



MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

DISCIPLINA DE METODOLOGIA CIENTÍFICA

Expansão automatizada de Léxicos para a Análise de Sentimentos por meio de programação evolucionária

Autor: Airton Bordin Junior
Orientador: Nádia Félix Felipe da Silva
Coorientador: Celso Gonçalves Camilo Junior

1 de junho de 2017

1 Apresentação

Airton Bordin Junior, bacharel em Ciência da Computação. Cursou os primeiros 3 anos do curso na Universidade Estadual do Oeste do Paraná (UNIOESTE) campus Foz do Iguaçu, e finalizou a graduação na faculdade Anglo Americano, na mesma cidade, no ano de 2011. Possui, também, graduação em Gestão Pública, pelo Instituto Federal de Santa Catarina, cursado por meio da UaB, no campus de Foz do Iguaçu. Atua na área da computação desde 2010. Já trabalhou como desenvolvedor de software, testador e, por fim, como Analista de Sistemas no Parque Tecnológico Itaipu, responsável pela área de TI do projeto de Segurança de Barragens. Lecionou 2 semestres no curso Técnico em Informática Para Internet do Pronatec, atuando nas disciplinas de Sistemas Operacionais e Segurança de Sistemas, e 1 semestre no curso de Ciência da Computação em uma faculdade local, lecionando as disciplinas de Processamento de Imagens e Sistemas Inteligentes. Sempre quis continuar estudando, e o mestrado era um objetivo a ser atingido. Por conta da falta de oportunidades na cidade onde morava (Foz do Iguaçu), acabei adiando um pouco essa meta. Hoje tenho a oportunidade e a honra de estar participando como aluno regular do programa de mestrado pela Universidade Federal de Goiás. No programa da UFG, estou incluído na linha de pesquisa de Inteligência Computacional. Pretendo aprofundar meu trabalho na área de Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases), mais precisamente com dados da área médica, alinhado com trabalhos de alguns professores da Universidade. Após minha graduação, cursei especialização em Redes de Computadores pela Universidade Federal Tecnológica do Paraná, campus de Cornélio Proença e MBA em Gerenciamento de Projetos pelo Centro Universitário Dinâmica das Cataratas, em Foz do Iguaçu.

- Nome, formação acadêmica e experiências passadas.
- Motivação para estudos mais avançados.
- Motivação para a escolha do tema a ser desenvolvido.
- Principais razões que o candidato é um ótimo nome para desenvolver o tema escolhido.

2 Resumo

Deve ter aproximadamente 300 palavras. Além disto, deve ter uma descrição breve de todo o projeto, atendendo a estas quatro áreas:

- O que você vai fazer? (o problema)
- Como será feito (metodologia)
- Resultados esperados? (apenas os mais relevantes)
- Qual a importância destes? (Conclusões / recomendações)

3 Introdução

- Fornece informações contextuais para o problema, identificando a **lacuna de investigação** nesta área.
- Apresenta os objetivos.
- Identifica as questões de pesquisa específicas.
- Opcionais (de acordo com a área de pesquisa):
 - Apresentar a metodologia de forma sucinta.
 - Apresentar os principais resultados e conclusões que se espera do projeto.
- Indicar a estrutura do resto do relatório.

3.1 Formato sugerido para uma introdução

- Introduza a área de pesquisa.
- Revise publicações chave.
- Identifique qualquer gap no conhecimento ou questões que você quer responder.
- Suas hipóteses.
- Seus objetivos, incluindo uma breve descrição da metodologia.
- Como e para quem a sua pesquisa será benéfica.

4 Descrição do Problema

Nesta seção, o proponente tem a oportunidade de discorrer livremente sobre o problema a ser estudado, de forma a descrever e justificar o problema aos possíveis leitores. A revisão bibliográfica, identificação de projetos semelhantes e possíveis "brechas" em trabalhos já realizados também são identificados aqui, de forma a justificar, através de períodos e outros aceitos pela comunidade científica, a existência do problema.

5 Objetivos

5.1 Objetivo geral

Aqui, descreve-se os objetivos gerais – resultados mais abrangentes – e específicos – especificação concisa de tudo que se espera atingir com a execução do trabalho, bem como possíveis produtos que possam ser gerados como resultados – do projeto proposto.

5.2 Objetivos específicos

- 1.
- 2.
3. ...

6 Revisao bibliografica

Nesta secao, que deve ocupar **no maximo** 3 paginas, devera ser apresentada a revisao bibliografica. O numero minimo de artigos e de 20 vinte (internacionais) e 10 (dez) nacionais.

7 Impacto Científico

A técnica de Análise de Sentimento fazendo uso de um léxico é uma das mais utilizadas na literatura. Mostra-se, portanto, essencial a obtenção de um conjunto de palavras consolidado, juntamente com as orientações semânticas respectivas. Uma palavra pode ter um significado e, conseqüentemente, uma polaridade diferente, dependendo do contexto ao qual está inserido. Um conjunto léxico inadequado leva a análises inconsistentes, prejudicando o resultado final do sistema. A solução proposta neste trabalho criará, de forma automatizada, léxicos para diferentes domínios, que poderão servir como entrada para diversos classificadores e sistemas de análise de sentimentos. Além disso, a técnica pode ser utilizada em outros idiomas, de forma a suprir uma carência de dicionários consistentes em linguagens pouco conhecidas. O conjunto de palavras gerado pela solução proposta poderá ser utilizado como *benchmark* para outros trabalhos na área, bem como ser expandido com outras técnicas adequadas.

8 Metodologia

A metodologia descreve a forma como serão desenvolvidas cada uma das etapas do processo. Estas devem ser as mesmas que foram apresentadas no cronograma de trabalho, descrito na seção 9.

8.1 Informações importantes

- Pergunta a ser respondida: Como você atingirá os objetivos da pesquisa?
- Divida o seu projeto de pesquisa em diversas etapas.
- Cada etapa deverá ter uma breve descrição informando ao leitor **como** a mesma será executada.
- Faça uma previsão de tempo de cada uma das etapas.
- Ao final de cada etapa, apresente um **marco físico**¹

8.2 Sugestão

- D1. **Revisão bibliográfica:** Nesta etapa do trabalho será feita uma revisão bibliográfica com vistas a identificar o estado da arte do problema que está sendo proposto. Importante registrar que esta revisão bibliográfica seguirá os moldes propostos por [Kitchenham, 2004]. Serão consultadas as bases de dados do *Portal da Capes*, *IEEEExplore* e *ACM Digital Library*.
- D2. **Estudo dos principais classificadores de sentimentos:** Nesta etapa será feita uma pesquisa sobre os principais classificadores disponíveis, preferencialmente livres e *open source*, utilizados para a análise de sentimentos, e que permitem a manipulação de seu dicionário. O objetivo principal desse passo é selecionar ferramentas que proporcionarão dados comparativos para teste da solução proposta.
- D3. **Análise dos principais léxicos disponíveis:** Busca pelos principais conjuntos de palavras, e suas respectivas polaridades, disponíveis para utilização. Esses léxicos servirão como base e material de testes para a solução.
- D4. **Recuperação das principais bases de opiniões anotadas disponíveis:** Busca e recuperação de bases de opiniões anotadas e consistentes, representando resultados confiáveis e corretos. Essas bases, já revisadas por especialistas humanos, servirão como parâmetro de correção da solução proposta, bem como serão utilizadas para o cálculo de erro dos resultados obtidos.
- D5. **Recuperação das principais bases de opiniões anotadas disponíveis:** Busca e recuperação de bases de opiniões anotadas e consistentes, representando resultados confiáveis e corretos. Essas bases, já revisadas por especialistas humanos, servirão como parâmetro de correção da solução proposta, bem como serão utilizadas para o cálculo de erro dos resultados obtidos.

8.3 Marcos físicos

- D1. Documento com a revisão bibliográfica.
- D2. Lista dos principais classificadores.
- D3. Lista dos principais léxicos.
- D4. Bases de opiniões recuperadas.

¹Relatório, artigo, projeto de sistema, etc.

	2017												2018												2019	
	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2		
Disciplinas																										
<i>Fim do 1º semestre</i>																										
<i>Fim do 2º semestre</i>																										
Dissertacao																										
D_1																										
D_2																										
D_3																										
D_4																										
D_5																										
D_6																										
D_7																										
<i>Relatorio 1</i>																										
<i>Relatorio 2</i>																										
<i>Relatorio 3</i>																										
<i>Qualificacao</i>																										
<i>Defesa</i>																										

- D1. Revisao bibliografica.
- D2. Estudo dos principais classificadores de sentimentos.
- D3. Análise dos principais léxicos disponíveis.
- D4. Recuperação das principais bases de opiniões anotadas disponíveis.
- D5. Implementação da solução.
- D6. Teste da solução com os classificadores selecionados.
- D7. Levantamento dos dados de testes e relatórios.

10 Resultados Esperados

Espera-se, com o presente trabalho, a criação de um processo automatizado de expansão de léxico dependente de domínio, fazendo uso de técnicas de algoritmos evolucionários. Nesse sentido, expansão significa tanto a criação e definição da orientação semântica de novas palavras, bem como a alteração das polaridades das palavras já existentes para um valor mais adequado ao domínio que trata o processo. Pela característica genérica da solução, a criação de diversos léxicos para vários domínios diferentes é limitada tão somente à escolha dos contextos específicos e à disponibilidade de dados anotados para teste da solução. Podemos citar, também, uma possível melhoria em algumas técnicas de Análise de Sentimentos que fazem uso de léxicos padrão, contribuindo assim para a evolução de outros sistemas de Mineração de Opiniões que usam a estratégia de dicionário. Os resultados parciais e finais do trabalho serão descritos em artigos científicos que serão submetidos à eventos na área, de forma a compartilhar o conhecimento e avanços alcançados pela técnica proposta.

10.1 Algoritmos

Será implementado um algoritmo que criará e/ou ampliará, de forma automatizada, um léxico para um domínio específico que será utilizado como entrada para um sistema classificador de Análise de Sentimentos. Esse software fará uso de técnicas de algoritmos bioinspirados, mais precisamente Programação Evolucionária, para a atribuição de valores sentimentais para cada palavra, de forma a maximizar a taxa de acerto ao ser processado por um classificador existente. Ao passo que o algoritmo é independente de domínio, pode ser utilizado, desde que haja dados de testes suficientes, para qualquer contexto desejado.

10.2 Artigos científicos

Quantidade	Qualis	Tipo	Nome
1	B1	Conferencia	FIE - Frontiers in Education
1	A1	Periodico	Computers in Education

11 Identificacao dos Participantes e Colaboradores

Aqui o candidato devera descrever se o projeto que esta sendo proposto faz parte de um projeto de pesquisa maior ou nao. Alem disto, deve descrever as possiveis colaboracoes (alunos de iniciacao cientifica, mestrado ou doutorado) que possam contribuir para o seu trabalho.

12 Referencias bibliograficas

Referências

[Kitchenham, 2004] Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(TR/SE-0401):28.