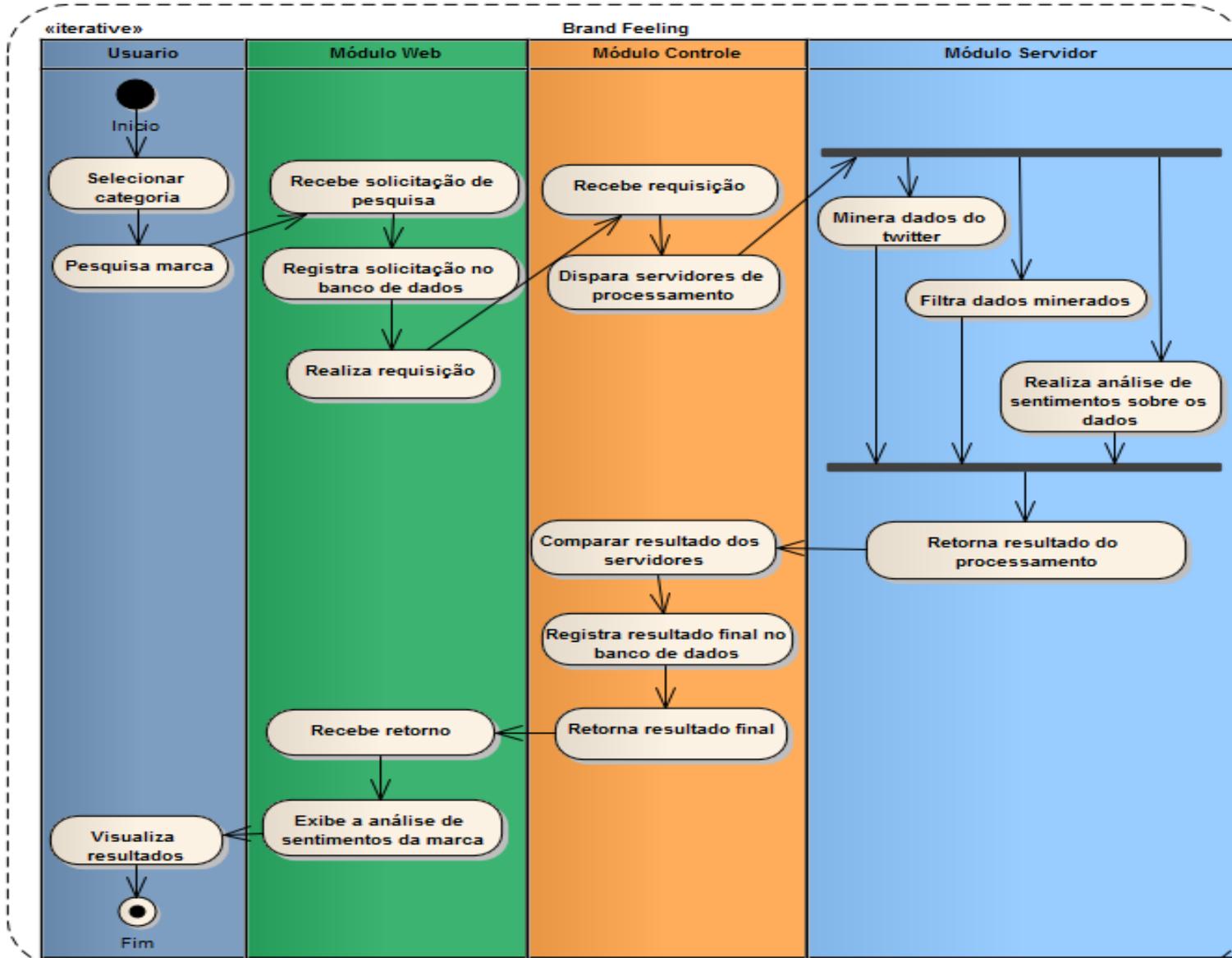


Diagrama de Atividades

act Diagrama de Atividades



Arquitetura

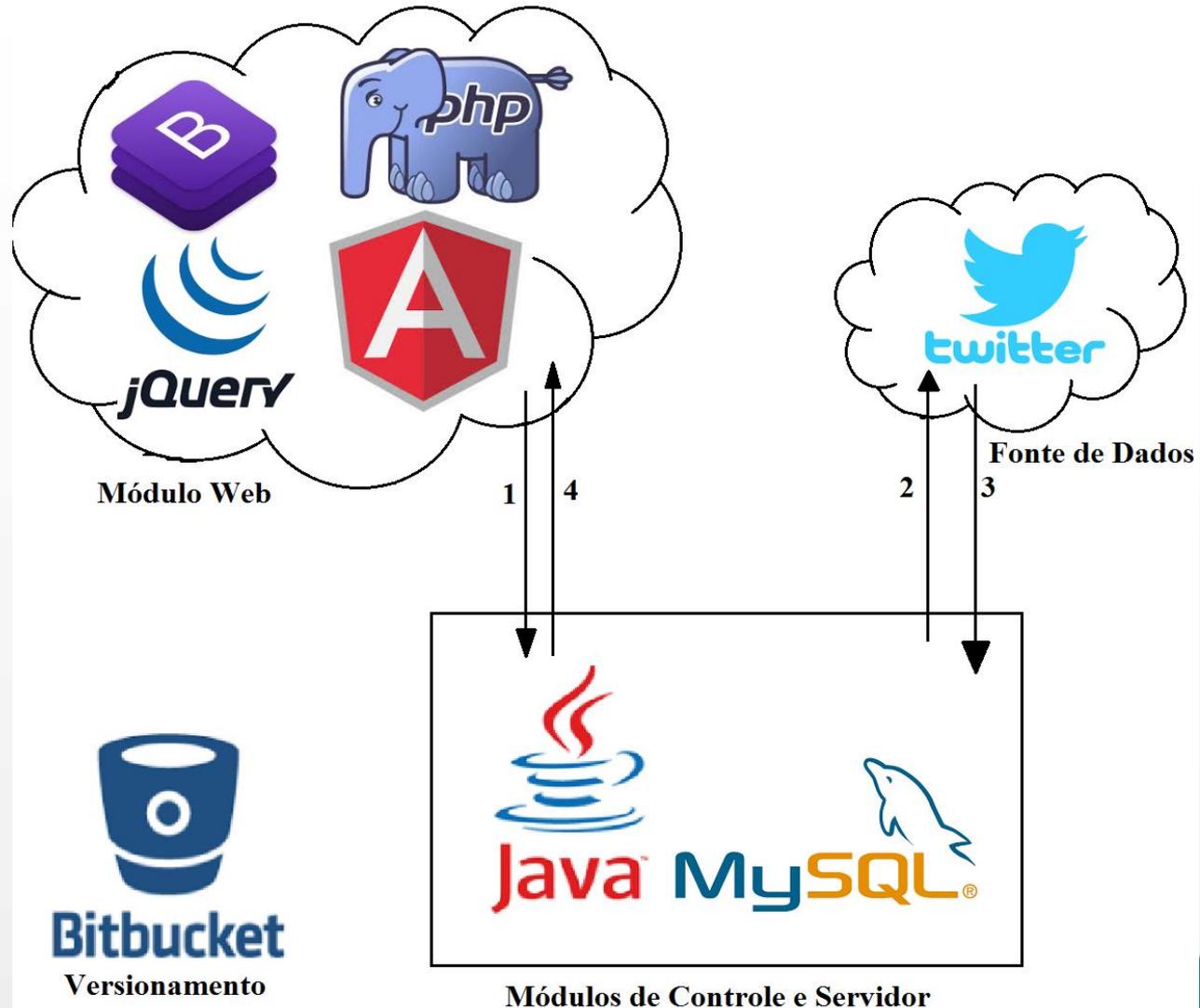


Diagrama de classes - Web

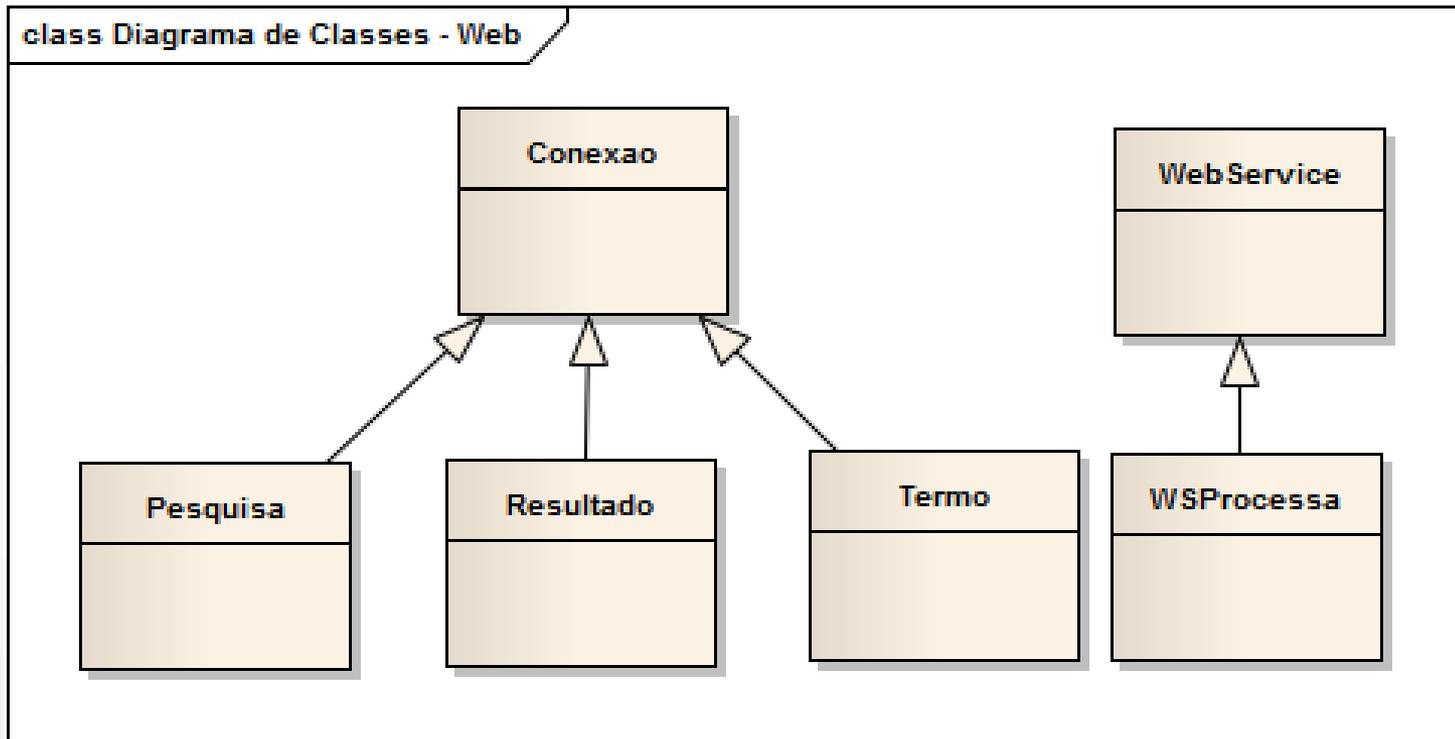


Diagrama de classes - Controle

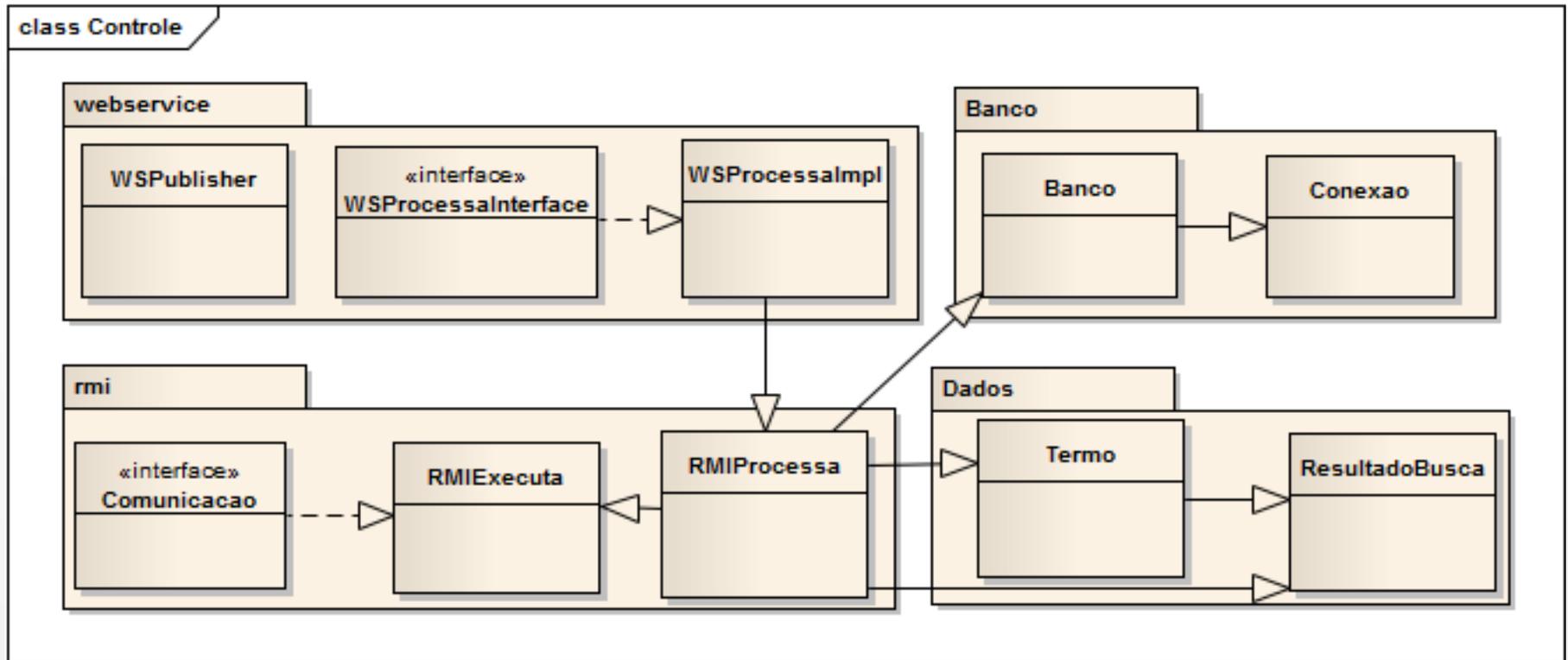
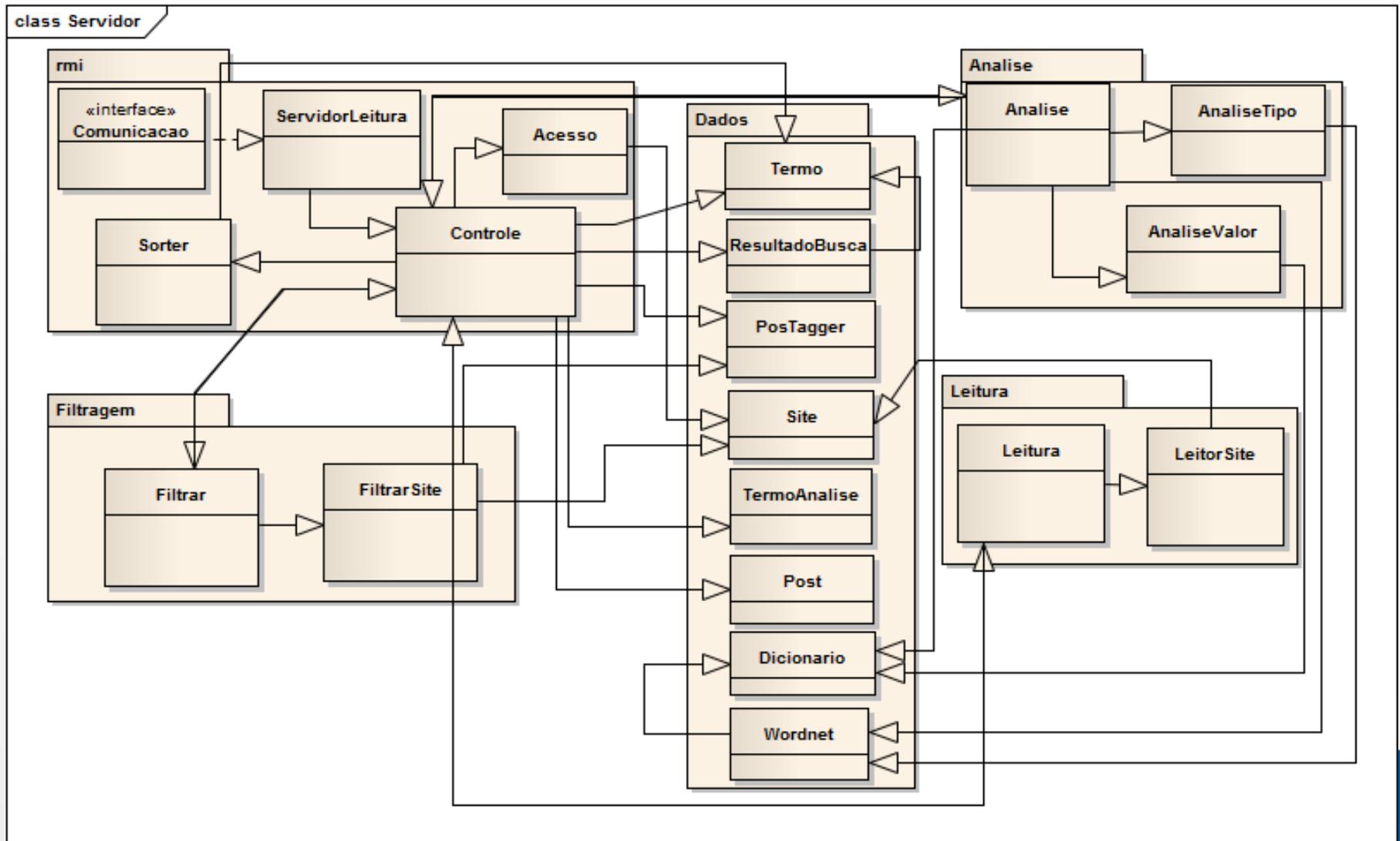


Diagrama de classes - Servidor



Implementação

- Tecnologias Utilizadas
 1. WordNet
 2. SentiWordNet
 3. MIT Java WordNet Interface (JWI)
 4. Stanford Log-Linear POS Tagger

Operacionalidade da Implementação

Menu

Brand Feeling

Brand Feeling

Sistema para analisar o sentimento dos usuários em relação a uma marca

ÚLTIMAS PESQUISAS *

Marca	Positividade	Ver
dell	68 %	ⓘ
asus	66 %	ⓘ
amazon	62 %	ⓘ
netflix	63 %	ⓘ
microsoft	62 %	ⓘ
microsoft	64 %	ⓘ
android	62 %	ⓘ
sony	56 %	ⓘ
eletronic arts	55 %	ⓘ
nintendo	61 %	ⓘ

* Correspondente as pesquisas anteriores realizadas pelos usuários

Informe uma marca

Adicione outra marca +

Categoria de pesquisa

Pesquisar 🔍

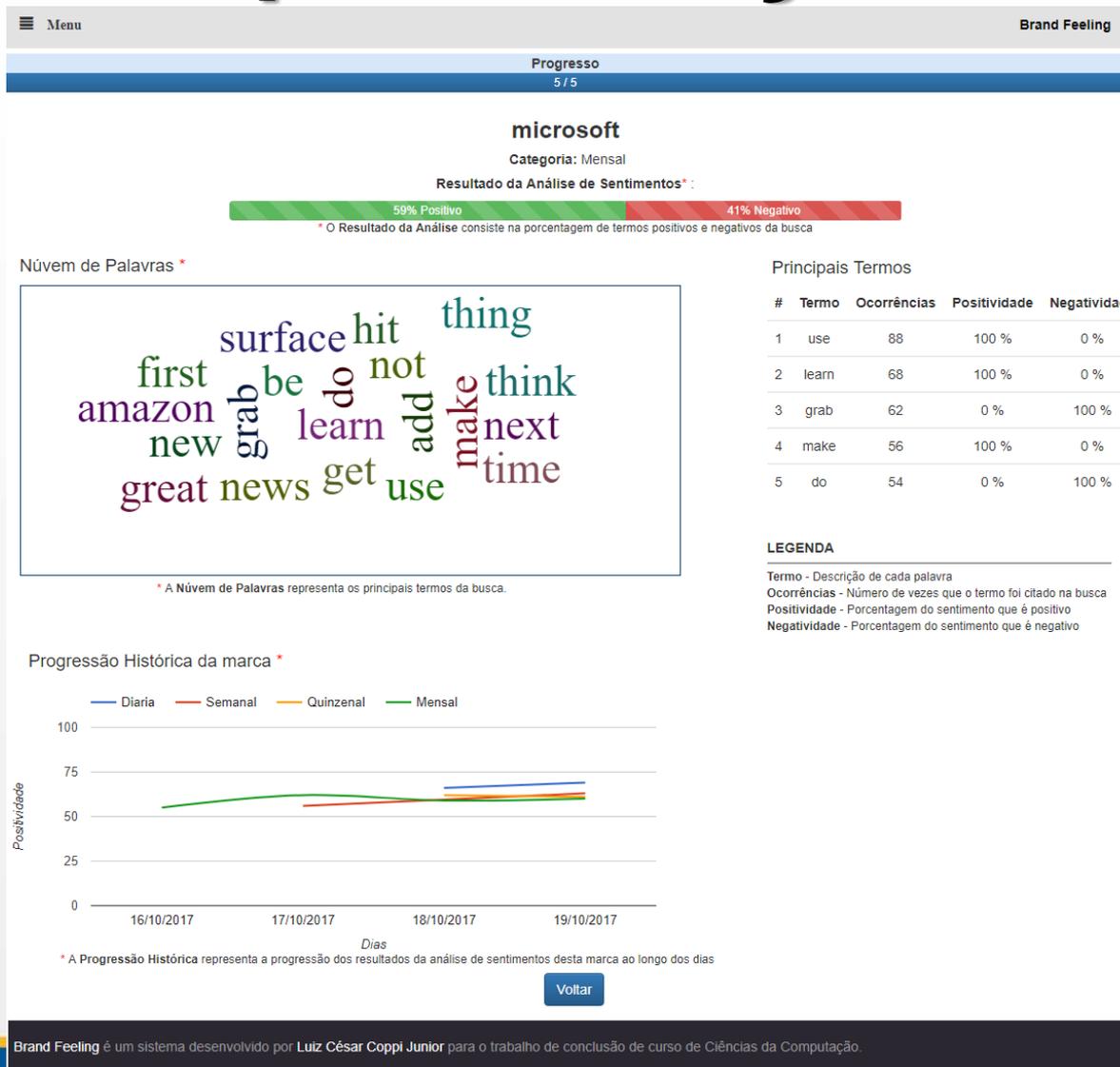
Limpar

HISTÓRICO DAS MARCAS *

#	Marca	Positividade		
			Positivo	Negativo
1	dell	68 %		
2	motorola	66 %		
3	asus	66 %		
4	samsung	64 %		
5	ubisoft	63 %		
6	netflix	63 %		
7	amazon	62 %		
8	facebook	62 %		
9	google	62 %		
10	android	62 %		

* Média das pesquisas realizadas para cada marca.

Operacionalidade da Implementação



Dificuldades e Limitações

- Limite no numero de postagens
- Apresentação dos resultados finais

Resultados e Discussões

- Resultados na busca pelo termo Microsoft

Categorias	Número de Termos	Número de Sites	Número de Postagens	Termo com maior recorrência	Positividade	Tempo em segundos (S)
Diária	65	2	48	add – 8 vezes	73 %	5,3 S
Semanal	284	16	402	new – 50 vezes	62 %	7,8 S
Quinzenal	456	32	740	new – 86 vezes	61 %	11,6 S
Mensal	669	62	1316	new – 163 vezes	60 %	18,2 S

Comparação Entre os Trabalhos Correlatos

Trabalhos	Trabalho Desenvolvido	Thelwall, Buckle e Paltoglou (2012)	Santos (2010)	Paltoglou e Thelwall (2012)
Características				
Objetivo do trabalho	Analisar o sentimento de usuários em relação a uma marca.	Verificar o desempenho do algoritmo SentiStrength	Analisar opinião na web	Validar uma abordagem baseada num dicionário de expressões
Geração de dados	Mineração de dados do Twitter	Extraídos de redes sociais	Utiliza um <i>crawler</i>	Extraídos do Twitter, MySpace e Digg
Utiliza algoritmo de aprendizado de máquina	✗	✗	<i>Support Vector Machine (SVM)</i>	Classificador não supervisionado criado pelos autores.
Algoritmo de análise de sentimentos utilizado	SentiWordNet 3.0	<i>SentiStrength 2.0</i>	Algoritmo próprio	Classificador baseado em um dicionário de expressões pré-definidas
Mostra resultados aos usuários	✓	✗	✓	✗

Conclusões

- Sistema desenvolvido
- Objetivos atingidos
- Tempo acima do esperado
- Evidencia a área para outro público
- Exemplo para o meio acadêmico
- Ferramenta para o público

Extensões

- Adicionar um algoritmo de aprendizado de máquina
- Criar um mecanismo para efetuar o tratamento de palavras mal escritas
- Efetuar a mineração de dados de outras redes sociais
- Disponibilizar no Módulo Web uma paginação na tela inicial
- Estudar uma abordagem para tratar a utilização de figuras de linguagem, como sarcasmo e ironia
- Implementar as etapas de análise semântica e análise pragmática do PLN

Demonstração