

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**Míria Luísa das Dores Ramos Bóbó**

**SASys: uma arquitetura de Análise de Sentimento usando abordagem lexical**

**Juiz de Fora**  
**2019**

**Míria Luísa das Dores Ramos Bóbó**

**SASys: uma arquitetura de Análise de Sentimento usando abordagem lexical**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação, da Universidade Federal de Juiz de Fora como requisito parcial a obtenção do grau de Mestre em Ciência da Computação. Área de concentração: Engenharia de Software/Banco de dados

Orientadora: Dra. Fernanda Cláudia Alves Campos

Coorientador: Dr. Victor Ströele de Andrade Menezes

**Juiz de Fora**

**2019**

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Bóbó, Míria Luísa das Dores Ramos .

SASys: uma arquitetura de Análise de Sentimento usando abordagem lexical / Míria Luísa das Dores Ramos Bóbó. -- 2019. 108 p. : il.

Orientadora: Fernanda Cláudia Alves Campos

Coorientador: Victor Ströele

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Ciência da Computação, 2019.

1. Análise de Sentimento. 2. Mineração de Opinião. 3. FrameNet. 4. Abordagem Lexical. I. Campos, Fernanda Cláudia Alves, orient. II. Ströele, Victor, coorient. III. Título.

**Míria Luísa das Dores Ramos Bóbó**

**SASys: uma arquitetura de Análise de Sentimento usando abordagem lexical**

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação, da Universidade Federal de Juiz de Fora como requisito parcial a obtenção do grau de Mestre em Ciência da Computação. Área de concentração: Engenharia de Software/Banco de dados

Aprovada em 02 de Setembro de 2019

**BANCA EXAMINADORA**

---

Profª. D.Sc. Fernanda Cláudia Alves Campos - Orientadora  
Universidade Federal de Juiz de Fora

---

Prof. D.Sc. Victor Stroële de Andrade Menezes – Coorientador  
Universidade Federal de Juiz de Fora

---

Profª. D.Sc. Rosa Maria Esteves Moreira da Costa  
Universidade do Estado do Rio de Janeiro

---

Prof. D.Sc. Tiago Timponi Torrent  
Universidade Federal de Juiz de Fora

---

Prof. D.Sc. José Maria Nazar David  
Universidade Federal de Juiz de Fora

*Dedico este trabalho às minhas avós e aos meus pais,  
que são a minha inspiração  
e aos meus sobrinhos que me fazem sonhar  
com um futuro melhor.*

## AGRADECIMENTOS

A Deus, pela sua graça e benevolência, por me permitir alcançar essa conquista e por ser sempre o meu sustento.

Aos meus pais, Francisco Bóbó e Maria Judith, por mais uma vez abdicarem dos seus sonhos para tornarem os meus realidade e por me presentear com a melhor herança que poderia ter: a educação. Vocês são os meus exemplos de vida e de superação.

Aos meus irmãos pelo apoio incondicional, motivação, companheirismo, paciência e zelo. De modo especial, agradeço à minha irmã gêmea Mayra, por ser minha companheira de estrada, por sofrer comigo em todas as minhas derrotas e se alegrar em todas as minhas conquistas. Por ter sempre uma palavra de incentivo, pelo cuidado, carinho e empatia. Por vocês (pais e irmãos) sempre apostarem em mim, o meu muito obrigada.

À minha orientadora, professora Fernanda Campos, por assumir o desafio de orientação deste trabalho, compartilhar sua experiência acadêmica e não aceitar nada menos que a excelência.

Ao meu coorientador, Victor Ströele, por sua mentoria e amizade desde o primeiro ano da graduação, por me incentivar a fazer o mestrado e por, mais uma vez, enxergar um potencial que eu desconhecia.

Aos amigos e colegas de mestrado por estarem comigo e transformarem as tristezas e desesperos em alegrias. À comunidade angolana de Juiz de Fora, de modo especial as meninas de JF, que sempre me ofereceram o calor de uma família. Às meninas da computação por me proporcionarem as melhores lembranças que poderia ter da UFJF.

Aos professores Ely Matos e Tiago Torrent por me proporcionarem a possibilidade de trabalhar com um projeto de tamanho porte, como a FrameNet Brasil.

Aos professores do PGCC por todo o ensinamento e de maneira especial ao professor Jairo de Souza que me iniciou no trajeto da ciência.

À UFJF pela infraestrutura e o apoio financeiro através do fornecimento de uma bolsa de estudos.

A todos que de alguma forma contribuíram para que esta dissertação se tornasse realidade, o meu muito obrigada!

*“A mente humana é como o pêndulo de um relógio  
que flutua entre a emoção e a razão”*

Augusto Cury

## RESUMO

A Análise de Sentimento ou Mineração de Opinião é uma subárea da mineração de texto que tem como objetivo detectar as opiniões, sentimentos ou emoções expressas em um texto. Ela é uma área multidisciplinar que engloba desde técnicas simples de Processamento de Linguagem Natural até algoritmos sofisticados de Aprendizagem de Máquina. Ela necessita de bases de treinamento ou léxicos que contemplem as peculiaridades do contexto em que é aplicada. Assim sendo, este trabalho se propõe a contribuir no campo de Mineração de Opinião apresentando uma arquitetura de Análise de Sentimento que usa a FrameNet como parte de sua abordagem lexical, com o objetivo de descobrir o estado emocional do autor do texto, através do emprego de diferentes tipos de dados. Esta dissertação também apresenta uma rede polarizada de *frames* desenvolvida a partir da estrutura tradicional da FrameNet, visando detectar a polaridade das palavras através da identificação dos *frames* evocados. Um estudo de caso foi conduzido com o apoio de especialistas e os resultados evidenciam a empregabilidade da FrameNet como uma abordagem lexical de análise de sentimento.

Palavras-chave: Análise de Sentimento, Mineração de Opinião, FrameNet, Abordagem Lexical.



## ABSTRACT

Sentiment Analysis or Opinion Mining is a subarea of text mining that aims to detect the opinions, feelings or emotions expressed in a text. It is a multidisciplinary area that ranges from simple Natural Language Processing to sophisticated Machine Learning algorithms. It needs training bases or lexicons that cover peculiarities of the context in which it is applied. Thus, this work proposes to contribute in the field of Opinion Mining presenting a Sentiment Analysis architecture that uses FrameNet as part of its lexical approach, with the objective of discovering the emotional state of the author of the text, through the use of different types of data. This dissertation also presents a polarized network of *frames* developed from the traditional framework of FrameNet, aiming to detect the polarity of the words through the identification of the evoked *frames*. A case study with real data was conducted with the support of experts and the results evidence the employability of FrameNet as a lexical approach to sentiment analysis.

Keywords: Sentiment Analysis, Opinion Mining, FrameNet, Lexical Approach.

## LISTA DE ILUSTRAÇÕES

Figura 1: Métodos de Análise de Sentimento.....	19
Figura 2: Relacionamentos do frame <i>Dar</i> com os outros <i>frames</i> .....	24
Gráfico 1 – Métodos de AS citados nas revisões e mapeamentos.....	30
Gráfico 2 – Contexto de aplicação de AS citados nas revisões e mapeamentos.....	31
Gráfico 3 – Fontes de dados citados nas revisões e mapeamentos.....	32
Figura 3: Fluxo de detecção de estado emocional.....	39
Figura 4: Arquitetura conceitual SASys.....	40
Figura 5: Grafo da rede de <i>frames</i> da FrameNet.....	46
Figura 6: Comunidades do grafo.....	47
Figura 7: Visualização em árvore de <i>frames</i> de uma das componentes conexas do grafo.....	49
Figura 8: Tela da <i>WebTool</i> da FrameNet, onde o <i>frame</i> <i>abandono</i> foi selecionado.....	50
Figura 9: Arquitetura SASys instanciada para o estudo de caso.....	59
Figura 10: Regras usadas na rotulação das frases por especialistas em EaD.....	60
Figura 11: Boxplots do acesso semanal dos alunos no AVA.....	72
Figura 12: Boxplots da média de acesso dos alunos de cada perfil motivacional.....	74
Figura 13: Formas de apresentação da mensagem motivacional: vídeo, texto e áudio.....	75

## LISTA DE TABELAS

Tabela 1 – Questões do mapeamento e respectivas justificativas.....	28
Tabela 2 – <i>String</i> de busca do mapeamento de estudos terciários.....	28
Tabela 3 – Lista dos artigos aceitos no estudo terciário .....	29
Tabela 4 – Exemplos de tipos de dados coletados na SASys.....	41
Tabela 5 – Exemplo de frases contidas na base de dados do YELP.....	45
Tabela 6 – Medidas de centralidade e o intervalo de valores.....	48
Tabela 7 – Exemplos de polaridades encontradas na rede de <i>frames</i> .....	51
Tabela 8 – Detecção do sentimento geral das frases da base de dados do YELP.....	52
Tabela 9 – Matriz de confusão do estudo observacional.....	53
Tabela 10 – Tabela de critérios de avaliação do estudo observacional.....	55
Tabela 11 – Questões de pesquisa estudo de caso.....	57
Tabela 12 – Amostra de frases rotuladas por especialistas em EaD.....	61
Tabela 13 – Detecção do sentimento dos comentários.....	62
Tabela 14 – Matriz de confusão.....	63
Tabela 15 – Medidas de desempenho da polaridade <i>Positiva</i> .....	63
Tabela 16 – Medidas de desempenho da polaridade <i>Negativa</i> .....	64
Tabela 17 – Medidas de desempenho da polaridade <i>Neutra</i> .....	64
Tabela 18 – Tabela de Critérios Gerais de avaliação.....	64
Tabela 19 – Tabela comparativa entre o rótulo dos professores e o rótulo do linguista.....	66
Tabela 20 – Amostra do conjunto de comentários dos alunos com novos rótulos, definidos pelo especialista e as polaridades detectadas pela rede de <i>frames</i> .....	67
Tabela 21 – Nova matriz de confusão.....	67
Tabela 22 – Medidas de desempenho da polaridade <i>Positiva</i> .....	68
Tabela 23 – Medidas de desempenho da polaridade <i>Negativa</i> .....	68
Tabela 24 – Medidas de desempenho da polaridade <i>Neutra</i> .....	68
Tabela 25 – Tabela de Critérios Gerais de avaliação.....	69
Tabela 26 – Amostra de dados dos alunos.....	71
Tabela 27 – Perfil de Motivacional e Estado Emocional.....	74

## **LISTA DE ABREVIATURAS E SIGLAS**

AS	Análise de Sentimento
AVA	Ambiente Virtual de Aprendizado
CSV	Comma-Separated Values
EaD	Ensino a Distância
MOOC	Massive Open Online Course
PLN	Processamento de Linguagem Natural
PM	Perfil Motivacional
SASys	Sentiment Analysis System

# SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
1.1 PROBLEMA .....	14
1.2 QUESTÕES DE PESQUISA .....	15
1.3 JUSTIFICATIVAS .....	15
1.4 OBJETIVOS .....	16
1.6 METODOLOGIA .....	16
1.7 CONTRIBUIÇÕES .....	17
1.8 ORGANIZAÇÃO .....	17
<b>2 PRESSUPOSTOS TEORICOS</b> .....	18
2.1 ANÁLISE DE SENTIMENTO .....	18
2.2 FRAMENET BRASIL .....	21
2.2.1 Estrutura semântica da FrameNet .....	22
2.2.2 Relacionamentos .....	23
2.2.3 Modelo conceitual da FrameNet .....	25
2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	26
<b>3 TRABALHOS RELACIONADOS</b> .....	27
3.1 MAPEAMENTO DA LITERATURA SOBRE ANÁLISE DE SENTIMENTO – UM ESTUDO TERCIÁRIO .....	27
3.1.1 Planejamento .....	27
3.1.2 Resultados .....	29
3.1.3 Considerações sobre o Mapeamento .....	32
3.2 OUTROS TRABALHOS RELACIONADOS .....	33
3.2.1 Lexicon-based methods for sentiment analysis (TABOADA et al., 2011) .....	33
3.2.2 Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks (CHAPLOT; RHIM; KIM, 2015) .....	34
3.2.3 Senti-n-gram: An n-gram lexicon for sentiment Analisys (DEY; JENAMANI; THAKKAR, 2018) .....	35
3.2.4 Creating a portugueses context sensitive lexicon for sentiment analisys MACHADO et al. (2018) .....	36
3.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	37
<b>4 ARQUITETURA SASYS</b> .....	38
4.1 ARQUITETURA CONCEITUAL .....	40
4.1.1 Extração de dados .....	40
4.1.2 Análise de Sentimentos .....	41
4.1.3 Identificação do estado emocional .....	43
4.2 ESTUDO OBSERVACIONAL .....	44
4.2.1 Rede de <i>frames</i> .....	45
4.2.2 Caracterização da rede de <i>frames</i> .....	49
4.2.3 Rede de <i>frames</i> polarizada .....	50
4.2.4 Resultados obtidos .....	52
4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	55
<b>5 ESTUDO DE CASO</b> .....	56
5.1 DEFINIÇÃO DO ESCOPO .....	56
5.2 CONDUÇÃO DO ESTUDO .....	58
5.2.1 Coleta de dados .....	58
5.2.2 Análise de Sentimento .....	59
5.2.2.1 Rotulação das Mensagens por especialistas em EaD .....	60

5.2.2.2 Melhorias no processo de detecção de polaridade .....	65
5.2.2.3 Rotulação das Mensagens por especialista da FrameNet .....	66
5.2.2.4 Considerações sobre os resultados obtidos na AS .....	69
5.2.3 <b>Deteção do estado emocional</b> .....	70
5.2.3.1 <i>Estado Emocional</i> .....	71
5.2.4 <b>Recomendação de mensagens à partir do perfil motivacional</b> .....	75
5.3 AMEAÇAS A VALIDADE .....	76
5.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	77
<b>6 CONSIDERAÇÕES FINAIS DO TRABALHO</b> .....	79
6.1 LIMITAÇÕES E TRABALHOS FUTUROS .....	80
<b>REFERÊNCIAS</b> .....	81
<b>APÊNDICE A - Conjunto de dados usados no protótipo da SASys</b> .....	86
<b>APÊNDICE B - Conjunto de dados usados no estudo de caso</b> .....	92

## 1 INTRODUÇÃO

A Análise de Sentimento é uma área multidisciplinar que tem sido largamente usada para mineração de texto e recuperação de informação. A grande adoção da Internet possibilitou a geração de conteúdo tanto por parte de quem disponibiliza algum tipo de informação, quanto de quem consome. O aparecimento de sites de instituições governamentais, educacionais ou comerciais (de fins lucrativos), sites de transparência, páginas de reclamações de produtos, blogs, fóruns e redes sociais online facilitou o compartilhamento de pensamentos, impressões, experiências, sentimentos e opiniões sobre objetos, lugares, momentos, notícias, preferências e serviços.

As mídias sociais têm um papel relevante na rápida propagação de informação (ABEDIN; BABAR; ABBASI, 2014). Nesses ambientes, as pessoas tendem a compartilhar seus pontos de vista sobre os mais variados assuntos e a confiar no feedback de outros usuários que tenham experimentado serviços e produtos, ou visitado lugares (AHMAD et al., 2018). Essas informações ajudam não só as pessoas na tomada de decisão sobre o que consumir, como também oferecem às empresas a oportunidade de obter informações sobre seu público e aceitação da sua marca, além de viabilizar a adoção de estratégias que as impactem positivamente (CIRQUEIRA et al., 2018; RANA; CHEAH; LETCHMUNAN, 2016). Existe um interesse econômico na extração dessas informações, contudo, a quantidade e a variedade de dados disponíveis tornam o trabalho bastante dispendioso, demorado e propenso a erros, quando realizado de forma manual (RANA; CHEAH; LETCHMUNAN, 2016; SOUZA et al., 2018). Assim sendo, a Análise de Sentimento tem sido amplamente adotada por permitir a descoberta de opinião ou sentimento de forma automática.

A Análise de Sentimento (AS) ou Mineração de Opinião é o processo de extrair pensamentos e percepções humanas de textos não estruturados (HEMMATIAN; SOHRABI, 2017). Ela identifica informações sobre o sentimento do autor do texto através de análise semântica (NASSIRTOUSSI et al., 2014). É um domínio multidisciplinar que pode aplicar técnicas de Linguística Computacional, Inteligência Artificial, Análise de Texto ou Processamento de Linguagem Natural, com o propósito de identificar automaticamente o sentimento do autor em relação ao texto escrito (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014), considerando as valências positiva, negativa e neutra (CIRQUEIRA et al., 2018).

A AS é amplamente usada em aplicações de pesquisa de mercado e recomendação de conteúdo (KUMAR; JAIN, 2015), marketing, política e compras on-line (FELDMAN, 2013),

educação (ALTRABSHEH; COCEA; FALLAHKHAIR, 2015; ALTRABSHEH; GABER; COCEA, 2013; LIMA; DE CASTRO, 2016), saúde e bem-estar (GOHIL; VUIK; DARZI, 2018), influenciadores digitais (BAMAKAN; NURGALIEV; QU, 2018), desastres naturais (ABEDIN; BABAR; ABBASI, 2014), entre outras áreas.

Por ser uma área multidisciplinar, ela tem atraído pesquisadores da academia e da indústria durante as duas últimas décadas (RANA; CHEAH; LETCHMUNAN, 2016) e suas soluções podem ser agrupadas em quatro grandes abordagens: Lexical, Aprendizado de Máquina e Híbrida. A abordagem lexical procura encontrar palavras carregadas de sentimento, o aprendizado de máquina reconhece padrões no texto e a abordagem híbrida, como próprio nome sugere, combina métodos lexicais e algoritmos de aprendizado de máquina, para a detecção de polaridade em textos.

Os algoritmos de aprendizado de máquina necessitam de bases de treinamento e de teste rotuladas, e a geração desse conjunto de dados requer anotadores humanos. Apesar desta abordagem gerar melhores resultados, ela tem sido trocada por métodos lexicais por ser mais dispendiosa (DEY; JENAMANI; THAKKAR, 2018; HEMMATIAN; SOHRABI, 2017). Os métodos lexicais usam listas compostas de palavras e suas respectivas polaridades, que são geradas uma única vez e são consultadas sempre que a análise de sentimento é efetuada. Estas listas geralmente se encontram disponíveis online e essa abordagem tem sido a escolha de alguns pesquisadores pelo seu baixo custo de implementação (HEMMATIAN; SOHRABI, 2017).

A FrameNet é um projeto de lexicografia computacional que tem como objetivo gerar soluções computacionais para problemas de Processamento de Linguagem Natural (SALOMÃO, 2009). Ela foi desenvolvida e é mantida por linguistas que usam *frames* para a descrição de significados das palavras. Os *frames* são estruturas cognitivas que definem situações, objetos ou eventos através de cenários (FILLMORE et al., 2003a). Eles formam uma rede complexa através dos seus relacionamentos semânticos, e neste trabalho são usados na análise de sentimento como método de detecção de polaridade das palavras de um texto, através do pareamento palavra-*frames* (onde o *frame* determina o sentimento da palavra).

## 1.1 PROBLEMA

Apesar do grande interesse em proporcionar a identificação automática do sentimento, a qualidade da detecção ainda é um problema na Mineração de Opinião (NASSIRTOUSSI et al., 2014).



Os algoritmos de aprendizado de máquina utilizados na AS apresentam bons resultados. Contudo, as bases de dados rotuladas usadas para o treinamento de seus classificadores são caras, demoradas e difíceis de obter (HEMMATIAN; SOHRABI, 2017). Além disso, os custos relativos a processamento e memória têm se mostrado um empecilho para adoção dessa abordagem como solução de AS. Por esses motivos, pesquisadores e a indústria têm recorrido a abordagens lexicais por serem rápidas e menos custosas de se implementar. Todavia, um dos maiores desafios das abordagens lexicais consiste em lidar com dados semi ou não-estruturados, que requerem técnicas de Processamento de Linguagem Natural (HEMMATIAN; SOHRABI, 2017). A dinamicidade da língua com o aumento do uso de jargões, gírias, símbolos, imagens e *emojicons* também dificulta o processo de identificação de polaridades e motiva a geração de novos léxicos. Estas características fazem com que o desempenho de abordagens lexicais seja inferior à das outras abordagens e uma forma de melhorar a qualidade de seus resultados seria usando solução de Linguística Computacional, como a FrameNet, que lida com jargões e contextos em que as palavras são empregadas.

Neste contexto, surge o problema a ser tratado neste trabalho que é identificar e adotar soluções da Linguística Computacional para melhorar o desempenho da abordagem lexical na análise de sentimento.

## 1.2 QUESTÕES DE PESQUISA

Dada a contextualização do tema e do problema, é apresentada, a seguir, a questão de pesquisa do presente trabalho:

*QP0: Uma arquitetura baseada na FrameNet como abordagem lexical é capaz de identificar o sentimento do autor de um texto?*

A partir da questão QP0 são derivadas duas questões de pesquisa secundárias:

*QPS1: É possível usar a FrameNet como uma solução lexical para a AS?*

*QPS2: A arquitetura proposta funciona adequadamente para a AS em textos?*

## 1.3 JUSTIFICATIVAS

A FrameNet pode ser usada como uma solução lexical por possuir um embasamento linguístico na criação e manutenção dos *frames*, que podem ser usados como léxico na

determinação do sentimento das palavras, através da semântica contida neles. Ela também possibilita a desambiguação de termos, pois os *frames* definem cenários específicos.

Além disso, os métodos lexicais de Análise de Sentimento, comumente encontradas na literatura, procuram extrair aspectos do texto ou determinar a orientação semântica das palavras. Com a arquitetura proposta neste trabalho, também é possível definir traçar o estado emocional momentâneo do autor do texto, através da variedade de dados coletados.

#### 1.4 OBJETIVOS

O presente trabalho tem como objetivos principais contribuir no campo de Mineração de Opinião apresentando a SASys, uma arquitetura de Análise de Sentimento que usa abordagem lexical, bem como propor o uso da FrameNet para detecção de sentimentos.

Como objetivos específicos, o trabalho se propõe a:

- I. Contextualizar a FrameNet, apresentado o estado-da-arte do projeto;
- II. Avaliar a estrutura da rede de *frames* da FrameNet para verificar a viabilidade do seu uso para AS;
- III. Avaliar o uso da FrameNet como uma solução lexical de AS através da condução de um estudo observacional;
- IV. Apresentar a arquitetura conceitual proposta;
- V. Avaliar a arquitetura proposta através de um estudo de caso.

#### 1.6 METODOLOGIA

A metodologia utilizada neste trabalho foi definida em 5 etapas: (i) estudo terciário da literatura sobre Análise de Sentimento, (ii) proposta de uma arquitetura de Análise de Sentimento baseada em uma abordagem lexical para detecção de sentimento e descoberta do estado emocional do autor do texto, (iii) desenvolvimento de uma rede de *frames* polarizada para a detecção de sentimento, (iv) validação da rede de *frames* através de um estudo observacional e (v) avaliação da arquitetura proposta através de um estudo de caso.

O estudo terciário da literatura foi realizado através de uma busca por mapeamentos e revisões sistemáticas da literatura sobre Análise de Sentimento, com o objetivo de revelar os métodos usados, áreas de aplicação bem como identificar lacunas de pesquisa na área.

A arquitetura proposta foi desenvolvida com o objetivo de fazer uso dos recursos oferecidos pela FrameNet para a análise de sentimento, bem como os dados disponíveis para a

descoberta do estado emocional do autor no momento da escrita do texto. A FrameNet possui um rico léxico com embasamento linguístico que é mantido e atualizado por especialistas. A arquitetura se baseia em uma abordagem lexical que utiliza os *frames* para a desambiguação de palavras e definição de sentimentos.

A rede polarizada de *frames* foi gerada com o objetivo de criar uma estrutura computacional que refletisse a estrutura conceitual da FrameNet e que contivesse a polaridade dos *frames*. Ela foi construída através da definição de uma rede complexa e uso de algoritmo de propagação de rótulo, que teve como entrada uma amostra aleatória de 266 *unidades lexicais* com sentimento, que correspondem a 0,2% do seu total.

Um estudo observacional foi realizado como o primeiro esforço de avaliação da rede polarizada de *frames*, que usou uma base de dados conhecida na literatura para avaliação de abordagens de análise de sentimento. As frases usadas neste estudo foram selecionadas através de amostragem aleatória que correspondia a 5% do total da base de dados.

A arquitetura proposta foi avaliada através de um estudo de caso com alunos da disciplina de Metodologia de Pesquisa Científica e Educacional de um curso à distância. Três especialistas avaliaram o desempenho da arquitetura. Os resultados apontam para a viabilidade da proposta através da assertividade apresentada na AS dos textos escritos pelos alunos.

## 1.7 CONTRIBUIÇÕES

Esta dissertação apresenta como contribuições: um mapeamento sistemático da literatura sobre AS na educação (BÓBÓ et al., 2019a), o estudo da estrutura da rede de *frames* da FrameNet Brasil (BÓBÓ et al., 2018), a rede de *frames* polarizada e a identificação do perfil emocional do aluno através da arquitetura SASys (BÓBÓ et al., 2019b).

## 1.8 ORGANIZAÇÃO

Este trabalho apresenta mais 4 capítulos, além desta introdução. O Capítulo 2 apresenta os pressupostos teóricos que embasaram este trabalho. O Capítulo 3 discorre sobre os trabalhos relacionados a essa pesquisa. A arquitetura SASys é descrita no capítulo 4. E no capítulo 5 são destacadas as contribuições e limitações da pesquisa através de um estudo de caso. O Capítulo 6 apresenta as considerações finais sobre a pesquisa, bem como trabalhos futuros.

## 2 PRESSUPOSTOS TEORICOS

Este capítulo discorre sobre a Análise de Sentimento, evidenciando as diversas metodologias empregadas na área, bem como apresenta a FrameNet, contextualizando o projeto e apresentando a sua estrutura semântica.

### 2.1 ANÁLISE DE SENTIMENTO

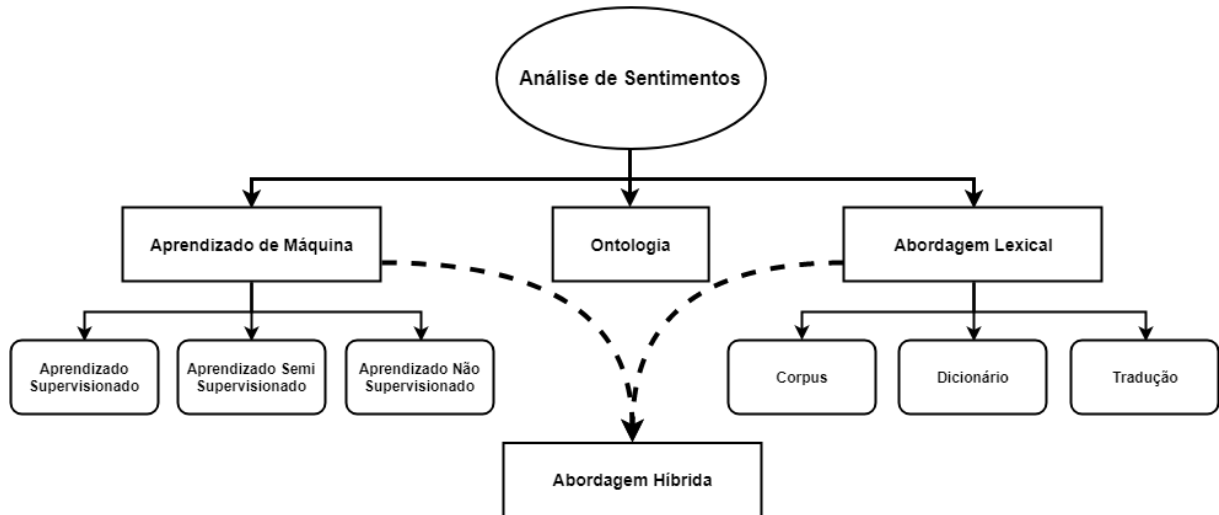
O paradigma gerado pela Web 2.0 fez com que uma vasta quantidade de dados, dos mais diversos assuntos, ficasse disponível e pronta a ser consumida. A proliferação das redes sociais facilitou o compartilhamento de opiniões (ALTRABSHEH; GABER; COCEA, 2013; KUMAR; BAADKAR; JOSHI, 2016). Segundo (KUMAR; BAADKAR; JOSHI, 2016) é possível verificar um contágio emocional em mídias sociais, sem interação face-a-face entre os usuários, através do conteúdo compartilhado. Esse conteúdo diz respeito a opinião, que pode ser explorada pela Análise de Sentimentos, para ser usada para aplicações de pesquisa de mercado e recomendação de conteúdo (KUMAR; JAIN, 2015), marketing, política e compras on-line (FELDMAN, 2013), educação (ALTRABSHEH; COCEA; FALLAHKHAIR, 2015; ALTRABSHEH; GABER; COCEA, 2013; LIMA; DE CASTRO, 2016), saúde e bem-estar (GOHIL; VUIK; DARZI, 2018), influenciadores digitais (BAMAKAN; NURGALIEV; QU, 2018), desastres naturais (ABEDIN; BABAR; ABBASI, 2014), entre outras áreas.

A Análise de Sentimento (AS) é uma subárea da Mineração de Dados que começou a ser explorada nos anos 2000 (HEMMATIAN; SOHRABI, 2017) e que tem atraído tanto pesquisadores da academia como da indústria (RANA; CHEAH; LETCHMUNAN, 2016). Ela pode ser definida como sendo o estudo computacional de opiniões, sentimentos e emoções expressas em um texto (ALTRABSHEH; GABER; COCEA, 2013) com relação a produtos, serviços, organizações, indivíduos, problemas, eventos ou tópicos. Ela visa extrair a polaridade ou a subjetividade do texto, geralmente de domínio específico, de forma automática (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014). Polaridade pode ser definida como sendo o julgamento que a palavra traz à mente (HEMMATIAN; SOHRABI, 2017) e ela pode ser positiva, negativa ou neutra. E a subjetividade representa fatos, emoções, visões ou crenças (ALTRABSHEH; GABER; COCEA, 2013).

Segundo Hemmatian e Sohrabi (2017), as soluções usadas em Análise de Sentimento podem ser agrupadas em três abordagens: Aprendizado de Máquina, Lexicais e Híbridas (que

é a junção das duas primeiras), como é ilustrado na Figura1. Contudo, existem estudos na literatura que fazem uso de ontologias (JI et al., 2016; XU; ZHAN; ZHU, 2008).

Figura 1: Métodos de Análise de Sentimento.



Fonte: Autora do texto, baseada em (HEMMATIAN; SOHRABI, 2017)

A abordagem de Aprendizado de Máquina usa algoritmos para reconhecimento de padrões no texto e eles podem ser divididos em:

- **Aprendizado supervisionado:** algoritmos que necessitam de um conjunto significativo de dados rotulados (conjunto de treinamento). Neste tipo de abordagem, as classes são previamente estabelecidas. Para cada dado do conjunto de treinamento, é atribuída uma classe, gerando assim a função de mapeamento que posteriormente é usada para classificar conjuntos de dados novos. Os algoritmos mais conhecidos são: Naive Bayes, Maximum Entropy, Support Vector Machine, Neural Networks e Decision tree (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014; HEMMATIAN; SOHRABI, 2017);
- **Aprendizado semi-supervisionado:** algoritmos que demandam menos esforço humano para a construção da função de mapeamento, pois não necessitam de um conjunto grande de treinamento. Neste tipo de abordagem nem todas as classes precisam ser previamente conhecidas. Os algoritmos mais usados se agrupam em quatro abordagens: Self-training, Co-training, Multi-view Learning, Generative Models e Graph-based Methods (HEMMATIAN; SOHRABI, 2017);

- Aprendizado não supervisionado: algoritmos baseados em agrupamento, que não necessitam de um conjunto de treinamento para gerar o resultado. Eles usam a estrutura do conjunto de dados para definir similaridades. Nesta abordagem, as classes (grupos) são encontradas pelos algoritmos automaticamente. Os algoritmos mais conhecidos são: k-means, Fuzzy C-means e LDA (HEMMATIAN; SOHRABI, 2017).

A abordagem Lexical é uma solução baseada em vocabulário de sentimento. Neste método, cada palavra possui um sentimento ou polaridade e pode ou não ter peso associado (probabilidade de pertencerem a uma classe). A polaridade é determinada analisando a orientação semântica da palavra (MACHADO; PARDO; RUIZ, 2018). Técnicas de Processamento de Linguagem Natural e Análise de Texto são usadas para identificar a estrutura sintática da frase e assim definir o sentimento de um documento ou parágrafo (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014; HEMMATIAN; SOHRABI, 2017). Segundo (SOUZA et al., 2011) a abordagem lexical se ramifica em:

- Corpus: usa a relação entre palavras e expressões encontrada em grandes corpus para determinar a polaridade. É de domínio específico por isso diferentes sentidos e conotação da mesma palavra podem não ser capturados;
- Dicionário ou léxico: explora as relações semânticas anotadas em recursos como dicionários de sinônimos e tesouros. É uma lista genérica de palavras e suas polaridades. Tem como vantagem o uso de uma lista formalmente codificada e validada por uma vasta base lexical. Contudo, gírias, jargões e outras especificidades de contextos podem não ser contemplados;
- Tradução ou multilíngue: usam recursos linguísticos disponíveis em outras línguas. Possuem a vantagem de disponibilizar uma solução de AS às línguas sem conjunto de dados disponível. Porém, têm de lidar com o desafio de traduzir palavras ou expressões de outras línguas, mantendo o sentido original.

A estratégia básica desta abordagem consiste em contar o número de palavras positivas e negativas e associar à frase a polaridade de maior quantidade (DING; LIU; YU, 2008). Caso o número de palavras positivas seja igual ao de palavras negativas, a frase recebe a polaridade neutra. Geralmente, soluções que usam essa abordagem focam nas palavras de forma individual, não considerando dependências entre elas (como por exemplo, grau

comparativo de adjetivos) e por esse motivo se mostram pouco eficientes. Para sanar tais lacunas, o uso de técnicas como *n-gram* têm se mostrado recorrentes neste tipo de abordagem.

Métodos híbridos são os que misturam abordagens lexicais e de Aprendizado de Máquina nas suas soluções (HEMMATIAN; SOHRABI, 2017). Essa junção acontece com o objetivo de aumentar a acurácia das soluções, usando abordagens lexicais como técnicas de pré-processamento para os algoritmos de aprendizado de máquina (ALTRABSHEH; GABER; COCEA, 2013).

As ontologias são as abordagens menos usadas na AS. Elas são empregadas para armazenamento de dados e inferência na identificação do sentimento (JI et al., 2016).

A extração de aspectos é uma das etapas da Análise de sentimento que tem sido largamente estudada pelos pesquisadores (HEMMATIAN; SOHRABI, 2017; RANA; CHEAH; LETCHMUNAN, 2016). Ela consiste em identificar tópicos ou entidades contidas no texto e atribuir-lhes polaridades (RANA; CHEAH; LETCHMUNAN, 2016). Estas entidades podem ser características de um produto, comida, hotel, entre outros. A extração de aspecto é usada quando se requer detalhes de análise, pois um texto de sentimento positivo, por exemplo, não contém necessariamente todas as entidades classificadas com polaridade positiva.

## 2.2 FRAMENET BRASIL

A FrameNet Brasil é um laboratório de Linguística Computacional, cuja missão é desenvolver soluções computacionais para problemas de Processamento de Linguagem Natural, usando Semântica de *frames*. O projeto central do laboratório é a manutenção da FrameNet de Fillmore, que foi criada em 1997, por Charles J. Fillmore, através do cruzamento da Semântica de *Frames* com a Lexicografia (geração de dicionário) (FILLMORE et al., 2003; SALOMAO, 2009). O projeto tem como objetivo organizar as descrições lexicográficas por *frames* e usar os dados extraídos de corpus para descobrir todas as funções semânticas e propriedades gramaticais das palavras que evocam os mesmos *frames*.

Segundo (SALOMAO, 2009), a Semântica de *Frames*, ou semântica da compreensão, foi gerada como uma abordagem para solucionar problemas da semântica lexical – dar significados as palavras. A ideia central é que os significados das palavras devem ser descritos em relação aos *frames* semânticos, que são “representações esquemáticas das estruturas conceituais e padrões de crenças, praticas, instituições, imagens, etc., que fornecem uma base para uma interação significativa em uma determinada comunidade de fala” (FILLMORE et

al., 2003). Neste sentido, os *frames* são pacotes de conhecimento que moldam e permitem que os humanos deem sentido as suas experiências (FILLMORE; BAKER, 2010). Os *frames* que são evocados se baseiam no conhecimento que temos sobre os fenômenos e sua associação com os valores culturais, por exemplo a frase:

*Mary foi convidada para a festa de Jack. Ela se perguntou se ele gostaria de um brinquedo.*

Os *frames* evocados na frase acima estão ancorados ao verbo *convidar* – que indica um relacionamento envolvendo um anfitrião, um convidado e uma ocasião - e ao substantivo *festa* – que evoca um evento social geralmente com um anfitrião, convidados e uma ocasião. O trecho *festa de Jack* remete a uma festa na qual o Jack é o anfitrião ou na qual Jack é celebrado. Não se verificam implicações linguísticas que evoquem diretamente o *frame* de aniversário, porém o substantivo *brinquedo*, a preocupação *se o Jack iria gostar de um* e os outros detalhes fornecidos pela linguagem, permitem que o leitor infira que o Jack é o aniversariante, que *brinquedo* é o presente de aniversário, que a Mary é a convidada e assim por diante.

### 2.2.1 Estrutura semântica da FrameNet

A FrameNet é constituída por *unidades lexicais*, *frames*, *elementos de frame* e relacionamentos, e nesta seção serão apresentados cada um destes elementos, de forma detalhada.

#### I) *Unidade Lexical (UL)*

Uma *Unidade Lexical - UL* é a palavra quando lhe é atribuída um dos seus significados, ou seja, é um emparelhamento de uma palavra com um dos significados de um *frame* (RUPPENHOFER et al., 2006). Uma palavra com quatro significados é tratada como quatro unidades lexicais e, na maioria dos casos, ela pode pertencer a mais de um *frame* (FILLMORE et al., 2004). Dizemos que a palavra *evoca* um *frame* quando o significado dela é baseado no *frame*.

A meta é que toda UL evoque um *frame*, porém, ela deve destacar algum elemento desse *frame* de forma particular (SALOMAO, 2009). Por exemplo, o *frame* *aplicar-calor* é o que descreve uma situação envolvendo cozinha, comida e um instrumento de aquecimento, e ele é evocado pelas palavras *assar*, *cozer*, *ferver*, *secar*, *borbulhar*, *corar*, *dourar*, *grelhar*, *queimar*, *vapor*, etc. Essas palavras são



chamadas de *Unidades Lexicais* e elas podem carregar sentimentos (a palavra *queimar* pode estar associada a um sentimento negativo, por exemplo).

## II) *Frame*

Um *Frame* é uma “estrutura conceitual que descreve um tipo particular de situação, objeto ou evento, juntamente com seus participantes e adereços” (RUPPENHOFER et al., 2006). É um sistema de conceitos relacionados de modo que para compreender qualquer um deles, é necessário compreender o sistema como um todo. Ele é composto por elementos que ajudam a completar o seu significado (*Elementos de Frame*), evocado por uma *Unidade Lexical* e se liga a outros *frames* com relacionamentos específicos.

## III) *Elemento de Frame (EF)*

Os *Elementos de Frame* – EF são atributos usados como etiquetas para as palavras ou frases que estão na construção gramatical com as UL que evocam o *frame* (FILLMORE; BAKER; SATO, 2004). São os papéis semânticos das entidades envolvidas em cada *frame* (FILLMORE; RUPPENHOFER; WRIGHT, 2003).

Os EF existem na estrutura do *frame* e podem ou não estar representados na frase em que o *frame* é evocado. Por exemplo o *frame* *Danificar* possui como principais EFs os atributos:

- *Danificar*: um *Agente* que afeta um *Paciente* de modo que este mude para um estado não-canônico:
  - *Agente*: A entidade consciente, intencional que resulta no dano ao *Paciente*;
  - *Paciente*: A entidade que é afetada pelo *Agente*, para que esteja danificada;
  - *Causa*: Um evento que leva ao dano do *Paciente*.

Atualmente a FrameNet cobre mais de 13.000 *Unidades Lexicais*, distribuídas em mais de 1.200 *Frames* e atestadas por mais de 200.000 frases anotadas (TORRENT et al., [s.l: s.n.]).

### 2.2.2 **Relacionamentos**

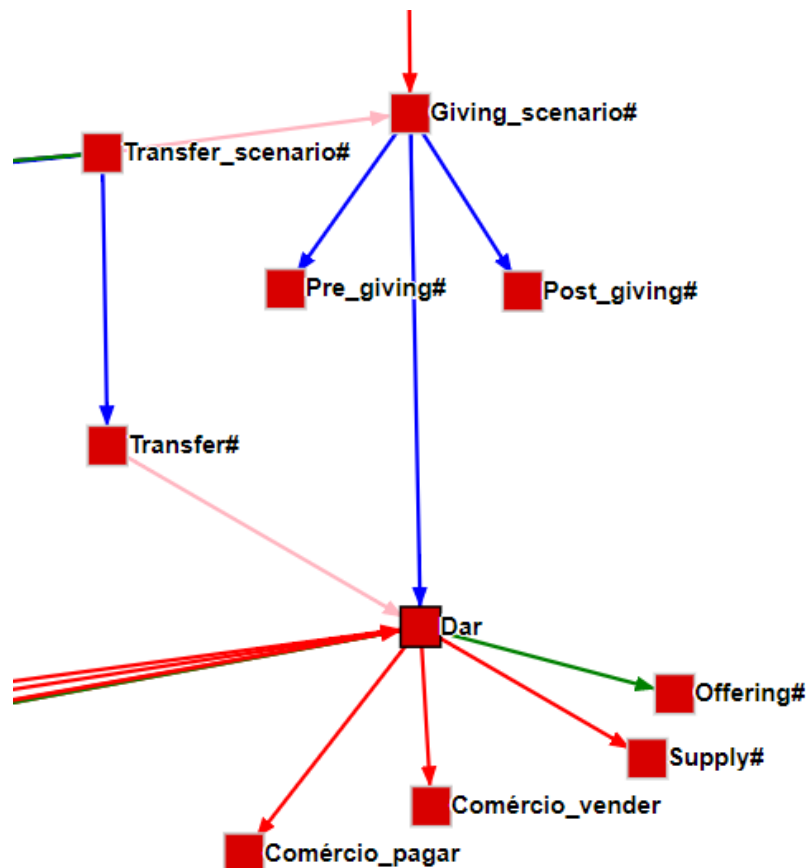
A FrameNet é uma rede de *frames* em que as ligações ocorrem através de relações específicas (*Herança, Perspectiva, Uso, Subframe e Precedência*). As relações são usadas para melhorar a compreensão dos *frames* e proporcionar robustez (visto que *frames* semanticamente

similares podem estar associados, apesar de estarem separados) (RUPPENHOFER et al., 2006). A Figura 2 ilustra relacionamentos entre *frames* e a seguir são descritos cada um deles.

a. *Herança*

É o relacionamento mais forte na FrameNet (RUPPENHOFER et al., 2006). Ocorre entre um *frame* pai e um *frame* filho, onde o *frame* filho herda todos ou parte dos *EFs* e *subframes* do pai. Os *EFs* do *frame* filho não têm, necessariamente, os mesmos nomes dos *EFs* do pai e podem ser adicionados outros, dadas as especificidades do *frame* filho. Por exemplo, o *frame* Fornecer herda do *frame* Dar e, além de possuir os *EFs* Tema e Destinatário, ele também especifica que o doador é um *Fornecedor* e que tem que existir o *Objetivo* do tema proposto.

Figura 2: Relacionamentos do *frame* Dar com os outros *frames*.



Fonte: *WebTool*<sup>1</sup>

b. *Uso*

É o relacionamento usado quando parte do cenário evocado pelo *frame* filho se refere a um *frame* pai. É um tipo de relacionamento de herança, onde o *frame* filho pode usar (herdar de)

<sup>1</sup> <http://webtool.framenetbr.ufjf.br/>

múltiplos *frames* pai. Por exemplo, o *frame* Oferta usa o *frame* Dar, já que o *Ofertante* oferece um *Tema* para um *Potencial\_Destinatário* somente quando o *Potencial\_Destinatário* aceita a oferta e que a *Transferência* ocorre.

#### c. *Perspectiva*

Este relacionamento é parecido ao de *Uso* e consiste em indicar, pelo menos, dois pontos de vista de um *frame* neutro. Por exemplo, o *frame* Dar é uma perspectiva do *frame* Transferência, uma vez que os verbos Dar e Receber expressam o ponto de vantagem da cena (do *Doador* ou do *Destinatário*, respectivamente).

#### d. *Subframe*

Esse relacionamento é usado para representar sub-eventos ou subpartes de um *frame* complexo. Os *subframes* comumente se referem a sequência de passos ou transações e podem ser descritos, separadamente, por *frames*. Por exemplo, o *frame* Dar juntamente com os *frames* Pré\_Doação e Pós\_Doação são *subframes* do *frame* Cenário\_Doação.

#### e. *Precedência*

Esse relacionamento ocorre apenas entre dois *subframes* de um *frame* complexo para especificar a sequência dos eventos de um certo cenário. Este é o único relacionamento que pode formar ciclos. No *frame* Cenário\_Doação, o relacionamento *Precedência* é usado para conectar o *frame* Pré\_Doação ao *frame* Dar e o *frame* Dar ao *frame* Pós\_Doação, pois eles devem ocorrer nessa ordem.

### 2.2.3 Modelo conceitual da FrameNet

A FrameNet procura capturar os *insights* humanos em estruturas semânticas, que são armazenadas em banco de dados relacional. O processo de adição do *frame* ao banco começa com a caracterização do *frame*, definidos o tipo de entidade, situação ou objeto que representa, escolhendo os rótulos dos EFs e a lista de ULs vinculadas a ele (FILLMORE; JOHNSON; PETRUCK, 2003).

Segundo Baker, Fillmore e Lowe (1998), quatro etapas de processamento foram necessárias para a geração do banco de dados da FrameNet:

- Preparação: descrições iniciais dos *frames*, ULs e EFs, e verificação do padrão

sintático de cada um;

- Extração de corpus: geração de boas frases de exemplo através de ferramentas de linguística computacional;
- Anotação: marcação manual dos EFs detectados no subcorpus e identificação de padrões de exemplos e frases com problemas;
- Entrada: adição de dos dados às tabelas do banco.

Os relacionamentos entre *frames* foram implementados no banco através da criação das tabelas *frame*, ULs, EFs e relacionamento. Essas tabelas procuram refletir a estrutura semântica do projeto (FILLMORE; BAKER; SATO, 2004).

Os linguístas são responsáveis pela manutenção e atualização da FrameNet. Eles escolhem os termos que farão parte da lista de ULs, examinam o uso dos EFs e determinam os contextos sintáticos e coloquiais do significado do *frame*, consultando dicionários de tesouros (FILLMORE et al., 2003).

A expansão do projeto inicial da FrameNet para outras línguas inclui o uso da estrutura semântica definida em Inglês. A FrameNet Brasil, procurou contribuir com o projeto inicial através de um léxico, um Construction, um banco de dados trilingue para o domínio do esporte e turismo e a WebTool<sup>2</sup> (TORRENT et al., [s.l: s.n.]).

### 2.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Nesse capítulo foram apresentados conceitos relacionados a Análise de Sentimento (AS) e a FrameNet Brasil, com o intuito de fornecer embasamento teórico para a compreensão desta dissertação.

A AS foi brevemente contextualizada, apresentado as soluções para cada tipo de abordagem que são frequentemente encontradas na literatura.

A estrutura semântica e o modelo conceitual da FrameNet Brasil foram também apresentados neste capítulo na tentativa de esclarecer o processo linguístico na geração do léxico, que viabilisa a adoção da FrameNet Brasil como solução lexical. Tanto a estrutura quanto o modelo foram validados no capítulo 4, ao ser examinada a estrutura da rede de *frames* através de análises topológicas.

Neste trabalho, a FrameNet Brasil que também foi alvo de estudo, será referenciada como apenas FrameNet.

---

<sup>2</sup> <http://webtool.framenetbr.ufjf.br/>

### 3 TRABALHOS RELACIONADOS

Nesta seção são apresentados artigos relacionados ao tema deste trabalho. O objetivo é discutir, de forma pontual, assuntos abordados na literatura que motivaram o desenvolvimento da arquitetura SASys.

Foi conduzido um estudo terciário sobre revisões e mapeamentos sistemáticos da literatura sobre análise de sentimento. Este estudo visa obter o estado-da-arte do tema e identificar oportunidades de pesquisa. Também são apresentados outros trabalhos relacionados à análise de sentimento que usam abordagem lexical e que exploram a aplicabilidade de alguns tipos de dados.

#### 3.1 MAPEAMENTO DA LITERATURA SOBRE ANÁLISE DE SENTIMENTO – UM ESTUDO TERCIÁRIO

Mapeamentos e Revisões Sistemáticas da Literatura são estudos secundários que tem como propósito prover a visão geral de uma determinada área de pesquisa (KITCHENHAM; BUDGEN; BRERETON, 2011). São estudos sistemáticos que se propõem identificar e classificar conteúdo relacionado a um tópico, evidenciando o seu estado-da-arte e ressaltando subtópicos necessitam de mais estudos primários (WOHLIN et al., 2012).

O estudo realizado nesta etapa da pesquisa foi restringido a um mapeamento da literatura que procura proporcionar uma visão geral dos trabalhos realizados no campo de Análise de Sentimento, através da busca por estudos secundários sobre o assunto. O protocolo utilizado se aproxima ao de um mapeamento sistemático da literatura, se distinguindo apenas na realização de uma busca *ad-hoc* num único motor de busca bibliográfico. As etapas do mapeamento são descritas nas seções que se seguem.

##### 3.1.1 Planejamento

Nesta etapa do mapeamento é definido o protocolo a ser utilizado, de modo a mitigar o viés inerente a pesquisa. O protocolo especifica as questões de pesquisa, a estratégia de busca, as fontes de estudo e os critérios de seleção (NAKAGAWA et al., 2017).

Por se basear em estudos terciários, as questões de pesquisa do mapeamento foram mais genéricas, de modo a apenas estruturar a área. Assim sendo, foram elaboradas 4 questões de pesquisas apresentadas na Tabela 1 com suas respectivas justificativas.

Tabela 1 – Questões do mapeamento e respectivas justificativas

	<b>Questão do Mapeamento</b>	<b>Justificativa</b>
QM1	Quais são as técnicas ou métodos de AS usados?	Esta questão pretende evidenciar o estado-da-arte da análise de sentimento, distinguindo as abordagens e soluções encontradas na literatura.
QM2	Em que contexto os estudos têm sido aplicados?	Com esta questão se pretende identificar a empregabilidade da AS em contextos reais.
QM3	Quais são as fontes de dados mais comuns?	Esta questão se preocupa em identificar as fontes de teste reais que, normalmente, estão vinculadas ao contexto da pesquisa.
QM4	Quais são as lacunas da área?	Esta questão pretende evidenciar as oportunidades de pesquisa no campo de AS.

A *string* de busca foi baseada em dois termos principais: *Sentiment Analysis* e *Systematic Review*. Contudo, ela foi incrementada de sinónimos que pudessem recuperar o maior número de artigos (Tabela 2). A *string* foi rodada no motor de busca *Scopus*<sup>3</sup>, buscando apenas por campos específicos do artigo (título, resumo e palavras-chave).

Tabela 2 – *String* de busca do mapeamento de estudos terciários

((“*Sentiment Analysis*” OR “*Opinion Mining*”) AND (“*Systematic Review*” OR “*Systematic Mapping*”))

Foram definidos alguns critérios de seleção visando analisar a relevância dos artigos recuperados das bases de bibliográficas (KITCHENHAM; BUDGEN; BRERETON, 2011; WOHLIN et al., 2012). Os critérios foram divididos em critérios de inclusão e de exclusão:

#### *Critério de Inclusão*

- Artigos que apresentem revisões ou mapeamentos da literatura sobre análise de sentimento;
- Estudos escritos em Inglês, Português ou Espanhol;

<sup>3</sup> <http://www.scopus.com>

- Trabalhos publicados em Conferências, Revistas ou Relatórios Técnicos que foram revisados por pares.

#### *Critério de Exclusão*

- Trabalhos que não apresentem revisões ou mapeamentos da literatura sobre análise de sentimento;
- Trabalhos escritos em outras línguas;
- Trabalhos publicados em Conferências, Revistas ou Relatórios Técnicos que não foram revisados por pares;
- Trabalhos que tenham versões mais recentes;

### 3.1.2 Resultados

Foram recuperados 14 trabalhos que foram inspecionados para a remoção de duplicatas, além da aplicação dos critérios de seleção. Obteve-se um conjunto de 10 artigos que foram lidos na íntegra, procurando responder as questões de pesquisa deste estudo terciário. A Tabela 3 apresenta a lista dos artigos aceitos

Tabela 3 – Lista dos artigos aceitos no estudo terciário

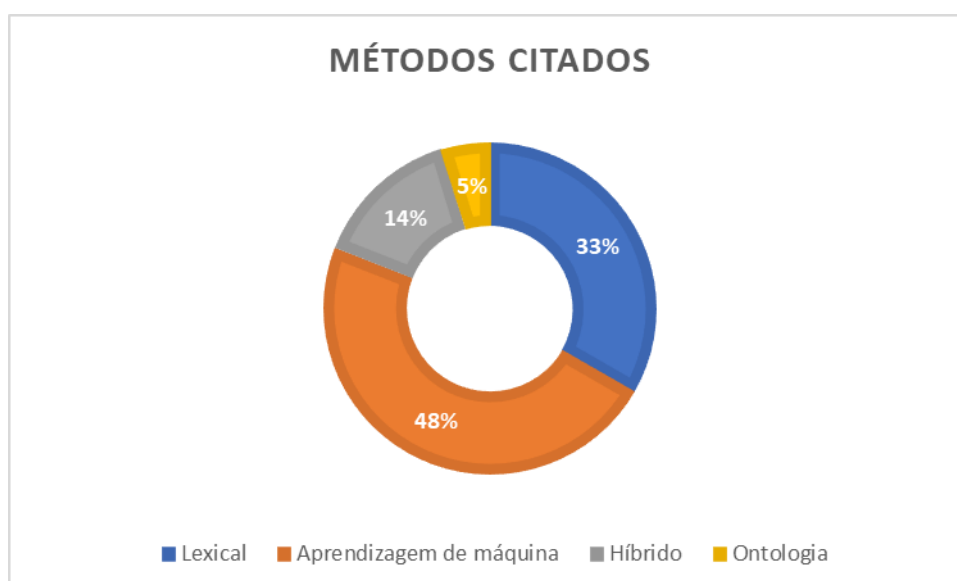
<b>Título</b>	<b>Referência</b>
Approaches to Automated Detection of Cyberbullying: A Survey	SALAWU, Semiu; HE, Yulan; LUMSDEN, Joanna. Approaches to automated detection of cyberbullying: A survey. <b>IEEE Transactions on Affective Computing</b> , 2017.
A Systematic Literature Review on Opinion Types and Sentiment Analysis Techniques	QAZI, Atika, et al. A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges. <b>Internet Research</b> , 2017, 27.3: 608-630.
A Survey on Classification Techniques for Opinion Mining and Sentiment Analysis	HEMMATIAN, Fatemeh; SOHRABI, Mohammad Karim. A survey on classification techniques for opinion mining and sentiment analysis. <b>Artificial Intelligence Review</b> , 2017, 1-51.
Characterization of the Use of Social Media in Natural Disasters: A Systematic Review	ABEDIN, Babak; BABAR, Abdul; ABBASI, Alireza. Characterization of the use of social media in natural disasters: a systematic review. In: <b>2014 IEEE Fourth International Conference on Big Data and Cloud Computing</b> . IEEE, 2014. p. 449-454.
Characterizing Opinion Mining: A Systematic Mapping Study of the Portuguese Language	SOUZA, Ellen, et al. Characterizing opinion mining: A systematic mapping study of the portuguese language. In: <b>International Conference on Computational Processing of the Portuguese Language</b> . Springer, Cham, 2016. p. 122-127.
Sentiment Analysis of Health Care Tweets: Review of the	GOHIL, Sunir; VUIK, Sabine; DARZI, Ara. Sentiment analysis of health care tweets: review of the methods used. <b>JMIR public health</b>

Methods Used	<b>and surveillance</b> , 2018, 4.2: e43.
Sentiment Analysis Using SVM: A Systematic Literature Review	AHMAD, Munir, et al. Sentiment Analysis using SVM: A Systematic Literature Review. <b>International Journal of Advanced Computer Science and Applications</b> , 2018, 9.2: 182-188.
Survey of Visual Sentiment Prediction for Social Media Analysis	Ji, Rongrong, et al. Survey of visual sentiment prediction for social media analysis. <b>Frontiers of Computer Science</b> , 2016, 10.4: 602-611.
Text Mining for Market Prediction: A Systematic Review	NASSIRTOUSSI, Arman Khadjeh, et al. Text mining for market prediction: A systematic review. <b>Expert Systems with Applications</b> , 2014, 41.16: 7653-7670.
Topic Modeling in Sentiment Analysis: A Systematic Review	RANA, Toqir Ahmad; CHEAH, Yu-N.; LETCHMUNAN, Sukumar. Topic modeling in sentiment analysis: a systematic review. <b>Journal of ICT Research and Applications</b> , 2016, 10.1: 76-93.

*QPI: Quais são as técnicas ou métodos de AS usados?*

Os artigos citaram um conjunto grande de algoritmos e técnicas de análise de sentimento que agruparam em abordagens: Aprendizagem de máquina (citado em 48% das revisões), Lexical (citado em 33% das revisões), Híbrido (citado em 14% das revisões) e Ontologias (citado em 5% das revisões), como pode ser visto no Gráfico 1. Foi recuperado um artigo que se propunha a fazer uma revisão sistemática de trabalhos de AS usando somente *Support Vector Machine (SVM)*, contudo, esse mapeamento evidenciou o uso de métodos lexicais como parte das soluções apresentadas pelos artigos e por este motivo foi classificado como um trabalho que cita métodos de aprendizagem de máquina e híbridos. Uma única revisão citou o uso de ontologias e essas se destinavam ao reconhecimento de expressões faciais.

Gráfico 1 – Métodos de AS citados nas revisões e mapeamentos

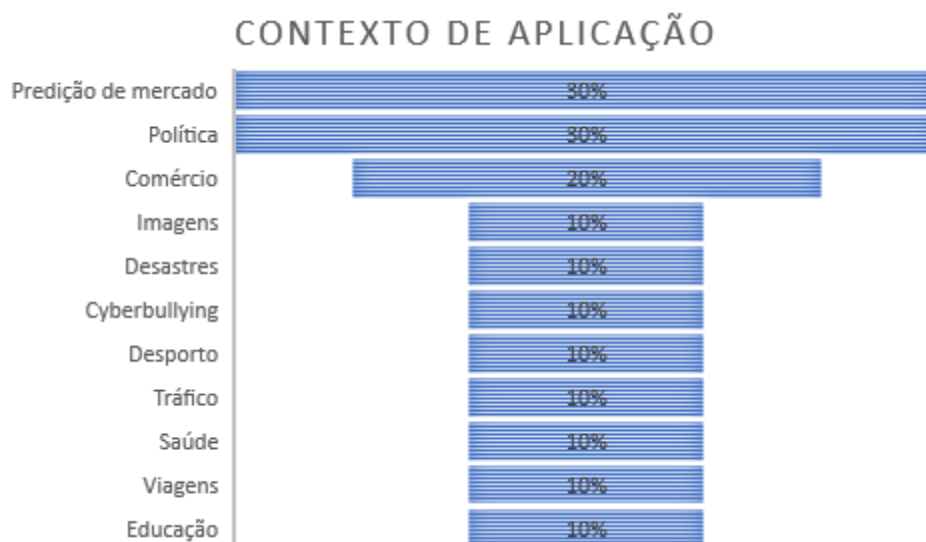




*QP2: Em que contexto os estudos têm sido aplicados?*

As aplicações mais comuns de AS encontradas na literatura são de predição de mercado (citado em 30% das revisões), política (citado em 30% das revisões) e comércio (citado em 20% das revisões). Estas foram as primeiras áreas a despertar o interesse dos pesquisadores por mineração de opinião. Contudo, a AS também tem sido empregada em trabalhos de identificação de desastres naturais, *cyberbullying*, desporto, saúde, viagens, educação, tráfico e até mesmo em imagens em redes sociais (Gráfico 2).

Gráfico 2 – Contexto de aplicação de AS citados nas revisões e mapeamentos



*QP3: Quais são as fontes de dados mais comuns?*

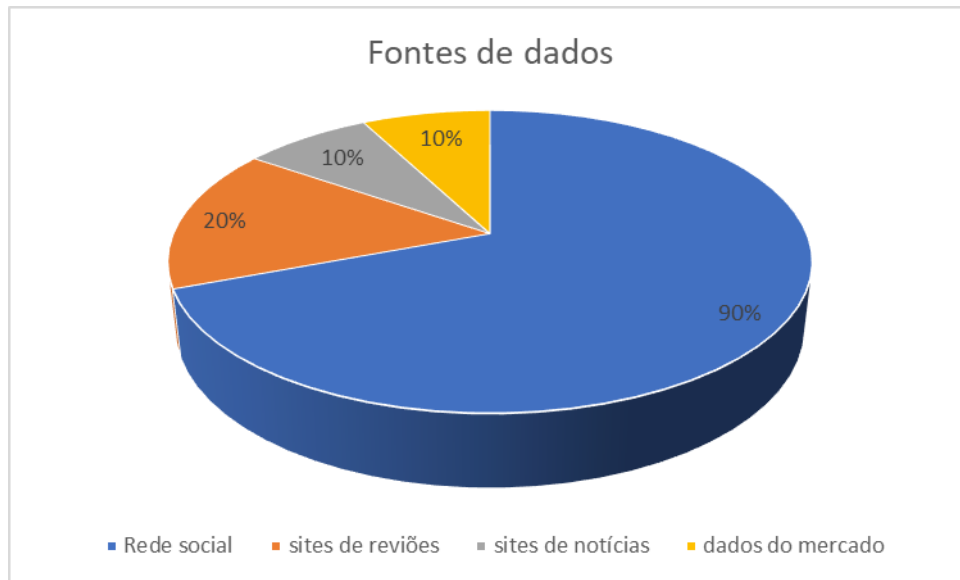
Baseado nas informações coletadas do conjunto de trabalhos recuperados da busca, os artigos de AS geralmente utilizam bases de dados de frases revisadas e rotuladas por humanos, extraídas de redes sociais (citado em 90% das revisões), sites de opiniões (citado em 20% das revisões), sites de notícias (citado em 10% das revisões) e dados de mercado (citado em 10% das revisões), como pode ser constatado no Gráfico 3. As redes sociais são largamente adotadas por possuírem conteúdos com grande teor emocional além de algumas já disporem de *APIs* que facilitam a coleta de dados.

*QP4: Quais são as lacunas da área?*

Um dos principais focos de revisões e mapeamentos sistemáticos é identificar subtópicos da área pesquisada que necessitam de mais estudos. Com base nesse princípio os estudos recuperados apontaram algumas áreas de AS que necessitam de mais foco como: a

falta de sistemas de AS especializados para as áreas pesquisadas (como por exemplo: desastres, saúde, educação ou AS em imagens), a falta ou a baixa qualidade de bases de dados de treinamento e léxico, e a pequena quantidade de trabalhos que abordam o uso do sentimento neutro.

Gráfico 3 – Fontes de dados citados nas revisões e mapeamentos



### 3.1.3 Considerações sobre o Mapeamento

As soluções de AS encontradas na literatura são agrupadas em duas abordagens principais: Aprendizagem de Máquina e Lexical. A primeira é largamente usada pelos pesquisadores, porém, constatou-se através das revisões e dos mapeamentos recuperados que alguns métodos lexicais são utilizados como pré-processamento para estes tipos de algoritmos, gerando assim soluções híbridas.

Os problemas de comércio, política e predição de mercado são os mais citados na literatura por serem os primeiros exemplos de aplicação de AS e por oferecerem bases de treinamento mais consolidadas. Contudo, a interdisciplinaridade da área tem aberto espaço para aplicações na área da educação, saúde, turismo, desporto, entre outros. As redes sociais são as fontes de dados de teste mais citadas nas revisões e mapeamentos, e elas possuem a facilidade de recuperação de dados através de *APIs*.

Soluções de AS tendem a melhorar o desempenho quando são construídas considerando o contexto de aplicação. Assim sendo, a falta de sistemas de AS especializados é apontado como uma das principais lacunas da área. A baixa qualidade dos léxicos e bases de

treinamentos de algoritmos também se revelou como uma lacuna da área que precisa ser solucionada.

O estudo terciário foi feito recuperando apenas revisões e mapeamentos sistemáticos de uma única fonte bibliográfica. Essa fonte, apesar de ser referenciada na literatura, não indexa todo o conteúdo disponível na Web. Por este motivo é possível que outros trabalhos relevantes não tenham sido contemplados por este mapeamento.

## 3.2 OUTROS TRABALHOS RELACIONADOS

Nesta seção são apresentados alguns artigos que não foram contemplados pelo estudo terciário e que exploram características da abordagem lexical que foram incorporadas na arquitetura proposta neste trabalho. Também são comparadas as soluções apresentadas pelos outros autores e a arquitetura SASys.

### 3.2.1 Lexicon-based methods for sentiment analysis (TABOADA et al., 2011)

Taboada *et. al* (2011) apresentam neste artigo o SO-CAL (Semantic Orientation CALculator) que é uma calculadora semântica capaz de identificar a polaridade das palavras contidas num texto, além da sua pontuação – intensidade a qual a palavra pertence a uma das polaridades.

Ao criarem o SO-CAL, os autores assumiram que as palavras possuem polaridades prévias, que é a orientação semântica da palavra independente do contexto, e que elas podem ser expressas em valores numéricos. Os autores também fizeram uso de várias estratégias e destacaram a aplicabilidade das negações e modificadores de valência na sua abordagem lexical, que são palavras capazes de alterar a polaridade das palavras.

O SO-CAL foi construído baseado em um dicionário desenvolvido pelos autores, anotando manualmente o sentimento das palavras e sua pontuação, variando de -5 (*extremamente negativa*) a +5 (*extremamente positiva*). As palavras com pontuação igual a zero foram excluídas do dicionário. Os modificadores de valência foram representados com valores percentuais e as negações com valores inteiros. A seguir são apresentados exemplos de frases classificadas pelo SO-CAL:

- a) *The performances were all really very good.*
- b) *She's not terrific but not terrible either.*

Na frase (a), o adjetivo *good* possui a polaridade positiva e a valência igual a +3 que somando aos intensificadores *really* (de valência igual a 25%) e *very* (de valência igual a

15%) geram o sentimento geral positivo de valência igual a 4.3 (cálculo da polaridade =  $[3 \times (100\% + 25\%) \times (100\% + 15\%)] = 4.3$ ).

No texto (b), as valências dos adjetivos *terrific* (+5) e *terrible* (-5) são alteradas pela negação, gerando um sentimento geral neutro (*She's not terrific* =  $-4 + 5 = 1$ , *not terrible* =  $+4 - 5 = -1$ ).

A proposta deste trabalho é similar à de Taboada *et. al.* (2011a) no uso de negações e modificadores de valência na AS. Contudo, a SO-CAL gera apenas a orientação semântica do texto ao passo que a SASys procura explorar informações sobre o indivíduo que escreveu o texto, através da variedade de dados que podem ser coletados.

### 3.2.2 Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks (CHAPLOT; RHIM; KIM, 2015)

Chaplot *et. al* (2015) disponibilizam neste trabalho um algoritmo capaz de prever a evasão de alunos de cursos online através de uma abordagem híbrida da AS.

Os cursos *MOOCs* (*Massive Open Online Courses*) têm uma grande aceitação devido ao seu fácil acesso, a flexibilidade de horário e não padronização de *background* dos alunos. Contudo, apesar da grande adesão, grande parte dos alunos desiste dos cursos por diversos fatores. Os autores deste artigo propõem o uso do método lexical na AS das mensagens dos alunos escritas em fóruns e redes neurais para a modelagem do problema de evasão. Os dados usados como parâmetro de entrada do algoritmo foram coletados de um dos cursos oferecidos pelo Coursera<sup>4</sup>:

- **ID do usuário:** identificador do usuário;
- **Semana do Curso:** Número de semanas desde o início do curso;
- **Semana do usuário:** Número de semanas desde que o aluno aderiu ao curso;
- **Número de cliques do usuário;**
- **Número de sessões de estudo realizadas pelo usuário;**
- **Número de páginas do curso visualizadas pelo usuário:** que incluem todas as páginas, exceto as palestras de vídeo;
- **Número de páginas do fórum visualizadas pelo usuário;**

---

<sup>4</sup> <https://www.coursera.org/>

- **O sentimento dos alunos** nas postagens do fórum.

Dado um post de um fórum, as frases eram passadas a um *post tagger* e *stemmer* e depois ao *SentiWordNet* que retornava as suas polaridades. A rede neural desenvolvida possuía 7 nós na camada de entrada (semana do curso, semana do usuário, número de cliques, número de sessões, número de visualizações de página, número de visualizações do fórum e sentimento do aluno) e um único nó na camada de saída que prevê a desistência do aluno.

Ao prever a desistência dos alunos os autores estavam focados em identificar os alunos que abandonariam o curso e assim sendo, concentraram esforços em minimizar a taxa de falsos negativos (quantidade de alunos que estão previsto em permanecer no curso, mas que na verdade o abandonariam); obtendo uma acurácia de 74%.

Chaplot *et. al* (2015) procuraram explorar a variedade de dados oferecida pelo ambiente de aprendizado do Coursera e com eles identificar os alunos com propensão a evasão. Sua metodologia foi incluída na SASys para a identificação do estado emocional do usuário, no estudo de caso.

### **3.2.3 Senti-n-gram: An n-gram lexicon for sentiment Analysis (DEY; JENAMANI; THAKKAR, 2018)**

Senti-N-Gram é um artigo que apresenta um *framework* para criar léxico *n-gram*. Segundo Dey, Jenamani e Thakkar (2018), é possível encontrar vários léxicos para a análise de sentimento constituídos apenas por palavras e suas polaridades, contudo, abordagens que utilizam as polaridades das palavras em conjunto com intensificadores ou negações tendem a ter um melhor desempenho. Sempre que essa técnica é implementada a polaridade da palavra é recalculada, por não existir um dicionário com esse sentimento já apurado, o que pode implicar em aumento no tempo de processamento.

Os autores propõem a criação automática de um dicionário de sentimentos *n-gram* através do uso de várias bases de dados de avaliações, onde a classificação numérica (de 1-5 pontos) substitui o julgamento humano para a determinação do sentimento da palavra. Eles usam como base a polaridade unitária das palavras extraídas de um dicionário público, depois utilizam uma lista de intensificadores e negadores criada por eles para gerar combinações *bigrams* e *trigrams* e suas respectivas polaridades.

Também é apresentado no artigo uma nova abordagem de análise de sentimento usando *n-gram*, onde o sentimento geral de uma avaliação é calculado tendo apenas a diferença entre o número de frases positivas e negativas. Caso o número de frases positivas

seja maior, a avaliação é considerada positiva, caso contrário é negativa. A polaridade neutra também foi considerada e ocorreu nos casos em que o número de frases com sentimentos positivos e negativos era o mesmo.

O uso de *n-gram* foi adicionado a SASys através das negações e os modificadores de valência. Porém, como os *frames* da FrameNet já descrevem as palavras no contexto específico e fazem todo o tratamento de desambiguação, a lista de negações e modificadores de valência foi gerada através deles, não havendo uma necessidade inicial de consultar outras fontes.

### **3.2.4 Creating a portugueses context sensitive lexicon for sentiment analisys MACHADO et al. (2018)**

Machado *et al* (2018) apresentaram uma solução de AS para textos em português. A sua proposta foi baseada em extração de aspectos, que consistem em detectar características de uma entidade descrita no texto.

O LexReli (Lexicon Reli) é um léxico, gerado pelos autores, que é especializado em identificar a polaridade de características extraídas em revisões de livros. O léxico é composto apenas de adjetivos e suas polaridades, e foi baseado no corpus ReLi, o único corpus português baseado em aspectos.

A detecção da polaridade dos adjetivos foi feita computando a frequência em que os adjetivos apareciam em frases positivas e negativas. Caso o número de ocorrências em frases positivas fosse maior, o adjetivo recebia a polaridade positiva; caso contrário, recebia a polaridade negativa. Se a diferença entre as ocorrências fosse menor que dois, o adjetivo era excluído do léxico.

Apesar do LexReli ser uma das poucas iniciativas de AS em português, a solução disponibilizada pelos autores é especializada em comentários sobre livros. E ao contrário da SASys, cujo foco está na detecção de polaridade de textos, a LexReli se concentra na extração de aspectos. Ademais, a SASys procura ser livre de contexto usando um léxico, construído por linguistas, que prima pelas polaridades prévias das palavras.

### 3.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

No presente capítulo foram apresentados um estudo terciário sobre revisões e mapeamentos sistemáticos da literatura sobre análise de sentimentos, bem como artigos que usam abordagem lexical e que se assemelham a proposta deste trabalho.

O estudo terciário teve o propósito prover a visão geral da área de AS através de questões de pesquisa que procuraram identificar as abordagens citadas na literatura, evidenciar a empregabilidade da AS em contextos reais, detectar fontes de dados e revelar oportunidades de pesquisa no campo de AS. Ele serviu de ponto de partida para a construção da arquitetura proposta nesta dissertação. Os outros trabalhos relacionados fundamentaram o uso de algumas características incorporadas na arquitetura, que são detalhadas no capítulo 4.

A abordagem lexical foi eleita como o método de AS utilizado pela arquitetura e a qualidade do léxico é uma das principais contribuições desta dissertação.

## 4 ARQUITETURA SASYS

Este capítulo apresenta a SASys (*Sentiment Analysis System*), uma arquitetura que procura detectar sentimentos em textos através de uma abordagem lexical e identificar o estado emocional do autor de um texto, num determinado momento, usando dados coletados e a orientação semântica das palavras. Ela usa técnicas lexicais e de uma rede de *frames* polarizada, derivada da FrameNet (BÓBÓ et al., 2018), para a caracterização do sentimento. A SASys se propõe a contribuir no campo da Mineração de Opinião oferecendo uma arquitetura que reflita os conceitos relevantes da área. A motivação de usar a rede de *frames* se dá pelo embasamento linguístico que a própria FrameNet fornece. Os *frames* indicam cenários específicos e contextualizados, além das palavras associadas a eles.

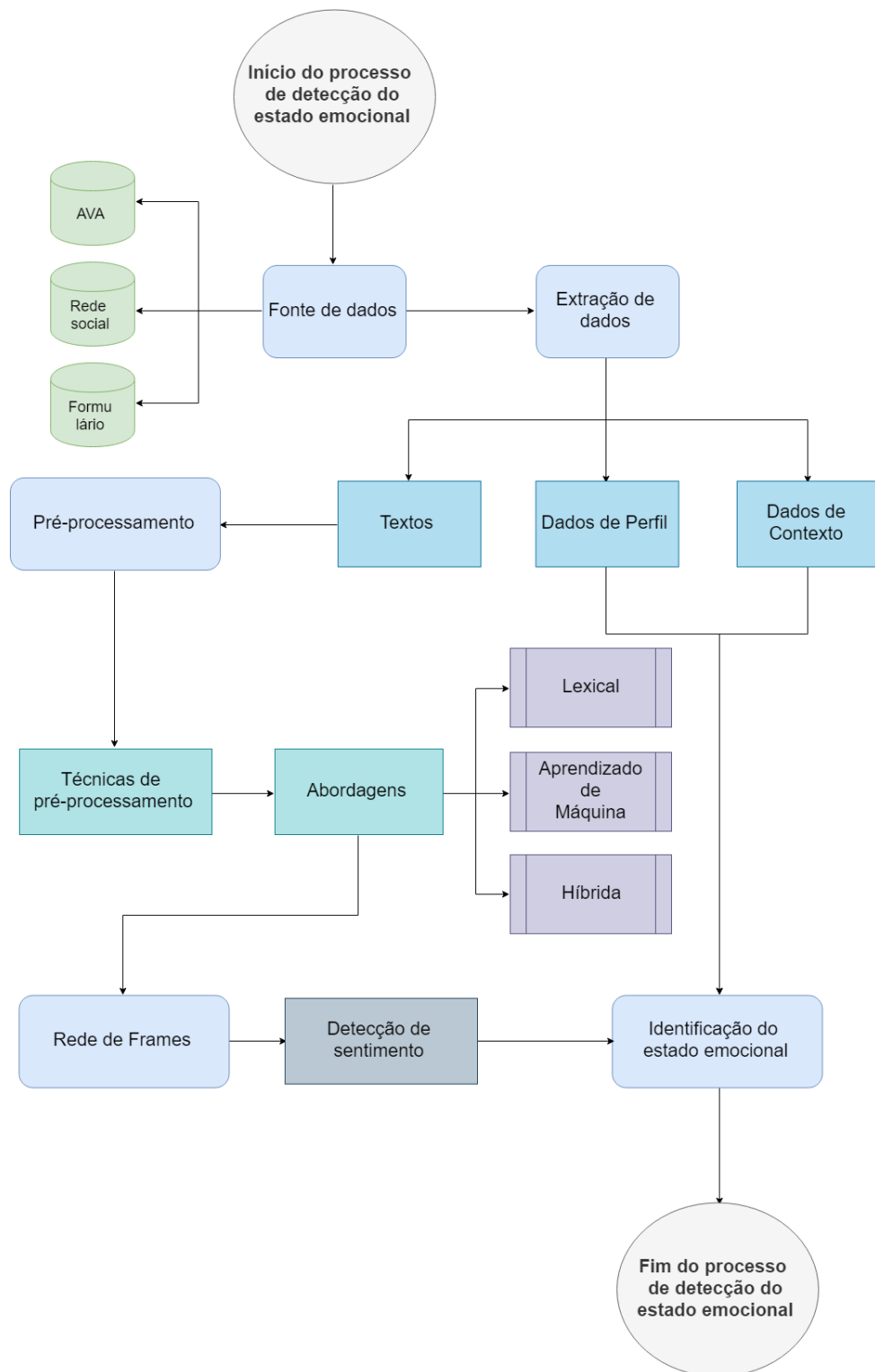
A SASys tem como objetivo principal fazer o uso da rede de *frames* para determinar o sentimento do autor em textos. Como objetivos secundários, ela tenciona:

- Identificar o estado emocional do autor do texto;
- Possibilitar o uso de mais de uma fonte de dados para a obtenção de uma AS mais precisa.

O fluxo de detecção do estado emocional, que está sendo usado na SASys, pode ser visualizado de forma simplificada na Figura 3. Ele se inicia com a extração de dados do autor de alguma fonte como por exemplo redes sociais, formulários ou Ambientes Virtuais de Aprendizagens (AVAs). Os dados são usados tanto para AS como para a definição do perfil do autor. Depois de coletados, são aplicadas técnicas de pré-processamento aos textos, como por exemplo, remoção de sinal de pontuação ou quebra de linha. Em seguida os textos são submetidos à rede complexa formada pelos *frames* da FrameNet, para a identificação das palavras com polaridade. Após essa etapa, a orientação semântica do texto é dada a partir do resultado de alguma das abordagens de AS (Lexical, Aprendizado de Máquina, Híbrida ou Ontológica). O fluxo termina com a identificação do estado emocional do autor do texto.



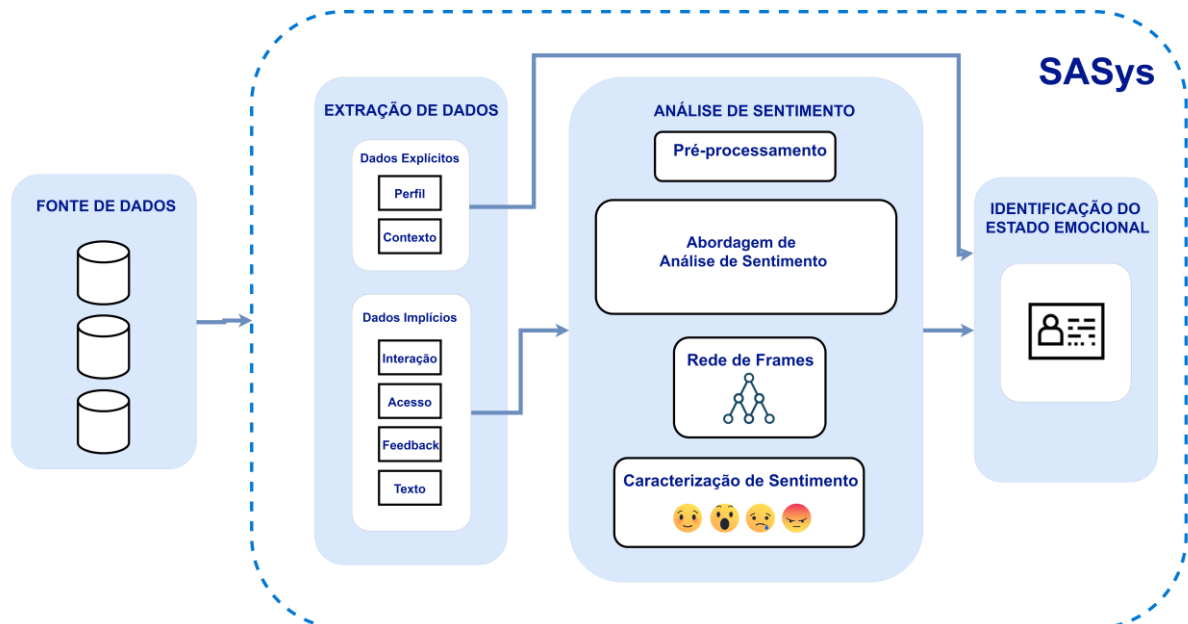
Figura 3 – Fluxo de detecção de estado emocional.



## 4.1 ARQUITETURA CONCEITUAL

A SASys é uma arquitetura composta por 3 camadas: extração de dados, análise de sentimento e identificação do estado emocional. A camada de extração de dados é encarregada de coletar dados, que podem ser provenientes de diversas fontes. A estratégia de AS é definida na segunda camada, onde técnicas de pré-processamento e algoritmos de AS são empregados, além do uso da rede de *frames* na caracterização do sentimento. A definição do perfil emocional é feita na última camada. A Figura 4 apresenta a SASys com seus componentes que serão detalhados nas seções que se seguem.

Figura 4 – Arquitetura conceitual SASys



### 4.1.1 Extração de dados

Esta camada é responsável por identificar os dados do autor do texto e coletar os mais relevantes para a AS e composição do perfil emocional. Ela engloba diferentes tipos de dados com o objetivo de enriquecer a análise.

Os dados coletados podem ser explícitos e implícitos. Os dados explícitos englobam dados de perfil como nome, idade e idioma, e dados de contexto que se referem à localização, dispositivo usado, entre outros. Estes dados são fornecidos pelo autor do texto de forma direta (através de um formulário, por exemplo) ou de forma indireta (como por exemplo, disponibilizado em uma rede social). Os dados implícitos são os dados que carregam maior

valor semântico. Eles podem ser dados de interação, acesso, *feedback* e textos. Os dados de interação se referem a dados qualitativos como: a página web mais acessada, dia de semana, entre outros. Os dados de acesso dizem respeito a dados quantitativos como número de vezes que visitou uma página web, horário e tempo dispensado no acesso. O *feedback* é a expressão do sentimento do autor usando recursos que não sejam texto, como por exemplo o *like*, *amei* ou *seguir* que são recursos de redes sociais que, geralmente, indicam que o usuário concorda com o conteúdo disponibilizado. Os textos são as informações utilizadas para a extração do sentimento que, quando agregados aos outros dados coletados, geram o estado emocional do autor do texto.

A variedade de dados propicia o enriquecimento da análise através de informações que não são textos, mas, que carregam valor semântico. A combinação de duas ou mais fontes de dados também viabiliza o uso de mais de uma abordagem de AS, permitindo que associações sejam feitas com os resultados obtidos, fornecendo assim a condição emocional do autor. A coleta pode ser feita tanto em tempo real como em dados históricos, não interferindo no estado emocional detectado para aquele conjunto de dados. A Tabela 4 exemplifica a diversidade de dados que podem ser usados pela SASys.

Tabela 4 – Exemplos de tipos de dados coletados na SASys

<b>Perfil</b>	<b>Contexto</b>	<b>Interação</b>	<b>Acesso</b>	<b>Feedback</b>	<b>Texto</b>
Nome	Localização	Páginas acessadas	Número de páginas acessadas	Curtir	Comentários
Idade	Dispositivos	Dia da semana	Número de acessos por dia	Seguir	Respostas
Idioma	Tópicos de interesse	Horário de visita	Número de tópicos	Inscrever	Postagens
Gênero	Preferências (mídias, notícias, política)	Tópico mais abordado	Tempo dispensado	Compartilhar	Perguntas

#### 4.1.2 Análise de Sentimentos

Depois de coletar os dados, é necessário o uso de alguma técnica de pré-processamento, antes de submeter os dados a qualquer tipo de solução de AS. Textos retirados da web podem

conter ruídos e partes sem valor semântico como *tags* HTML, *scripts* e publicidade (HADDI; LIU; SHI, 2013). O pré-processamento é a primeira etapa da AS e é empregado para limpar o texto inicial de dados não relevantes para a análise (HADDI; LIU; SHI, 2013; VIJAYARANI; ILAMATHI; NITHYA, 2015). Ele tem o propósito de modelar os dados de acordo com a entrada do algoritmo usado, além de reduzir a inconsistência, viabilizando o aumento da acurácia do método (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014).

As etapas de pré-processamento mais comuns incluem remoção de *tags*, espaços em branco e *stop words*, expansão de abreviações, identificação de palavras em caixa alta, de sinais de exclamação e interrogação, negações e *emoticons*, *stemming*, *lemmatization* e *TF-IDF* (*Term Frequency - Inverse Document Frequency*) (HADDI; LIU; SHI, 2013; VIJAYARANI; ILAMATHI; NITHYA, 2015).

As *stop words* são as palavras que mais se repetem no texto e, por este motivo, são removidas. Elas, normalmente, não são palavras-chave para a análise e sua eliminação no texto diminui de forma significativa o vocabulário usado nas soluções de AS (VIJAYARANI; ILAMATHI; NITHYA, 2015). As *stop words* mais comuns são artigos, preposições e pronomes.

O *Stemming* consiste em reduzir a palavra a sua raiz. É usado com o objetivo de remover afixos das palavras, reduzindo o vocabulário do texto e economizando tempo na análise (VIJAYARANI; ILAMATHI; NITHYA, 2015). Por exemplo, as palavras *infeliz*, *infelizmente*, *felizes* e *felizardo* seriam todas reduzidas à palavra *feliz*. Contudo, a remoção brusca dos prefixos e sufixos podem gerar resultados não tão bons, como é o caso da palavra *felicitações*, que seria reduzida a *feli*.

O *Lemmatization* é o processo de reduzir a palavra ao seu lema. Ao contrário do *stemming*, que encolhe bruscamente as palavras até um tamanho predefinido, o *lemmatization* transforma plurais em singulares e flexões verbais no infinitivo do verbo (CAMBRIA et al., 2017). Por exemplo, as palavras *louvado*, *louváveis* e *louvar-se* seriam reduzidas ao lema *louvar*.

O *TF-IDF* (*Term Frequency - Inverse Document Frequency*) é a medida estatística que revela a importância da palavra em relação ao conjunto de textos coletados. O valor do *TF-IDF* de um termo aumenta proporcionalmente à quantidade de vezes que ele aparece na base de palavras (VIJAYARANI; ILAMATHI; NITHYA, 2015). Ele é bastante usado para identificação de *stop words*.

A escolha da abordagem de AS a ser utilizada pode ser influenciada pelo tipo de dado coletado, custo, rapidez e acurácia do método. Segundo (HEMMATIAN; SOHRABI, 2017)

os algoritmos de aprendizado de máquinas são rápidos e tendem a possuir acurácias elevadas; todavia, necessitam de envolvimento humano, detêm custos elevados devido à carência de conjuntos de treinamento suficientemente grandes e o tempo de treinamento dos algoritmos. Já os métodos lexicais não requerem envolvimento humano, são dependentes do tamanho do léxico, têm acurácia limitada, porém apresentam baixo custo de implementação.

Todas as abordagens podem ser usadas em conjunto com a rede de *frames* para a determinação do sentimento geral do texto. Dado o texto a ser analisado, as frases são passadas à rede de *frames* polarizada, que identifica os *frames* evocados pelas *unidades lexicais* e retorna suas polaridades: *positiva*, *negativa* ou *neutra*. A saída da rede de *frames* pode ser usada como parâmetro de entrada do método de AS escolhido.

Para as abordagens de aprendizado de máquina, as *unidades lexicais* podem ser substituídas por suas polaridades na modelagem dos algoritmos e o sentimento geral do texto será determinado pelo próprio método. Nas abordagens lexicais, a rede de *frames* se comporta como o léxico a ser consultado, onde as palavras são as *unidades lexicais* que evocam os *frames* da frase e o sentimento é a polaridade do *frame*. O sentimento geral pode ser calculado tendo em conta os sentimentos individuais das *unidades lexicais* e os transformadores de valência presentes nas frases. Os transformadores de valência são termos que alteram a polaridade da palavra mais próxima que possua sentimento (TABOADA et al., 2011). Eles podem ser intensificadores (como por exemplo *muito*, *absolutamente*, *demasiado*, sinais de exclamação e interrogação repetidos), redutores (*pouco*, *menos*, *quase*, *apenas*) e negações (*não*, *jamais*, *nunca*).

#### **4.1.3 Identificação do estado emocional**

A identificação da polaridade do texto é uma informação relevante que, combinada com outros dados, pode revelar o estado emocional do autor e até traçar um possível perfil emocional do mesmo. Além de saber como o autor se sentia enquanto escrevia o texto, distinguir a sua idade, gênero, localização, condição social, horário que escreveu, com quem interagiu, dispositivo que usou, entre outros, pode agregar valor semântico à análise (SHAPIRO et al., 2017). Usuários de redes sociais, por exemplo, podem influenciar ou serem influenciados pelos sentimentos de outros e esse contágio emocional pode ser utilizado para formação de tendências nos negócios e marketing, liderança em corrente política e conscientização de problemas sociais e ambientais (BAMAKAN; NURGALIEV; QU, 2019).

Na arquitetura proposta, a detecção do estado emocional é feita depois de gerado o sentimento geral do texto, usando também os dados de perfil e de contexto. Esses dados podem ser submetidos a métodos estatísticos e o resultado obtido, combinado com a AS, fornecem o *status* do autor, no momento da escrita, que pode ser utilizado para compor o perfil do usuário em uma aplicação ou até para recomendação de material personalizado.

## 4.2 ESTUDO OBSERVACIONAL

Para avaliar a viabilidade da arquitetura SASys, foi desenvolvido um estudo observacional com o objetivo de medir a acurácia do uso da rede de *frames* da FrameNet como uma abordagem lexical para a AS.

O primeiro esforço de implementação da SASys foi feito desenvolvendo a rede polarizada de *frames* e usando uma base<sup>5</sup> de dados com avaliações de produtos, filmes e restaurantes (KOTZIAS et al., 2015). A camada de Análise de Sentimento foi o foco de implementação nesta fase, deixando as camadas de Extração de Dados e de Identificação do Estado Emocional para serem avaliadas, posteriormente, no estudo de caso.

A base de dados escolhida é composta por 3.000 frases rotuladas, retiradas de três grandes websites: Amazon<sup>6</sup>, IMDb<sup>7</sup> e YELP<sup>8</sup>. Ela contém três arquivos textos em inglês, cada um correspondendo ao respectivo website, contendo 500 frases positivas e 500 frases negativas.

Para a avaliação de desempenho da rede polarizada de *frames* foi escolhido o arquivo do YELP, uma base de dados de avaliações de restaurantes. O arquivo possui comentários curtos, escritos na língua inglesa, onde os usuários procuraram descrever de forma sucinta suas experiências nos restaurantes que visitaram. As frases são apenas positivas ou negativas, não possuindo comentários rotulados como neutros. As frases não passaram por nenhum tipo de pré-processamento (como *stop words*, *lemmatization* ou *stemming*) e não foram traduzidas para o Português, pois a FrameNet é uma ferramenta multilíngue.

O arquivo no formato texto foi transformando no formato *csv* (*comma-separated values*), contendo apenas as frases e os rótulos no formato numérico: 0 para negativo e 1 para positivo. Foi selecionada uma amostra aleatória de 50 frases do arquivo para a base de teste (correspondendo a 5% da base), sendo 25 frases rotuladas com polaridade positiva e 25 com

---

<sup>5</sup> <https://www.kaggle.com/marklvl/sentiment-labelled-sentences-data-set/data>

<sup>6</sup> <https://www.amazon.com/>

<sup>7</sup> <https://www.imdb.com/>

<sup>8</sup> <https://www.yelp.com/>

polaridade negativa. A Tabela 5 apresenta algumas das frases encontradas na base de dados do YELP (no Apêndice A se encontra o conjunto de frases selecionadas, as *unidades lexicais* identificadas e as polaridades geradas).

Tabela 5 – Exemplo de frases contidas na base de dados do YELP. Sentimento positivo=1, sentimento negativo=0

Frases	Rótulo
Wow... Loved this place	1
Crust is not good	0
Not tasty and the texture was just nasty	0
Stopped by during the late May bank holiday off Rick Steve recommendation and loved it	1
The potatoes were like rubber and you could tell they had been made up ahead of time being kept under a warmer	0

As *unidades lexicais* foram identificadas e passadas à rede de *frames* polarizada que detectou os *frames* e retornou as suas respectivas polaridades. Assim sendo, dado o conjunto de frases  $F=\{f_0, \dots, f_m\}$ , onde  $m$  é o total de frases a serem analisadas, para todo  $f_i$  foram extraídas as *unidades lexicais*, gerando o conjunto  $UL(f_i)=\{ul_0, \dots, ul_n\}$ , que foi passado a rede de *frames* que, por sua vez, retornava o sentimento associado ao *frame* que a *unidade lexical* evocava, onde  $n$  é o número de unidades lexicais da frase  $i$ .

Nesta fase, as polaridades foram apenas somadas, gerando o sentimento geral de cada frase, conforme a Equação (1), onde  $P(f_i)$  é a polaridade da frase  $f_i$ , e  $RF(ul_j)$  é a polaridade individual da *unidade lexical*  $ul_j$  detectada pela rede de *frames*. Os resultados obtidos são apresentados, de forma detalhada, na seção 4.2.4.

$$P(f_i) = \sum_{j=0}^{|UL(f_i)|} RF(ul_j) \quad (1)$$

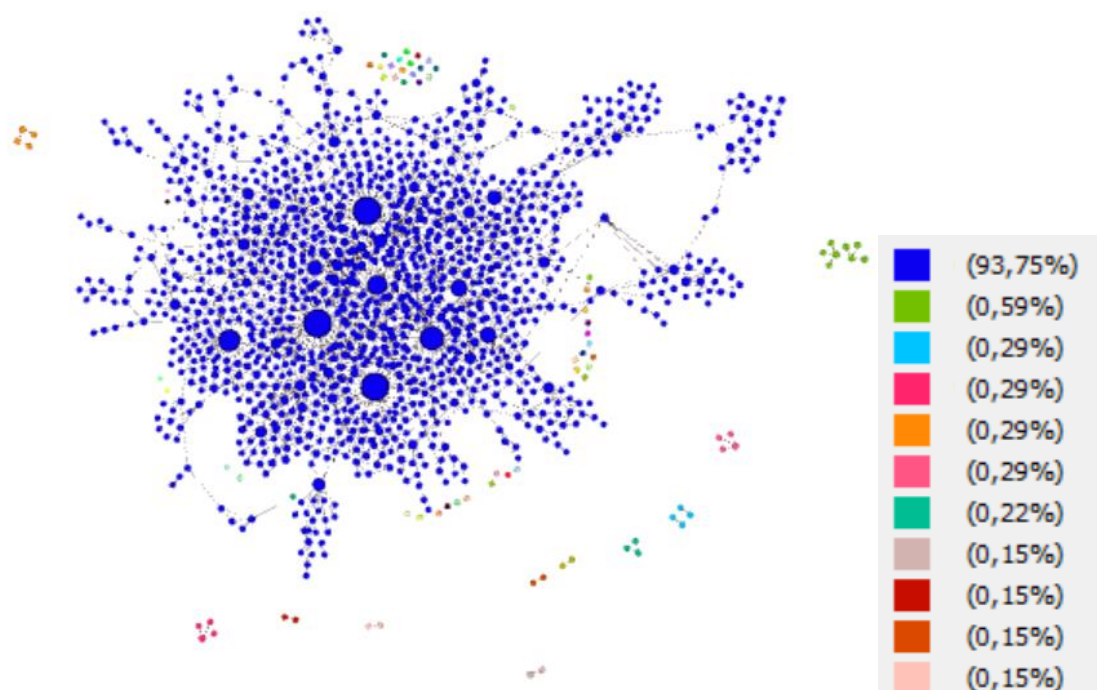
#### 4.2.1 Rede de *frames*

A rede de *frames*, desenvolvida para este trabalho, é um grafo  $G$  formado pelos *frames* e relacionamentos da FrameNet Brasil, denotado pelo par  $G\{V,E\}$ , onde  $V=\{v_1, v_2, \dots, v_n\}$  é o conjunto de nós ou vértices e  $E=\{e_1, e_2, \dots, e_m\}$  é o conjunto de arestas. O grafo é dirigido, pois

os relacionamentos indicam uma certa hierarquia na rede, e contém 1.359 vértices e 1.960 arestas.

Existem *frames* mais genéricos que possuem muitos *frames* conectados a eles, gerando assim núcleos bem definidos no grafo. Esses mesmos *frames* podem ser enxergados como pais das árvores geradas por suas conexões, onde quanto maior o nível do nó na árvore mais específico é o *frame*. Assim sendo, o grafo da FrameNet pode ser considerado como uma floresta de *frames* em que: os pais das árvores descrevem cenários mais gerais, os nós folha cenários mais específicos, os vértices podem se relacionar uns com os outros como irmãos e a altura da árvore pode ser um indicativo de quão bem estudado foi o assunto em questão. A Figura 5 ilustra o grafo da FrameNet.

Figura 5 – Grafo da rede de *frames* da FrameNet. O tamanho do vértice foi determinado pelo seu grau. As cores assinalam as componentes conexas.



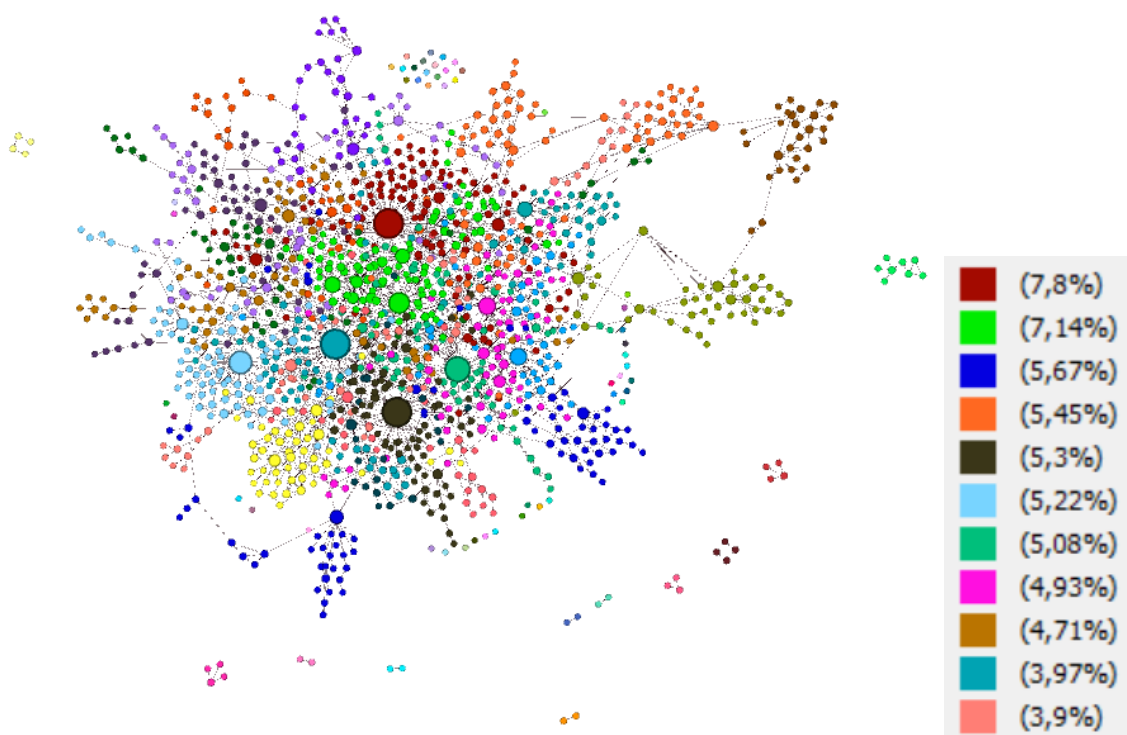
Foram aplicados, à rede de *frames*, algumas métricas de redes complexas com o objetivo de caracterizar a rede e identificar possíveis anomalias em sua estrutura. Dois algoritmos de agrupamento foram usados e as análises feitas a partir deles são descritas a seguir. Para analisar a conectividade da rede considerou-se a detecção de Componentes Conexas e de Comunidades, utilizando algoritmos fornecidos pelo *Gephi*<sup>9</sup>.

<sup>9</sup> <https://gephi.org/>



Para detecção de componentes conexas aplicou-se o algoritmo *Componentes Conexas* (WASSERMAN; FAUST, 1994). Foram identificadas 60 componentes conexas, sendo que a maior engloba mais de 90% dos vértices (1.274 *frames*). Nessa componente se encontram os vértices com maior grau e as árvores de *frames* mais altas. Das 59 componentes restantes 48 possuem apenas um vértice, o que indica que alguns *frames* estão sem relacionamentos na FrameNet. Estes *frames* representam cenários muito específicos ou que foram pouco estudados pelos linguistas, por isso se encontram desconectados do restante da rede. A Figura 5 ilustra as componentes do grafo, onde as cores distinguem os vértices pertencentes à mesma componente.

Figura 6 – Comunidades do grafo.



Para detecção de comunidades aplicou-se o algoritmo de *Modularidade*, que consiste em agrupar os vértices de acordo com o valor da modularidade da partição (BLONDEL et al., 2008) (usou-se o valor de resolução = 10 como parâmetro de entrada do algoritmo no *Gephi*). Foram detectadas 82 comunidades no grafo. Estas comunidades são grupos com ligações muito densas entre os vértices participantes. Os dois maiores grupos contêm dois dos vértices de maior grau e possuem mais de 90 membros, respectivamente. Como pode ser observado na Figura 6, os vértices desconexos constituem também comunidades compostas de menos de 10 *frames*.

Medidas de centralidade revelam a posição de um nó em relação aos outros e ao próprio grafo (COSTENBADER; VALENTE, 2003). Elas foram aplicadas na principal componente conexa do grafo, com o objetivo de detectar os vértices de maior importância no grafo. A Tabela 6 lista essas medidas e o intervalo em que seus valores ocorrem.

Tabela 6 – Medidas de centralidade e o intervalo de valores. O *Closeness* possui os valores contidos entre [0,1]

<b>Intervalo</b>	<b>Grau</b>	<b>Grau de Entrada</b>	<b>Grau de Saída</b>	<b>Closeness</b>	<b>Betweenness</b>
Menor	1	0	0	0.10858	0.0
Maior	54	6	53	1.0	2950.5

O grau do nó indica a quantidade de relacionamentos que o *frame* possui. Por ser um grafo dirigido, o valor do grau é a soma do grau de entrada e o de saída. O grau de saída foi usado para revelar vértices que poderiam apresentar uma certa relevância no grafo, por possuírem valores maiores que o grau de entrada (como pode ser verificado na Tabela 5). Os vértices com maior grau de saída são os que possuem muitos relacionamentos de *herança* e que, por essa razão, podem ser considerados como vértices críticos, visto que a sua remoção implicaria na fragmentação da rede de *frames*, bem como do grafo.

O *Closeness* está associado à posição do nó na rede e a propagação de sentimento. Ele mede a quantidade de saltos que um nó dá, em média, para alcançar todos os restantes no grafo, ou seja, quão central o nó é em comparação com os demais (COSTENBADER; VALENTE, 2003). Vértices com *Closeness* alto são, geralmente, adjacentes a um número considerável de nós (FREEMAN, 1978), porém, na rede de *frames* o inverso acontece, ou seja, os vértices com um grau elevado são os que possuem o *Closeness* baixo. Por ser um caso atípico à realidade computacional, essa medida demandou uma análise mais acentuada, por parte de um especialista que verificou o valor semântico dos *frames* e os respectivos valores do *Closeness*. Não foi possível verificar uma relação que explicasse tais ocorrências, porém constatou-se que ela não representa risco semântico à FrameNet.

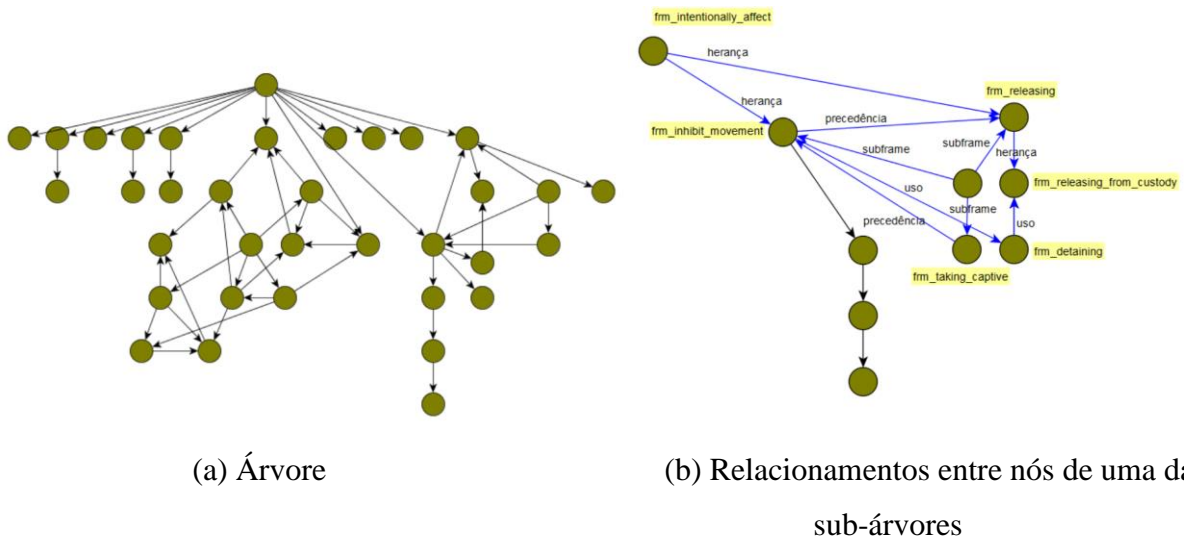
O *Betweenness* está associado ao fluxo de informação e detecta gargalos na rede. Ele calcula a frequência em que um nó aparece no caminho geodésico de pares de nós do grafo (FREEMAN, 1978), ou seja, ele reconhece os nós *ponte* do grafo (COSTENBADER; VALENTE, 2003). Na rede de *frames* é possível encontrar um número considerável de nós com valores altos, inclusive, treze deles possuem *Betweenness* maior que 1.000. Esses vértices são os que ligam grandes núcleos da rede e por isso aparecem em muitos caminhos

geodésicos da rede. Ou seja, eles são pontes entre árvores da rede de *frames*. Esta medida se revelou relevante tanto a nível topológico quanto semântico, pois, a remoção de nós com *Betweenness* alto implicaria em perda de informação na rede física e em perda de significado no *frame*.

#### 4.2.2 Caracterização da rede de *frames*

A estrutura de dados que mais se aproxima da rede de *frames* é a árvore, apesar de suas peculiaridades. Computacionalmente, os relacionamentos entre os *frames* violam as restrições de uma árvore: filhos com mais de um pai, irmãos com ligações diretas, ciclos entre nós de uma sub-árvore, entre outros. Porém, semanticamente essas infrações podem fazer sentido para a descrição do assunto específico. A Figura 7(b) ilustra um exemplo de *frames* irmãos se relacionando sem passar, necessariamente, pelo *frame* pai: os *frames* *inibir\_o\_movimento* e *liberar\_herdam* do *frame* *afetar\_intencionalmente*, porém eles se ligam pela relação de *precedência* que define a ordem em que os fatos ocorrem.

Figura 7 – Visualização em árvore de *frames* de uma das componentes conexas do grafo.



As medidas e os algoritmos apresentados foram empregados com o objetivo de caracterizar o grafo da FrameNet e revelar pontos que favorecessem o uso da mesma na AS. Constatou-se alguns casos incomuns a nível topológico que foram encaminhados a um dos linguistas da FrameNet, contudo, verificou-se que os mesmos fazem sentido semanticamente e não oferecem nenhum risco no que tange ao uso da rede de *frames* como um método de AS.

Os algoritmos de detecção de componentes conexas e de comunidades revelaram um subgrafo contendo 90% dos vértices ligados, majoritariamente, pelo relacionamento de *herança*. Esta informação foi usada na geração da rede polarizada de *frames* que é descrita na seção a seguir.

### 4.2.3 Rede de *frames* polarizada

A rede *frames* de polaridade foi criada utilizando os *frames*, *unidade lexicais* e o relacionamento de *herança*, por ser o maior da rede, por conectar grande parte dos nós, por permitir a transitividade e por carregar informações semânticas. A transitividade foi usada na propagação do sentimento na rede.

Figura 8 – Tela da *WebTool* da FrameNet, onde o *frame* abandono foi selecionado.

The screenshot shows the FrameNet WebTool interface. On the left, a tree view lists various frames, with 'Abandono' selected. The main area displays the details for the 'Abandono' frame, including its definition, examples, core elements (FE Core), and lexical units. The definition states: 'Um **Agente** deixa para trás um **Tema** efetivamente tornando-o fora de seu controle ou da segurança normal como propriedade da pessoa.' The core elements include 'Agente (Agent)' and 'Tema (Theme)'. The lexical units listed are 'abandonado.a', 'abandonar.v', 'abandono.n', 'deixar.v', and 'esquecer.v'.

Fonte: *WebTool*

Inicialmente definiu-se, de forma manual, o sentimento de algumas *unidades lexicais* que foram as sementes do algoritmo de *Label Propagation* (ZHU; GHAHRAMANI; LAFFERTY, 2003) utilizado na proliferação das polaridades na rede. As *unidades lexicais* foram escolhidas como detentoras iniciais de sentimento por serem o pareamento entre o *frame* e um dos seus significados. O sentimento atribuído a cada *unidade lexical* foi extraído da descrição do *frame* que ela evoca, encontrada na *WebTool*<sup>10</sup>, que é o sistema web de

<sup>10</sup> <http://webtool.framenetbr.ufjf.br/index.php/webtool/report/frame/main>

gerenciamento de *frames* da FrameNet. Por exemplo, a *unidade lexical* *deixar* possui sentimento negativo quando evoca o *frame* *abandono* (Figura 8).

O conjunto de dados ainda é um problema na área de AS por ser difícil de ser encontrado, dispendioso ou demorado de ser obtido e, muitas vezes, requer envolvimento de especialistas (HEMMATIAN; SOHRABI, 2017). Por esse motivo foi definida uma amostra aleatória simples de 266 *unidades lexicais*, que representa 0,2% do total, que foi usada como *semente* do algoritmo.

O algoritmo *Label Propagation* é um algoritmo de aprendizado de máquina semi-supervisionado que foi o escolhido por exigir menos esforço humano no que tange a quantidade inicial de *unidades lexicais* com sentimento requeridas pelo algoritmo. O algoritmo, como o próprio nome diz, é um propagador de rótulos e trata do problema de classificação de nós. Ele é empregado em problemas onde não existe um conjunto inicial grande de nós rotulados, mas é possível usar informações topológicas para prever as etiquetas restantes do grafo. Ele parte do princípio de que nós conectados deveriam ser parecidos e usa informações da estrutura da rede para determinar o grau de similaridade entre eles.

Nesse trabalho foram definidas três polaridades: *positiva* - palavras que evocam um sentimento positivo, *negativa* - palavras que evocam um sentimento negativo, e *neutra* - palavras que não fazem um julgamento prévio. Elas foram estabelecidas nas *unidades lexicais* e se propagaram pelos *frames*, fazendo o caminho inverso dos seus relacionamentos. A polaridade neutra foi usada porque a FrameNet possui uma quantidade considerável de *unidades lexicais* que evocam cenários sem sentimentos. A Tabela 7 apresenta exemplos de *unidades lexicais* e sentimentos encontrados na rede polarizada obtida.

Tabela 7 – Exemplos de polaridades encontradas na rede de *frames*

<b>Unidade Lexical</b>	<b>Frame evocado</b>	<b>Polaridade</b>
Embelezado	Abundante em	Positiva
Sabedoria	Consciência	Positiva
Empoeirado	Abundante em	Negativa
Horrendo	Estética	Negativa
Pousada	Acomodação	Neutra
Frio	Temperatura ambiente	Neutra

O algoritmo foi implementado utilizando a linguagem de programação Python e a biblioteca *NetworkX*<sup>11</sup> que permite criar, manipular e usar funções de redes complexas. Depois de obtida a rede polarizada, a mesma foi armazenada no banco de dados em grafo *Neo4j*<sup>12</sup>, que permite consultas em grafos.

#### 4.2.4 Resultados obtidos

Dado o conjunto de 50 frases rotuladas, a AS foi feita identificando *unidades lexicais* contidas nas frases, que foram passadas à rede polarizada de *frames*, que por sua vez retornou o sentimento geral do texto (Equação 1). A Tabela 8 exemplifica o processo de detecção do sentimento geral do texto, onde a primeira coluna apresenta uma amostra do conjunto de frases usadas no estudo observacional, a segunda coluna mostra as *unidades lexicais* detectadas em cada frase com as respectivas polaridades dos *frames* que elas evocam (os *frames* foram omitidos nesta tabela) e a última coluna exhibe o sentimento geral do comentário.

Tabela 8 – Detecção do sentimento geral das frases da base de dados do YELP

<b>Frases</b>	<b>Unidades lexicais detectadas (polaridade individual)</b>	<b>Polaridade final detectada pela rede de frames</b>
Great brunch spot.	great ( <i>positiva</i> )	Positiva
But now I was completely grossed out.	completely ( <i>positiva</i> )	Positiva
The pizza selections are good.	good ( <i>positiva</i> )	Positiva
I've had better bagels from the grocery store.	better ( <i>positiva</i> )	Positiva
I think not again	not ( <i>negativa</i> )	<u>Negativa</u>
Our server was very nice, and even though he looked a little overwhelmed with all of our needs, he stayed professional and friendly until the end.	nice ( <i>positiva</i> ), friendly ( <i>positiva</i> )	Positiva
Also there are combos like a burger, fries, and beer for 23 which is a decent deal.	decent ( <i>positiva</i> )	Positiva
The selection was probably the worst I've seen in Vegas.....there was none.	probably ( <i>neutra</i> ), worst ( <i>negativa</i> )	Negativa
I think this restaurant suffers from not trying hard enough.	not ( <i>negativa</i> ), hard ( <i>negativa</i> ), enough ( <i>neutra</i> )	Negativa

<sup>11</sup> <https://networkx.github.io/>

<sup>12</sup> <https://neo4j.com/>

Os critérios de avaliação comumente usados na área de mineração de opinião para mensurar a eficiência da solução proposta são: acurácia, precisão, *recall* e *F1-score* (DEY; JENAMANI; THAKKAR, 2018; HEMMATIAN; SOHRABI, 2017). Eles se baseiam em quatro medidas:

- *TP* – *True Positive (verdadeiro positivo)* – que se refere à quantidade de frases positivas que foram classificadas corretamente;
- *TN* – *True Negative (verdadeiro negativo)* – que diz respeito à quantidade de frases negativas classificadas corretamente;
- *FP* – *False Positive (falso positivo)* – que se refere à quantidade de frases negativas que foram classificadas, incorretamente, como positivas;
- *FN* – *False Negative (falso negativo)* – que diz respeito à quantidade de frases positivas que foram classificadas, incorretamente, como negativas.

A partir dessas medidas é gerada a matriz de confusão e calculados os critérios de avaliação, que estão descritos nas Tabelas 9 e 10. Na matriz de confusão, a coluna *rótulo* corresponde a polaridade correta das frases e a coluna *rede de frames* diz respeito a classificação feita pela rede polarizada de *frames*. Cada célula da matriz corresponde ao valor de uma das medidas (Tabela 9 (b)), e os critérios de avaliação são calculados com os resultados obtidos.

Tabela 9 – Matriz de confusão do estudo observacional. Linha: rótulo, Coluna: rede de *frames*

Tabela97 (a) – Matriz de confusão modelo			Tabela 9 (b) – Matriz de confusão com valores reais		
Rótulo	Rede de Frames		Rótulo	Rede de Frames	
	Positiva	Negativa		Positiva	Negativa
Positiva	<i>TP</i>	<i>FN</i>	Positiva	24	1
Negativa	<i>FP</i>	<i>TN</i>	Negativa	6	19

A acurácia indica a quantidade de acertos que a abordagem teve em relação ao conjunto de dados, ou seja, revela a quantidade de frases que foram classificadas corretamente. Ela é a razão entre a quantidade de frases identificadas de forma acertada e o

conjunto de frases da base (HEMMATIAN; SOHRABI, 2017) (Equação 2). A rede de *frames* classificou 43 frases de forma correta, sendo que teve maior facilidade em identificar a polaridade positiva do que a negativa.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

A precisão mede a exatidão da abordagem (DEY; JENAMANI; THAKKAR, 2018). Ela é a razão entre a quantidade de frases positivas classificadas corretamente e o conjunto de frases identificadas de forma correta pela abordagem (HEMMATIAN; SOHRABI, 2017) (Equação 3). Como já foi mencionado, a rede de *frames* conseguiu maior acerto nas frases de polaridade positiva, categorizando de forma acertada 24 frases das 25 dessa classe. De forma geral, a rede obteve uma precisão de 80%.

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3)$$

O *Recall* mede a completude da abordagem (DEY; JENAMANI; THAKKAR, 2018). Ele é a fração de frases positivas classificadas corretamente que podem ser recuperadas da base de dados (Equação 4). A rede de *frames* polarizada alcançou um *recall* de 96%.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

O *F1-score* é a média harmônica da precisão e do *recall* (DEY; JENAMANI; THAKKAR, 2018) (Equação 5). A rede de *frames* obteve um *F1-score* de 87%.

$$F1 - score = 2 * \frac{\text{Precisão} * \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (5)$$

Como pode ser verificado na Tabela 8, a rede de *frames* obteve um desempenho aceitável para uma abordagem lexical (HEMMATIAN; SOHRABI, 2017). Ela utilizou a metodologia *unigram* para a obtenção do sentimento geral das frases, ou seja, não usou nenhuma informação de dependência entre as palavras. Apesar da abordagem simplificada e da quantidade limitada de frases usadas no teste inicial, os resultados obtidos motivaram o uso de *features* que pudessem melhorar o desempenho da rede polarizada em contextos reais com frases longas, como, por exemplo, o uso de transformadores de valência.



Tabela 10 – Tabela de critérios de avaliação do estudo observacional

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
86	80	96	87

Os resultados obtidos nos critérios de avaliação deste estudo observacional responderam a primeira questão de pesquisa secundária desta dissertação (*QPS1: É possível usar a FrameNet como uma solução lexical para a AS?*), pois os mesmos comprovam que a FrameNet pode ser usada como uma solução de Linguística Computacional para a AS e consequentemente como uma solução lexical.

#### 4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

O presente capítulo apresentou a arquitetura SASys, que visa detectar sentimento das frases através do uso de uma rede de *frames* polarizada da FrameNet e identificar o estado emocional do autor de um texto por meio do resultado obtido da AS e dos dados coletados. Ela viabiliza o uso de vários tipos e fontes de dados, permitindo deste modo o enriquecimento da AS e identificando um possível estado emocional da pessoa que escreve o texto, que pode ser usado tanto para composição de perfis de aplicações externas quanto para recomendação de material especializado.

A rede de *frames* usada é um grafo composto por elementos da FrameNe: *frames*, *elementos de frames* e *unidade lexicais* como vértices e o relacionamento de *herança* como arestas. Este relacionamento foi escolhido por carregar informações semânticas da rede, além de possuir a propriedade da transitividade. O grafo é direcionado e informações da estrutura da rede foram usadas na propagação do sentimento, através do algoritmo de *Label Propagation*.

Foi dirigido um estudo observacional da AS por meio da rede polarizada de *frames* visando constatar a potencialidade da FrameNet como uma abordagem lexical, e os resultados obtidos responderam a primeira questão de pesquisa secundária desta dissertação (*QPS1*).

Enxergou-se a necessidade de avaliar a arquitetura SASys em um ambiente mais abrangente, que pudesse validar todas as suas camadas e, por esse motivo, ela foi submetida a um estudo de caso, com dados de contexto real, que é descrito no capítulo que se segue.

## 5 ESTUDO DE CASO

Segundo (WOHLIN et al., 2012) existem três estratégias de investigação, aplicadas de acordo com o propósito da avaliação: *surveys*, estudo de caso e experimento. O estudo de caso é mais adequado para a avaliação da SASys, por se tratar de uma pesquisa empírica, realizada em um contexto real, que deseja investigar a adequação da arquitetura ao modelo de extração de dados e de identificação de estado emocional proposto.

O estudo de caso permite descrições detalhadas dos fenômenos, normalmente baseados em fontes de dados diversas, e assegura que a investigação e o entendimento do problema sejam feitos em profundidade (DRESCH; LACERDA; JÚNIOR, 2015). A formalização do estudo de caso contribui para a avaliação das questões de pesquisa formuladas, e verificação do artefato desenvolvido.

Para a avaliação da SASys, foi escolhido o cenário de Educação a Distância (EaD) por ser uma área de aplicação da AS em ascensão (ALTRABSHEH; COCEA; FALLAHKHAIR, 2014; BÓBÓ et al., 2019; RANI; KUMAR, 2017; SHAPIRO et al., 2017). Existe um apelo forte para o aumento da qualidade da educação e alguns estudos têm mostrado que informações sobre o estado emocional do aluno são importantes, pois impactam o desempenho acadêmico de forma direta (FENG et al., 2016; LEI et al., 2015; PONG-INWONG; RUNGWORAWUT, 2014).

O *feedback* dos alunos tem sido usado em Ambientes Virtuais de Aprendizagens (AVAs) com o propósito de melhorar a experiência dos mesmos (SHAPIRO et al., 2017). Para tal, os dados podem ser coletados em questionários (PONG-INWONG; RUNGWORAWUT, 2014), comentários em fóruns (CUMMINS; BURD; HATCH, 2010; SHAPIRO et al., 2017) e *logs* de acesso (SPATIOTIS et al., 2016); e técnicas de AS podem ser empregadas visando auxiliar no processo de aprendizado personalizado, que supram necessidades pontuais dos alunos.

### 5.1 DEFINIÇÃO DO ESCOPO

O estudo de caso foi conduzido no contexto de EaD (Ensino a Distância), com o propósito de avaliar a arquitetura SASys na extração de dados e no reconhecimento do estado emocional do aluno, através do uso da rede de *frames*, além de validar a aplicabilidade da FrameNet como abordagem lexical em mais de um domínio. Ele foi realizado em uma turma da disciplina de Metodologia de Pesquisa Científica e Educacional de um curso da área de

Computação, da Universidade Federal de Juiz de Fora, no primeiro semestre de 2018. A SASys foi instanciada num sistema de recomendação educacional que tem como objetivo manter os alunos do curso motivados, retardando a evasão. Por este motivo, foi necessário a definição de um perfil motivacional, baseado no estado emocional detectado.

O escopo de avaliação foi baseado no método GQM (Goal/Question/Metric), proposto por Basili e Rombach (BASILI; ROMBACH, 1988), e é descrito como:

*“Coletar e analisar textos e acessos de alunos em um AVA, com o objetivo de identificar o perfil motivacional dos alunos através do estado emocional detectado pela SASys, no contexto de EaD”.*

A partir deste escopo foram definidas três questões de pesquisa, apresentadas na Tabela 11 juntamente com suas justificativas.

Tabela 11 – Questões de pesquisa do estudo de caso

	<b>Questão</b>	<b>Justificativa</b>
<b>QPEC1</b>	<i>É possível usar a rede de frames polarizada nos textos dos alunos de uma disciplina on-line para AS?</i>	Esta questão pretende avaliar a aplicabilidade da rede polarizada de frames em diferentes domínios;
<b>QPEC2</b>	<i>É possível determinar o estado emocional do aluno, através dos dados coletados?</i>	Com esta questão pretende-se avaliar todas as camadas da arquitetura SASys;
<b>QPEC3</b>	<i>É possível identificar os alunos com tendência à evasão em uma disciplina on-line?</i>	Esta questão visa evidenciar um dos propósitos do uso da AS no contexto educacional.

A disciplina foi ministrada a 58 alunos com apoio do AVA *Moodle*. As atividades foram passadas semanalmente e executadas em grupo, com um número máximo de três participantes. Esta última característica teve um papel determinante na análise dos dados, pois os grupos eram, majoritariamente, liderados por um único aluno, que se encarregava de fazer o envio das atividades no AVA. Essa informação poderia ser usada como justificativa da baixa interação de alguns alunos com o AVA, visto que já possuíam um colega que os representava; contudo, esta teoria não foi verificada por não ser o foco do estudo de caso.

Cada atividade possuía um fórum de dúvidas específico, que foi usado para a análise de sentimento. As mensagens dos alunos eram agrupadas pelos fóruns e, conseqüentemente,

em semanas, por esse motivo definiu-se janelas de tempo de 7 dias seguidos nas análises de interação com o ambiente de aprendizagem.

A base de dados para a AS consistiu dos comentários coletados dos fóruns no *Moodle*, visto que os alunos não possuíam uma rede social através da qual interagissem sobre o conteúdo da disciplina de Metodologia de Pesquisa Científica e Educacional; ou seja, apesar dos alunos usarem mídias sociais, cada um as utilizava apenas para fins pessoais e não houve unanimidade na eleição de uma mídia que fosse usada pela maioria dos alunos.

Essa base de dados foi rotulada por professores com anos de experiência em EaD com o objetivo de comparar a *expertise* humana em identificar alunos com sentimentos à abordagem proposta neste trabalho. Neste estudo de caso, a atenção foi canalizada no sentimento negativo, pois, pressupõe-se que ele seja um dos elementos que contribui no aumento da propensão à evasão escolar.

A rede polarizada de *frames* foi evoluída para uma versão que usa transformadores de valência. Essa versão foi usada para detecção dos sentimentos dos alunos e, posteriormente, ela passou pela avaliação de um linguista da FrameNet que gerou a sua versão melhorada. Todas etapas do estudo de caso estão descritas nas seções que se seguem.

## 5.2 CONDUÇÃO DO ESTUDO

### 5.2.1 Coleta de dados

A arquitetura foi instanciada tendo em conta as particularidades do estudo de caso (Figura 9). Os dados foram todos coletados dentro do ambiente *Moodle*, utilizando a opção de exportação disponível no AVA. Os alunos responderam a um questionário *on-line*<sup>13</sup> que identificava o dispositivo mais usado para o acesso ao AVA, as suas preferências em relação à forma de apresentação do conteúdo (texto, áudio ou vídeo) e a rede social mais usada. Também foram extraídos os textos dos alunos e dados de acesso ao ambiente de ensino, que foram exportados em planilhas, respectivamente, e que receberam processamentos distintos.

Todos os dados foram coletados na tentativa de enriquecer a análise e reconhecer associações que promovessem descoberta de algum conhecimento como, por exemplo, alunos que acessam o AVA frequentemente e que comentam nos fóruns tendem a ter sentimentos negativos, mas nem todos estão desmotivados em relação ao curso.

---

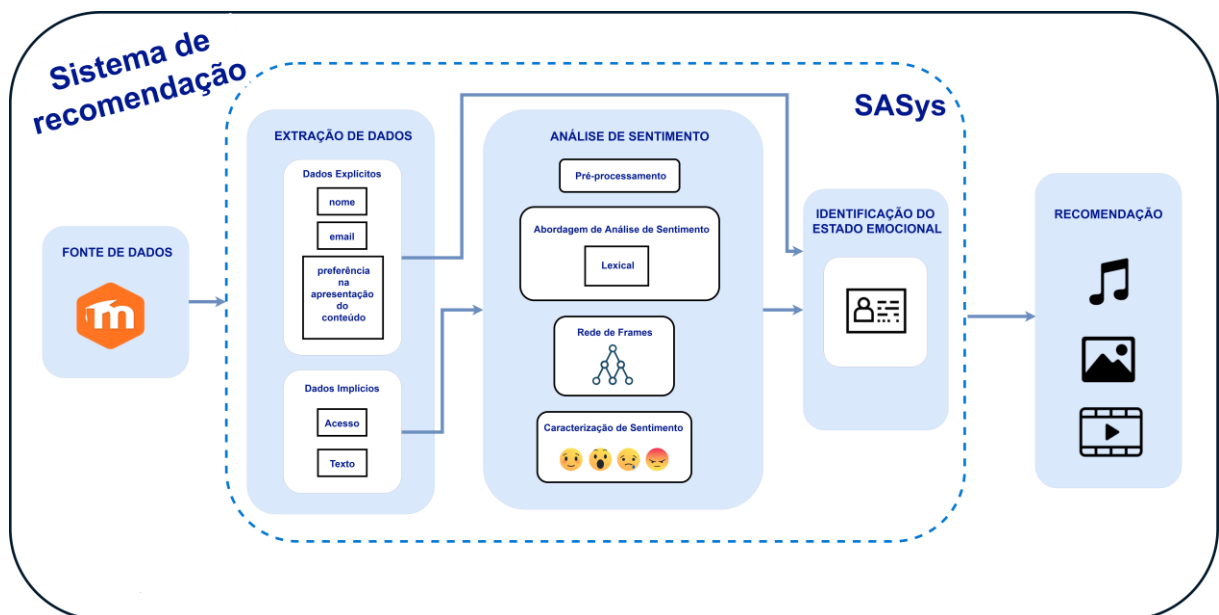
<sup>13</sup> <https://forms.gle/tRZwPmTKqhcWmRxF6>

Não foi usada nenhuma rede social na etapa de coleta de dados, pois não houve um consenso entre os alunos na eleição da mídia social mais usada. Nenhuma das redes sociais elencadas por eles continha um grupo ou lugar virtual formado por eles, onde os mesmos expusessem assuntos relacionados ao curso. Neste contexto, a coleta de dados deveria ser feita em mais de uma rede social e a AS seria baseada em aspectos, que não é o foco deste trabalho.

### 5.2.2 Análise de Sentimento

Os fóruns de dúvidas de cada atividade foram analisados a fim de identificar o estado emocional dos alunos. Eles foram escolhidos por conterem mensagens com teor emocional e por serem o veículo de comunicação entre os alunos, o professor e o tutor. Foram consideradas apenas as frases escritas pelos alunos e ignoradas todas as respostas do professor e tutor da disciplina.

Figura 9 – Arquitetura SASys instanciada para o estudo de caso



Foram coletadas 27 mensagens de 10 fóruns, que constituíam o semestre de aulas da disciplina. Algumas semanas não tiveram atividades por serem semanas de avaliações presenciais.

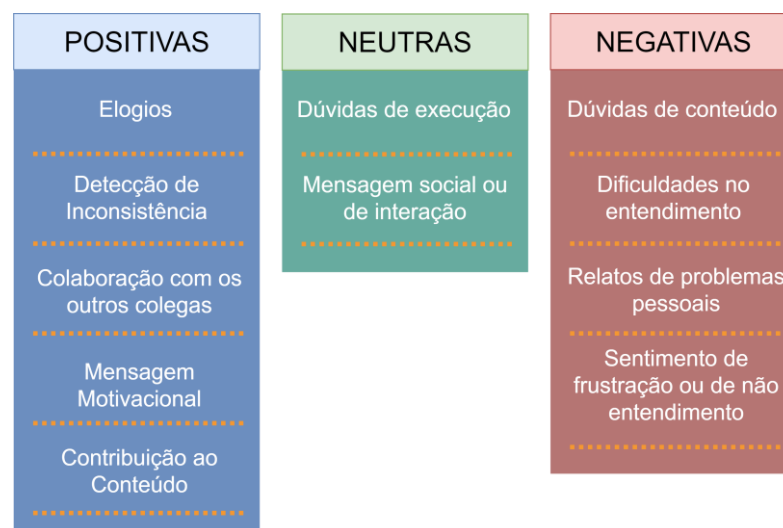
### 5.2.2.1 Rotulação das Mensagens por especialistas em EaD

A rotulação das mensagens ocorreu em duas etapas. Na primeira predominou a *expertise* de dois profissionais da área de educação. Essa classificação foi usada para avaliar o desempenho da rede de *frames* na detecção de sentimento em frases em português. Com a classificação gerada pelos mesmos, foi possível identificar os perfis motivacionais dos alunos.

Na segunda etapa um linguista especialista da FrameNet utilizou de processamento de linguagem natural para sugerir melhorar o desempenho da rede de *frames* (vide seção 5.2.2.3).

Na primeira etapa do estudo de caso, as mensagens foram rotuladas por dois professores, um com mais de 15 anos e outro com 6 anos de experiência em EaD, nas três polaridades: positivas, negativas e neutras. Estes docentes utilizaram sua *expertise* para identificar características, dentro do contexto educacional que qualificassem as frases em uma das três categorias (Figura 10). Segundo os professores, as frases positivas deveriam conter elogios, detecção e correção de inconsistência (correção de algum conteúdo da disciplina, demonstrando cooperação por parte do aluno), colaboração entre os alunos e todo tipo de conteúdo motivacional. As negativas seriam dúvidas sobre o conteúdo apresentado na aula, relatos de dificuldades na vida pessoal que interferissem no desempenho acadêmico, demonstração de frustrações ou de desorientação. E as neutras seriam dúvidas de execução de alguma tarefa dentro do AVA ou mensagens de cunho social.

Figura 10 – Regras usadas na rotulação das frases por especialistas em EaD



Das 27 mensagens postadas nos fóruns, 2 delas foram classificadas pelos docentes como positivas, 13 como negativas e 12 neutras. A Tabela 12 mostra um exemplo de cada

categoria (a tabela completa se encontra no Apêndice B). Somente os alunos que publicaram nos fóruns tiveram seus perfis motivacionais detectados. Dentre todos os alunos, sete foram responsáveis pela geração de todo conteúdo negativo, sendo que três deles tiveram o sentimento recorrente, ou seja, publicaram mais de uma vez mensagens com essa polaridade.

Tabela 12 – Amostra de frases rotuladas por especialistas em EaD

<b>Frase</b>	<b>Rótulo</b>
A etapa vence no dia 29, correto? Pergunto pois dentro do link da tarefa fala que vence no dia 22, e na parte da etapa fala que vence no dia 22	Positiva
Estou fazendo as tarefas, mas estou muito em dúvida se estou indo pelo caminho certo, tenho muitas dificuldades neste trabalho	Negativa
Boa tarde Professora, a tarefa sobre o objetivo dessa semana pode ser feita em meia lauda?	Neutra

A AS foi feita considerando as polaridades retornadas pela rede de *frames* com os transformadores de valência. Como pré-processamento, foram ignoradas todas as saudações (ex.: boa noite, bom dia, olá) contidas nas frases. Por não serem curtos, os comentários foram divididos em sentenças, que foram passadas a rede de *frames* e que retornou suas polaridades. Essas polaridades foram somadas gerando assim o sentimento geral do comentário. A operação de soma foi uma operação binária onde o sentimento geral do comentário completo era determinado pelas seguintes condições:

- Se as duas frases têm sentimentos iguais, *então* sentimento recorrente;
- Se uma das frases tem sentimento negativo, *então* sentimento negativo;
- Se uma das frases tem sentimento neutro, *então* prevalece o outro sentimento.

A Tabela 13 ilustra o processo de identificação do sentimento geral do texto, onde a primeira coluna apresenta uma amostra do conjunto de frases usadas neste estudo de caso, a segunda coluna mostra as *unidades lexicais* detectadas em cada frases com as respectivas polaridades dos *frames* que elas evocam, a terceira coluna apresenta a polaridade parcial do texto evidenciando os valores a serem somados, e a última coluna exhibe o sentimento geral do comentário.

Tabela 13 – Detecção do sentimento dos comentários. As polaridades parciais são descritas no formato numérico onde a *positiva* = 1, *neutra* = 0 e *negativa* = -1

<b>Comentário</b>	<b>Unidade Lexical evocada (polaridade)</b>	<b>Polaridade parcial</b>	<b>Sentimento gerado pela rede de frames</b>
No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Preciso ( <i>neutra</i> ), escolher ( <i>neutra</i> )	[0,0]	Neutra
na tarefa consta que o resumo deve ser em uma página. Gostaria de saber se a referência, caso seja necessário, posso colocá-la em uma segunda página. Se possível, só para conferir, poderia informar as dimensões usadas nas margens padrão. Obrigada.	deve( <i>neutra</i> ), saber( <i>positiva</i> ), se(reductor não alcançado), posso( <i>positiva</i> ), se(reductor alcançado) possível( <i>positiva</i> -> <i>negativa</i> ), poderia( <i>positiva</i> )	[0,+1,0,+1, -1,+1]	Positiva
Tive um infortúnio não consegui enviar a tarefa dentro da hora solicitada (23:55)... Agora são 23:58h... Conforme o orientação no guia do aluno as atividades com entrega no atraso são avaliadas com desconto de 50% no valor da nota, porem o link para enviar a atividade não se encontra disponível. Como faço pra poder enviar a tarefa solicitada depois do prazo?	tive( <i>neutra</i> ), não(negação alcançada) consegui( <i>positiva</i> -> <i>negativa</i> ), não(negação alcançada) se(reductor alcançado) disponível ( <i>positiva</i> -> <i>negativa</i> ), poder( <i>positiva</i> )	[0,-1,-2,+1]	Negativa



A Tabela 14 apresenta a matriz de confusão obtida das medidas *TP*, *TN*, *FP* e *FN*. Neste estudo de caso consideramos a polaridade neutra e por este motivo, a matriz de confusão foi gerada usando 3 rótulos (positivo, negativo e neutro). A coluna *rede de frames* diz respeito aos resultados obtidos da classificação feita pela rede polarizada de *frames*.

Tabela 14 – Matriz de confusão. Linha: rótulo dos especialistas em EaD, Coluna: rede de *frames*

<b>Rótulo</b>	<b>Rede de <i>frames</i></b>		
	Positiva	Negativa	Neutra
Positiva	1	0	1
Negativa	4	7	2
Neutra	3	5	4

Para uma melhor compreensão e obtenção dos valores dos critérios de avaliação, a matriz de confusão foi subdividida em 3 outras matrizes, correspondendo cada uma as polaridades consideradas neste estudo de caso (Tabelas 15, 16 e 17). A acurácia, precisão, *recall* e *F1-score* foram calculados, separadamente, para cada uma das polaridades e somados, gerando assim o valor global de cada um dos critérios.

Tabela 15 – Medidas de desempenho da polaridade *Positiva*

Tabela 15 (a) – Matriz de Confusão da classe *Positiva*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Positiva	Outras
Positiva	1	1
Outras	7	18

Tabela 15 (b) – Tabela de Critérios de avaliação da classe *Positiva*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
70	50	13	20

Tabela 16 – Medidas de desempenho da polaridade *Negativa*Tabela 16 (a) – Matriz de Confusão da classe *Negativa*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Negativa	Outras
Negativa	7	6
Outras	5	9

Tabela 16 (b) – Tabela de Critérios de avaliação da classe *Negativa*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
59	54	58	56

Tabela 17 – Medidas de desempenho da polaridade *Neutra*Tabela 17 (a) – Matriz de Confusão da classe *Neutra*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Neutra	Outras
Neutra	4	8
Outras	3	12

Tabela 17 (b) – Tabela de Critérios de avaliação da classe *Neutra*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
59	33	57	42

Para a obtenção dos valores referentes ao desempenho geral da rede de *frames* na AS em frases em português, tendo em vista os rótulos gerados por especialistas em EaD, foram calculadas as médias de cada critério de avaliação, que são apresentadas na Tabela 18.

Tabela 18 – Tabela de Critérios Gerais de avaliação

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
63	46	42	39

A rede de *frames* conseguiu uma acurácia de 62%, com precisão de 46% e recall de 42%, valores razoáveis para uma abordagem lexical (HEMMATIAN; SOHRABI, 2017). Ela foi melhor em identificar as polaridades *positiva*, tendo uma acurácia de 70%. A precisão obteve seu maior resultado na polaridade negativa (54%), pois foi a polaridade que menos obteve *falsos positivos*. O *recall* está associado à quantidade de amostras da categoria em relação a base, por isso a polaridade *positiva* apresenta o valor mais baixo.

Dados os resultados obtidos, a rede polarizada de *frames* foi analisada por um linguista, visando entender e melhorar o processo de detecção de polaridade, com o objetivo de melhorar seu desempenho na AS. As correções sugeridas por ele são descritas na seção que se segue.

### 5.2.2.2 Melhorias no processo de detecção de polaridade

A rede de *frames* polarizada avaliada por um doutor em Linguística, que é um especialista da FrameNet, com o objetivo de melhorar seu desempenho na detecção de sentimento. O especialista verificou todos os *frames* retornados da rede polarizada de *frames* que foram evocados do conjunto de frases dos alunos, identificou as polaridades atribuídas de forma incorreta e sugeriu melhorias que foram incorporadas ao modelo de AS da arquitetura SASys:

#### a) Aumento do alcance dos transformadores de valência

A estratégia usada era baseada no *3-gram* (DEY; JENAMANI; THAKKAR, 2018), onde três era o número de palavras alcançadas por negações, intensificadores ou redutores de valência. Contudo, verificou-se que os transformadores influenciam a sentença, e não apenas a palavra mais próxima, e por isso o seu alcance foi estendido até o sinal de pontuação mais próximo, abrangendo e influenciando todas as outras palavras no caminho. Por exemplo a frase:

*Boa tarde! Gostaria de saber se a minha atividade 03 está correta, tenho medo de continuar e também não fazer esta tarefa corretamente.*

Na abordagem inicial a conjunção condicional *se*, que tem um sentimento negativo, não alcançaria o *frame* *correta* que continuaria com sua polaridade positiva. Na nova abordagem esse sentimento muda, pois a conjunção passa a alcançar todos os *frames* antes da vírgula.

#### b) Palavras que expressam dever ou necessidade têm sentimento negativo e as que expressam possibilidade ou escolha têm sentimento positivo

A distinção foi feita devido as palavras *dever*, *precisar*, *ter que*, *que* foram consideradas neutras quando deveriam ser negativas e *poder*, *escolher*, *possibilidade*, que também estavam sendo tratados como palavras neutras quando possuíam sentimento positivo.

c) As aspas representam citação e não devem ser usadas na análise de sentimento. As citações são recorrentes no contexto educacional e devem ser desconsideradas da análise, porém, as aspas também podem ser usadas para intensificar o sentimento do autor. Na primeira abordagem elas eram desconsideradas, utilizando apenas o texto que elas comportavam na análise. Na nova estratégia, foi definido o número 4 como a quantidade mínima de palavras para que a sentença fosse considerada uma citação.

### 5.2.2.3 Rotulação das Mensagens por especialista da FrameNet

Depois da análise minuciosa do procedimento usado para a detecção de polaridade, o especialista da FrameNet gerou novos rótulos para as frases escritas pelos alunos, tendo em conta uma metodologia com embasamento linguístico. Os comentários coletados dos fóruns do Moodle foram lidos e classificados pelo especialista e a Tabela 19 apresenta uma amostra das frases com os novos rótulos, comparando com os rótulos gerados pelos professores especialistas em EaD (tabela completa Apêndice B). Grande parte das frases rotuladas como neutras pelos professores especialistas adquiriram outras polaridades, quando rotuladas pelo linguista.

Tabela 19 – Tabela comparativa entre o rótulo dos professores e o rótulo do linguista

<b>Comentário</b>	<b>Rótulo professores</b>	<b>Rótulo linguista</b>
No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Neutro	Negativo
Professora, Pesquisei sobre o tema objetos de aprendizagem e encontrei: "De acordo com o Learning Objects Metadata Workgroup, objetos de aprendizagem (Learning Objects) podem ser definidos por 'qualquer entidade, digital ou não digital, que possa ser utilizada, reutilizada ou referenciada durante oaprendizado suportado por tecnologias'.". Seria uma sequência didática para lecionar?	Negativo	Negativo
Não sou boa em escrever, discutir um texto, eu fiz as atividades anteriores, seria bom se tivessem corrigido para ver como eu fui, bem ou muito mau. Obrigada!	Neutro	Positivo

Uma nova AS dos comentários dos alunos foi obtida usando a rede polarizada de *frames* com a metodologia atualizada, implementando as melhorias sugeridas pelo especialista da FrameNet. A Tabela 20 apresenta uma amostra dos comentários rotulados pelo especialista da FrameNet e a nova classificação obtida da rede polarizada de *frames*, considerando o aumento do alcance dos transformadores de valência, mudanças nas polaridades de palavras que expressam necessidade, dever, possibilidade ou escolha e desconsiderando citações (vide tabela completa no Apêndice B).

Tabela 20 – Amostra do conjunto de comentários dos alunos com novos rótulos, definidos pelo linguista e as polaridades detectadas pela nova metodologia da rede de *frames*.

<b>Comentário</b>	<b>Rótulo do Especialista</b>	<b>Deteção da rede de <i>frames</i></b>
No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Negativa	Neutra
A etapa vence dia 29 correto? pergunto pois dentro do link da tarefa fala que vence dia 22, e na parte da etapa fala que vende dia 22.	Neutra	Positiva
Professora Fernanda, Essa atividade tem que ser feita dentro do Template que já estamos usando??	Negativa	Negativa

A nova rotulação gerou conjuntos de 5 frases positivas, 18 negativas e 4 neutras. Com base nesses conjuntos, foi obtida a nova matriz de confusão foi gerada com o objetivo de quantificar o desempenho da nova metodologia de AS da rede de *frames*. A Tabela 21 mostra os resultados obtidos.

Tabela 21 – Nova matriz de confusão. Linha: rótulo do especialista da FrameNet, Coluna: rede de *frames*

<b>Rótulo</b>	<b>Rede de <i>frames</i></b>		
	Positiva	Negativa	Neutra
Positiva	5	0	0
Negativa	0	14	4
Neutra	1	0	3

Assim como a seção 5.2.2.1, a matriz de confusão foi subdividida em 3 outras matrizes, correspondendo cada uma as polaridades consideradas neste estudo de caso, a fim de se obter os resultados dos critérios de avaliação (Tabelas 22, 23 e 24). A acurácia, precisão, *recall* e *F1-score* foram calculados, separadamente, para cada uma das polaridades. Para a obtenção do desempenho geral da nova metodologia de AS da rede polarizada de *frames* foram calculadas as médias de cada critério de avaliação.

Tabela 22 – Medidas de desempenho da polaridade *Positiva*

Tabela 22 (a) – Matriz de Confusão da classe *Positiva*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Positiva	Outras
Positiva	5	0
Outras	1	21

Tabela 22 (b) – Tabela de Critérios de avaliação da classe *Positiva*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
96	83	100	90

Tabela 23 – Medidas de desempenho da polaridade *Negativa*

Tabela 23 (a) – Matriz de Confusão da classe *Negativa*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Negativa	Outras
Negativa	14	4
Outras	0	9

Tabela 23 (b) – Tabela de Critérios de avaliação da classe *Negativa*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
85	100	78	87

Tabela 24 – Medidas de desempenho da polaridade *Neutra*

Tabela 24 (a) – Matriz de Confusão da classe *Neutra*

<b>Rótulo</b>	<b>Rede de Frames</b>	
	Neutra	Outras
Neutra	3	1
Outras	4	19

Tabela 24 (b) – Tabela de Critérios de avaliação da classe *Neutra*

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
81	43	75	55

Utilizando com base a rotulação do linguista, a rede de *frames* obteve um desempenho geral bom na classificação das frases dos alunos, com pode ser constatado na Tabela 25. Ela obteve sua maior precisão na detecção de frases negativas porque não possui nenhum *falso positivo* nesta categoria, ou seja, apesar de não identificar todos os comentários negativos da base, ela não categorizou, de forma errônea, frases de outros sentimentos como negativas. Suas maiores acurácia, *recall* e *f1-score* foram de 95%, 100% e 90% respectivamente, nas frases positivas, pois alcançou maior completude nesta classe, ou seja, ela foi capaz de reconhecer todos os comentários positivos presentes na base.

Tabela 25 – Tabela de Critérios Gerais de avaliação

Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
88	75	84	78

#### 5.2.2.4 Considerações sobre os resultados obtidos na AS

A AS foi realizada uma única vez na base de comentários de alunos, porém, seus rótulos são resultado de aplicação do conhecimento de profissionais de áreas distintas, gerando assim desempenhos diferentes.

A AS foi feita na versão original dos comentários coletados, retirando apenas as saudações que se encontravam no início das frases. Erros ortográficos não foram corrigidos e abreviações não foram substituídas por seus significados.

A rotulação feita pelos professores levou em consideração parâmetros da educação como tipos de dúvidas (de execução de uma tarefa ou de conteúdo ministrado em aula), tópicos (motivacionais, social ou de colaboração) e manifestações (elogios, frustrações ou dificuldades pessoais). Estas informações são baseadas em aspectos do contexto educacional e por este motivo a SASys obteve um desempenho não satisfatório, pois sua abordagem não se baseia em aspectos; ou seja, a rede polarizada de *frames* se limita na identificação do sentimento do texto através da orientação semântica das palavras e não na identificação de tópicos ou características do contexto que possam gerar sentimento as frases. Por outro lado, os rótulos atribuídos pelo linguista foram livres de contexto, utilizando-se apenas da semântica das palavras para determinar o sentimento contidos nos comentários dos alunos.

Com a implementação das melhorias sugeridas pelo especialista da FrameNet, a rede polarizada de *frames* apresentou um bom desempenho na AS de textos dos alunos

matriculados em uma disciplina *on-line*, respondendo assim a primeira questão de pesquisa deste estudo de caso (*QPCE1: É possível usar a rede de frames polarizada nos textos dos alunos de uma disciplina on-line para AS?*). Tanto no contexto gastronômico como educacional, a rede de frames obteve uma acurácia maior que 80% (86% e 88% respectivamente), indicando o seu uso como solução lexical em diferentes domínios.

Buscando responder as outras questões de pesquisa deste trabalho, os dados coletados dos alunos foram analisados e agregados aos resultados da AS para compor seus perfis motivacionais. A análise dos dados e a definição dos perfis estão detalhadas nas seções que se seguem.

### **5.2.3 Detecção do estado emocional**

A SASys foi instanciada num sistema de recomendação educacional visando identificar alunos com baixa motivação, através da detecção do seu estado emocional, procurando recomendar conteúdo aderente ao seu perfil. Para tal, foi necessário a definição de perfis motivacionais dos alunos.

Motivação vem do latim, *Movere*, que significa ir em direção à ou ser movido à alguma coisa (RYAN; DECI, 2000). Ela é definida por Afonso e Leal (2009) como sendo o que move as pessoas a agirem, pensarem e se desenvolverem.

Segundo Ryan e Deci (2000), a motivação possui níveis (quão motivado) e orientações (tipos de motivações) mas, de um modo geral, uma pessoa motivada é a que está energizada ou estimulada a um determinado fim, enquanto que a pessoa que não possui impulso a agir é denominada desmotivada. Uma teoria bastante conhecida sobre motivação é a Teoria da Autodeterminação (DECI; RYAN, 2008; RYAN; DECI, 2000) que distingue a motivação em dois tipos: intrínseca e extrínseca.

A motivação intrínseca está ligada a realização de uma tarefa por ser interessante, agradável ou por promover algum tipo de satisfação (RYAN; DECI, 2000). Ela é definida por Guimarães e Boruchovitch (2004) como sendo a tendência natural em buscar novidade, desafio e em exercitar as próprias capacidades. Na educação, ela é uma fonte natural de aprendizagem e criatividade que produz resultados de alta qualidade (RYAN; DECI, 2000).

A motivação extrínseca está ligada a realização de uma atividade que gera satisfação através de fatores externos (SEGURADO, 2015). A pessoa motivada extrinsecamente é impulsionada pelas recompensas da ação. Para Guimarães e Boruchovitch (2004), um aluno motivado possui características como persistência, engajamento e entusiasmo na realização de



tarefas desafiadoras, Para além disso, o orgulho pelos resultados de seu desempenho e o sentimento de superação de suas habilidades e conhecimentos prévios também são vistos como elementos motivadores.

Os perfis motivacionais definidos no estudo de caso apresentado nesta dissertação foram baseados na motivação extrínseca, através da análise dos dados de acesso dos alunos; pois com eles é possível identificar alunos engajados com a disciplina, por meio da análise da frequência de acesso ao AVA.

### 5.2.3.1 Estado Emocional

Os dados de acesso dos alunos ao AVA foram usados para definir os perfis motivacionais que junto com o resultado da AS determina o estado emocional dos mesmos. Os *logs* dos alunos foram coletados e a Tabela 26 apresenta uma amostra da consolidação dos dados de acesso de seis alunos no AVA, na qual cada linha contém a quantidade de acesso que o aluno teve em cada semana de aula. A última linha exibe a média de ações por aluno.

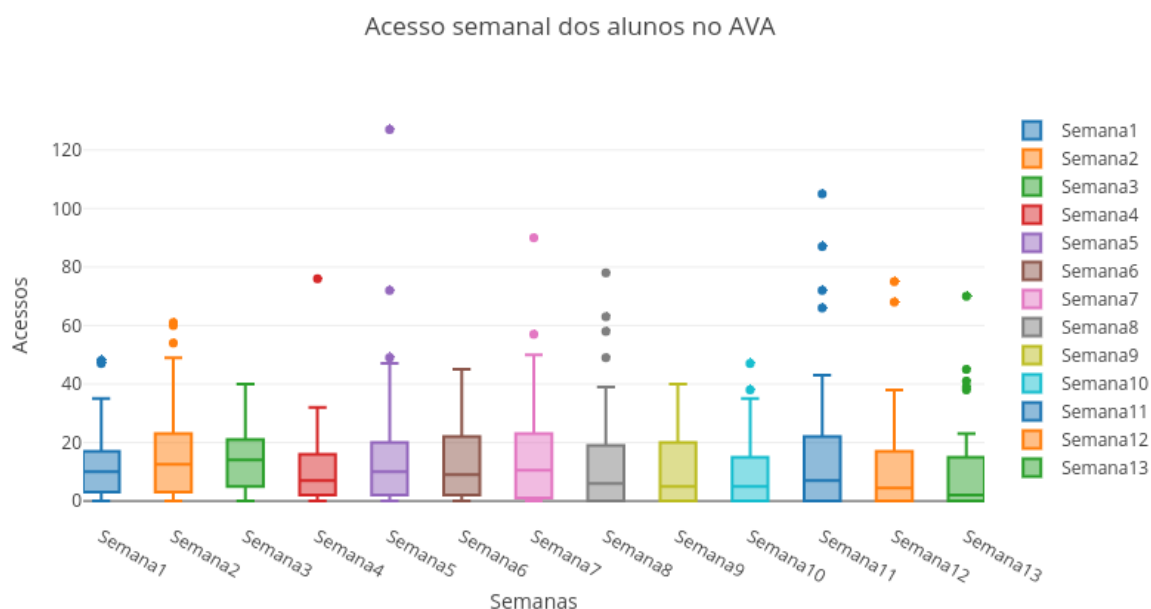
Tabela 26 – Amostra de dados dos alunos

	<b>A1</b>	<b>A2</b>	<b>A3</b>	<b>A4</b>	<b>A5</b>	<b>A6</b>
<b>S1</b>	2	11	48	15	6	5
<b>S2</b>	25	17	60	28	0	15
<b>S3</b>	10	19	24	0	9	11
<b>S4</b>	22	11	20	18	2	1
<b>S5</b>	3	19	72	15	7	11
<b>S6</b>	43	33	36	30	10	9
<b>S7</b>	19	41	90	18	6	57
<b>S8</b>	16	19	34	15	0	4
<b>S9</b>	5	30	21	11	8	6
<b>S10</b>	3	14	47	24	0	19
<b>S11</b>	5	38	105	21	12	14
<b>S12</b>	0	8	68	34	0	37
<b>S13</b>	0	13	70	45	0	17
<b>Média</b>	12	21	53	21	5	16

Foi observado que há uma grande variação na média de acessos entre os alunos e entre as semanas. Neste sentido, foi feita uma análise estatística buscando compreender melhor o comportamento dos alunos.

A Figura 11 mostra a distribuição de acessos de todos os alunos por semana utilizando Boxplot. Todos os gráficos têm o mínimo em zero por existir, pelo menos, um aluno que não acessou com o ambiente na respectiva semana. A grande maioria dos *outliers* representa os alunos líderes dos grupos. A partir da Semana7, o valor da mediana decaiu, dando indícios do aumento do número de alunos com pouca interação e, conseqüentemente, da existência de alunos desmotivados e com risco de evasão.

Figura 11 – Boxplots do acesso semanal dos alunos no AVA



A análise dos dados coletados gerou 4 Perfis Motivacionais (PM) que estratificam os alunos de acordo ao tipo de acesso e interação na disciplina:

- PM1 – Entusiasmo – Alunos com um número grande de ações no AVA. Estes alunos tendem a ser mais participativos, a realizar todas as atividades propostas no curso e possuem sentimentos positivos ou neutros. A média de acesso dos alunos deste grupo foi de 20 ações por semana;
- PM2 – Moderação e Dúvida – Alunos com um número razoável de ações no AVA. Este grupo abrange dois estados emocionais: Moderação, estudantes que frequentam a

disciplina com regularidade, tendem a realizar boa parte das atividades propostas e raramente comentam em fóruns; e Dúvida, estudantes que interagem com o fórum reportando dúvidas de execução de tarefa ou de conteúdo. De forma generalizada, alunos deste grupo apresentam sentimentos neutros e negativos. A média de acesso dos alunos deste grupo foi entre 10 e 19 ações por semana;

- PM3 – Desânimo – Alunos com pouca interação (alerta de evasão). Estes alunos tendem a ficar vários dias sem acessar o AVA e a realizar um número reduzido de atividades da disciplina. Seus comentários são de desorientação e frustração, indicando o sentimento negativo. A média de acesso deste grupo foi menos de 10 ações por semana;
- PM4 – Alunos com interação quase nula (iminência de evasão) - Alunos que interagiram com o AVA e que, depois de um período, o seu número de ações ficou nulo e, por este motivo, não possuem comentários para a AS. A média de acesso dos alunos deste grupo foi menos de 5 ações por semana.

Com base no Perfil Motivacional dos alunos e na AS dos comentários, foram detectados 3 estados emocionais dos alunos em relação à disciplina:

- *Satisfeito*: os alunos PM1 e PM2 que possuem sentimento positivo;
- *Neutro*: os alunos PM1 e PM2 que possuem sentimento negativo ou neutro;
- *Insatisfeitos*: alunos PM3 com sentimento negativo;

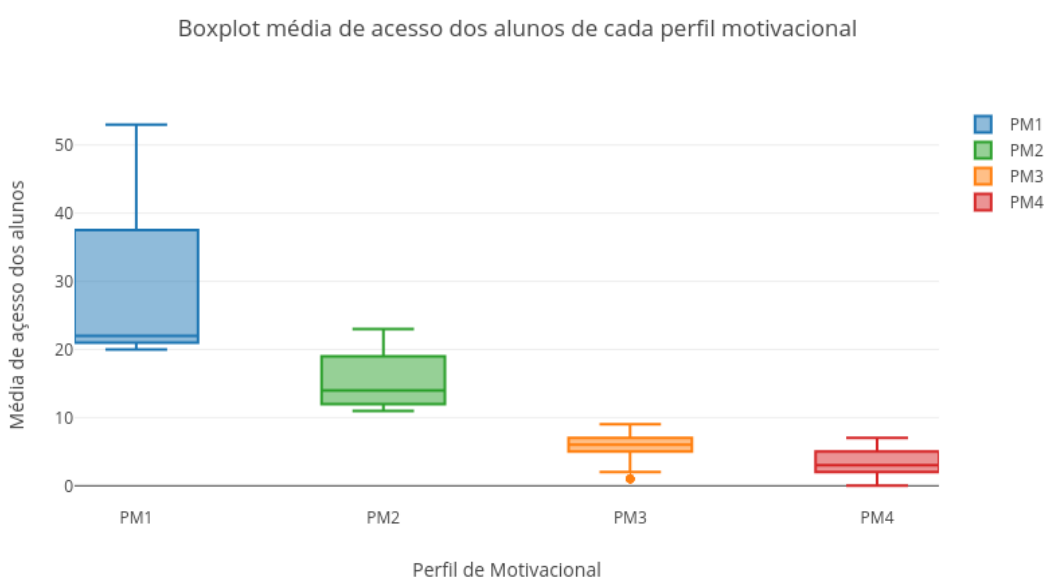
Assim sendo, para responder à segunda questão de pesquisa do estudo de caso (*QPEC2: É possível determinar o estado emocional do aluno através dos dados coletados?*) é apresentada a Tabela 27, que evidencia as combinações de perfis motivacionais e AS que geram os estados emocionais dos alunos. Os alunos com o estado emocional *Satisfeito* são os únicos que tiveram comentários positivos e que, devido a média de acesso deste grupo ser a maior, foi possível inferir que os mesmos estavam engajados e, conseqüentemente, satisfeitos com a disciplina. Os alunos com o estado emocional *Neutro* são os que mais interagiram no AVA e que, por este motivo, comentaram nos fóruns de dúvidas. A média de acesso deste grupo dá indícios de que eles permanecem na disciplina, esclarecendo dúvidas. A polaridade negativa de seus comentários foi anulada com o engajamento na disciplina, classificando assim como neutros. Os alunos *Insatisfeitos* são os que além de possuírem comentários negativos, têm a menor média de acesso ao AVA.

Tabela 27 – Perfil de Motivacional e Estado Emocional

Perfil Motivacional	AS nos comentários	Estado emocional
PM1 e PM2	Positivo	Satisfeito
PM1 e PM2	Neutro	Neutro
PM1 e PM2	Negativo	Neutro
PM3	Negativo	Insatisfeito

Para responder a terceira questão de pesquisa deste estudo de caso (*QPEC3: É possível identificar alunos com tendência à evasão em uma disciplina on-line?*) foram considerados, primordialmente, os perfis motivacionais dos alunos. Os alunos com o PM1 e PM2, com sentimento negativo e estado emocional neutro foram considerados alunos em estado de alerta de evasão porque, apesar de se mostrarem interessados com a disciplina, possuem inquietações que podem levá-los a desmotivação. Os alunos com o PM3 e PM4 geralmente possuem sentimentos negativos e são identificados como alunos com tendência de evasão e que necessitam de uma atenção especial por parte do professor. Além disso, foi estabelecido um limite motivacional, através da mediana de acessos de cada PM, capaz de identificar, com alguma antecedência, alunos desmotivados.

Figura 12 – Boxplots da média de acesso dos alunos de cada perfil motivacional



A Figura 12 mostra um gráfico onde é possível observar a distribuição dos alunos nos respectivos PMs. Cada boxplot foi gerado com as médias de interação dos alunos. As medianas dos boxplots foram usadas como o limite motivacional de cada perfil: mediana do PM1 = 22; PM2 = 14; PM3 = 6 e PM4 = 3. Todo aluno com média de acesso abaixo da mediana do seu respectivo PM e com sentimento negativo também é considerado desmotivado e com risco de evasão.

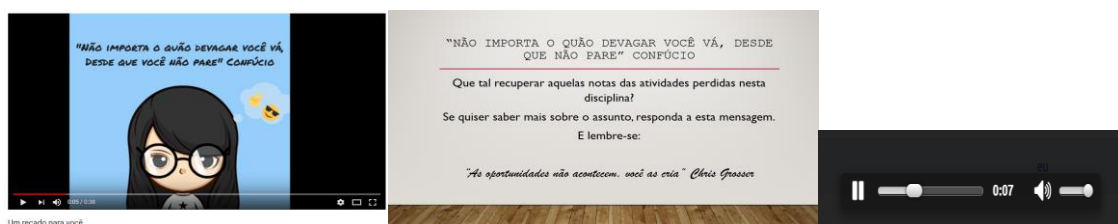
#### 5.2.4 Recomendação de mensagens à partir do perfil motivacional

A partir do perfil motivacional dos alunos identificados como desmotivados, foi desenvolvido um sistema de recomendação para envio de mensagens aderentes às suas preferências buscando reverter a tendência à evasão.

Os alunos responderam a um questionário *on-line*<sup>14</sup> que foi enviado na 11ª semana de aulas, através do *Moodle*, com o objetivo de detectar suas preferências. O questionário foi composto de 7 perguntas, que procuravam descobrir o dispositivo mais usado para o acesso ao AVA, preferências em relação ao tipo de conteúdo (texto, áudio ou vídeo), tamanho do conteúdo (curto ou longo) e a rede social mais usada.

Apenas 28% da turma respondeu o questionário, ou seja, foram coletadas respostas de 16 alunos dentro dos quais 8 deles faziam parte do PM1, 7 do PM2 e 1 do PM3. Foram selecionados dois alunos de cada PM para a recomendação do conteúdo motivacional. Essa seleção considerou o sentimento dos alunos bem como o limite motivacional e a nota; ou seja, alunos que responderam o questionário, que possuíam média de acesso abaixo do limite motivacional e com notas baixas. No PM3, apenas um aluno foi selecionado por ser o único que respondeu ao questionário.

Figura 13 – Formas de apresentação da mensagem motivacional: vídeo, texto e áudio<sup>15</sup>



<sup>14</sup> <https://forms.gle/tRZwPmTKqhcWmRxF6>

<sup>15</sup> <https://www.youtube.com/watch?v=z8BvxujzK5w&feature=youtu.be>

Foram criadas mensagens motivacionais para recomendação. As formas de apresentação das mensagens foram definidas considerando as preferências dos alunos em relação ao tipo e tamanho de material didático, sendo gerados de três tipos de mensagens: vídeo curto, texto curto e áudio curto, como é apresentado na Figura 13.

Os 5 alunos identificados com risco de evasão receberam as mensagens motivacionais, através do *Moodle*, de acordo às suas preferências. Essas mensagens foram encaminhadas durante o semestre letivo, visando incentivar os alunos a retomarem as atividades da disciplina.

Foi possível coletar *feedback* positivo em relação a recomendação de material, pois todos os alunos responderam as mensagens motivacionais mostrando interesse em recuperar as atividades não realizadas, exceto o aluno do PM3.

Os alunos voltaram a acessar o AVA e enviaram os trabalhos pendentes ou em atraso, dando indícios de que a identificação precoce de alunos desmotivados através da SASys e a recomendação de conteúdo motivacional podem retardar o processo de evasão.

### 5.3 AMEAÇAS A VALIDADE

O estudo de caso aqui apresentado teve como objetivo avaliar a arquitetura SASys, em todas as suas camadas. Contudo, ele foi conduzido pela própria pesquisadora, o que pode trazer algum viés ao resultado obtido, visto que a mesma tem plena consciência das questões de pesquisa, dos dados coletados e do processo a ser realizado.

Outra ameaça a validade consiste nas áreas de atuação dos especialistas que rotularam as frases do conjunto de dados, pois são extremamente ligadas ao contexto da pesquisa e podem enviesar os resultados alcançados.

A quantidade de dados analisados também constitui uma ameaça a validade deste estudo de caso, por apenas estar na casa de dezenas.

O contato com o especialista da FrameNet para rotulação das frases e avaliação da metodologia de AS foi feito apenas uma vez e não havendo um retorno, por parte da pesquisadora, para a validação do desempenho do método atualizado de AS.

Os resultados do estudo de caso possuem a característica de serem muito específicos do contexto aplicado, não podendo ser generalizados.

## 5.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

A arquitetura SASys foi instanciada no contexto da educação para avaliar seu desempenho na detecção do estado emocional dos alunos. Para tal, foram utilizados dados de acesso ao AVA, textos de fóruns e questionário para determinação do estado emocional.

Os comentários dos alunos foram retirados de fóruns de dúvidas o que implicou em um número considerável de frases negativas e neutras. Os textos eram de teor educacional e por isso não continham abreviações ou *emotions*. Os erros ortográficos foram mantidos como forma de avaliação do desempenho da rede de *frames* polarizada.

A rotulação dos textos foi feita por profissionais de áreas distintas: pelos professores com anos de experiência em EaD e um linguista especialista da FrameNet. Os professores consideraram aspectos do contexto educacional como tipos de dúvidas, tópicos e manifestações para rotular as frases, o que proporcionou um desempenho razoável para a abordagem lexical proposta neste trabalho.

O linguista utilizou-se apenas semântica das palavras para rotular os comentários dos alunos, o que resultou no melhor desempenho da rede polarizada de *frames*. A arquitetura SASys alcançou uma acurácia de 88%, precisão de 75%, *recall* 84 de e *f1-score* de 78%.

Revisando as questões de pesquisa deste estudo de caso temos que:

*QPCE1: É possível usar a rede de frames polarizada nos textos dos alunos de uma disciplina on-line para AS?*

Com base no resultado do desempenho obtido pela rede polarizada de *frames* podemos afirmar que ela pode ser usada para detecção de sentimento nos textos de alunos de uma disciplina *on-line*.

*QPEC2: É possível determinar o estado emocional do aluno através dos dados coletados?*

Os dados coletados através do AVA e do formulário de preferências foram usados para determinar os perfis motivacionais dos alunos, que combinados com a AS determinaram os respectivos estados emocionais; sendo eles *satisfeito*, *neutro* e *insatisfeito*.

*QPEC3: É possível identificar alunos com tendência à evasão em uma disciplina on-line?*

Com base nos perfis motivacionais foi possível traçar critérios que qualificavam alunos como desmotivados ou em risco de evasão. Esses alunos foram notificados pelo AVA e receberam recomendação de material motivacional, visando retardar o processo de evasão dos mesmos.

Assim sendo, o estudo de caso foi capaz de passar por todas as camadas da SASys, validando o seu uso para o contexto da educação. Contudo, os resultados são preliminares e pesquisas precisam ser feitas com um conjunto maior de dados.



## 6 CONSIDERAÇÕES FINAIS DO TRABALHO

A Análise de Sentimento ou Mineração de Opinião permite a descoberta de opinião ou sentimento de forma automática. Ela tem um impacto econômico por possibilitar que as instituições conheçam seus públicos e obtenham informações sobre aceitação de seus produtos e serviços. Por este motivo ela tem atraído pesquisadores da academia e indústria.

Esta dissertação se propôs a contribuir nesta área, apresentando a arquitetura SASys que usa a FrameNet como parte de sua abordagem lexical. A SASys visa descobrir o estado emocional do autor de um texto, através da identificação da orientação semântica das palavras, usando a rede de *frames* polarizada da FrameNet, e a coleta de outros dados.

A FrameNet foi escolhida como solução de AS da arquitetura proposta, por possuir embasamento linguístico na criação e manutenção de *frames*, que são estruturas cognitivas que definem palavras através de cenários e que por este motivo seriam também capazes de determinar o sentimento das mesmas.

A principal questão de pesquisa desta dissertação se dispunha a investigar se *uma arquitetura baseada na FrameNet como abordagem lexical é capaz de identificar o sentimento do autor de um texto*. Sua resposta pode ser obtida respondendo as suas questões secundárias:

*QPS1: É possível usar a FrameNet como uma solução lexical para a AS?*

Para responder a essa questão de pesquisa foi conduzido um estudo sobre a estrutura física e semântica da FrameNet e gerada uma rede polarizada de *frames* para ser usada na análise de sentimento.

A rede polarizada de *frames* é uma rede complexa que foi construída usando *frames* e agregando sentimentos a eles mediante o uso do algoritmo de *Label Propagation*. Esses sentimentos foram retirados da própria FrameNet, através da ferramenta *WebTool*, que permitiu a desambiguação de termos por meio da diferenciação do uso de cada um deles em cenários distintos.

Um estudo observacional foi conduzido no contexto gastronômico, com frases curtas escritas na língua inglesa e um estudo de caso no contexto educacional, com frases longas escritas na língua portuguesa. Em ambos estudos, a rede de *frames* obteve uma acurácia acima de 80%, corroborando no uso da FrameNet como solução lexical de AS.

*QPS2: A arquitetura proposta funciona adequadamente para a AS em textos?*

No estudo de caso apresentado no capítulo 5, a SASys foi usada para determinar o perfil motivacional dos estudantes, visando identificar os alunos com risco de evasão. Este

perfil foi gerado pelos dados coletados no AVA e em conjunto com o resultado da AS, foi usado para recomendação de material motivacional, evidenciando o funcionamento adequado da arquitetura.

No estudo observacional, rede polarizada de *frames* obteve uma acurácia de 86%, precisão de 80%, *recall* de 96% e *F1-score* de 87%, resultados bons para uma abordagem lexical. Do mesmo modo, no estudo de caso, a rede polarizada de *frames* alcançou seu desempenho máximo com uma acurácia de 88%, precisão de 75%, *recall* de 84% e *F1-score* de 78%. Tais resultados dão indícios do uso da FrameNet como solução lexical para mais de um domínio.

Assim sendo, os resultados obtidos nesta pesquisa dão indícios de que uma arquitetura de AS usando a FrameNet pode sim obter bom desempenho para uma solução de abordagem lexical.

## 6.1 LIMITAÇÕES E TRABALHOS FUTUROS

Neste trabalho foram identificadas algumas limitações, entre elas:

- Uma pequena amostra de *unidades lexicais* com sentimento (0,2% do seu total) foi usada, como semente para o algoritmo de propagação de rótulo. Este número pode impactar no desempenho da rede polarizada de *frames* e estudos precisam ser feitos para determinar o número mínimo de sementes para o algoritmo.
- A identificação (desambiguação) do *frame* evocado pela *unidade lexical* contida na frase a ser analisada é feita pela API da FrameNet, estando a AS da arquitetura sujeita ao resultado obtido da API.
- O processo de geração de *frames* ainda é empírico, podendo impactar no desempenho da rede polarizada de *frames*.

Foram conduzidos dois estudos em contextos diferentes para a validação da SASys, porém outras avaliações precisam ser conduzidas em condições e contextos diferentes, com maior quantidade de dados, maior quantidade de frases a serem analisadas, uso de *emojis* e imagens, entre outros.

Como trabalhos futuros se pretende estudar quão impactante pode ser o aumento da amostra de *unidades lexicais* com sentimentos usada na propagação de rótulos e o emprego de outros algoritmos de propagação de rótulos. Também se tenciona estudar o impacto do uso de redes sociais no desempenho da SASys e a detecção de emoções através da rede polarizada de *frames*.

## REFERÊNCIAS

- ABEDIN, Babak; BABAR, Abdul; ABBASI, Alireza. Characterization of the use of social media in natural disasters: a systematic review. In: **2014 IEEE Fourth International Conference on Big Data and Cloud Computing**. IEEE, 2014. p. 449-454
- AFONSO, José de Abreu; LEAL, Isabel P. Escala de motivação: adaptação e validação da Motivation Scale (MS) de Rempel, Holmes & Zanna. **Psicologia, Saúde & Doenças**, v. 10, n. 2, p. 249-266, 2009.
- AHMAD, M. et al. Sentiment Analysis using SVM: A Systematic Literature Review. **International Journal of Advanced Computer Science and Applications**, 2018. p. 182-188
- ALTRABSHEH, Nabeela; COCEA, Mihaela; FALLAHKHAIR, Sanaz. Sentiment analysis: towards a tool for analysing real-time students feedback. In: **2014 IEEE 26th international conference on tools with artificial intelligence**. IEEE, 2014. p. 419-423.
- ALTRABSHEH, Nabeela; GABER, Mohamed Medhat; COCEA, Mihaela. SA-E: sentiment analysis for education. In: **International Conference on Intelligent Decision Technologies**. 2013. p. 353-362.
- BAKER, Collin F.; FILLMORE, Charles J.; LOWE, John B. The berkeley framenet project. In: **Proceedings of the 17th international conference on Computational linguistics- Volume 1**. Association for Computational Linguistics, 1998. p. 86-90.
- BAMAKAN, Seyed Mojtaba Hosseini; NURGALIEV, Ildar; QU, Qiang. Opinion leader detection: A methodological review. **Expert Systems with Applications**, v. 115, p. 200-222, 2019.
- BASILI, Victor R.; ROMBACH, H. Dieter. The TAME project: Towards improvement-oriented software environments. **IEEE Transactions on software engineering**, v. 14, n. 6, p. 758-773, 1988.
- BLONDEL, Vincent D. et al. Fast unfolding of communities in large networks. **Journal of statistical mechanics: theory and experiment**, v. 2008, n. 10, p. P10008, 2008.
- BÓBÓ, Míria et al. Análise da estrutura da rede de frames da FrameNet Brasil. In: **7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)**. SBC, 2018.
- BÓBÓ, Míria et al. Análise de Sentimento na Educação: Um Mapeamento Sistemático da Literatura, **Simpósio Brasileiro de Informática na Educação (SBIE) 2019a**.
- BÓBÓ, Míria et al. Identificação do Perfil Emocional do Aluno Através de Análise de Sentimento: Combatendo a Evasão Escolar, **Simpósio Brasileiro de Informática na Educação (SBIE) 2019b**.
- CAMBRIA, Erik et al. Sentiment analysis is a big suitcase. **IEEE Intelligent Systems**, v. 32, n. 6, p. 74-80, 2017.

- CHAPLOT, Devendra Singh; RHIM, Eunhee; KIM, Jihie. Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks. In: **AIED Workshops**. 2015. p. 54-57.
- CIRQUEIRA, Douglas et al. A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. In: **2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)**. IEEE, 2018. p. 746-749.
- COSTENBADER, Elizabeth; VALENTE, Thomas W. The stability of centrality measures when networks are sampled. **Social networks**, v. 25, n. 4, p. 283-307, 2003.
- CUMMINS, Stephen; BURD, Liz; HATCH, Andrew. Using feedback tags and sentiment analysis to generate sharable learning resources investigating automated sentiment analysis of feedback tags in a programming course. In: **2010 10th IEEE International Conference on Advanced Learning Technologies**. IEEE, 2010. p. 653-657.
- DECI, Edward L.; RYAN, Richard M. Self-determination theory: A macrotheory of human motivation, development, and health. **Canadian psychology/Psychologie canadienne**, v. 49, n. 3, p. 182, 2008.
- DEY, Atanu; JENAMANI, Mamata; THAKKAR, Jitesh J. Senti-N-Gram: An n-gram lexicon for sentiment analysis. **Expert Systems with Applications**, v. 103, p. 92-105, 2018.
- DING, Xiaowen; LIU, Bing; YU, Philip S. A holistic lexicon-based approach to opinion mining. In: **Proceedings of the 2008 international conference on web search and data mining**. ACM, 2008. p. 231-240.
- DRESCH, Aline; LACERDA, Daniel Pacheco; JÚNIOR, José Antonio Valle Antunes. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. Bookman Editora, 2015.
- FELDMAN, Ronen. Techniques and applications for sentiment analysis. **Communications of the ACM**, v. 56, n. 4, p. 82-89, 2013.
- FENG, Yunping et al. The assessment of the points reward mechanism in online course forum. In: **2016 International Conference on Progress in Informatics and Computing (PIC)**. IEEE, 2016. p. 722-727.
- FILLMORE, Charles J. et al. FrameNet in action: The case of attaching. **International journal of lexicography**, v. 16, n. 3, p. 297-332, 2003.
- FILLMORE, Charles J.; BAKER, Collin. A frames approach to semantic analysis. In: **The Oxford handbook of linguistic analysis**. 2010.
- FILLMORE, Charles J.; BAKER, Collin F.; SATO, Hiroaki. FrameNet as a "Net". In: **LREC**. 2004.
- FILLMORE, Charles J.; JOHNSON, Christopher R.; PETRUCK, Miriam RL. Background to framenet. **International journal of lexicography**, v. 16, n. 3, p. 235-250, 2003.

FILLMORE, Charles J. et al. FrameNet in action: The case of attaching. **International journal of lexicography**, v. 16, n. 3, p. 297-332, 2003.

FREEMAN, Linton C. Centrality in social networks conceptual clarification. **Social networks**, v. 1, n. 3, p. 215-239, 1978.

GOHIL, Sunir; VUIK, Sabine; DARZI, Ara. Sentiment analysis of health care tweets: review of the methods used. **JMIR public health and surveillance**, v. 4, n. 2, p. e43, 2018.

GUIMARÃES, Sueli Édi Rufini et al. O estilo motivacional do professor e a motivação intrínseca dos estudantes: uma perspectiva da teoria da autodeterminação. **Psicologia: reflexão e crítica**, v. 17, n. 2, p. 143-150, 2004.

HADDI, Emma; LIU, Xiaohui; SHI, Yong. The role of text pre-processing in sentiment analysis. **Procedia Computer Science**, v. 17, p. 26-32, 2013.

HEMMATIAN, Fatemeh; SOHRABI, Mohammad Karim. A survey on classification techniques for opinion mining and sentiment analysis. **Artificial Intelligence Review**, p. 1-51, 2017.

JI, Rongrong et al. Survey of visual sentiment prediction for social media analysis. **Frontiers of Computer Science**, v. 10, n. 4, p. 602-611, 2016.

KITCHENHAM, Barbara A.; BUDGEN, David; BRERETON, O. Pearl. Using mapping studies as the basis for further research—a participant-observer case study. **Information and Software Technology**, v. 53, n. 6, p. 638-651, 2011.

KOTZIAS, Dimitrios et al. From group to individual labels using deep features. In: **Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. ACM, 2015. p. 597-606.

KUMAR, Alok; JAIN, Renu. Sentiment analysis and feedback evaluation. In: **2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)**. IEEE, 2015. p. 433-436.

KUMAR, Vinayshekhar Bannihatti; BAADKAR, Tanisha Rajesh; JOSHI, Vinay. CRYPTANITE: A New look to the world of Social Networks using Deep learning. In: **2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)**. IEEE, 2016. p. 358-364.

LEI, Chi-Un et al. Advancing MOOC and SPOC development via a learner decision journey analytic framework. In: **2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)**. IEEE, 2015. p. 149-156.

LIMA, Ana Carolina ES; DE CASTRO, Leandro N. Predicting temperament from Twitter data. In: **2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)**. IEEE, 2016. p. 599-604.

MACHADO, Mateus Tarcinalli; PARDO, Thiago AS; RUIZ, Evandro Eduardo Seron. Creating a Portuguese context sensitive lexicon for sentiment analysis. In: **International**

**Conference on Computational Processing of the Portuguese Language.** Springer, Cham, 2018. p. 335-344.

NAKAGAWA, Elisa Yumi et al. **Revisão sistemática da literatura em Engenharia de Software: teoria e prática.** Elsevier Brasil, 2017.

NASSIRTOUSSI, Arman Khadjeh et al. Text mining for market prediction: A systematic review. **Expert Systems with Applications**, v. 41, n. 16, p. 7653-7670, 2014.

PONG-INWONG, Chakrit; RUNGWORAWUT, Wararat Songpan. Teaching senti-lexicon for automated sentiment polarity definition in teaching evaluation. In: **2014 10th International Conference on Semantics, Knowledge and Grids.** IEEE, 2014. p. 84-91.

RANA, Toqir Ahmad; CHEAH, Yu-N.; LETCHMUNAN, Sukumar. Topic modeling in sentiment analysis: a systematic review. **Journal of ICT Research and Applications**, v. 10, n. 1, p. 76-93, 2016.

RANI, Sujata; KUMAR, Parteek. A sentiment analysis system to improve teaching and learning. **Computer**, v. 50, n. 5, p. 36-43, 2017.

RUPPENHOFER, Josef et al. **FrameNet II: Extended theory and practice.** 2006.

RYAN, Richard M.; DECI, Edward L. Intrinsic and extrinsic motivations: Classic definitions and new directions. **Contemporary educational psychology**, v. 25, n. 1, p. 54-67, 2000.

SALOMAO, Maria Margarida Martins. **FrameNet Brasil: um trabalho em progresso.** **Calidoscópico**, v. 7, n. 3, p. 171-182, 2009.

SEGURADO, Rita Martins. **Gestão de recompensas e perfil motivacional em profissionais de TI: papel na prevenção da rotatividade externa.** 2015. Tese de Doutorado.

SHAPIRO, Heather B. et al. Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. **Computers & Education**, v. 110, p. 35-50, 2017.

SOUZA, Ellen et al. Characterising text mining: a systematic mapping review of the Portuguese language. **IET Software**, v. 12, n. 2, p. 49-75, 2017.

SOUZA, Marlo et al. Construction of a portuguese opinion lexicon from multiple resources. In: **Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology.** 2011.

SPATIOTIS, Nikolaos et al. Sentiment Analysis for the Greek Language. In: **Proceedings of the 20th Pan-Hellenic Conference on Informatics.** ACM, 2016. p. 2.

TABOADA, Maite et al. Lexicon-based methods for sentiment analysis. **Computational linguistics**, v. 37, n. 2, p. 267-307, 2011.

TORRENT, T. T. et al. **A flexible tool for an enriched FrameNet: the FrameNet Brasil Webtool.** [s.l: s.n.].

VIJAYARANI, S.; ILAMATHI, Ms J.; NITHYA, Ms. Preprocessing techniques for text mining-an overview. **International Journal of Computer Science & Communication Networks**, v. 5, n. 1, p. 7-16, 2015.

WASSERMAN, Stanley et al. **Social network analysis: Methods and applications**. Cambridge university press, 1994.

WOHLIN, Claes et al. **Experimentation in software engineering**. Springer Science & Business Media, 2012.

XU, Liting; ZHAN, Yongzhao; ZHU, Miao. A guide based on situation and emotion analysis in ontology-based E-Learning system. In: **2008 3rd IEEE Conference on Industrial Electronics and Applications**. IEEE, 2008. p. 2417-2421.

## APÊNDICE A - Conjunto de dados usados no protótipo da SASys

Tabela A1 – Tabela contendo 50 frases selecionadas da base de dados YELP, de forma aleatória. As polaridades estão representadas pelo valor numérico, onde a *positiva* = 1 e *negativa* = 0

Número	Frases	Rótulos
1	Great brunch spot.	1
2	But now I was completely grossed out.	0
3	The pizza selections are good.	1
4	I've had better bagels from the grocery store.	0
5	I think not again	0
6	Our server was very nice, and even though he looked a little overwhelmed with all of our needs, he stayed professional and friendly until the end.	1
7	Also there are combos like a burger, fries, and beer for 23 which is a decent deal.	1
8	The selection was probably the worst I've seen in Vegas.....there was none.	0
9	I think this restaurant suffers from not trying hard enough.	0
10	Now this dish was quite flavourful.	1
11	This greedy corporation will NEVER see another dime from me!	0
12	The black eyed peas and sweet potatoes... UNREAL!	1
13	All in all, I can assure you I'll be back.	1
14	I checked out this place a couple years ago and was not impressed.	0
15	Be sure to order dessert, even if you need to pack it to-go - the tiramisu and cannoli are both to die for.	1
16	Fantastic service here.	1
17	Definitely a turn off for me & i doubt I'll be back unless someone else is buying.	0
18	It was absolutely amazing.	1
19	Food was great and so was the service!	1
20	We were sat right on time and our server from the get go was FANTASTIC!	1
21	Worse of all, he humiliated his worker right in front of me..Bunch of horrible name callings.	0
22	Never had anything to complain about here.	1
23	Nargile - I think you are great.	1
24	So we went to Tigerlilly and had a fantastic afternoon!	1
25	They have horrible attitudes towards customers, and talk down to each one when customers don't enjoy their food.	0
26	For service, I give them no stars.	0
27	To summarize... the food was incredible, nay, transcendent... but nothing brings me joy quite like the memory of the pneumatic condiment dispenser.	1
28	The sweet potato fries were very good and seasoned well.	1
29	The seasonal fruit was fresh white peach puree.	1
30	If you want healthy authentic or ethnic food, try this place.	1
31	They had a toro tartare with a cavier that was extraordinary and I liked the thinly sliced wagyu with white truffle.	1
32	I will not return.	0
33	Anyway, this FS restaurant has a wonderful breakfast/lunch.	1



34	My husband said she was very rude... did not even apologize for the bad food or anything.	0
35	I can take a little bad service but the food sucks.	0
36	this was a different cut than the piece the other day but still wonderful and tender s well as well flavored.	1
37	Sorry, I will not be getting food from here anytime soon :(	0
38	Not much seafood and like 5 strings of pasta at the bottom.	0
39	The ripped banana was not only ripped, but petrified and tasteless.	0
40	2 times - Very Bad Customer Service !	0
41	Would not go back.	0
42	Honeslty it didn't taste THAT fresh.)	0
43	The potatoes were like rubber and you could tell they had been made up ahead of time being kept under a warmer.	0
44	The fries were great too.	1
45	I was shocked because no signs indicate cash only.	0
46	A great touch.	1
47	They never brought a salad we asked for.	0
48	This was like the final blow!	0
49	Ample portions and good prices.	1
50	It's too bad the food is so damn generic.	0

Tabela A2 – Tabela identificando as *unidades lexicais* contidas nas frases, suas polaridades individuais e a polaridade final da frase

Número	Frases	Unidades lexicais detectadas ( <i>polaridade individual</i> )	Polaridade final detectada pela rede de <i>frames</i>
1	Great brunch spot.	great ( <i>positiva</i> )	Positiva
2	But now I was completely grossed out.	completely ( <i>positiva</i> )	Positiva
3	The pizza selections are good.	good ( <i>positiva</i> )	Positiva
4	I've had better bagels from the grocery store.	better ( <i>positiva</i> )	Positiva
5	I think not again	not ( <i>negativa</i> )	Positiva
6	Our server was very nice, and even though he looked a little overwhelmed with all of our needs, he stayed professional and friendly until the end.	nice ( <i>positiva</i> ), friendly ( <i>positiva</i> )	Positiva
7	Also there are combos like a burger, fries, and beer for 23 which is a decent deal.	decent ( <i>positiva</i> )	Positiva
8	The selection was probably the worst I've seen in Vegas.....there was none.	probably ( <i>neutra</i> ), worst ( <i>negativa</i> )	Negativa
9	I think this restaurant suffers from not trying hard enough.	not ( <i>negativa</i> ), hard ( <i>negativa</i> ), enough ( <i>neutra</i> )	Negativa
10	Now this dish was quite flavourful.	flavourful ( <i>positiva</i> )	Positiva
11	This greedy corporation will NEVER see another dime from me!	never ( <i>negativa</i> )	Negativa

12	The black eyed peas and sweet potatoes... UNREAL!	black ( <i>neutra</i> ), sweet ( <i>positiva</i> )	Positiva
13	All in all, I can assure you I'll be back.	assure ( <i>positiva</i> )	Positiva
14	I checked out this place a couple years ago and was not impressed.	not ( <i>negativa</i> )	Negativa
15	Be sure to order dessert, even if you need to pack it to-go - the tiramisu and cannoli are both to die for.	sure ( <i>positiva</i> )	Positiva
16	Fantastic service here.	fantastic ( <i>positiva</i> )	Positiva
17	Definitely a turn off for me & i doubt I'll be back unless someone else is buying.	doubt ( <i>negativa</i> )	Negativa
18	It was absolutely amazing.	amazing ( <i>positiva</i> )	Positiva
19	Food was great and so was the service!	great ( <i>positiva</i> )	Positiva
20	We were sat right on time and our server from the get go was FANTASTIC!	fantastic ( <i>positiva</i> )	Positiva
21	Worse of all, he humiliated his worker right in front of me..Bunch of horrible name callings.	worse ( <i>negativa</i> ), horrible ( <i>negativa</i> )	Negativa
22	Never had anything to complain about here.	never ( <i>negativa</i> ), complain ( <i>negativa</i> )	Negativa
23	Nargile - I think you are great.	great ( <i>positiva</i> )	Positiva
24	So we went to Tigerlilly and had a fantastic afternoon!	fantastic ( <i>positiva</i> )	Positiva
25	They have horrible attitudes towards customers, and talk down to each one when customers don't enjoy their food.	horrible ( <i>negativa</i> )	Negativa
26	For service, I give them no stars.	no ( <i>negativa</i> )	Negativa
27	To summarize... the food was incredible, nay, transcendent... but nothing brings me joy quite like the memory of the pneumatic condiment dispenser., I,,	Incredible ( <i>positiva</i> ), like ( <i>positiva</i> )	Positiva
28	The sweet potato fries were very good and seasoned well.	sweet ( <i>positiva</i> ), good ( <i>positiva</i> )	Positiva
29	The seasonal fruit was fresh white peach puree.	fresh ( <i>positiva</i> )	Positiva
30	If you want healthy authentic or ethic food, try this place.	healthy ( <i>positiva</i> )	Positiva
31	They had a toro tartare with a cavier that was extraordinary and I liked the thinly sliced wagyu with white truffle. ,	extraordinary ( <i>positiva</i> ), like ( <i>positiva</i> )	Positiva
32	I will not return.	not ( <i>negativa</i> )	Negativa
33	Anyway, this FS restaurant has a wonderful breakfast/lunch.	wonderful ( <i>negativa</i> )	Positiva
34	My husband said she was very rude... did not even apologize for the bad food or anything.	rude ( <i>negativa</i> ), not ( <i>negativa</i> ), bad ( <i>negativa</i> )	Negativa
35	I can take a little bad service but the food sucks.	little ( <i>neutra</i> ), bad ( <i>negativa</i> )	Negativa

36	this was a different cut than the piece the other day but still wonderful and tender s well as well flavored.	different ( <i>negativa</i> ), wonderful ( <i>positiva</i> ), flavored ( <i>positiva</i> )	Positiva
37	Sorry, I will not be getting food from here anytime soon :(	not ( <i>negativa</i> )	Negativa
38	Not much seafood and like 5 strings of pasta at the bottom.	not ( <i>negativa</i> ), like ( <i>positiva</i> )	Negativa
39	The ripped banana was not only ripped, but petrified and tasteless.	not ( <i>negativa</i> ), tasteless ( <i>negativa</i> )	Negativa
40	2 times - Very Bad Customer Service !	bad ( <i>negativa</i> )	Negativa
41	Would not go back.	not ( <i>negativa</i> )	Negativa
42	Honeslty it didn't taste THAT fresh.)	taste ( <i>positiva</i> ), fresh ( <i>positiva</i> )	Positiva
43	The potatoes were like rubber and you could tell they had been made up ahead of time being kept under a warmer.	like ( <i>positiva</i> ), warm ( <i>neutra</i> )	Positiva
44	The fries were great too.	great ( <i>positiva</i> )	Positiva
45	I was shocked because no signs indicate cash only.	shocked ( <i>negativa</i> ), no ( <i>negativa</i> )	Negativa
46	A great touch.	great ( <i>positiva</i> )	Positiva
47	They never brought a salad we asked for.	never ( <i>negativa</i> )	Negativa
48	This was like the final blow!	like ( <i>positiva</i> )	Positiva
49	Ample portions and good prices.	ample ( <i>neutra</i> ), good ( <i>positiva</i> )	Positiva
50	It's too bad the food is so damn generic.	bad ( <i>negativa</i> )	Negativa

Tabela A3 – Tabela comparativa entre o rótulo da base de dado YELP e a polaridade detectada pela rede de *frames* polarizada

Número	Frases	Rótulos	Polaridade final detectada pela rede de <i>frames</i>
1	Great brunch spot.	Positiva	Positiva
2	But now I was completely grossed out.	Negativa	Positiva
3	The pizza selections are good.	Positiva	Positiva
4	I've had better bagels from the grocery store.	Negativa	Positiva
5	I think not again	Negativa	Positiva
6	Our server was very nice, and even though he looked a little overwhelmed with all of our needs, he stayed professional and friendly until the end.	Positiva	Positiva
7	Also there are combos like a burger, fries, and beer for 23 which is a decent deal.	Positiva	Positiva
8	The selection was probably the worst I've seen in Vegas.....there was none.	Negativa	Negativa
9	I think this restaurant suffers from not trying hard enough.	Negativa	Negativa
10	Now this dish was quite flavourful.	Positiva	Positiva

11	This greedy corporation will NEVER see another dime from me!	Negativa	Negativa
12	The black eyed peas and sweet potatoes... UNREAL!	Positiva	Positiva
13	All in all, I can assure you I'll be back.	Positiva	Positiva
14	I checked out this place a couple years ago and was not impressed.	Negativa	Negativa
15	Be sure to order dessert, even if you need to pack it to-go - the tiramisu and cannoli are both to die for.	Positiva	Positiva
16	Fantastic service here.	Positiva	Positiva
17	Definitely a turn off for me & i doubt I'll be back unless someone else is buying.	Negativa	Negativa
18	It was absolutely amazing.	Positiva	Positiva
19	Food was great and so was the service!	Positiva	Positiva
20	We were sat right on time and our server from the get go was FANTASTIC!	Positiva	Positiva
21	Worse of all, he humiliated his worker right in front of me..Bunch of horrible name callings.	Negativa	Negativa
22	Never had anything to complain about here.	Positiva	Negativa
23	Nargile - I think you are great.	Positiva	Positiva
24	So we went to Tigerlilly and had a fantastic afternoon!	Positiva	Positiva
25	They have horrible attitudes towards customers, and talk down to each one when customers don't enjoy their food.	Negativa	Negativa
26	For service, I give them no stars.	Negativa	Negativa
27	To summarize... the food was incredible, nay, transcendent... but nothing brings me joy quite like the memory of the pneumatic condiment dispenser.	Positiva	Positiva
28	The sweet potato fries were very good and seasoned well.	Positiva	Positiva
29	The seasonal fruit was fresh white peach puree.	Positiva	Positiva
30	If you want healthy authentic or ethnic food, try this place.	Positiva	Positiva
31	They had a toro tartare with a cavier that was extraordinary and I liked the thinly sliced wagyu with white truffle.	Positiva	Positiva
32	I will not return.	Negativa	Negativa
33	Anyway, this FS restaurant has a wonderful breakfast/lunch.	Positiva	Positiva
34	My husband said she was very rude... did not even apologize for the bad food or anything.	Negativa	Negativa
35	I can take a little bad service but the food sucks.	Negativa	Negativa
36	this was a different cut than the piece the other day but still wonderful and tender s well as well flavored.	Positiva	Positiva
37	Sorry, I will not be getting food from here anytime soon :(	Negativa	Negativa
38	Not much seafood and like 5 strings of pasta at the bottom.	Negativa	Negativa
39	The ripped banana was not only ripped, but petrified and tasteless.	Negativa	Negativa
40	2 times - Very Bad Customer Service !	Negativa	Negativa
41	Would not go back.	Negativa	Negativa

42	Honeslty it didn't taste THAT fresh.)	Negativa	Positiva
43	The potatoes were like rubber and you could tell they had been made up ahead of time being kept under a warmer.	Negativa	Positiva
44	The fries were great too.	Positiva	Positiva
45	I was shocked because no signs indicate cash only.	Negativa	Negativa
46	A great touch.	Positiva	Positiva
47	They never brought a salad we asked for.	Negativa	Negativa
48	This was like the final blow!	Negativa	Positiva
49	Ample portions and good prices.	Positiva	Positiva
50	It's too bad the food is so damn generic.	Negativa	Negativa

## APÊNDICE B - Conjunto de dados usados no estudo de caso

Tabela B1 – Tabela contendo as 27 frases coletadas dos fóruns de dúvidas da disciplina Metodologia de Pesquisa Científica e Educacional ministrada através do Moodle no primeiro semestre de 2018. As frases foram rotuladas por professores experientes em Educação a Distância

Num	Mensagem	Rótulo professores
1	Boa noite, pra a entrega da tarefa 1, etapa 2(fazer um resumo de pesquisa quantitativa e qualitativa) precisa de capa como um trabalho?	Neutra
2	Boa noite, No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Negativa
3	O exemplo de cada tipo de pesquisa deve ser dado com uma pesquisa de própria autoria ou uma pesquisa já previamente realizada por outro autor?	Neutra
4	Bom dia, na tarefa consta que o resumo deve ser em uma página. Gostaria de saber se a referência, caso seja necessário, posso colocá-la em uma segunda página. Se possível, só para conferir, poderia informar as dimensões usadas nas margens padrão. Obrigada.	Neutra
5	Boa tarde! Estou em dúvida em relação às normas da referências bibliográficas. No arquivo "Template.doc" disponibilizado em nossa disciplina está diferente do manual de normalização também disponibilizado por vocês no espaço da nossa disciplina. Conforme enviado abaixo, os sobrenomes não estão em caixa alta e não há títulos em negrito: Boulic, R. and Renault, O. (1991) "3D Hierarchies for Animation", In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons Ltd., England. Oliveira A., Murta, L., Braganholo V. Uso de Inferência na Compreensão das Modificações em Documentos Semiestruturados. In Simpósio Brasileiro de Banco de Dados (SBBD), 2012.	Neutra
6	Olá Pessoal, Vamos fazer um grupo de WhatsApp, principalmente para definir os grupos para esta atividade. Ou caso já possuem, me adicionem. Sou de BH e estou vinculado ao polo de Timóteo. Aguardo, obrigado.	Positiva
7	Professora, Pesquisei sobre o tema objetos de aprendizagem e encontrei: "De acordo com o Learning Objects Metadata Workgroup, objetos de aprendizagem (Learning Objects) podem ser definidos por 'qualquer entidade, digital ou não digital, que possa ser utilizada, reutilizada ou referenciada durante oaprendizado suportado por tecnologias'.". Seria uma sequência didática para lecionar?	Negativa
8	Estou fazendo as tarefas, mas estou muito em dúvida se estou indo pelo caminho certo, tenho muitas dificuldades nestes trabalhos.	Negativa
9	Oi! Não sou boa em escrever, discutir um texto, eu fiz as atividades anteriores, seria bom se tivessem corrigido para ver como eu fui, bem ou muito mau. Obrigada!	Negativa
10	Boa Tarde! Gostaria de saber se minha atividade 03 está correta, tenho medo de continuar e também não fazer esta tarefa corretamente.	Negativa

11	Boa tarde Professora Fernanda, a tarefa sobre o objetivo dessa semana pode ser feita em meia lauda?	Neutra
12	Olá Professora Fernanda, Essa atividade tem que ser feita dentro do Template que já estamos usando??	Neutra
13	Boa noite... Tive um infortúnio não consegui enviar a tarefa dentro da hora solicitada (23:55)... Agora são 23:58h... Conforme o orientação no guia do aluno as atividades com entrega no atraso são avaliadas com desconto de 50% no valor da nota, porem o link para enviar a atividade não se encontra disponível. Como faço pra poder enviar a tarefa solicitada depois do prazo?	Neutra
14	Professor Francisco Ferreira boa tarde, enviei de forma errônea a tarefa 5 da etapa 7, quando do envio inverti os arquivos o que pode ser feito? Acabei de conferir e constatar o erro. Preciso de ajuda nesta questão.	Neutra
15	Boa Tarde! Teria como corrigir a tarefa 05 para ver como fui? Nas outras eu não me saí muito bem, e fiz algumas alterações nesta. Obrigada!	Negativa
16	Boa tarde professora. Na etapa 7 nos objetivos escrevi um texto único que abrange tanto o objetivo geral, quanto os específicos. Nesta etapa 8, você pede para que estes objetivos sejam separados em subtítulos, é obrigatório que eles se apresentem assim ou posso deixar em um texto único, conforme já havia feito? Obrigada.	Neutra
17	Estou usando o OFFICE 16, a numeração do template começa 1.4, não estou sabendo mudar. É para deixar.	Neutra
18	Teria possibilidade de reabrir a atividade desta semana, como foi feriado estes dias, eu me perdi um pouco. Desculpe o incômodo e obrigada!	Neutra
19	Não entendi muito bem como serão feitas essas referências. Esta semana foi muito complicado para mim, tenho consulta médica hoje, não vou conseguir ver a explicação e fazer o trabalho hoje, não consigo prorrogar o prazo para amanhã? Qualquer coisa posto o atestado da data de hoje do médico. Desculpe o incômodo.	Negativa
20	Estou fazendo o trabalho, mas estou bem confusa em relação a ordem dele, me formei tem muito tempo, o ensino mudou muito, fiquei muito tempo parada, aprendi a fazer trabalhos onde contém Capa, Introdução, Conclusão e Bibliografia, e nunca fui boa em escrever. Então eu vejo os modelos, mas tenho dificuldade em montar. Desculpe.	Negativa
21	A etapa vence dia 29 correto? pergunto pois dentro do link da tarefa fala que vence dia 22, e na parte da etapa fala que vende dia 22.	Positiva
22	Boa tarde Professora Fernanda Campos, estou com muita dificuldade em formatar o meu trabalho no template; principalmente sumário e índice, de forma a admitir minha pouca maleabilidade no word.	Negativa
23	Boa tarde professora, poderia, por favor, explicar melhor o item que fala da ferramenta de documentação de dados, não consegui compreender que ferramentas seriam estas. Obrigada.	Negativa
24	No parte de orientações sobre a metodologia foi colocado o seguinte item "A pesquisa utilizar-se-á de xxxxxxxxxxxxxx (ferramenta de documentação de dados)..." Eu quero saber quais são essas "Ferramentas de documentação de dados"?	Negativa
25	Olá professora, Posso colocar o cronograma depois das referências bibliográficas no trabalho que estamos desenvolvendo? Desse modo, faço as considerações e acrescento o cronograma no mesmo arquivo.	Neutra

26	Professora, agora fiquei confusa. Você quer que já escrevemos uma parte das considerações finais ou só os resultados que esperamos encontrar? Você fala para enviar o pdf só com as duas atividades, então é para mandar em um arquivo separado ou junto com o que já escrevemos (que está em desenvolvimento)?	Negativa
27	Boa noite professora, tudo bem ? Quando você fala "elaborar o cronograma de atividades a serem completadas" no caso seria correções ainda pendentes ou algo a ser acrescentado ? não entendi muito bem...	Negativa

Tabela B2 – Tabela identificando as *unidades lexicais* contidas nas frases, suas polaridades individuais e a polaridade final da frase

Num	Mensagem	unidades Lexicais (polaridade)	Polaridade parcial	Polaridade final
1	pra a entrega da tarefa 1, etapa 2(fazer um resumo de pesquisa quantitativa e qualitativa)precisa de capa como um trabalho ?	precisa(neutra)	[0]	Neutra
2	No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	preciso(neutra), escolher(neutra)	[0,0]	Neutra
3	O exemplo de cada tipo de pesquisa deve ser dado com uma pesquisa de própria autoria ou uma pesquisa já previamente realizada por outro autor?	deve(neutra)	[0]	Neutra
4	na tarefa consta que o resumo deve ser em uma página. Gostaria de saber se a referência, caso seja necessário, posso colocá-la em uma segunda página. Se possível, só para conferir, poderia informar as dimensões usadas nas margens padrão. Obrigada.	deve(neutra), saber(positiva), se(redutor não alcançado), posso(positiva), se(redutor alcançado) possível(positiva -> negativa), poderia(positiva)	[0,+1,0,+1,-1,+1]	Positiva



5	<p>Estou em dúvida em relação às normas da referências bibliográficas. No arquivo "Template.doc" disponibilizado em nossa disciplina está diferente do manual de normalização também disponibilizado por vocês no espaço da nossa disciplina. Conforme enviado abaixo, os sobrenomes não estão em caixa alta e não há títulos em negrito: Boulic, R. and Renault, O. (1991) "3D Hierarchies for Animation", In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley &amp; Sons Ltd., England. Oliveira A., Murta, L., Braganholo V. Uso de Inferência na Compreensão das Modificações em Documentos Semiestruturados. In Simpósio Brasileiro de Banco de Dados (SBBDD), 2012.</p>	<p>dúvida(negativa), disponibilizado(neutra), diferente(negativa), disponibilizado(neutra), não(negação não alcançada -&gt; negativa), não(negação não alcançada -&gt; negativa)</p>	<p>[-1,0,-1,0,-1,-1]</p>	<p>Negativa</p>
6	<p>Pessoal, Vamos fazer um grupo de WhatsApp, principalmente para definir os grupos para esta atividade. Ou caso já possuem, me adicionem. Sou de BH e estou vinculado ao polo de Timóteo. Aguardo, obrigado.</p>	<p>principalmente(neutra), definir(neutra), possuem(neutra)</p>	<p>[0,0,0]</p>	<p>Neutra</p>
7	<p>Professora, Pesquisei sobre o tema objetos de aprendizagem e encontrei: "De acordo com o Learning Objects Metadata Workgroup, objetos de aprendizagem (Learning Objects) podem ser definidos por 'qualquer entidade, digital ou não digital, que possa ser utilizada, reutilizada ou referenciada durante o aprendizado suportado por tecnologias' ". Seria uma sequência didática para lecionar?</p>	<p>podem(positiva), não(negação não alcançada -&gt; negativa), possa(positiva)</p>	<p>[+1,-1,+1]</p>	<p>Positiva</p>

8	Estou fazendo as tarefas, mas estou muito em dúvida se estou indo pelo caminho certo, tenho muitas dificuldades nestes trabalhos.	muito(intensificador alcançado) dúvida(negativa), se(reductor não alcançado), certo(positiva), muitas(intensificador alcançado) dificuldades(negativa)	[-2,0,+1,-2]	Negativa
9	Não sou boa em escrever, discutir um texto, eu fiz as atividades anteriores, seria bom se tivessem corrigido para ver como eu fui, bem ou muito mau. Obrigada!	não(negação alcançada) boa(positiva -> negativa), bom (positiva), se(reductor alcançado) corrigido (positiva -> negativa), bem(positiva), muito(intensificador alcançado) mau(negativa)	[-1,+1,-1,+1,-2]	Negativa
10	Gostaria de saber se minha atividade 03 está correta, tenho medo de continuar e também não fazer esta tarefa corretamente.	saber(positiva), se(reductor não alcançado -> neutra), correta(positiva), medo(negativa), não(negação alcançada) corretamente(positiva -> negativa)	[+1,0,+1,-1,-1]	Neutra
11	Professora Fernanda, a tarefa sobre o objetivo dessa semana pode ser feita em meia lauda?	pode(positiva)	[+1]	Positiva
12	Professora Fernanda, Essa atividade tem que ser feita destro do Template que já estamos usando??	tem(neutra)	[0]	Neutra
13	Tive um infortúnio não consegui enviar a tarefa dentro da hora solicitada (23:55)... Agora são 23:58h... Conforme o orientação no guia do aluno as atividades com entrega no atraso são avaliadas com desconto de 50% no valor da nota, porem o link para enviar a atividade não se encontra disponível. Como faço pra poder enviar a tarefa solicitada depois do prazo?	tive(neutra), não(negação alcançada) consegui(positiva -> negativa), não(negação alcançada) se(reductor alcançado) disponível (positiva ->negativa), poder(positiva)	[0,-1,-2,+1]	Negativa
14	Professor Francisco Ferreira enviei de forma errônea a tarefa 5 da etapa 7,quando do envio inverti os arquivos o que pode ser feito? Acabei de conferir e constatar o erro. Preciso de ajuda nesta questão.	errônea(negativa), pode(positiva), erro(negativa), preciso(neutra)	[-1,+1,-1,0]	Negativa
15	Teria como corrigir a tarefa 05 para ver como fui? Nas outras eu não me saí muito bem, e fiz algumas alterações nesta. Obrigada!	teria(neutra), corrigir(positiva), não(negação alcançado) muito(intensificador alcançado) bem(positiva -> negativa)	[0,+1,-2]	Negativa
16	professora. Na etapa 7 nos objetivos escrevi um texto único que abrange tanto o objetivo geral, quanto os específicos.	separados(neutra), obrigatório(negativa), se(reductor não alcançado), posso(positiva)	[0,-1,0,+1]	Neutra

	Nesta etapa 8, você pede para que estes objetivos sejam separados em subtítulos, é obrigatório que eles se apresentem assim ou posso deixar em um texto único, conforme já havia feito? Obrigada.			
17	Estou usando o OFFICE 16, a numeração do template começa 1.4, não estou sabendo mudar. É para deixar.	não(negação alcançada) sabendo(neutra -> negativa)	[-1]	Negativa
18	Teria possibilidade de reabrir a atividade desta semana, como foi feriado estes dias, eu me perdi um pouco. Desculpe o incômodo e obrigada!	teria(neutra), possibilidade(neutra), perdi(negativa), pouco(redutor não alcançado)	[0,0,-1,0]	Negativa
19	Não entendi muito bem como serão feitas essas referências. Esta semana foi muito complicado para mim, tenho consulta médica hoje, não vou conseguir ver a explicação e fazer o trabalho hoje, não consigo prorrogar o prazo para amanhã? Qualquer coisa posto o atestado da data de hoje do médico. Desculpe o incômodo.	não(negação alcançada) muito(intensificador) bem(positiva -> negativa), muito(intensificador alcançado) complicado(negativa), tenho(neutra), não(negação alcançada) conseguir(positiva -> negativa), não(negação alcançada) consigo(positiva -> negativa)	[-2,-2,0,-1,-1]	Negativa
20	Estou fazendo o trabalho, mas estou bem confusa em relação a ordem dele, me formei tem muito tempo, o ensino mudou muito, fiquei muito tempo parada, aprendi a fazer trabalhos onde contém Capa, Introdução, Conclusão e Bibliografia, e nunca fui boa em escrever. Então eu vejo os modelos, mas tenho dificuldade em montar. Desculpe.	bem(positiva), confusa(negativa), muito(intensificador não alcançado), muito(intensificador não alcançado), muito(intensificador alcançado) parada(neutra -> neutra), nunca(negação alcançada) boa(positiva -> negativa), tenho(neutra), dificuldade(negativa)	[+1,-1,0,0,0,-1,0,-1]	Negativa
21	A etapa vence dia 29 correto? pergunto pois dentro do link da tarefa fala que vence dia 22, e na parte da etapa fala que vende dia 22.	correto(positiva)	[+1]	Positiva
22	Professora Fernanda Campos, estou com muita dificuldade em formatar o meu trabalho no template; principalmente sumário e índice, de forma a admitir minha pouca	muita(intensificador alcançado) dificuldade(negativa), principalmente(neutra), pouca(redutor não alcançado)	[-2,0,0]	Negativa

	maleabilidade no word.			
23	professora, poderia, por favor, explicar melhor o item que fala da ferramenta de documentação de dados, não consegui compreender que ferramentas seriam estas. Obrigada.	poderia(positiva), melhor(positiva), não(negação alcançada) consegui(positiva -> negativa), compreender(positiva)	[+1,+1,-1,+1]	Positiva
24	No parte de orientações sobre a metodologia foi colocado o seguinte item "A pesquisa utilizar-se-á de xxxxxxxxxxxxxx (ferramenta de documentação de dados)..." Eu quero saber quais são essas "Ferramentas de documentação de dados"?	saber(positiva)	[+1]	Positiva
25	professora, Posso colocar o cronograma depois das referências bibliográficas no trabalho que estamos desenvolvendo? Desse modo, faço as considerações e acrescento o cronograma no mesmo arquivo.	posso(positiva)	[+1]	Positiva
26	Professora, agora fiquei confusa. Você quer que já escrevemos uma parte das considerações finais ou só os resultados que esperamos encontrar? Você fala para enviar o pdf só com as duas atividades, então é para mandar em um arquivo separado ou junto com o que já escrevemos (que está em desenvolvimento)?	confusa(negativa), separado(neutra)	[-1,0]	Negativa
27	professora, tudo bem ? Quando você fala "elaborar o cronograma de atividades a serem completadas" no caso seria correções ainda pendentes ou algo a ser acrescentado ? não entendi muito bem...	bem(positiva), completadas(positiva), correções(positiva), não(negação alcançada) muito(intensificador alcançado) bem(positiva -> negativa)	[+1,+1,+1,-2]	Positiva

Tabela B3 – Tabela comparativa entre o rótulo dos professores e a polaridade detectada pela rede de *frames* polarizada

Num	Mensagem	Rótulo Professores	Rede de Frames
-----	----------	--------------------	----------------

1	pra a entrega da tarefa 1, etapa 2(fazer um resumo de pesquisa quantitativa e qualitativa)precisa de capa como um trabalho ?	Neutra	Neutra
2	No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Negativa	Neutra
3	O exemplo de cada tipo de pesquisa deve ser dado com uma pesquisa de própria autoria ou uma pesquisa já previamente realizada por outro autor?	Neutra	Neutra
4	na tarefa consta que o resumo deve ser em uma página. Gostaria de saber se a referência, caso seja necessário, posso colocá-la em uma segunda página. Se possível, só para conferir, poderia informar as dimensões usadas nas margens padrão. Obrigada.	Neutra	Positiva
5	Estou em dúvida em relação às normas da referências bibliográficas. No arquivo "Template.doc" disponibilizado em nossa disciplina está diferente do manual de normalização também disponibilizado por vocês no espaço da nossa disciplina. Conforme enviado abaixo, os sobrenomes não estão em caixa alta e não há títulos em negrito: Boulic, R. and Renault, O. (1991) "3D Hierarchies for Animation", In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons Ltd., England. Oliveira A., Murta, L., Braganholo V. Uso de Inferência na Compreensão das Modificações em Documentos Semiestruturados. In Simpósio Brasileiro de Banco de Dados (SBBBD), 2012.	Neutra	Negativa
6	Pessoal, Vamos fazer um grupo de WhatsApp, principalmente para definir os grupos para esta atividade. Ou caso já possuem, me adicionem. Sou de BH e estou vinculado ao polo de Timóteo. Aguardo, obrigado.	Positiva	Neutra
7	Professora, Pesquisei sobre o tema objetos de aprendizagem e encontrei: "De acordo com o Learning Objects Metadata Workgroup,objetos de aprendizagem (Learning Objects) podem ser definidos por 'qualquer entidade, digital ou não digital, que possa ser utilizada, reutilizada ou referenciada durante oaprendizado suportado por tecnologias'.". Seria uma sequência didática para lecionar?	Negativa	Positiva
8	Estou fazendo as tarefas, mas estou muito em dúvida se estou indo pelo caminho certo, tenho muitas dificuldades nestes trabalhos.	Negativa	Negativa
9	Não sou boa em escrever, discutir um texto, eu fiz as atividades anteriores, seria bom se tivessem corrigido para ver como eu fui, bem ou muito mau. Obrigada!	Negativa	Negativa
10	Gostaria de saber se minha atividade 03 está correta, tenho medo de continuar e também não fazer esta	Negativa	Neutra

	tarefa corretamente.		
11	Professora Fernanda, a tarefa sobre o objetivo dessa semana pode ser feita em meia lauda?	Neutra	Positiva
12	Professora Fernanda, Essa atividade tem que ser feita dentro do Template que já estamos usando??	Neutra	Neutra
13	Tive um infortúnio não consegui enviar a tarefa dentro da hora solicitada (23:55)... Agora são 23:58h... Conforme o orientação no guia do aluno as atividades com entrega no atraso são avaliadas com desconto de 50% no valor da nota, porem o link para enviar a atividade não se encontra disponível. Como faço pra poder enviar a tarefa solicitada depois do prazo?	Neutra	Negativa
14	Professor Francisco Ferreira enviei de forma errônea a tarefa 5 da etapa 7, quando do envio inverti os arquivos o que pode ser feito? Acabei de conferir e constatar o erro. Preciso de ajuda nesta questão.	Neutra	Negativa
15	Teria como corrigir a tarefa 05 para ver como fui? Nas outras eu não me saí muito bem, e fiz algumas alterações nesta. Obrigada!	Negativa	Negativa
16	professora. Na etapa 7 nos objetivos escrevi um texto único que abrange tanto o objetivo geral, quanto os específicos. Nesta etapa 8, você pede para que estes objetivos sejam separados em subtítulos, é obrigatório que eles se apresentem assim ou posso deixar em um texto único, conforme já havia feito? Obrigada.	Neutra	Neutra
17	Estou usando o OFFICE 16, a numeração do template começa 1.4, não estou sabendo mudar. É para deixar.	Neutra	Negativa
18	Teria possibilidade de reabrir a atividade desta semana, como foi feriado estes dias, eu me perdi um pouco. Desculpe o incômodo e obrigada!	Neutra	Negativa
19	Não entendi muito bem como serão feitas essas referências. Esta semana foi muito complicado para mim, tenho consulta médica hoje, não vou conseguir ver a explicação e fazer o trabalho hoje, não consigo prorrogar o prazo para amanhã? Qualquer coisa posto o atestado da data de hoje do médico. Desculpe o incômodo.	Negativa	Negativa
20	Estou fazendo o trabalho, mas estou bem confusa em relação a ordem dele, me formei tem muito tempo, o ensino mudou muito, fiquei muito tempo parada, aprendi a fazer trabalhos onde contém Capa, Introdução, Conclusão e Bibliografia, e nunca fui boa em escrever. Então eu vejo os modelos, mas tenho dificuldade em montar. Desculpe.	Negativa	Negativa
21	A etapa vence dia 29 correto? pergunto pois dentro do link da tarefa fala que vence dia 22, e na parte da etapa fala que vende dia 22.	Positiva	Positiva

22	Professora Fernanda Campos, estou com muita dificuldade em formatar o meu trabalho no template; principalmente sumário e índice, de forma a admitir minha pouca maleabilidade no word.	Negativa	Negativa
23	professora, poderia, por favor, explicar melhor o item que fala da ferramenta de documentação de dados, não consegui compreender que ferramentas seriam estas. Obrigada.	Negativa	Positiva
24	No parte de orientações sobre a metodologia foi colocado o seguinte item "A pesquisa utilizar-se-á de xxxxxxxxxxxxxx (ferramenta de documentação de dados)..." Eu quero saber quais são essas "Ferramentas de documentação de dados"?	Negativa	Positiva
25	professora, Posso colocar o cronograma depois das referências bibliográficas no trabalho que estamos desenvolvendo? Desse modo, faço as considerações e acrescento o cronograma no mesmo arquivo.	Neutra	Positiva
26	Professora, agora fiquei confusa. Você quer que já escrevemos uma parte das considerações finais ou só os resultados que esperamos encontrar? Você fala para enviar o pdf só com as duas atividades, então é para mandar em um arquivo separado ou junto com o que já escrevemos (que está em desenvolvimento)?	Negativa	Negativa
27	professora, tudo bem ? Quando você fala "elaborar o cronograma de atividades a serem completadas" no caso seria correções ainda pendentes ou algo a ser acrescentado ? não entendi muito bem...	Negativa	Positiva

Tabela B4 – Diferença nos *frames* detectados pela rede de *frames* polarizada sem e com as melhorias

Num	Rede de Frames sem melhorias - <i>frame</i> (polaridade)	Rede de Frames com melhorias - <i>frame</i> (polaridade)
1	precisa(neutra)	precisa(negativa)
2	preciso(neutra), escolher(neutra)	preciso(negativa), escolher(neutra)
3	deve(neutra)	deve(negativa)
4	deve(neutra), saber(positiva), se(redutor não alcançado), posso(positiva), se(redutor alcançado) possível(positiva -> negativa), poderia(positiva)	deve(negativa), saber(positiva), se(redutor não alcançado), posso(positiva), se(redutor alcançado) possível(positiva -> negativa), poderia(positiva)
5	dúvida(negativa), disponibilizado(neutra), diferente(negativa), disponibilizado(neutra), não(negação não alcançada -> negativa), não(negação não alcançada -> negativa)	dúvida(negativa), disponibilizado(neutra), diferente(negativa), disponibilizado(neutra), não(negação não alcançada -> negativa), não(negação não alcançada -> negativa)

6	principalmente(neutra), definir(neutra), possuem(neutra)	principalmente(neutra), definir(neutra), possuem(neutra)
7	podem(positiva), não(negação não alcançada -> negativa), possa(positiva)	NULL
8	muito(intensificador alcançado) dúvida(negativa), se(redutor não alcançado), certo(positiva), muitas(intensificador alcançado) dificuldades(negativa)	muito(intensificador alcançado) dúvida(negativa), se(redutor alcançado) certo(positiva -> negativa), muitas(intensificador alcançado) dificuldades(negativa)
9	não(negação alcançada) boa(positiva -> negativa), bom (positiva), se(redutor alcançado) corrigido (positiva -> negativa), bem(positiva), muito(intensificador alcançado) mau(negativa)	não(negação alcançada) boa(positiva -> negativa), bom (positiva), se(redutor alcançado) corrigido (positiva -> negativa), bem(positiva), muito(intensificador alcançado) mau(negativa)
10	saber(positiva), se(redutor não alcançado -> neutra), correta(positiva), medo(negativa), não(negação alcançada) corretamente(positiva -> negativa)	saber(positiva), se(redutor alcançado) correta(positiva -> negativa), medo(negativa), não(negação alcançada) corretamente(positiva -> negativa)
11	pode(positiva)	pode(positiva)
12	tem(neutra)	tem(negativa)
13	não(negação alcançada) consegui(positiva -> negativa), não(negação alcançada) se(redutor alcançado) disponível (positiva ->negativa), poder(positiva)	tive(negativa), não(negação alcançada) consegui(positiva -> negativa), não(negação alcançada) se(redutor alcançado) disponível (positiva ->negativa), poder(positiva)
14	errônea(negativa), pode(positiva), erro(negativa), preciso(neutra)	errônea(negativa), pode(positiva), erro(negativa), preciso(negativo)
15	teria(neutra), corrigir(positiva), não(negação alcançado) muito(intensificador alcançado) bem(positiva -> negativa)	teria(negativa), corrigir(positiva), não(negação alcançado) muito(intensificador alcançado) bem(positiva -> negativa)
16	separados(neutra), obrigatório(negativa), se(redutor não alcançado), posso(positiva)	separados(neutra), obrigatório(negativa), se(redutor alcançado) posso(positiva -> negativa)
17	não(negação alcançada) sabendo(neutra -> negativa)	não(negação alcançada) sabendo(neutra -> negativa)
18	teria(neutra), possibilidade(neutra), perdi(negativa), pouco(redutor não alcançado)	teria(negativa), possibilidade(neutra), perdi(negativa), pouco(redutor não alcançado)



19	não(negação alcançada) muito(intensificador) bem(positiva -> negativa), muito(intensificador alcançado) complicado(negativa), tenho(neutra), não(negação alcançada) conseguir(positiva -> negativa), não(negação alcançada) consigo(positiva -> negativa)	não(negação alcançada) muito(intensificador) bem(positiva -> negativa), muito(intensificador alcançado) complicado(negativa), tenho(negativa), não(negação alcançada) conseguir(positiva -> negativa), não(negação alcançada) consigo(positiva -> negativa)
20	bem(positiva), confusa(negativa), muito(intensificador não alcançado), muito(intensificador não alcançado), muito(intensificador alcançado) parada(neutra -> neutra), nunca(negação alcançada) boa(positiva -> negativa), tenho(neutra), dificuldade(negativa)	bem(positiva), confusa(negativa), tem (negativa), muito(intensificador não alcançado), muito(intensificador não alcançado), muito(intensificador alcançado) parada(neutra -> neutra), nunca(negação alcançada) boa(positiva -> negativa), tenho(negativa), dificuldade(negativa)
21	correto(positiva)	correto(positiva)
22	muita(intensificador alcançado) dificuldade(negativa), principalmente(neutra), pouca(redutor não alcançado)	muita(intensificador alcançado) dificuldade(negativa), principalmente(neutra), pouca(redutor não alcançado)
23	poderia(positiva), melhor(positiva), não(negação alcançada) consegui(positiva -> negativa), compreender(positiva)	poderia(positiva), melhor(positiva), não(negação alcançada) consegui(positiva -> negativa), compreender(positiva)
24	saber(positiva)	saber(positiva)
25	posso(positiva)	posso(positiva)
26	confusa(negativa), separado(neutra)	confusa(negativa), separado(neutra)
27	bem(positiva), completadas(positiva), correções(positiva), não(negação alcançada) muito(intensificador alcançado) bem(positiva -> negativa)	bem(positiva), completadas(positiva), correções(positiva), não(negação alcançada) muito(intensificador alcançado) bem(positiva -> negativa)

Tabela B5 – Diferença de polaridades finais detectadas pela rede de *frames* polarizada sem e com as melhorias

Num	Rede de Frames sem melhorias	Rede de Frames com melhorias
1	Neutra	Negativa
2	Neutra	Negativa
3	Neutra	Negativa
4	Positiva	Positiva
5	Negativa	Negativa
6	Neutra	Neutra
7	Positiva	Neutra

8	Negativa	Negativa
9	Negativa	Negativa
10	Neutra	Neutra
11	Positiva	Positiva
12	Neutra	Negativa
13	Negativa	Negativa
14	Negativa	Negativa
15	Negativa	Negativa
16	Neutra	Neutra
17	Negativa	Negativa
18	Negativa	Negativa
19	Negativa	Negativa
20	Negativa	Negativa
21	Positiva	Positiva
22	Negativa	Negativa
23	Positiva	Positiva
24	Positiva	Positiva
25	Positiva	Positiva
26	Negativa	Negativa
27	Positiva	Positiva

Tabela B6 – Tabela comparativa entre o rótulo dos professores e o rótulo do linguista

<b>Mensagem</b>	<b>Rótulo professores</b>	<b>Rótulo linguista</b>
No resumo e preciso escolher dois artigos é fazer um resumo sobre esses um com pesquisa quantitativa e outro qualitativa, citando as fontes, ou exemplos de vários modelos pesquisa quantitativa e qualitativa na área da computação?	Neutro	Negativo
Professora, Pesquisei sobre o tema objetos de aprendizagem e encontrei: "De acordo com o Learning Objects Metadata Workgroup, objetos de aprendizagem (Learning Objects) podem ser definidos por 'qualquer entidade, digital ou não digital, que possa ser utilizada, reutilizada ou referenciada durante oaprendizado suportado por tecnologias! ". Seria uma sequência didática para lecionar?	Negativo	Negativo
Estou fazendo as tarefas, mas estou muito em dúvida se estou indo pelo caminho certo, tenho muitas dificuldades nestes trabalhos.	Neutro	Negativo
Não sou boa em escrever, discutir um texto, eu fiz as atividades anteriores, seria bom se tivessem corrigido para ver como eu fui, bem ou muito mau. Obrigada!	Neutro	Positivo
Gostaria de saber se minha atividade 03 está correta, tenho medo de continuar e também não fazer esta tarefa corretamente.	Neutro	Negativo

Teria como corrigir a tarefa 05 para ver como fui? Nas outras eu não me saí muito bem, e fiz algumas alterações nesta. Obrigada!	Positivo	Neutro
Não entendi muito bem como serão feitas essas referências. Esta semana foi muito complicado para mim, tenho consulta médica hoje, não vou conseguir ver a explicação e fazer o trabalho hoje, não consigo prorrogar o prazo para amanhã? Qualquer coisa posto o atestado da data de hoje do médico. Desculpe o incômodo.	Negativo	Positivo
Estou fazendo o trabalho, mas estou bem confusa em relação a ordem dele, me formei tem muito tempo, o ensino mudou muito, fiquei muito tempo parada, aprendi a fazer trabalhos onde contém Capa, Introdução, Conclusão e Bibliografia, e nunca fui boa em escrever. Então eu vejo os modelos, mas tenho dificuldade em montar. Desculpe.	Negativo	Negativo
Professora Fernanda Campos, estou com muita dificuldade em formatar o meu trabalho no template; principalmente sumário e índice, de forma a admitir minha pouca maleabilidade no word.	Negativo	Negativo
professora, poderia, por favor, explicar melhor o item que fala da ferramenta de documentação de dados, não consegui compreender que ferramentas seriam estas. Obrigada.	Negativo	Negativo
No parte de orientações sobre a metodologia foi colocado o seguinte item "A pesquisa utilizar-se-á de xxxxxxxxxxxxxx (ferramenta de documentação de dados)..." Eu quero saber quais são essas "Ferramentas de documentação de dados"?	Neutro	Positivo
Professora, agora fiquei confusa. Você quer que já escrevemos uma parte das considerações finais ou só os resultados que esperamos encontrar? Você fala para enviar o pdf só com as duas atividades, então é para mandar em um arquivo separado ou junto com o que já escrevemos (que está em desenvolvimento)?	Neutro	Negativo
professora, tudo bem ? Quando você fala "elaborar o cronograma de atividades a serem completadas" no caso seria correções ainda pendentes ou algo a ser acrescentado ? não entendi muito bem...	Neutro	Negativo

Tabela B7 – Diferença polaridades finais detectadas pela rede de *frames* polarizada inicial, com as melhorias e apenas com transformadores

Num	Rede de Frames	Rede de Frames com melhorias	Rede de frames apenas transformadores
1	Neutra	Negativa	Neutra
2	Neutra	Negativa	Neutra
3	Neutra	Negativa	Neutra
4	Positiva	Positiva	Positiva

5	Negativa	Negativa	Negativa
6	Neutra	Neutra	Neutra
7	Positiva	Neutra	Neutra
8	Negativa	Negativa	Negativa
9	Negativa	Negativa	Negativa
10	Neutra	Neutra	Negativa
11	Positiva	Positiva	Positiva
12	Neutra	Negativa	Neutra
13	Negativa	Negativa	Negativa
14	Negativa	Negativa	Negativa
15	Negativa	Negativa	Negativa
16	Neutra	Neutra	Negativa
17	Negativa	Negativa	Negativa
18	Negativa	Negativa	Negativa
19	Negativa	Negativa	Negativa
20	Negativa	Negativa	Negativa
21	Positiva	Positiva	Positiva
22	Negativa	Negativa	Negativa
23	Positiva	Positiva	Positiva
24	Positiva	Positiva	Positiva
25	Positiva	Positiva	Positiva
26	Negativa	Negativa	Negativa
27	Positiva	Positiva	Positiva