



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**INFERÊNCIA DE PERSONALIDADE A PARTIR DE
TEXTOS DE REDE SOCIAL UTILIZANDO UM LÉXICO
AFETIVO EM PORTUGUÊS BRASILEIRO**

ANTONIO ALIBERTE DE ANDRADE MACHADO

São Cristóvão

2016

ANTONIO ALIBERTE DE ANDRADE MACHADO

**INFERÊNCIA DE PERSONALIDADE A PARTIR DE
TEXTOS DE REDE SOCIAL UTILIZANDO UM LÉXICO
AFETIVO EM PORTUGUÊS BRASILEIRO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da computação da Universidade Federal de Sergipe (PROCC-UFS), como requisito parcial para obtenção do título de Mestre em Ciência da computação.

Área de Concentração: Computação Inteligente

Orientadora: Prof. Dr^a Maria Augusta Silveira Netto Nunes
Co-Orientador: Prof. Dr. Thiago Alexandre Salgueiro Pardo

São Cristóvão

2016

ANTONIO ALIBERTE DE ANDRADE MACHADO

**INFERÊNCIA DE PERSONALIDADE A PARTIR DE
TEXTOS DE REDE SOCIAL UTILIZANDO UM LÉXICO
AFETIVO EM PORTUGUÊS BRASILEIRO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da computação da Universidade Federal de Sergipe (PROCC-UFS), como requisito parcial para obtenção do título de Mestre em Ciência da computação.

Área de Concentração: Computação
Inteligente

BANCA EXAMINADORA

Prof.^a Dr.^a Maria Augusta Silveira Netto Nunes, Presidente
Universidade Federal de Sergipe (PROCC/UFS)

Prof. Dr. Rogério Patrício Chagas do Nascimento, Membro
Universidade Federal de Sergipe (PROCC/UFS)

Prof. Dr. Henrique Nou Schneider, Membro
Universidade Federal de Sergipe (PPGED/UFS)

Prof.^a Dr.^a Magali Teresinha Longhi, Membro
Universidade Federal do Rio Grande do Sul (PPGIE/UFRGS)

Dedico este trabalho ao meu Deus por ter me dado equilíbrio e concentração nos momentos mais difíceis e por ter me guiado e orientado quando mais necessitei, sem a sua ajuda nada disso seria possível.

Agradecimentos

No fechar deste ciclo de aprendizagem gostaria de particularizar os agradecimentos a todos aqueles que, de forma direta ou indireta, fizeram parte desta dissertação ajudando-me a ultrapassar as dificuldades surgidas, tornando mais simples a passagem pelas diferentes fases, contribuindo para que esta dissertação se tornasse realidade.

Em primeiro lugar, quero agradecer a Deus por ter me guiado e me dado à força necessária para superar todas as barreiras e obstáculos.

A minha esposa, Suely Cristina Silva Souza e minha filha, Amanda Cristina Souza por estarem sempre do meu lado nestes anos de estudo. Agradeço a paciência e compreensão que tiveram comigo, por entenderem quando eu estava centrado e pensando somente no objeto ou fazendo outras atividades no decorrer do mestrado. Amo vocês!

Aos meus pais, Antonio Aliberte Machado e Lídia Conceição Rebelo de Andrade por entenderem as minhas ausências e me ensinarem a ser um ser humano correto e honesto.

A meu irmão, Antonio Sinval de Andrade Machado pela compreensão e entendimento nos momentos que me fiz ausente.

A minha orientadora, a professora Dr^a Maria Augusta Silveira Netto Nunes, a quem devo orientação científica desta dissertação, pelos seus ensinamentos, pelo seu apoio e confiança depositada, pelas críticas e sugestões, por toda paciência em entender as minhas limitações, por ter me mostrado a maneira correta de se fazer ciência e por me ensinar que sempre temos algo a melhorar.

Ao meu co-orientador professor Dr. Thiago Alexandre Salgueiro Pardo, pelas ajudas e orientações constantes ao longo da concretização desta dissertação e pelo apoio que em muito contribuiu para que todo este processo se traduzisse numa aprendizagem estimulante e enriquecedor.

A professora Dr^a Magalí Teresinha Longhi pelos incentivos, críticas e elogios. Pelas orientações, direções e caminhos. A senhora é uma das responsáveis por essa conquista.

A meu amigo Natan do Nascimento, pela ajuda e apoio incansável. Meu amigo, você foi fundamental neste sonho. Obrigado!

Aos meus amigos e companheiros professores do IFS, Alan Sá, Ana Carla, Davi Carnaúba, Jaziel Lobo, Rafael Jacaúna, Marcos Aurélio e Saulo Galileo pelos incentivos e conselhos.

Aos meus amigos da Fundação Aperipê, Jefferson Andrade, Robson Santos, Humberto Callado e Zailson Júnior, obrigado por me compreenderem e me incentivarem.

A todos aqueles que fazem parte da coordenação do PROCC, em especial ao professor Dr. Rogério Patrício Chagas do Nascimento, por sempre estar sempre a disposição a solucionar todas as pendências relacionadas ao curso e guiar todos os alunos a bons congressos e caminhos.

Aos meus companheiros de mestrado, em especial a Rafael Santos e Clebeson Canuto, pelos momentos de discussões e debates nas soluções das atividades propostas.

Enfim, uma pesquisa não se faz sozinho, como também não se constrói uma história sem personagens. Alguns são coadjuvantes; outros, atores principais. Alguns colaboram com a trama; outros, criam conosco o enredo. Esta dissertação faz parte da minha história e foi minuciosamente desenhada e vivenciada em todos os seus capítulos.

*“Eu acredito na intuição e na inspiração.
A imaginação é mais importante que o
conhecimento. O conhecimento é limitado,
enquanto a imaginação abraça o mundo inteiro,
estimulando o progresso, dando à luz à
evolução. Ela é, rigorosamente falando, um
fator real na pesquisa científica.”*

*Albert Einstein
Sobre Religião Cósmica e Outras Opiniões e
Aforismos (1931)*

Resumo

A presente dissertação apresenta estudos sobre a correlação de informações léxicas em textos em Português com características de personalidade do modelo *Big Five* e as facetas do IPIP-NEO. Discorre-se, especialmente, sobre o uso das classes de sentimentos afetivos do léxico *LIWC*. O objetivo principal é correlacionar os fatores do *Big Five* e as facetas do IPIP-NEO dos questionários IPIP-NEO 120 e TIPI com *posts* da rede social *Facebook*. Para isso, foi construído o *Personalitatem Lexicon*. O resultado do experimento mostra que a inferência de personalidade feita pelos questionários é mais preciso pelo mesmo conter perguntas e respostas específicas para aferir tal característica. Já a inferência feita através de *posts* é mais susceptível a ruídos porque nem todas as situações vividas são expostas na rede social. Contudo, a inferência de personalidade por *posts* é possível, mas os resultados da inferência são retratados daquele momento.

Palavras-chave: Personalidade; *Big Five*; Facetas; Léxico.

Abstract

This master thesis studies on the correlation of lexical information in texts in Portuguese with model personality feature and Big Five facets of IPIP-NEO. It elaborates especially on the use of classes of affective feelings LIWC dictionary. The main goal is to correlate the factors of the Big Five and facets of IPIP-NEO of IPIP-NEO questionnaires 120 and TIPI with posts of Facebook social network. For this, the Personalitatem Lexicon was built. The result of the experiment shows that the personality of inference made by questionnaires is more accurate for the same contain questions and answers specific to assess such a feature. Already the inference made by posts is more susceptible to noise because not all experienced situations are exposed on the social network. However, the personality of inference through posts is possible, but its results are depicted at the time.

Keywords: Personality; Big Five; Facets, Lexicon

Lista de Figuras

| | |
|--|----|
| Figura 1: Modelo Hierárquico de Personalidade | 23 |
| Figura 2: Estrutura de construção da base Wordnet AffectBR | 39 |
| Figura 3: Pesquisadores que publicaram mais artigos..... | 57 |
| Figura 4: Universidades que mais publicaram | 57 |
| Figura 5: Países da universidade do autor | 58 |
| Figura 6: Quantidade de artigos publicados por ano. | 59 |
| Figura 7: Tela de cadastro/acesso | 68 |
| Figura 8: Questão 1 do questionário IPIP-NEO 120 (JOHNSON, 2000b) | 70 |
| Figura 9: Resumo do Resultado do questionário IPIP-NEO 120 de um participante. ... | 70 |
| Figura 10: Questão 1 do questionário TIPI (JOHNSON, 2000b) | 71 |
| Figura 11: Resumo do Resultado do questionário TIPI de um participante..... | 71 |

Lista de Abreviaturas e Siglas

| | |
|-----------------|--|
| LIWC | <i>Linguistic Inquiry e Word Count</i> |
| CA | Computação Afetiva |
| IA | Inteligência Artificial |
| PLN | Processamento de Linguagem Natural |
| NILC | <i>Interinstitutional Center for Computational Linguistics</i> |
| IPIP-NEO | <i>Neo International Personality Item Pool</i> |
| OCC | <i>Ortony, Clore e Collins</i> |
| 16-PF | <i>16 Personality Factors Questionnaire</i> |
| ROCI-II | <i>Rahim Organizational Conflict Inventory – II</i> |
| MMPI | <i>Minnesota Multiphasic Personality Inventory</i> |
| NEO-PI-R | <i>Revised NEO Personality Inventory</i> |
| FFPI | <i>Five Factor Personality Inventory</i> |
| BFQ | <i>Big Five Questionnaire</i> |
| GPI | <i>Global Personality Inventory</i> |
| ANEW | <i>Affective Norms for English Words</i> |
| SAM | <i>Self Assessment Manikin</i> |
| TCC | <i>The cognitive and Communication Technologies</i> |
| QP | Questões de Pesquisa |
| OP | <i>Opinion Mining</i> |

| | |
|-------------|---------------------------------|
| SA | <i>Sentiment Analyses</i> |
| SVM | <i>Support Vector Machines</i> |
| NB | <i>Naive Bayes</i> |
| IFS | Instituto Federal de Sergipe |
| UFS | Universidade Federal de Sergipe |
| UNIT | Universidade Tiradentes |
| TPF | Traço de Personalidade e Faceta |

Lista de Símbolos

| | |
|-----------|---|
| r | Coeficiente de Correlação de Pearson |
| n | Número de elementos no vetor |
| x | Vetor de valores dos dados reais |
| y | Vetor dos valores dos dados obtidos |
| i | Representa o i -ésimo elemento do vetor |
| \bar{x} | Média dos valores do vetor x |
| \bar{y} | Média dos vetores de y . |
| w | Palavras marcadas como afetivas |
| p | Escala de significação da palavra |
| qw | Quantidade de palavras afetivas |

Lista de Tabelas

| | |
|---|----|
| Tabela 1: Dimensões do modelo <i>Big Five</i> | 30 |
| Tabela 2: Facetas do IPIP-NEO..... | 31 |
| Tabela 3: Classificação utilizando o léxico <i>EmoSenticNet</i> | 41 |
| Tabela 4: Questões de Pesquisa..... | 45 |
| Tabela 5: Estudos primários retornados a partir das etapas propostas | 47 |
| Tabela 6: Artigos analisados | 48 |
| Tabela 7: Códigos atribuídos a cada fator do <i>Big Five</i> | 62 |
| Tabela 8: Códigos atribuídos a cada faceta do IPIP-NEO..... | 62 |
| Tabela 9: Exemplo de palavras com facetas e pesos | 64 |
| Tabela 10: Quantidade de participantes (homens e mulheres) | 68 |
| Tabela 11 - Valores de referência para a interpretação do coeficiente de correlação Pearson. | 73 |
| Tabela 12: Correlações entre IPIP-NEO 120, TIPI e <i>posts</i> do <i>Facebook</i> | 74 |
| Tabela 13: Correlação entre Facetas do IPIP-NEO 120 e <i>posts</i> do <i>Facebook</i> | 75 |

Sumário

| | |
|--|----|
| 1 INTRODUÇÃO..... | 17 |
| 1.1 Contextualização | 17 |
| 1.2 Motivação | 20 |
| 1.3 Problema..... | 20 |
| 1.4 Hipótese..... | 20 |
| 1.5 Objetivos..... | 20 |
| 1.6 Estruturação do Documento | 21 |
| 2 COMPUTAÇÃO AFETIVA | 22 |
| 2.1 Subjetividade em Texto | 23 |
| 2.1.1 Afeto | 24 |
| 2.1.2 Sentimento (<i>Sentiment</i>)..... | 24 |
| 2.1.3 Emoção | 25 |
| 2.1.4 Estado de Ânimo (Humor) | 27 |
| 2.2 Personalidade..... | 27 |
| 2.3 Traços de personalidade..... | 28 |
| 2.4 <i>Big Five</i> | 29 |
| 2.5 Formas de Inferir Personalidade..... | 31 |
| 2.6 Considerações Finais | 33 |
| 3 LÉXICOS | 34 |
| 3.1 Léxicos Computacionais..... | 34 |
| 3.1.1 <i>MRC Psycholinguistic DataBase</i> | 36 |
| 3.1.2 <i>WordNet</i> | 36 |
| 3.1.3 <i>VerbNet</i> | 36 |
| 3.2 Léxicos Afetivos | 37 |
| 3.2.1 <i>LIWC</i> | 37 |
| 3.2.2 <i>ANEW</i> | 38 |
| 3.2.3 <i>SentiWordNet</i> | 38 |
| 3.2.4 <i>Wordnet AffectBR</i> | 39 |
| 3.2.5 <i>SentiStrength</i> | 40 |
| 3.2.6 <i>SenticNet</i> | 40 |
| 3.2.7 <i>EmoSenticNet</i> | 41 |
| 3.2.8 <i>ANEW-Br</i> | 41 |
| 3.2.9 <i>OpinionLexicon</i> | 42 |
| 3.2.10 <i>WordnetAffectBR_adapt</i> | 42 |
| 3.2.11 <i>VerbNet.Br</i> | 43 |

| | |
|--|-----|
| 3.2.12 <i>Reli-Lex</i> | 43 |
| 3.3 Considerações Finais | 44 |
| 4 TRABALHOS RELACIONADOS | 45 |
| 4.1 Escopo de Busca e Questões de Pesquisa..... | 45 |
| 4.2 Estudos Primários | 46 |
| 4.3 Artigos relevantes do banco de dados Scopus | 47 |
| 4.4 Resultados e Discussões | 47 |
| 4.5 Considerações Finais | 59 |
| 5 CONSTRUÇÃO DO <i>PERSONALITATEM LEXICON</i> | 60 |
| 5.1 Algoritmo de Correlação Facetas <i>LIWC</i> | 63 |
| 5.2 Exemplos | 64 |
| 5.2.1 Exemplo Listagem_de_palavras como matriz..... | 64 |
| 5.2.2 Exemplo de Palavras (calculado_sobreposição_de_significado)..... | 64 |
| 5.3 Considerações Finais | 66 |
| 6 EXPERIMENTOS E ANÁLISES DE RESULTADOS..... | 67 |
| 6.1 Participantes | 67 |
| 6.2 <i>Personalitatem Inventory</i> | 67 |
| 6.3 Análises de Resultados | 72 |
| 6.5 Considerações Finais | 76 |
| 7 CONSIDERAÇÕES FINAIS | 77 |
| 7.1 Contribuições..... | 78 |
| 7.2 Limitações | 78 |
| 7.3 Trabalhos Futuros | 79 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 80 |
| Apêndice A - Referências Bibliográficas do Mapeamento Sistemático | 93 |
| Apêndice B – Termos de Consentimento Livre, Esclarecido e Condições de Uso | 99 |
| Apêndice C - Instruções para responder os questionários de pesquisa | 101 |
| Apêndice D - Questões IPIP-NEO 120 | 102 |
| Apêndice E - Questões TIPI | 105 |

1 INTRODUÇÃO

Esta dissertação parte da possibilidade da inferência de personalidade através de textos na rede social *Facebook*. Para conseguir inferir personalidade, serão abordadas algumas etapas como, por exemplo, aplicações de questionários e análises dos *posts* dos usuários participantes. Para realização das análises, será proposto a construção de um léxico afetivo, construído a partir do *LIWC - Linguistic Inquiry e Word Count*, que através deste, será possível inferir personalidade.

A seguir, descreve-se o contexto dos estudos realizados para o desenvolvimento da dissertação e como consequência apresentam-se os resultados.

1.1 Contextualização

Diante dos estudos (JOHNSON, 2000a; 2000b) realizados sobre personalidade, formas de inferência de personalidade vêm chamando atenção dos pesquisadores. Essas formas são representadas, por exemplo, usando o modelo *Big Five*, pelos traços de personalidade e recentemente com surgimento de facetas (JOHNSON, 2000b). Na Ciência da Computação, a área da Computação Afetiva (CA) (do inglês, *Affective Computing*) agrega fatores afetivos em dispositivos computacionais para reconhecer, modelar e responder às emoções e personalidade (NUNES, 2008), podendo expressar afetividade através da interface computacional (PICARD, 1997). No entanto, há poucos trabalhos em CA voltados para a identificação de traços de personalidade.

O computador, entre outras finalidades, é utilizado para o diálogo e a interação entre pessoas e dessas com sistemas computacionais, instigando estudos no ramo da Inteligência Artificial (IA). No caso da CA, um sub-ramo da IA, os sistemas computacionais analisam a afetividade do usuário para determinar a personalidade, o sentimento e as emoções.

Autores como Allport (1937), Schultz (1990), Nunes (2008), Lisetti (2002), Liu *et al.* (2003), Ma *et al.* (2005), Wang *et al.* (2005) estudam formas de reconhecimento da personalidade. Lisetti (2002), por exemplo, cita que o modelo hierárquico de personalidade é formado por afeto, estado de ânimo e emoção de um usuário. Já Liu *et al.* (2003), Ma *et al.* (2005) e Wang *et al.* (2005) pesquisam o propósito de reconhecer e expressar emoções por meio de padrões. Esses padrões são observados através das expressões faciais, ações de comportamento corporal e sociológico, entonação vocal e sinais fisiológicos (PICARD, 1997).

Outra forma de reconhecimento da afetividade em sistemas computacionais é a análise automática dos documentos escritos em linguagem natural. Para isso, são

utilizadas técnicas conhecidas como Análise de Sentimento (do inglês, “*Sentiment Analysis*”) ou Mineração de Opinião (do inglês, “*Opinion Mining*”), fundamentados em categorias de emoção, com frequência classificadas em “positiva” ou “negativa” (GREGORY *et al.* 2006).

Pesquisas na área da análise de linguagem natural empregam processos de classificação automática dos aspectos subjetivos do discurso (textos ou conversas), tais como emoções e sentimentos. Essas pesquisas reúnem as áreas da CA e do Processamento de Linguagem Natural (PLN) para desenvolver métodos, técnicas e recursos que, integrados, tornam sistemas computacionais capazes de manipular o significado desses aspectos em um discurso. A CA se baseia nas teorias da Psicologia para incorporar afetividade em sistemas computacionais de modo a torná-los mais personalizados. A PLN apresenta uma característica multidisciplinar ao integrar Linguística e Ciência da Computação com a finalidade de apresentar soluções computacionais para o reconhecimento e representação de informações contidas em textos.

Na PLN, há grandes esforços na construção de léxicos computacionais. Segundo Specia e Nunes (2004), léxicos computacionais são recursos lexicais criados, geralmente de forma manual, para o tratamento computacional. São também chamadas de bases de dados lexicais (*Lexical Databases*). De acordo com Jurafsky (2000), um léxico pode ser entendido como uma estrutura altamente sistemática que define o significado das palavras e como elas podem (e devem) ser usadas. Cada palavra, de um léxico, é identificada como um lexema. O lexema consiste do formato ortográfico e fonológico com uma forma de representação de significado.

Os autores, Liu *et al.* (2003), Ma *et al.* (2005), apresentam o uso de um léxico computacional para reconhecimento de emoções baseadas em texto. Da mesma forma, Pasqualotti e Vieira (2008) e Longhi (2011) desenvolveram seus trabalhos baseados em léxicos afetivos na identificação de afetividade a partir de mensagens registradas em ambientes virtuais. Ortony *et al.* (1987) explicam e apontam a importância de um léxico afetivo não conter somente palavras que se referem diretamente a emoções, a exemplo de amor e ódio, mas também, conter outras categorias de palavras, como por exemplo, raiva, tristeza, alegria, orgulho, vergonha. Sendo assim, outras formas de reconhecimento, através de léxicos afetivos, estariam surgindo para fins de pesquisas e explorações, a exemplo da personalidade.

Para entender a importância da inferência de personalidade para uso em sistemas computacionais, se faz necessário compreender o que é personalidade (FUNDER, 2001). De acordo com Schultz (1990), a origem em latim da palavra personalidade vem de *Persona*, ou da máscara usada por um ator para a encenação de uma peça teatral. Schultz ainda amplia sua definição descrevendo personalidade como “um conjunto permanente e exclusivo de características identificáveis nas ações/interações do indivíduo em diferentes situações”.

Autores como McCrae e John (1992), Costa e McCrae (1992), Johnson (2000b), Hendricks *et al.* (2002), Barbaranelli e Caprara (2002) e Schimit *et al.* (2002) criaram questionários para inferir personalidade. Tais experiências, de alguma maneira, colaboram com a sistemática e o desenvolvimento da CA (PASQUALI, 1999, 2001; WECHSLER, 1999, 2001), podendo ser replicadas em sistemas computacionais. No trabalho de Pasquali (2001), por exemplo, é oferecido um extenso manual de orientações sobre as técnicas de exame psicológico. O autor aborda o conceito dos questionários, a história, os tipos, os fundamentos científicos, os parâmetros psicométricos, os princípios éticos e os questionários de personalidade comercializados no Brasil. Diferentemente de Pasquali, Wechsler (1999, 2001) estrutura orientações fundamentais que porventura são desconhecidas de alguns profissionais para uma boa aplicação, além de discutir alguns princípios de construção de questionários psicológicos.

As análises destes questionários foram feitas de maneira semiautomática ou automática dependendo do algoritmo ou técnica utilizada pelos autores. Um fato importante é que todos eles utilizam léxicos, corpus, corpora, database ou dicionários, todos afetivos, próprios ou já existentes, em diversas línguas, menos em português.

A personalidade não é somente um aspecto visível e físico de uma pessoa. Ela é relativamente sólida e, muitas vezes, previsível, porém não é impreterivelmente rigorosa e fixa. A personalidade, geralmente, permanece estável por um período de 45 anos iniciando na fase adulta, relata Soldz e Vaillant (1998) e medida através de instrumentos conhecidos por questionários. Uma abordagem bastante interessante é a abordagem de traços de personalidade que permite diferenciar psicologicamente pessoas usando traços mensuráveis, conceituáveis e possíveis de serem implementados em computadores (NUNES e HU, 2012; NUNES e CAZELLA, 2011).

A personalidade é um complexo de atributos que caracterizam um indivíduo. Goldberg (1992) formaliza a personalidade, segundo a teoria de traços, em cinco fatores conhecidos como *Big Five*. Os cinco fatores que o *Big Five* contempla são: Abertura, Neuroticismo, Extroversão, Socialização e Realização (ver detalhes na seção 2.4).

Dentre os cinco fatores, o Neuroticismo desempenha um papel importante nas redes sociais. Ao estudar redes sociais, Kanfer e Tanaka (1993) relatam a existência de interações de indivíduos considerados “inseguros”. Van Zalk *et al.* (2011) comenta que esses indivíduos são socialmente ansiosos, têm menos amigos na sua rede e tendem a escolher aqueles que também são socialmente ansiosos. Sendo assim, os traços de personalidade indicam o grau de adequação às atividades impostas, ou seja, se a pessoa não for flexível, por exemplo, terá dificuldade de se adequar à determinada situação (REEVES e NASS, 1996).

Observa-se que os traços de personalidade inferidos a partir de questionários pelo *Big Five* vêm sendo largamente utilizados nas pesquisas da CA (JOHN e

SRIVASTAVA, 1999; GOLDBERG *et al.* 2006; NUNES, 2008). Entretanto, os trabalhos relacionados ao reconhecimento dos traços de personalidade ou de aspectos afetivos a partir de textos em língua portuguesa estão apenas começando (PASQUALOTTI e VIEIRA, 2008; LONGHI, 2011; BALAGE FILHO *et al.* 2013). Existem poucos léxicos em português do Brasil, voltados a essa finalidade, de forma a comparar os resultados das análises.

1.2 Motivação

Vários foram os fatores que conduziram à presente dissertação. Num primeiro momento, pode ser destacado o interesse do pesquisador, oriundo da área da Ciência da Computação, em CA como uma área de tomada de decisão computacional. Foram analisadas as dificuldades enfrentadas por um professor para motivar seu aluno. Com a leitura e estudo em ambientes virtuais baseados em traços de personalidade (LONGHI, 2011) nota-se que ao traçar o perfil de um aluno, o professor consegue traçar uma melhor rota de ensino/aprendizagem. Além disso, estudos como Schultz (1990), Lisetti (2002) e Nunes (2008) têm comprovado como a personalidade tem sido fundamental e determinante na tomada de decisão.

1.3 Problema

Como visto acima, Allport (1937), Schultz (1990) e Johnson (2000a; 2000b) são exemplos de psicólogos que estudam formas de inferência de personalidade. O detalhe que chama atenção é que eles utilizam *corpus*, corpora, léxicos ou database. A maioria dessas bases de dados foram desenvolvidas na língua inglesa. Sendo assim, o problema de pesquisa é a inexistência de formas de inferência de personalidade a partir de bases de dados desenvolvidas em português brasileiro.

1.4 Hipótese

Como hipótese da dissertação, acredita-se que, a partir de um léxico afetivo, é possível, inferir personalidade a partir de textos de rede social.

1.5 Objetivos

Como objetivo geral, é proposto analisar formas de como identificar traços de personalidade de um sujeito a partir da mineração da subjetividade de textos.

Em linhas gerais, esta dissertação tem três objetivos específicos a serem cumpridos:

- Readaptar o léxico *LIWC*¹ (PENNEBAKER e KING, 1999; TAUSCZIK e PENNEBAKER, 2010), traduzido pelo grupo do NILC², para que o mesmo contenha aspectos afetivos e características de personalidade, baseadas no modelo do *Big Five* e nas facetas do IPIP-NEO (*Neo-International Personality Item Pool*) (JOHNSON, 2005). Essa readaptação dará origem a um léxico afetivo chamado de *Personalitatem Lexicon*.
- Validar o *Personalitatem Lexicon* a partir da participação de usuários da rede social *Facebook*.

1.6 Estruturação do Documento

O conteúdo desta dissertação está estruturado em outros 6 capítulos, além deste. No próximo capítulo, são apresentados os conceitos utilizados nesta dissertação, que incluem definição de CA, termos subjetivos como traços de personalidade, afeto, sentimento, emoção e humor. No caso de traços de personalidade, os conceitos e as características de cada um deles são descritos a partir do modelo *Big Five* como também é evidenciado as formas de como inferi-los.

No capítulo 3, os léxicos computacionais e afetivos são descritos a partir de trabalhos encontrados na literatura.

No capítulo 4 foi desenvolvido um mapeamento sistemático. A base de dados para as investigações foi a SCOPUS. Para as buscas dos trabalhos relacionados foram montadas questões de pesquisa. Foi montada uma tabela com as respostas das questões de pesquisa que foram analisadas posteriormente.

No capítulo 5, descreve-se a metodologia da readaptação do léxico *LIWC*. É apresentado todos os passos de reconstrução, modos e métodos utilizados. Além disso, apresenta-se o algoritmo de correlação utilizado com exemplos de classificações.

No capítulo 6 são descritos os experimentos, os métodos avaliativos e as análises dos resultados. Neste capítulo é apresentado a forma de escolha dos participantes, detalhes do portal onde estão disponíveis os questionários IPIP-NEO 120, 300 e o TIPI e as correlações dos resultados dos questionários com as extrações dos *posts* do *Facebook* dos participantes.

Por fim, no capítulo 7, são apresentadas as considerações finais da presente dissertação de mestrado, juntamente com as contribuições, limitações e trabalhos futuros.

¹ *LIWC* – Termo definido na seção 3.2.1.

² NILC - *Interinstitutional Center for Computational Linguistics – Grupo de Pesquisa de São Carlos – SP, Brasil.*

2 COMPUTAÇÃO AFETIVA

O campo da Inteligência Artificial (IA) que trata formas de inferir personalidade em computadores é chamada de CA. Segundo Hassin *et al.* (2004) o surgimento da CA se deu pela necessidade de fazer com que os computadores pudessem interagir de forma direta, como pensar, receber e transmitir a personalidade das pessoas. Eles enfatizam a CA como uma linha de pesquisa que analisa a possibilidade de como sistemas computacionais podem identificar, classificar e contestar a personalidade humana além de agrupar conhecimentos em outras áreas, como por exemplo, a psicologia e a ciência cognitiva.

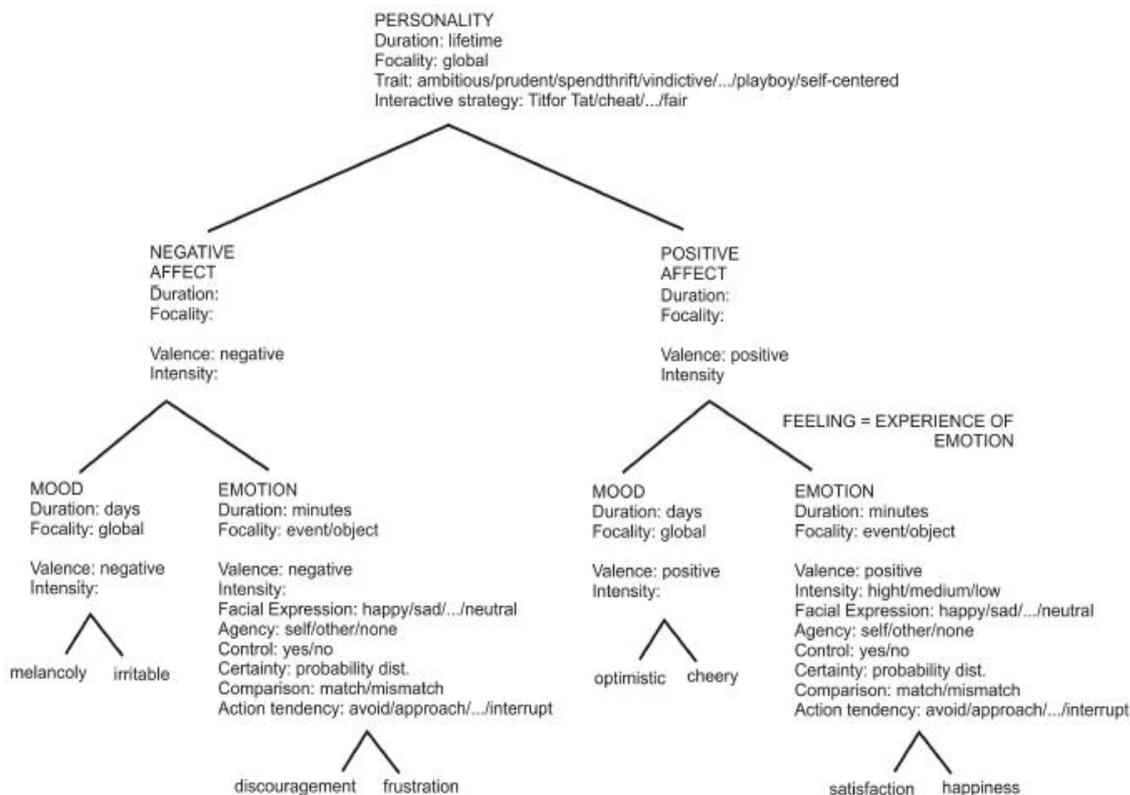
Cientistas como Picard (1997), Trappl *et al.* (2003), Thagard (2006) e Nunes *et al.* (2009) comprovam a importância da CA em sistemas computacionais para modelar e implementar formas de inferir personalidade em computadores. Para tal, a CA passou a investigar como os computadores poderiam modelar, reconhecer e responder aos comportamentos humanos e, assim, como expressá-los através de uma interface/interação computacional. A finalidade de se promover essa caracterização afetiva é colaborar para o aumento da consistência, coerência e credibilidade das reações e respostas computacionais providas durante a interação humana via interface humano-computador (PICARD, 1997).

Aperfeiçoando os estudos, Reeves e Nass (1996) analisam os comportamentos das pessoas por meio de agentes computacionais. Para fundamentar e criar esses agentes, os autores utilizam os modelos de psicólogos, como Ortony *et al.* (1990), Roseman *et al.* (1990), Damásio (1996), Scherer (2000a) e Roseman (2001), sendo assim, estes agentes são criados contendo características emotivas. Para tal, os psicólogos tem usado a CA para aperfeiçoar tais características, como personalidade e emoção. Essas características contribuem para a consistência, coerência e previsão da reação emocional em respostas de computadores. A finalidade deste aperfeiçoamento é fazer com que estes agentes interajam como se fossem humanos. Para Rousseau e Hayes-Roth (1996) a personalidade de um agente desenvolve motivações e acredita-se que isso é interessante para quem os utiliza. André *et al.* (2000) concorda quando afirma que um agente, para ser mais convincente, precisa incorporar um modelo de personalidade.

Ryff e Keyes (1995) proporcionaram um modelo composto por seis elementos, (auto avaliação, crescimento pessoal, sentido de vida, relações positivas com outras pessoas, domínio do ambiente e autonomia) distintos e com múltiplas dimensões, que avaliam a personalidade e o perfil de uma pessoa. Já Lisetti (2002) monta um esquema, baseado no modelo proposto por Ortony *et al.* (1990) conhecido por modelo OCC, para representar a personalidade. Esse esquema também pode ser utilizado na construção de agentes artificiais socialmente inteligentes. Na Figura 1 é apresentada que a personalidade é composta por afeto, emoção e humor. A personalidade, que está no

topo, mostra que agentes com personalidades diferentes podem experimentar todas as possíveis emoções. Já afeto, emoção e humor, que tem relação com ações e metas dos agentes, estão na parte inferior, e não implicam necessariamente na personalidade.

Figura 1: Modelo Hierárquico de Personalidade



Fonte: Lisetti (2002)

A personalidade humana é considerada uma característica psicológica importante no processo de tomada de decisão (DAMÁSIO, 1994), (DAMÁSIO, 1999), (SIMON, 1983), (GOLEMAN, 1995), (PAIVA, 2000), (PICARD, 1997), (PICARD, 2000), (SCHERER, 2000a), (ROSEMAN, 2001), (PICARD 2002), (TRAPPL, 2003) e (THAGARD, 2006). Como personalidade implica em afeto, emoção e humor, cientistas em CA como, por exemplo, Reeves e Nass (1996) e Bercht (2006), foram incorporando características de personalidade na modelagem dos seus agentes, tornando-os mais reais passíveis de melhores análises e busca por melhores resultados.

2.1 Subjetividade em Texto

A subjetividade é uma característica individual e “não é aberta à observação objetiva ou de verificação” (WIEBE, 1990). Sendo assim, a subjetividade tem uma grande importância no contexto e é muito difícil de ser estudada, comparada e interpretada, pois a maioria dos estudos da área tem sido direcionada somente para detecção de valência (positiva, neutra e negativa) (PANG e LEE, 2008). As pessoas são

capazes de distinguir linguisticamente (e, em particular, léxico) o que e como está sendo abordado um assunto no corpo de um texto. Contudo, a habilidade de identificar, computacionalmente, a subjetividade em um texto é tarefa árdua, mas fundamentalmente importante para classificar um texto de forma precisa. Segundo Munezero *et al.* (2014) e Lisetti (2002), a subjetividade humana está fortemente associada aos termos afeto, sentimento, emoção, personalidade e estado de ânimo.

2.1.1 Afeto

Segundo Batson *et al.* (1992) afeto é um termo mais geral e o descreve como sendo um conceito mais primitivo que emoções. Exemplifica sua tese ao dizer que “o afeto está presente em um latido de cão ou no choro de uma criança”. Sendo assim, Shouse (2005), diz que o afeto é a forma de como o corpo se prepara para receber uma ação em determinada situação. É uma experiência com intensidade não consciente, ou seja, o afeto tem uma função importante em relação ao corpo. O autor exemplifica afeto citando que uma criança não tem habilidades de linguagem nem uma história de experiências anteriores para que possa extrair sensações que percorrem em seu corpo. Portanto, a criança lida com intensidades (afetos positivos ou negativos).

Para Rousseau e Hayes-Roth (1996), um indivíduo é único e possui características psicológicas próprias, nas formas como pensa, age, reage, onde lhe dão um estilo. Assim, Massumi (2002) complementa que o afeto acontece sempre antes ou fora da consciência, ou seja, é a reação do corpo à ação sobre uma determinada circunstância.

Já Demos (1995) diz que o afeto é composto por um conjunto de respostas correlacionadas que envolvem os músculos faciais, as vísceras, do sistema respiratório, do esqueleto, as alterações do fluxo sanguíneo, que atuam juntos para produzir estímulos intensivos que possam ir de encontro aos estímulos do organismo. Ele complementa dizendo que para uma criança afeto é emoção. Já para o adulto afeto é o que faz sentir um sentimento, o que determina a intensidade (quantidade) de um sentimento.

2.1.2 Sentimento (*Sentiment*)

Gordon (1981) conceitua sentimento como um padrão que é construído através de gestos expressivos, sensações e significados culturais organizados ao redor de uma relação com um objeto social, geralmente outra pessoa ou grupo de pessoas (conhecidas ou não).

Damásio (2004) apresenta sentimento como um comportamento invisível, exceto para o seu proprietário. O sentimento é constituído resquícios de manifestações

emocionais, baseado na falsa ideia de que o sentimento ocorre primeiro e, em seguida se manifestam as emoções, confirmando assim a anteposição da emoção ao sentimento.

Diante disso, Friedenbergl e Silverman (2005) citam que o sentimento corresponde a uma experiência subjetiva a uma emoção reafirmando a hipótese de Wierzbicka (1999), quando o autor define que sentimento é subjetivo e está relacionado com o que acontece no corpo do indivíduo. Porém, ele afirma que sentimento tem um conceito universal, ou seja, em todas as línguas a palavra “sentir” é parte integrante de uma pessoa, isto é, em todas as culturas as pessoas atribuem sentimentos a outras pessoas, bem como para si mesmos.

Shouse (2005), por sua vez, conceitua como uma sensação de algo que foi verificado em relação a experiências anteriores. É pessoal e biográfico, porque cada pessoa tem um conjunto de sensação a partir de quando interpreta e rotula seu sentimento.

Portanto, Cattell (2006) define sentimentos como uma alienação neuropsíquica que reage emocionalmente e cognitivamente a um determinado objeto ou situação. O autor ressalta ainda que os sentimentos no indivíduo são aflorados por interesses e sensações, de sua parte, valores e objeto. Sendo assim, o sentimento, chega a possuir um poder em uma pessoa, mais ou menos duradouro, é capaz de provocar reações relativamente intensas e frequentes, positivas ou negativas (MURRAY e MORGAN, 1945).

2.1.3 Emoção

Kleinginna e Kleinginna (1981) definiram formalmente emoção como um conjunto complexo de interações entre os fatores subjetivos e objetivos, mediados por sistemas neural e hormonal, que pode dar origem a experiências afetivas, tais como sentimentos de prazer, desprazer e excitação; gerar processos cognitivos, como afetos, emocionalmente relevantes; ajustes fisiológicos generalizados; e levar a um comportamento, muitas das vezes, mas nem sempre é expressivo e objetivo. Já para Kirouac (1994) a emoção é vista como um estado supérfluo e não científico durante muito tempo até quando surgem estudos empíricos e teóricos aceitando os estados internos como variáveis explicativas do comportamento.

Segundo Damásio (1996) as emoções não são um “luxo”. Elas têm funções de comunicação de significados e de orientação cognitiva. Damásio concorda com Ames (1990) e Izard (1984) que as emoções influenciam diretamente em tarefas cognitivas, portanto, as funções biológicas dos estados afetivos de tendências comportamentais, são aderidas pela CA como modelo para permitir que os sistemas computacionais fiquem mais inteligentes e adaptados ao homem.

Scherer (2000b) fala que emoção pode ser considerada como um estado afetivo breve, de alta intensidade e de resposta sincronizada a um evento. A escolha afetiva e as classificações da emoção não devem ser comparadas nem analisadas da mesma maneira, pois emoções têm intensidade e classificações distintas. O autor complementa a definição de emoção como fatos coordenados em vários componentes orgânicos (*appraisal* cognitivo, reações fisiológicas, motivação para ação, expressão motora e sensação subjetiva) em resposta a eventos internos e externos de grande importância para o organismo.

Thoits (1989) cita que as emoções são pré-definidas culturalmente. Com isso, Shouse (2005) levanta a hipótese das emoções serem projeções ou exhibições de um sentimento, fato que represente uma “falsa verdade” ou fingimento e exemplifica que quando um indivíduo expressa emoções, às vezes essa transmissão pode ser uma exposição do seu lado interno, que é feita de forma rotulada a fim de somente cumprir uma expectativa social. Portanto, as emoções podem ser vistas como expressões de afeto e/ou sentimentos.

Contudo, Dolan (2002) afirma que as emoções possuem três características a partir de uma perspectiva psicológica. A primeira é que as emoções já nascem com indivíduos. A segunda, é que as emoções são muito difíceis de serem controladas. E a terceira é que emoções são muito mais difíceis de serem mudadas do que pensamentos. O autor fala que é mais fácil mudar pensamentos que emoções, pois os pensamentos têm um impacto mais global em nosso comportamento e são passageiros, ao contrário das emoções.

Longhi (2011) cita em sua tese que, para representar emoções em sistemas computacionais, os modelos psicológicos de Roseman e OCC são os mais utilizados. O modelo de Roseman (ROSEMAN *et al.* 1990 e ROSEMAN, 2001) diferencia as emoções em positivas e negativas em que é possível inferir 20 tipos de emoções. Já o modelo OCC³ (ORTONY *et al.* 1990), está estruturado em 22 tipos de emoções onde as palavras que identificam as emoções são: “*happy for*”, “*resentment*”, “*gloating*”, “*pity*”, “*joy*”, “*distress*”, “*pride*”, “*shame*”, “*admiration*”, “*reproach*”, “*love*”, “*hate*”, “*hope*”, “*fear*”, “*satisfaction*”, “*fears-confirmed*”, “*relief*”, “*disappointment*”, “*gratification*”, “*remorse*”, “*gratitude*” e “*anger*”. O modelo presume que os estados emocionais de um indivíduo são observados a partir de seus padrões, objetivos e preferências. Para Bercht (2001), esse modelo é o mais usado na avaliação e sintetização de emoções em sistemas computacionais, principalmente para modelar emoções básicas⁴ (alegria, tristeza, surpresa, medo, nojo).

³ Recebeu esse nome pelas letras iniciais dos seus autores, Ortony, Clore e Collins.

⁴ Essas emoções são constituídas pelo sistema motivacional primário do comportamento (IZARD; ACKERMAN, 2000), isto é, estão relacionados a impulsos fisiológicos, por exemplo, sede, fome, sono, etc.

2.1.4 Estado de Ânimo (Humor)

Longhi (2011) fala que o estado de ânimo pode ser compreendido como a capacidade do indivíduo apreciar ou expressar, perceber o que é divertido ou cômico; ou, segundo Dorsch *et al.* (2008), para designar as mudanças na estrutura psicológica provocadas por influências que são originadas interna ou externamente ao corpo humano.

Entretanto, Scherer (2000b) diz que o humor é um estado afetivo de baixa intensidade, de longa duração, difuso, e atua de forma subjetiva no ser vivo. É visto como um episódio que a primeira vista tem relativamente uma breve resposta, e no segundo momento, como um estado afetivo generalizado e sem causa aparente. Como exemplos, o autor cita a depressão, a serenidade e a irritação.

2.2 Personalidade

De acordo com Schultz (1990) a palavra personalidade tem étimo latino, derivando de “*persona*”, que significa máscara utilizada por atores em peça teatral. Ele diz que personalidade é como um conjunto duradouro e único com características que não tem qualquer chance de resposta a situações diferentes. Sendo assim, compreende-se que a personalidade são características externas e aparentes, ou seja, a imagem que se transmite as pessoas.

Allport (1937) afirma que existem, pelo menos, cinquenta significados diferentes para o termo personalidade, considerando-o como uma organização dinâmica dos sistemas bio-sociais⁵ que determinam a adaptação única do indivíduo ao mundo. Portanto, a personalidade é muito mais complexa do que se imagina, pois nela contém um conjunto de características de seres humanos que vão além de atributos superficiais.

Maddi (1980) define personalidade como um conjunto consolidado de características e tendências que determinam as semelhanças e dissemelhanças no comportamento psicológico das pessoas (ações, sentimentos e pensamentos), que tem continuidade no tempo. No entanto, a personalidade não é rígida e igual; ela se mantém normalmente estável num período de 45 anos, período que começa a maioridade (SOLDZ e VAILLANT, 1998).

Singer (1986) conceitua personalidade como a forma singular do indivíduo se expressar e reagir a um estímulo, firmado de uma estrutura básica, genética e de experiências de vida, principalmente da infância. Para Cook (1984), esses estímulos concentram-se na impressão em que um indivíduo causa, supondo-se que as pessoas

⁵ Sistemas bio-sociais incluem traços, hábitos, valores e motivos, cujas diferenças individuais são parcialmente hereditárias, resultado da aprendizagem e experiência (social). Estes sistemas estariam organizados dinamicamente e ativamente com o ambiente.

têm personalidades variadas, visto que o mesmo indivíduo pode causar variadas impressões em pessoas diferentes. Contudo, Burger (2000) define personalidade como padrões de comportamento consistentes e processos intrapessoais originários dentro do indivíduo.

Para Kleinmuntz (1967), a personalidade é definida como uma organização única de aspectos que caracterizam e influenciam a forma como um indivíduo interage com o meio que o rodeia. Para o autor, a personalidade é influenciada pelo tipo de ambiente em que se está inserido e pelas várias situações que a pessoa vive. Sendo assim, Buss e Finn (1987) afirmam que a personalidade afeta a forma de como as pessoas se adaptam ou reagem às condições ambientais. Johnson (1994) concorda com Buss e Finn quando diz que a personalidade de um indivíduo influencia em seu desempenho no trabalho, em sua saúde, em seus relacionamentos e outros eventos importantes na vida, sendo assim úteis para fornecer informações e previsões sobre indivíduos. Portanto, a personalidade é mais do que uma simples aparência física, ela é de fato estável e previsível (NUNES, 2008).

2.3 Traços de personalidade

Inicialmente, os traços de personalidade foram estudados e definidos por Allport e Allport (1921) e Allport (1927) que analisou a personalidade baseado em pessoas saudáveis. O autor criou 17.953 traços que descrevem a personalidade de um indivíduo por acreditar que cada ser humano é único, conseqüentemente Allport (1960) definiu traço de personalidade como um sistema neuropsíquico (peculiar ao indivíduo), generalizado e focalizado, com a capacidade de tornar muitos estímulos funcionalmente equivalentes, de iniciar e guiar formas coesas (equivalentes) de comportamento adaptativo e expressivo.

Feldman e Feinman (1992) concordam com Allport quando citam que traços são os aspectos básicos da personalidade diferenciando as pessoas umas das outras e que são consistentes ao longo do tempo. Porém Barkhuss e Csank (1999) ressaltam que os traços têm intensidades diferentes. Por exemplo, pessoas com características agressivas têm níveis de agressividade diferentes.

A teoria dos traços de personalidade estuda como definir amostras habituais de pensamento, comportamento e emoção (KASSIN, 2003). De fato, os traços são proporcionalmente estáveis ao passar do tempo. Por exemplo, algumas pessoas são extrovertidas e outras não, ou seja, os indivíduos são diferentes, cada um com sua característica e nível de intensidade.

Na opinião de Eysenck (1953), os traços da personalidade são combinações de características cognitivas, afetivas e físicas. Sendo assim, os traços são referidos

normalmente como “etiquetas/rótulos da personalidade” ou adjetivos descritivos que identificam um indivíduo.

Goldberg (1981) cita que traços de personalidade são características que não podem ser medidas com rigor e salienta que se um traço de personalidade for relevante, ou seja, capaz de causar diferenças individuais significativas, ele será notado. Sendo assim, uma expressão ou uma simples palavra acabará sendo arquitetada para delinear esse traço.

Cattell (1945) propõe um subconjunto dos traços de Allport criando 4.500 traços. Com isso Goldberg (1990) correlacionou esses traços com 171 escalas após realizar uma análise empírica. O autor reduziu cerca de 99% desses itens identificando 12 fatores de personalidade. Desses, apenas cinco fatores foram replicáveis. Como resultado, o modelo “*Big Five*”, para definição de traços de personalidade, foi criado.

2.4 *Big Five*

Goldberg (1990) iniciou o processo de criação do modelo *Big Five* essencialmente para simplificar e organizar os traços de personalidade (NUNES, 2008). Porém, de acordo com John e Srivastava (1999), a estrutura *Big Five* não implica que as diferenças de personalidade podem ser reduzidas a apenas cinco fatores. No entanto, estes cinco fatores representam a personalidade no nível mais amplo de abstração e cada fator resume um grande número de características de personalidade distintas, ou seja, mais específicas.

Silva e Nakano (2011) falam que no Brasil os cinco fatores do modelo *Big Five* têm sido chamados de Abertura à experiência, Neuroticismo, Extroversão, Socialização e Realização, ainda que a literatura internacional aponte algumas divergências em relação aos nomes, como por exemplo, em Urquijo (2001) onde os fatores socialização, realização e neuroticismo são chamados de escrupulosidade, agradabilidade e instabilidade emocional, respectivamente. Já McCrae (1993) cita que o fator abertura à experiência tem sido chamado de intelecto. Nunes *et al.* (2009) fala que os fatores socialização e realização são chamados de conscienciosidade e amabilidade, respectivamente. Embora existam divergências na forma como são chamados alguns fatores, as definições são consensuais e apontam para características semelhantes.

Na Tabela 1 estão descritos os fatores do *Big Five* utilizados nesta pesquisa apontando suas características.

Tabela 1: Dimensões do modelo *Big Five*

| <i>Big Five</i> | Características |
|-------------------------------|--|
| Abertura à Experiência | Este fator não está de fato diretamente relacionado com inteligência, mas refere-se aos comportamentos exploratórios e reconhecimento da importância em ter novas experiências. Pessoas com essa característica são imaginativas, criativas, curiosas, divertem-se com valores não convencionais e com novas ideias, ou seja, experienciam uma gama ampla de emoções mais intensamente do que pessoas fechadas (baixas em Abertura). As pessoas que tem nível baixo em Abertura tendem a serem convencionais nas suas atitudes e crenças, conservadores nas suas preferências, categóricos e rigorosos nas suas crenças; também tendem a serem menos responsivos emocionalmente (COSTA e WIDIGER, 1993). |
| Neuroticismo | Este fator refere-se ao nível crônico de ajustamento de instabilidade e emoções. Alto Neuroticismo identifica pessoas que são propensas a sofrerem psicologicamente, sendo assim apresentar altos níveis de ansiedade, falta de moderação, depressão, vulnerabilidade, auto-percepção e raiva. Neuroticismo também apresenta um baixo nível onde inclui a baixa tolerância e um estado de calma quando o indivíduo está pressionado. (COSTA e WIDIGER, 1993). |
| Extroversão | Este fator refere-se à intensidade e quantidade de interações interpessoais preferidas, nível de atividade, necessidade de estimulação e capacidade de alegrar-se. Indivíduos que são altos em Extroversão tendem a ser ativos, falantes, sociáveis, afetuosos e otimista; enquanto que indivíduos baixos em Extroversão tendem a ser mais reservados, indiferentes, sóbrios, independentes e quietos. Introversão não significa ser pessimistas ou infeliz, mas eles não são dados a estados de espíritos de pessoas animadas, o que caracteriza os extrovertidos (COSTA e WIDIGER, 1993). |
| Socialização | Este fator representa o grau de organização, persistência, controle e motivação em alcançar objetivos. Pessoas que são altas em Socialização tendem a ser organizadas, trabalhadoras, decididas, pontuais, escrupulosas, ambiciosas e perseverantes; por outro lado, pessoas que são baixas em socialização tendem a não ter objetivos claros, são preguiçosas, descuidadas, negligentes e hedonistas (COSTA e WIDIGER, 1993). |
| Realização | Realização é uma dimensão interpessoal e refere-se aos tipos de interações que uma pessoa apresenta ao longo de um contínuo que se estende da compaixão ao antagonismo. Indivíduos que são altos em Realização tendem a ser confiáveis, bondosos, afáveis, generosos, altruístas e prestativos, adora ajudar aos outros e tem compaixão pelos desabrigados. Aqueles que são baixos em Realização tendem a ser pessoas cínicas, não cooperativas e irritáveis, podendo também ser pessoas manipuladoras e vingativas. (COSTA e WIDIGER, 1993). |

Fonte: Costa e Widiger, 1993.

Contudo, mesmo com a criação do modelo *Big Five* e que ele represente grande eficiência na representação da estrutura de personalidade, Johnson (2000a) criou “facetas” para minuciar detalhes de cada fator.

De acordo com Johnson (2000b), as facetas para cada fator do *Big Five* foram detalhadas para uso em um questionário chamado IPIP-NEO, conforme mostra a Tabela 2. Os termos em inglês são os originais do Johnson, enquanto as traduções foram feitas baseados nos autores Costa e Widiger (1993).

Tabela 2: Facetas do IPIP-NEO

| Big Five | Openness to Experience (Abertura à Experiência) | Neuroticism (Neuroticismo) | Extraversion (Extroversão) | Conscientiousness (Socialização) | Agreeableness (Realização) |
|----------|--|--|---|--------------------------------------|-------------------------------|
| FACETAS | Imagination (Imaginação) | Anxiety (Ansiedade) | Activity-Level (Nível de Atividade) | Achievement – Striving (Empenho) | Trust (Confiança) |
| | Artistic Interests (Interesses Artísticos) | Immoderation (Falta de Moderação) | Assertiveness (Assertividade) | Orderliness (Ordem) | Morality (Moralidade) |
| | Intellect (Intelecto) | Depression (Depressão) | Gregariousness (Gregarismo) | Self-Discipline (Auto Disciplina) | Modesty (Modéstia) |
| | Emotionality (Emotividade) | Self-Consciousness (Auto Percepção) | Friendliness (Amigabilidade) | Self-Efficacy (Auto Eficácia) | Cooperation (Cooperação) |
| | Liberalism (Liberalismo) | Anger (Raiva) | Excitement-Seeking (Procura por excitação) | Dutifulness (Senso de dever) | Sympathy (Compaixão) |

Fonte: Johnson, 2000b

Nunes (2008) fala que para extrair traços de personalidade a partir de sistemas computacionais a partir do modelo *Big Five* e suas respectivas facetas, usa-se questionários. Esses questionários foram criados por psicólogos (BUTCHER, 1989), (GOLDBERG, 1990) e (JOHNSON, 2000a) e podem ser respondidos gratuitamente na web. Eles podem variar no quantitativo de perguntas, sendo mais curtos ou mais longos. O número de perguntas está relacionando diretamente com os traços de personalidade que desejam ser extraídos, quanto maior o número de perguntas, maior será a precisão na resposta.

2.5 Formas de Inferir Personalidade

Cattell (1945), Butcher (1989), Costa e McCrae (1988), Comrey (1988) e Goldberg (1990), são exemplos de psicólogos que pesquisaram e analisaram diversas formas de como se inferir personalidade. Para tal, os psicólogos desenvolveram questionários de personalidade (questionários), a exemplo do 16-PF⁶ (CATTELL, 1945), ROCI-II⁷ (RAHIM, 1983), o MMPI⁸ (BUTCHER, 1989), as escalas de Murray (COSTA e MCCRAE, 1988), as escalas de Comrey (COMREY, 1988) e o *Big Five* (GOLDBERG, 1990). Esses questionários podem ser respondidos, manual ou eletronicamente e seus resultados, geralmente, revelam algumas pistas de personalidade.

⁶ 16 Personality Factors Questionnaire.

⁷ Rahim Organizational Conflict Inventory–II

⁸ Minnesota Multiphasic Personality Inventory.

Os questionários do *Big Five* são os mais conhecidos e os mais utilizados, podendo ser citados, por exemplo, o NEO-PI-R⁹ (MCCRAE e JOHN, 1992), (COSTA e MCCRAE, 1992), IPIP-NEO¹⁰ (JOHNSON, 2000b), “FFPI¹¹” (HENDRINKS *et al.* 2002), BFQ¹² (BARBARANELLI e CAPRARA, 2002) e o GPI¹³ (SCHIMIT *et al.* 2002) e BFI-10¹⁴ (RAMMSTEDT e JOHN, 2007). Cada um desses questionários tem uma quantidade de itens propostos a serem respondidos e as suas definições de facetas específicas. Nunes (2008) afirma que depois de analisados cada questionário, existe a hipótese de o número de questões refletirem na precisão das respostas, pois quanto maior o número de questões presentes no questionário, mais detalhado fica o resultado.

O IPIP-NEO (JOHNSON, 2000b) foi criado após o autor analisar as formas de inferências avaliadas por Goldberg (1990). Johnson optou em desenvolver o IPIP-NEO ao estudar e observar o NEO-PI-R (COSTA e MCCRAE, 1992) que é um questionário muito conhecido, sólido e um dos mais bem validados do mundo (Johnson, 2000a) e principalmente porque é baseado nos fatores do *Big Five*. O IPIP-NEO foi utilizado e validado por Johnson (JOHNSON, 2000b) e (JOHNSON, 2005). Detalhes em Nunes (2008).

Seu formato original (inglês) é formado por 300 questões com pontuações em escala de cinco pontos. As pontuações são associadas em valores numéricos de 1 a 5 variando de acordo com as respostas dos usuários (JOHNSON, 2000b). As perguntas são divididas da seguinte maneira: Cada fator do *Big Five* possui um conjunto de 60 questões. Como cada fator corresponde a 6 facetas, foram desenvolvidas 10 questões para cada faceta. O autor também desenvolveu outro questionário, este com 120 questões, seguindo a mesma metodologia adotada.

Os questionários foram e estão disponibilizados em sua versão original (inglês) e em uma versão adaptada para o português no *Personalitatem Inventory*¹⁵. O objetivo do questionário IPIP-NEO (disponibilizado com permissão do Dr. John A. Johnson) é tentar inferir traços de personalidade. Depois de respondido, são feitas análises, de forma específica, é gerado um relatório com as pistas de personalidade pessoais relacionados com os cinco fatores do *Big Five*. Uma observação que é importante ressaltar é que o questionário não revela informações ocultas, privadas e/ou secretas sobre o indivíduo e, também não é capaz de inferir ou avaliar qualquer distúrbio ou psicológico ou psiquiátrico grave.

⁹ *Revised NEO (Neuroticism-Extroversion-Openness) Personality Inventory*

¹⁰ *NEO (Neuroticism-Extroversion-Openness) International Personality Item Pool*

¹¹ *Five Factor Personality Inventory*

¹² *Big Five Questionnaire*

¹³ *Global Personality Inventory* do autor Schimit *et al.*, 2002.

¹⁴ *Big Five 10*

¹⁵ Portal do *Personalitatem*: <http://personalitatem.ufs.br/>

2.6 Considerações Finais

Neste capítulo foram apresentados os principais conceitos da CA, termos subjetivos como afeto, sentimentos (*sentiment*), emoção e estado de ânimo. Também foram apresentados conceitos sobre personalidade, traços de personalidade, *Big Five* e as formas de inferir personalidade. Este passo é relevante para entender os métodos e formas de aplicações criadas pelos psicólogos como Johnson (2000a) e Costa e McCrae (1992).

Como o propósito desta dissertação é a readaptação de um léxico, contendo aspectos afetivos, para que seja possível a sua implementação em qualquer ferramenta computacional, os conceitos apresentados são fundamentais para o entendimento dos termos (diferenças x semelhanças) e assim poder aplicá-los para os objetivos propostos.

3 LÉXICOS

Os léxicos são indispensáveis na análise, processamento e geração da língua natural. Segundo Zavaglia (2006), léxicos para serem usados em PLN devem possuir informações adequadas e codificadas para que o algoritmo ou programa desenvolvido possa compreendê-lo e executá-lo. Já Trask (2008) conceitua léxico como um acervo de palavras que integram a língua, ou seja, é o vocabulário de uma língua.

Para Aston e Burnard (1998) um léxico é derivado do exame de um *corpus*, que por sua vez é definido no dicionário de inglês *Oxford* como um “órgão, coleção de escritas”. O plural mais usualmente utilizado de *corpus* é *corpora*¹⁶. Não existe um tamanho mínimo ou máximo para *corpora*, ou qualquer tipo de especificação do que ele deve conter. Sendo assim, o termo *corpus* está relacionado com algum recurso linguístico (SINCLAIR, 1996).

A construção de um léxico de maneira manual é tarefa árdua devido grande volume de informações e a quantidade de tempo que se gasta para a realização das etapas. Para tal, existe um grande esforço na criação de léxicos com ajuda de técnicas computacionais, a exemplo, aprendizado de máquina. (PALMER *et al.*, 2010). Outro método para a construção de léxicos computacionais parte da análise e aperfeiçoamento dos léxicos já existentes.

3.1 Léxicos Computacionais

Muniz e Nunes (2004) definem léxico computacional, ou dicionário, como a estrutura principal da maioria dos sistemas que englobam PLN. Azeredo (2008) cita que um léxico é constituído por um vocabulário com expressão oral e escrito, que foi usado no processo de comunicação e conquistado culturalmente através do tempo. Sua estrutura é formada por um banco de dados contendo palavras ou um conjunto de palavras isoladas.

Evans e Kilgarriff (1995) citam que o processo de desenvolvimento de léxicos e de *databases* com informações lexicais, até meados dos anos 80, era realizado sem dificuldades através de um modelo de elaboração. Com isso, os dados, utilizados na construção do banco, ficariam indisponíveis para possíveis modificações, melhoramentos. Sendo assim, outras formas e métodos começaram a serem estudados a fim de reutilizar esses dados e, conseqüentemente, como diminuir esforços para o desenvolvimento de novas aplicações.

¹⁶ A frequência e aceitação de outras formas de plural da palavra *corpus* têm sido muito debatido na lista de discussão eletrônica CORPORA. (ASTON e BURNARD, 1998: 63-73) dedicam dez páginas para a questão.

No Brasil, a pesquisa em léxicos está concentrada em centros mais voltados à linguística Computacional (SARDINHA, 2000). Léxicos, em português e em outras línguas, proporcionaram o acúmulo de uma extensa obra, cujos principais são: (1) Sinclair (1996), o trabalho é pioneiro na área de léxico. O autor traçou os caminhos da maioria das pesquisas em linguística de *corpus* realizadas até hoje; e (2) Francis e Hunston (1996) desenvolveram a primeira “gramática do léxico” que descreve, de modo amplo e profundo, os padrões verbais da língua inglesa a partir de um *corpus*, seguindo o princípio básico da identificação de colocações recorrentes por computador. Os autores continuaram os estudos, e no seu segundo volume (FRANCIS e HUNSTON, 1998) foi dedicado aos substantivos e adjetivos e ainda formularam a teórica dos princípios seguidos nas gramáticas.

Como dito anteriormente, o estudo e avanço das técnicas para construção de léxicos são muito complexos. Uma das maiores dificuldades é a falta de recursos linguísticos computacionais. Muniz e Nunes (2004) dividem os recursos computacionais em dois grupos: os que apresentam conhecimento linguístico, mas não classificam de forma automática, a exemplos de *corpus* e dicionários eletrônicos; e os que processam algum resultado pré-definido como, por exemplo, os analisadores semânticos. Neste contexto, os léxicos são fundamentais, pois em processamento de textos eles serão utilizados para realizar as análises lexicais. Sendo assim, é importante salientar que para o desenvolvimento das análises léxicas é necessário conhecer a abordagem léxica. De acordo com De Raad *et al.* (1998) a abordagem léxica é um conjunto de constructos universais e comuns que podem ser identificados pela descrição dos traços de personalidade.

Identificar palavras em textos para analisá-las não é tão simples. Pegando-se uma palavra de um texto qualquer e analisando-a quase não há significado. Entretanto Garcia (1982) afirma que cada palavra tem um eixo significante constante e estável servindo de base para construção de adjetivos, substantivos, verbos e advérbios. Esse eixo é chamado de lexema. Longhi *et al.* (2010) comenta que o lexema pode ser analisado como uma palavra ou parte da palavra observando o seu sentido, sendo denotativo ou conotativo. Para Garcia (1982) o sentido denotativo é como referência, ou seja, diz respeito aos traços semânticos (não subjetivo) contendo característica específica ou geral. Por exemplo, lexemas: casa, chapéu e carro são palavras denotativas. Já a palavra conotativa tem característica virtual e constituição subjetiva, ou seja, tem sentido afetivo. Palavras com sentido conotativo mostram o comportamento de um indivíduo como: repulsa; tranquilidade ou pânico; apreço ou desprezo. (AZEREDO, 2008). Para mais detalhes ver em Longhi *et al.* (2010).

A seguir apresenta-se uma descrição e análises dos principais léxicos computacionais, nas línguas inglesa e portuguesa, utilizados e citados mais comumente na literatura.

3.1.1 *MRC Psycholinguistic DataBase*

MRC é um dicionário criado por Coltheart (1981). Na versão inicial, a base foi composta por 98.538 palavras com categorias psicolinguísticas distintas. Na segunda versão foram incluídas 52.299 novas palavras o que resultou num total de 150.837 palavras. No geral, existe uma complementação das versões, o que deixa ambas semelhantes e contendo 26 propriedades psicolinguísticas diferentes para as palavras (WILSON, 1988). Localizado online¹⁷, onde estão disponíveis vários arquivos incluindo o dicionário de palavras e o programa para as avaliações, o dicionário foi desenvolvido com propósito de PLN e tarefas de IA, onde exige descrição psicológica e linguística das palavras.

3.1.2 *WordNet*

George A. Miller iniciou o projeto *WordNet*¹⁸ em meados da década de 1980 na Universidade de *Princeton* e teve continuidade por Christiane Fellbaum. O dicionário, desenvolvido na língua inglesa, é composto por palavras substantivas, verbos e advérbios. Essas palavras são agrupadas em conjuntos de sinônimos cognitivos, chamado de *synsets*, mantendo uma estrutura de níveis semânticos e morfológicos. Recentemente, o léxico contém 152.059 palavras e 115.424 *synsets*. A ideia principal era realizar uma relação entre as palavras sinônimas que continham no banco de dados. Esse grupo de palavras sinônimas foi organizado nos *synsets*, onde cada *synset* representaria uma definição que seria validada para todas as palavras do conjunto (SCARTON, 2013). Para o português foi desenvolvida uma versão pelo linguista (DIAS-DA-SILVA *et al.* 2008) chamada de *WordNetbr*. Na versão portuguesa existem 44.000 palavras (substantivos, verbos, advérbios e adjetivos) e 18.500 *synsets*.

3.1.3 *VerbNet*

Léxico desenvolvido por Karin-Kipper (2005), na língua inglesa, traz circunscrições dos verbos motivados pelo trabalho de Levin (1993). O léxico também é composto por informações semânticas e sintáticas o que deriva em uma coletânea de categorias verbais e análises sintáticas. Inicialmente, partindo das análises de Levin, foram criadas 191 classes que abrigou 4.656 verbos. Posteriormente foram incluídas novas classes, chegando a 274 com cobertura para 5.800 verbos (PALMER, 2010). Scarton e Aluísio (2012) tiveram objetivo de criar um léxico seguindo as mesmas características, mas para o português brasileiro e assim criaram o *VerbNet.br*. O trabalho dos autores analisa a possibilidade de uma tradução para o português, de forma semiautomática, dos verbos do léxico *VerbNet* utilizando um *cross-linguístico*, recurso que possibilitaria a herança dos recursos semânticos, deixando o léxico traduzido para o português de forma mais semelhante possível com o inglês.

¹⁷ <http://www.psych.rl.ac.uk/>

¹⁸ <https://wordnet.princeton.edu/>

3.2 Léxicos Afetivos

Ortony *et al.* (1987) afirmam que léxicos afetivos não contêm somente termos ligados à emoção, mas possuem outros termos e outras condições afetivas (afeto, estado de ânimo e sentimento). Termos como “afeto” e “emoção” são usados, às vezes como sinônimos. A distinção é feita quando o termo afeto refere-se a qualquer coisa cujo valor da valência é positivo ou negativo. Afeto tem uma categoria mais ampla comparada à emoção. As emoções são causadas por tipos de condições afetivas, mas nem todas as condições afetivas são emoções, como explicado na seção 2.1.3.

No início dos estudos em léxicos afetivos, Averill (1975) analisou a base de Allport e Odbert (1938) que foram selecionadas e julgadas em ter conotações afetivas. A finalidade do estudo foi desenvolver um método, chamado de “semântica”, com propósito de mapear um universo de palavras com características afetivas, mas o autor reconhece que nem todas as palavras incluídas na análise contém afetividade, justificando que qualquer divisão entre conceitos afetivos é, necessariamente, vaga e arbitrária (AVERILL, 1975).

Não existe um modelo pré-definido para a construção de um léxico afetivo. A maioria dos trabalhos criaram/definem etapas e metas a partir de estudos para conseguir chegar ao objetivo. Outros trabalhos pegam léxicos já desenvolvidos e implementados e fazem melhoramentos e extensões. A seguir apresentam-se uma descrição de alguns léxicos afetivos.

3.2.1 LIWC

O *LIWC* foi construído e validado na língua inglesa na década de 90 por Pennebaker e King (1999). Sua composição é feita por um programa que faz a leitura de textos, onde são buscadas as palavras que serão analisadas posteriormente por um dicionário. O dicionário, que é o coração do programa, contém uma base de palavras que foram atribuídas a categorias específicas. As palavras e as categorias foram definidas e julgadas por avaliadores. Quando o *LIWC* foi criado, o objetivo era fazer com que o computador fosse capaz de calcular a porcentagem de palavras com características positivas e negativas, dentro de um texto. Com o sucesso do experimento, o autor expandiu os estudos, fazendo com que o *LIWC* passasse a analisar palavras com outras características como, por exemplo, raiva, tristeza, afeto, entre outras. Para realizar as novas avaliações, avaliadores analisavam as palavras e as quais categorias fariam parte. A lista de palavras foi atualizada seguindo o conjunto de regras: (1) uma palavra permanece na lista de categoria se dois dos três avaliadores concordarem pela permanência; (2) uma palavra é excluída da lista de categoria se, pelo menos, dois dos três avaliadores concordarem pela exclusão; e (3) uma palavra é acrescentada à lista de categorias se dois dos três avaliadores concordarem pela inclusão. Então, todo esse processo foi repetido por um grupo separado de três avaliadores. As porcentagens finais de concordância entre os avaliadores para esta

segunda fase de classificação variou de 93% a 100%. No total, mais de 100 milhões de palavras foram analisadas. Algumas categorias de palavras foram excluídas e outras foram adicionadas. Para mais detalhes sobre o processo e os resultados específicos, consulte Pennebaker *et al.* (1997).

No Brasil, existe o grupo NILC – Grupo de Pesquisa de São Carlos – SP, que realizou a tradução do *LIWC* para o português brasileiro. Esse dicionário foi construído a partir de vários dicionários bilíngues Português-Inglês, por 3 equipes, uma do NILC, através da tradutora Mônica Martins, uma da empresa Checon Pesquisa, coordenado pela pesquisadora Rosângela Checon e outra da Unisinos, através do trabalho do Prof. Rove Chishman. As conjugações foram colocadas de forma automática usando o dicionário Unitex-PB do NILC e as categorias do dicionário foram listadas automaticamente. Como resultado do trabalho, a equipe criou o *Brazilian Portuguese LIWC 2007 Dictionary*. Observa-se que não foi feita a revisão do trabalho manual de tradução e nem sua validação.

3.2.2 ANEW

O léxico *ANEW - Affective Norms for English Words*, foi desenvolvido por Bradley e Lang (1999) na língua inglesa. O objetivo dos autores foi construir um dicionário a partir de palavras afetivas que tivessem características de emoção. Para isso, as atividades foram subdivididas em três grandes dimensões. A primeira variando de agradável a desagradável, a segunda de calmo a exaltado e a terceira de dominado a controlado. As avaliações das dimensões foram feitas por uma escala chamada de *Self-Assessment Manikin* (SAM) desenvolvida e cedida por Bradley e Lang em 1980. Esta escala é representada por figuras de bonecos os quais expressam variações entre as dimensões específicas. A representação da valência varia de sorridente (agradável) a uma figura descontente (desagradável); o que representa alerta varia de uma figura ativada (estimulado) a uma figura inerte (relaxado). A opção pelo SAM como escala de avaliação foi baseada em suas propriedades psicométricas e na tentativa de manter o método equivalente. Essa escala tem correlações com as dimensões atribuídas e ao fazer as análises das palavras, são gerados gráficos com os resultados compreendidos por escalas e que retratam diferentes valores para cada dimensão. No total foram avaliadas e analisadas 1044 palavras.

3.2.3 SentiWordNet

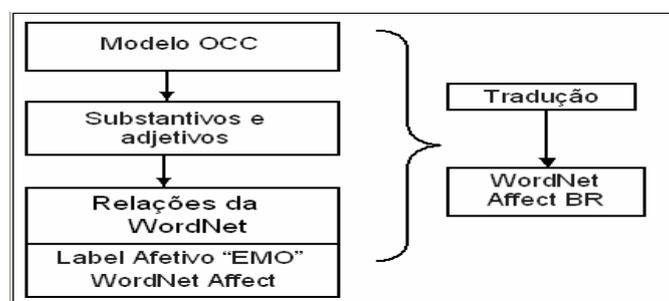
Esse léxico foi desenvolvido por Esuli e Sebastiani (2006b), na língua inglesa. O *SentiWordNet* foi fruto da adaptação da classificação do *Synset*, método de decisão e termos com valência (positiva, negativa ou neutra) (ESULI e SEBASTIANI, 2005; 2006a). Tais métodos são baseados em formações de conjuntos com classificadores ternários, ou seja, cada um dos métodos é capaz de decidir se um *Synset* é positivo, negativo ou neutro. Cada classificador ternário difere uns dos outros no conjunto de

treinamento e no dispositivo de aprendizagem utilizado para treiná-lo, produzindo diferentes resultados de classificação dos *Synsets Wordnet*. Os placares de um *synsets* são determinados pela proporção dos classificadores ternários. Esses classificadores atribuem uma *label* correspondente. Se todos os classificadores atribuírem uma mesma *label* para um *Synset*, essa *label* terá pontuação máxima para o *Synset* e se houver divergência entre os classificadores, terá pontuação proporcional ao número de classificadores, veja (ESULI e SEBASTIANI, 2006b) para uma descrição mais detalhada como os classificadores foram treinados.

3.2.4 *Wordnet AffectBR*

O Léxico *Wordnet AffectBR* foi adaptado por Pasqualotti e Vieira (2008) na língua portuguesa. A sua construção se deu a partir de outros léxicos: o *Wordnet*, *Base Affect* e o *Wordnet Affect*. *Wordnet* é um léxico composto por palavras e *Synsets* que é um conjunto de sinônimos que compõem a estrutura da *Wordnet* e descritos por um glossário. As palavras são organizadas pelo seu significado, com isso é formada uma matriz lexical bidimensional quanto as suas relações (VALITUTTI *et al.* 2004). A *Base Affect* é um recurso linguístico, criado de forma manual, que foi desenvolvida na Itália por pesquisadores do grupo TCC - *The cognitive and Communication Technologies* (<http://tcc.itc.it/>), e tem como principal composição as classes gramaticais da *Wordnet* e outras informações lexicais, semânticas e afetivas. As informações lexicais e semânticas apresentam a classe gramatical a qual a palavra pertence e correlaciona com as línguas inglesa e italiana. Já as informações afetivas dizem respeito às teorias de emoções (modelo OCC, introduzido na seção 2.1.3) e são baseados nos conceitos de avaliação cognitiva (ORTONY *et al.* 1990) e nas teorias das emoções básicas (ELLIOT, 1992), representado pela valência positiva ou negativa. Portanto, para ser implementado o *Wordnet AffectBR*, os *Synsets* precisam ter ligação com o *Wordnet*, ou seja, precisaria implementar um “*ID Synset*” (código de localização do *Synset* na base *Wordnet*) com o estado afetivo das emoções (campo EMO), baseando-se em ORTONY *et al.* (1990), conforme mostrado pela Figura 2.

Figura 2: Estrutura de construção da base *Wordnet AffectBR*



Fonte: Pasqualotti e Vieira, 2008.

3.2.5 *SentiStrength*

O *SentiStrength* foi criado por Thelwall *et al.* (2010) na língua inglesa, com objetivo de identificar sentimentos em textos curtos. O léxico classifica, de forma automática, até 16.000 textos por segundo com precisão como se fosse efetuado por um humano.

O *SentiStrength*, disponível online¹⁹, executa a classificação no decorrer de como são feitas as inserções das palavras, ou seja, a classificação é feita em um intervalo de 1 a 5 para as palavras positivas, e -5 a -1 para as palavras negativas (GARCIA e SCHWEITZER, 2011). Utilizando a ferramenta, simulou-se a frase “*I really love you but dislike your cold sister*” e como resultado, a ferramenta mostra que: *I really love* [3] [+1 booster word] *you but dislike* [-3] *your cold* [-2] *sister*, sendo assim, o resultado da análise aponta pontuação 4, sentimento positivo forte e pontuação -5, sentimento negativo forte. Observa-se que as pontuações, que são atribuídas a cada palavra, são decorrentes do que já foi previamente classificado, analisado e avaliado, estão presentes no léxico. As palavras que aparecem no texto, e por ventura não existam no léxico (SENTISTRENGTH, 2015), não são analisadas e conseqüentemente não interferem no resultado da classificação final. Além desta classificação, o analisador ainda oferece uma classificação única, ou seja, fazendo uma análise no geral (texto inteiro). A análise é feita pegando o maior positivo e o maior negativo e efetua a diferença. No exemplo, o resultado da frase para a classificação única é 1.

Outro fator importante do classificador é a possibilidade da análise de textos em diversos idiomas, entre eles o português, mas não com a precisão que se tem para o inglês.

3.2.6 *SenticNet*

O Léxico foi criado pelo laboratório MIT *Media* da Universidade de *Stirling* e uma empresa *Sitekit Solutions Ltd*, em 2009. Desenvolvido na língua inglesa o léxico vem sendo utilizado através de aplicações inteligentes, que vão da mineração de dados até a interação humano-computador, para a detecção de emoções via textos. O principal objetivo é fazer com que a informação conceitual e afetiva, transmitida pela linguagem natural, seja mais facilmente acessível às máquinas.

Para atingir o objetivo, o léxico proporciona significativamente a inferência de polaridade no texto por meio de informações de senso comum (*sentic computic*), *web* semântica ou técnicas de CA e não por meio semântico, ou seja, trechos do texto com emoções, baseada de forma explícita, como por exemplo: bom, excelente, agradável, melhor (termos positivos) ou ruim, infeliz, errado, pior (termos negativos) (CAMBRIA *et al.* 2010).

¹⁹ <http://sentistrength.wlv.ac.uk/>

O léxico está na sua versão 3.0, está disponível online²⁰, e utiliza um método de retorno de variáveis de sentimento que está relacionado com cada conceito encontrado em uma mensagem do texto, como o vetor *Sentic* e a pontuação de polaridade (positivo e negativo). O vetor é constituído por uma variação de emoções que estão agrupadas em quatro dimensões: simpatia, sensibilidade, aptidão e atenção. A pontuação de polaridade (positivo e negativo) é um valor real (-1, 1), equivalente a valores de polaridade proporcionados por outros métodos. (CAMBRIA *et al.* 2014)

3.2.7 *EmoSenticNet*

Léxico desenvolvido por Poria *et al.*(2013) na língua inglesa, reúne o grupo de emoções do *WordNetAffect* (raiva, tristeza, desgosto, alegria, medo e surpresa) para as definições do *sentic computic* do *SenticNet* propiciando a polaridade das definições dos conjuntos do *WordNetAffect*, demonstrado na Tabela 3.

Tabela 3: Classificação utilizando o léxico *EmoSenticNet*

| Conceito | Raiva | Tristeza | Desgosto | Alegria | Medo | Surpresa |
|---------------------|-------|----------|----------|---------|------|----------|
| <i>Peace</i> | 0 | 0 | 0 | 1 | 0 | 0 |
| <i>Indifference</i> | 0 | 0 | 1 | 0 | 0 | 0 |
| <i>Impatience</i> | 1 | 0 | 0 | 0 | 0 | 0 |
| <i>Flurry</i> | 1 | 1 | 0 | 0 | 0 | 0 |
| <i>Where</i> | 0 | 0 | 1 | 0 | 0 | 1 |
| <i>Emergency</i> | 0 | 0 | 0 | 0 | 1 | 0 |

Fonte: <http://www.gelbukh.com/emosenticnet/>

O léxico é considerado uma extensão do *WordNetAffect* porém possui um dicionário com cerca de 13.000 palavras, com os já presentes na lista do *WordNetAffect*. O léxico está disponível em <http://www.gelbukh.com/emosenticnet/> onde é possível encontrar a base de palavras, trabalhos publicados e download para uso. O seu uso é exclusivo para fins acadêmicos.

3.2.8 *ANEW-Br*

Este léxico foi adaptado por Kristensen *et al.* (2011) a partir da tradução do *ANEW* (BRADLEY e LANG, 1999) para o português brasileiro, dando origem ao *ANEW-Br*. Após a tradução para o português, houve uma tradução reversa (Cada uma das palavras foi traduzida novamente para inglês) por pelo menos duas pessoas com fluência na língua. Esse procedimento deu origem a duas listas de palavras, uma em português feita da tradução original e outra em inglês pela tradução reversa (tradução do

²⁰ <http://sentic.net/>

português). Posteriormente, um grupo de avaliadores independentes verificou a adequação das palavras traduzidas para o português brasileiro, utilizando como critério a equivalência semântica entre as duas traduções. Essa tarefa gerou concordância entre os avaliadores. Para as avaliações e análises foi mantido o método original que utiliza a escala *Self-Assessment Manikin* (SAM).

3.2.9 *OpinionLexicon*

O *OpinionLexicon* foi desenvolvido por Souza *et al.* (2011) na língua portuguesa. A técnica de construção é composta na aplicação de três métodos: (1) Baseado no *corpus* do Turney (2002), O *corpus* é composto de resenhas de filmes escritos em português e textos jornalísticos falando sobre diversos temas. No total resultaram em 1317 documentos com cerca de um milhão de palavras; (2) Baseado em semelhanças de palavras (KAMPS *et al.* 2004). Nesse método, os autores utilizaram uma função para calcular o menor caminho entre sinônimos e antônimos de palavras e; (3) Baseado em um sistema de tradução automática *on-line* ao invés de um dicionário bilíngue (MIHALCEA *et al.* 2007). Neste trabalho os autores utilizaram um mecanismo de tradução *online* do *Google*. Todas as expressões e palavras traduzidas foram utilizadas. Aquelas que o sistema de tradução não traduziu, devido à alta presença de variação linguística ou quaisquer outros motivos, como erros comuns no léxico original, foram descartados por revisão manual. O autor, em seguida, une os três métodos e dá origem a um grande léxico para o português brasileiro, mantendo os métodos que foram aplicados em seus originais.

3.2.10 *WordnetAffectBR_adapt*

O *WordnetAffectBR_adapt*²¹ foi adaptado por Longhi (2011) na língua portuguesa. Seu desenvolvimento foi uma ampliação do *WordnetAffectBR* (seção 3.2.4) proposto por Pasqualotti e Vieira (2008). O *WordAffectBR_adapt* contém 2194 registros de léxico afetivo, 100 *emoticons*²² e 139 interjeições, agrupados nos quadrantes e subquadrantes correspondentes. Já os advérbios (que totalizam 103) são necessários para modular a intensidade do termo afetivo. Ainda existem as *stopwords* (313 registradas), que a autora fala que são palavras de baixo valor semântico (artigos, preposições, pronomes e alguns verbos e advérbios), mas com alta frequência. As abreviações (238 totalizadas), sempre que identificadas, são desprezadas no processo de mineração. Um ponto positivo dessa ferramenta é que ela pode ser incorporada a qualquer outro ambiente virtual de aprendizagem, contanto que sejam feitas as adaptações necessárias na forma de recuperação dos textos.

²¹ Disponível em www.nuted.ufrgs.br/roodaafeto/relatorios/wordnet affect br_adapt.txt

²² Ícones criados a partir de uma sequência de pontuação, como por exemplo, :- e :-)

3.2.11 *VerbNet.Br*

O *VerbNet.Br* foi desenvolvido por Scarton (2013) na língua portuguesa de forma semiautomática para diminuir o tempo gasto e os possíveis erros como se fosse feito de forma manual. O método de criação é de forma genérica, ou seja, pode ser utilizado e reestruturado em outras línguas, além do português do Brasil. Portanto, a criação da *VerbNet.Br* foi fundamentando em quatro etapas: uma manual e três automáticas. A primeira etapa (manual) consiste da tradução das alternâncias sintáticas das classes da *VerbNet* para o Português. Na segunda etapa buscaram-se as alternâncias sintáticas dos verbos em português, utilizando-se corpora (Lácio-Ref (ALUÍSIO *et al.* 2004); PLN-BR-FULL (BRUCKSCHEN *et al.* 2008) e Revista FAPESP (AZIZ e SPECIA, 2011)) e a ferramenta de extração de frames de subcategorização desenvolvida por Zanette (2010). Na terceira etapa foram definidos os verbos candidatos a membros das classes através dos alinhamentos entre os recursos existentes. Por fim, a quarta etapa consistiu da combinação das três anteriores para selecionar os verbos da *VerbNet.Br*.

3.2.12 *Reli-Lex*

O *Reli-Lex* foi adaptado por (FREITAS, 2013) na língua portuguesa, a partir do *corpus Reli* (FREITAS *et al.* 2012). O léxico é composto por palavras e expressões compostas no *corpus* e sua seleção foi feita manualmente. Além de expressões e palavras, foram elaboradas listas que estavam relacionadas à inversão das polaridades (estruturas negativas). Para as classes das palavras, foram considerados adjetivos, verbos e negações. Para cada classe, foi criada uma lista de entradas com os lemas, obtidos de forma automática através de um analisador morfossintático chamado “Palavras” (BICK, 2000). Palavras que não tinham a presença de opinião foram descartadas.

Dois critérios guiaram a inclusão de uma entrada no léxico. De acordo com o primeiro, as palavras ou expressões precisariam ter a presença de opinião e polaridade. O segundo critério precisaria ainda apresentar uma estabilidade relativa quanto ao tipo de polaridade (e, por isso, a importância da verificação em outros corpora).

O léxico foi construído de forma manual pela preocupação que os autores tiveram em eliminar palavras que fossem portadoras de opinião/polaridade, mas somente em algum texto específico. Por isso, a construção do léxico não incidiu somente nas análises feitas no *corpus*, mas abrangeu a consulta a diversos corpora, por exemplo, *corpus* “Floresta” (AFONSO *et al.* 2002) e o *corpus* “*Reli*” (FREITAS, 2012).

3.3 Considerações Finais

Neste capítulo foram apresentados conceitos de léxicos computacionais e afetivos. Em léxicos computacionais foram analisados o *MRC Psycholinguistic DataBase*, o *WordNet* e o *VerbNet*, apresentando-os detalhadamente como foram compostos para serem adaptados ou diferenciados.

Dos léxicos afetivos foram analisados o *LIWC*, o *ANEW*, o *SentiWordNet*, o *Wordnet AffectBR*, o *SentiStrength*, o *SenticNet*, o *EmoSenticNet*, o *ANEW-Br*, *OpinionLexicon*, o *WordnetAffectBR_adapt*, o *VerbNet.br* e o *Reli-Lex*. Nesses léxicos foram apresentados a forma e os métodos utilizados para a sua construção ou adaptação. Foram também disponibilizados os endereços eletrônicos de alguns, onde é possível utilizá-los.

4 TRABALHOS RELACIONADOS

Para a construção deste capítulo, utilizou-se o estudo de mapeamento sistemático (PETERSEN *et al.* 2008). Ele foi escolhido por fornecer uma visão geral do estado da arte em alguma área.

Para alcançar este objetivo, seguiu-se um conjunto de etapas: (i) o primeiro foi para selecionar o escopo da pesquisa, considerando-se algumas questões de pesquisa a serem colocados sobre o assunto desejado; (ii) o segundo passo foi a criação dos estudos preliminares, apresentando a base de dados bibliográficos a ser pesquisada, bem como os parâmetros de pesquisa e filtros; depois, (iii) na etapa três, foram selecionados os artigos relevantes usando critérios de inclusão e exclusão; e, finalmente, (iv) no passo quatro, foram analisadas e os resultados discutidos.

4.1 Escopo de Busca e Questões de Pesquisa

Nesta seção, definiram-se as questões de pesquisa de interesse, que são apresentados na Tabela 4. De acordo com Petersen *et al.* (2008), as questões de pesquisa fornecem uma visão geral sobre a área de pesquisa mapeados pela identificação da quantidade de artigos e o tipo de pesquisa. O objetivo do mapeamento de pesquisa é encontrar artigos correlatos relacionados à mineração afetiva (dados afetivos) em textos. Foram estabelecidas questões de pesquisa (Qp) sobre o tema para a condução de busca dos artigos.

Tabela 4: Questões de Pesquisa

| | |
|-------------|---|
| Qp1 | Existe alguma palavra afetiva (sentimento, emoção, afeto e personalidade) em que os artigos foram baseados? |
| Qp2 | A qual teoria psicológica essas palavras afetivas são efetivamente relacionadas? |
| Qp3 | Existe menção a algum banco de dados (léxico, corpus, dicionário ou database própria)? |
| Qp4 | Qual o nome/título do banco de dados? |
| Qp5 | Esse banco de dados é afetivo? |
| Qp6 | Qual o idioma do banco de dados? |
| Qp7 | Quem é o autor do banco de dados? |
| Qp8 | Qual técnica computacional, ferramenta ou método foi utilizado para minerar os dados afetivos? |
| Qp9 | Quem são os autores dos artigos? |
| Qp10 | Quais as universidades que tem publicado neste campo? |

Qp11 Quais os países dos autores dos artigos?

Qp12 Em que ano os artigos foram publicados?

Este método permite a criação de um cenário sobre a área de cultivo com base nas questões de pesquisa.

4.2 Estudos Primários

Para fornecer uma visão geral sobre a área pesquisada, foi definido um conjunto de *strings*²³ com o objetivo de selecionar e filtrar os artigos para os estudos primários. De acordo com Petersen *et al.* (2008), "os estudos primários são identificados usando cadeias de pesquisa em bases de dados científicos ou navegar manualmente através anais de conferências relevantes ou publicações de revistas". Para realizar a pesquisa para os estudos primários, foi utilizada a base de dados bibliográfica SCOPUS²⁴. Acessada através do Portal de Periódicos da CAPES²⁵, abrange as bases de dados mais importantes em Ciência da Computação, como a Elsevier, Springer, IEEE e ACM. O SCOPUS foi acessado de fevereiro a março de 2016.

A fim de conseguir filtrar um maior número de artigos foram utilizadas nas buscas palavras-chave primárias e secundárias. As primárias referem-se à "*affective mining*" e "*personality mining*". As secundárias, "*emotion mining*", "*sentiment mining*", e "*feeling mining*" e "*sentiment analysis*".

A SCOPUS trabalha filtrando artigos a partir de uma *string* única ou uma coleção delas. Como foram definidas palavras-chave primárias e secundárias, a *string* utilizada para a pesquisa foi: "ABS (affect* AND "minin*" AND text) OR ABS (personalit* AND "minin*" AND text) OR ABS (sentiment* AND "minin*" AND text) OR ABS (emotion *AND "minin*" AND text) OR ABS (feeling AND "minin*" AND text) OR ABS (sentiment AND "analys*" AND text) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (EXCLUDE (DOCTYPE , "cr")) AND (LIMIT-TO (LANGUAGE , "English")) AND (EXCLUDE (DOCTYPE , "bk"))".

Essas *strings* foram definidas considerando-se as palavras que são igualmente utilizadas para definir afetividade, de acordo com Munezero et al. (2014) e Lisetti (2002). Além disso, usou-se a palavra "análise" como sinônimo de "mineração", quando usado em conjunto com o sentimento, porque os cientistas na área costumam usar essa terminologia, como descrito anteriormente.

²³ Palavra ou conjunto de palavras

²⁴ <https://www.elsevier.com/solutions/scopus>

²⁵ Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. <http://www.periodicos.capes.gov.br/>

4.3 Artigos relevantes do banco de dados Scopus

Nesta etapa da metodologia, foram definidos os critérios de inclusão e exclusão, a fim de manter apenas os artigos relevantes nos estudos primários. Estes artigos foram definidos com base em 3 fases.

Na primeira, fase 1, foram incluídos todos os documentos disponíveis na SCOPUS filtrados pelas strings que combinavam com os resumos dos trabalhos. Os documentos também foram filtrados pela área de Ciência da Computação e Inglês. Depois disso, retornaram-se 596 artigos da SCOPUS.

Na segunda, fase 2, foram excluídos os artigos de comentários de conferências, documentos que eram apenas resumos, e artigos que não estavam disponíveis para download. Depois disso, a SCOPUS retornou 448 artigos, 75,2% a partir dos artigos originais.

Finalmente, na fase 3, foram analisados todos os *abstracts*²⁶ desses 448 artigos. Em seguida, foram excluídos os artigos que considerados fora de escopo de pesquisa. Após a exclusão, foram selecionados e analisados 106 daqueles 596 artigos iniciais, o que corresponde a 17,8%.

Tabela 5: Estudos primários retornados a partir das etapas propostas

| Base de Dados Científica | Quantidade | | | | |
|--------------------------|------------|--------|------|--------|------|
| | Fase 1 | Fase 2 | % | Fase 3 | % |
| SCOPUS | 596 | 448 | 75,2 | 106 | 17,8 |

4.4 Resultados e Discussões

Nesta seção apresentam-se as análises sobre os artigos selecionados, criando uma visão geral sobre “*affective mining*” e “*personality mining*”.

Depois de analisar esses 106 trabalhos, foi construída a Tabela 6. Esta tabela oferece uma visão geral sobre as perspectivas no sentido de responder às perguntas de investigação propostos na Tabela 4.

²⁶ Resumos

Tabela 6: Artigos analisados

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27, 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|----------------------------|-----------------------|-----------------------------------|---------------------------------|-------------------|---------------------------------|---|-----------------------|--|---|--------------------|--------------------|
| 2 | <i>Sentiment e Emotion</i> | Não relaciona | Léxico | <i>SentiWordNet</i> | Não | Inglês | Esuli & Sebastiani | SVM | Khan, F.H., Qamar, U., Bashir, S. | <i>National University of Sciences and Technology</i> | Paquistão | 2016 |
| 10 | <i>Sentiment e Emotion</i> | Não relaciona | <i>Database Twitter</i> | <i>Stocktwits</i> | Não | Inglês | Nasseri, A.A., Tucker, A., De Cesare, S. | DT e C4.5 | Nasseri, A.A., Tucker, A., De Cesare, S. | <i>Brunel University</i> | Reino Unido | 2016 |
| 12 | <i>Sentiment</i> | Não relaciona | Léxico | <i>SentiWordNet</i> | Sim | Inglês | Esuli & Sebastiani | RF, LR e SVM | Alahmadi, D.H., Zeng, X.-J. | <i>University of Manchester</i> | Reino Unido | 2015 |
| 13 | <i>Sentiment</i> | Não relaciona | Léxico | <i>SenticNet e SentiWordNet</i> | Sim | Inglês | Cambria, E., Havasi, C. and Hussain, A.; Esuli e Sebastiani | SVM | Musto, C., Semeraro, G., Lops, P., Gemmis, M.D. | <i>Universita degli Studi di Bari Aldo</i> | Itália | 2015 |
| 15 | <i>Sentiment</i> | <i>Feldman (2013)</i> | Database | <i>FactNet</i> | Não | Inglês | Noferesti, S., Shamsfard, M. | PR e FM | Noferesti, S., Shamsfard, M. | <i>Shahid Beheshti University</i> | Iran | 2015 |
| 16 | <i>Sentiment</i> | Não relaciona | Léxico | Próprio | Não | Inglês | Ajitha, P., Gunasekaran, G. | Não apresenta | Ajitha, P., Gunasekaran, G. | <i>Sathyabama University</i> | Índia | 2015 |
| 17 | <i>Sentiment e Mood</i> | Não relaciona | Database fóruns | Próprio | Não | Inglês | Kazi, F., Joshi, S., Machchhar, S. | SVM, NB e DT | Kazi, F., Joshi, S., Machchhar, S. | <i>Faculty of PG Studies</i> | Índia | 2015 |
| 19 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Micro-Wnop</i> | Não | Inglês | Cerini et al. | SVM, NB, C45, DT, ME | Li, J., Fong, S., Zhuang, Y., Khoury, R. | <i>University of Macau</i> | Macau | 2015 |
| 21 | <i>Sentiment</i> | Não relaciona | <i>Database Facebook, Twitter</i> | Próprio | Não | Inglês, Alemão Tcheco e Polonês | Petz, G., Karpowicz, M., Fürschuß, H., Střiteský, V., Holzinger, A. | Não apresenta | Petz, G., Karpowicz, M., Fürschuß, H., (...), Střiteský, V., Holzinger, A. | <i>University of Applied Sciences Upper</i> | Áustria | 2015 |

²⁷ Qp1 – Existe alguma palavra afetiva (sentimento, emoção, afeto e personalidade) em que os artigos foram baseados?

Qp2 – A qual teoria psicológica essas palavras afetivas são efetivamente relacionadas?

Qp3 – Existe menção a algum banco de dados (léxico, corpus, dicionário ou database própria)?

Qp4 – Qual o nome/título do banco de dados?

Qp5 – Esse banco de dados é afetivo?

Qp6 – Qual o idioma do banco de dados?

Qp7 – Quem é o autor do banco de dados?

Qp8 – Qual técnica computacional (algoritmo), ferramenta ou método foi utilizado para minerar os dados afetivos?

Qp9 – Quem são os autores dos artigos?

Qp10 – Quais as universidades que tem publicado neste campo?

Qp11 – Quais os países dos autores dos artigos?

Qp12 – Em que ano os artigos foram publicados?

²⁸ Siglas das ferramentas, técnicas e métodos: Nearest neighbour (KNN), Naive Bayes (NB), Linear Regression (LR), Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), REPTree (REP), Precision Recall (PR), F-measure (FM), Decision Tree (DT), Random Forest (RF), Max Entropy (ME), Accuracy (AC), Airline Quality Rating (AQR), Random (RD), Bag of word (BOW), Combined Sentiment Topic (CST), Programação Não Linear (NLP), Learning Vector Quantization (LVQ), Probabilistic Latent Semantic Analyses (PLSA)

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|-------------------------------|---|----------------------------------|---|-------------------|--------------------|--|----------------------|--|--|--------------------|--------------------|
| 31 | <i>Sentiment</i> | Não relaciona | <i>Database Collection</i> | Próprio | Não | Inglês | Vinodhini, G., Chandrasekaran, R.M. | Ferramenta própria | Vinodhini, G., Chandrasekaran, R.M. | <i>Annamalai University</i> | Índia | 2015 |
| 32 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | <i>Sentence polarity e Large movie review</i> | Não | Inglês | Bo Pang e Lillian Lee; e Andrew Maas | NB e ME | Zhang, P., He, Z. | <i>Chongqing University</i> | China | 2015 |
| 35 | <i>Sentiment</i> | <i>Feldman (2013)</i> | <i>Database</i> | Próprio | Não | Inglês | Ravi, K., Ravi, V., Gautam, C. | SVM e LR | Ravi, K., Ravi, V., Gautam, C. | <i>University of Hyderabad</i> | Índia | 2015 |
| 43 | <i>Sentiment e Mood</i> | Não relaciona | Léxico | Conversas de voz | Não | China | Hegde, Y., Padma, S.K. | AC, PR e FM | Hegde, Y., Padma, S.K. | <i>SJCE</i> | Índia | 2015 |
| 54 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter e Amazon</i> | <i>Large movie review</i> | Não | Espanha | Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E. | AC e PR | Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E. | <i>University of Castilla</i> | Espanha | 2015 |
| 59 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Italiano | Candelieri, A., Archetti, F. | SVM | Candelieri, A., Archetti, F. | <i>University of Milano</i> | Itália | 2015 |
| 61 | <i>Sentiment e Opinion</i> | Não relaciona | Dicionário <i>Sports</i> | Próprio | Não | Arábia | Khasawneh, R.T., Wahsheh, H.A., Alsmadi, I.M., Ai-Kabi, M.N. | PR e FM | Khasawneh, R.T., Wahsheh, H.A., Alsmadi, I.M., Ai-Kabi, M.N. | <i>Jordan University of Science and Technology</i> | Jordânia | 2015 |
| 64 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Tailandês e Inglês | Polpinij, J. | NB e SVM | Polpinij, J. | <i>Maharakham University</i> | Tailândia | 2015 |
| 65 | <i>Opinion</i> | Não relaciona | <i>Database Facebook</i> | Próprio | Não | Inglês | Walha, A., Ghazzi, F., Gargouri, F. | Não apresenta | Walha, A., Ghazzi, F., Gargouri, F. | <i>Sfax University</i> | Tunísia | 2015 |
| 66 | <i>Sentiment e Opinion</i> | Não relaciona | Léxico | Próprio | Sim | Inglês | Muhammad, A., Wiratunga, N., Lothian, R. | SVM, NB e LR | Muhammad, A., Wiratunga, N., Lothian, R. | <i>Robert Gordon University</i> | Reino Unido | 2015 |
| 68 | <i>Personality (Big Five)</i> | <i>Cattell (1943) e Goldberg (1990)</i> | <i>Database Web</i> | Próprio | Sim | Inglês | Poddar, S., Kattagoni, V., Singh, N. | Binary | Poddar, S., Kattagoni, V., Singh, N. | <i>IIT Hyderabad</i> | Índia | 2015 |
| 73 | <i>Sentiment</i> | Não relaciona | <i>Database TripAdvisor</i> | Próprio | Não | Chinês | Cherif, W., Madani, A., Kissi, M. | SVM | Cherif, W., Madani, A., Kissi, M. | <i>Faculty of Sciences</i> | Marrocos | 2015 |
| 75 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês | Alsaffar, A., Omar, N. | KNN | Alsaffar, A., Omar, N. | <i>University Kebangsaan Malaysia</i> | Malásia | 2015 |
| 77 | <i>Sentiment</i> | <i>Feldman (2013)</i> | Léxico | Próprio | Sim | Inglês | Tang, D., Qin, B., Liu, T. | Não apresenta | Tang, D., Qin, B., Liu, T. | <i>Harbin Institute of Technology</i> | China | 2015 |

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|----------------------------|-----------------------|----------------------------|---------------------------------------|-------------------|-------------------|---|----------------------|---|---|--------------------|--------------------|
| 78 | <i>Sentiment e Opinion</i> | <i>Feldman</i> (2013) | <i>Database</i> | Próprio | Sim | Inglês | Thakur, D., Singh, J. | Não apresenta | Thakur, D., Singh, J. | <i>Chitkara University</i> | Índia | 2015 |
| 87 | <i>Sentiment</i> | Não relaciona | <i>Database documentos</i> | Próprio | Não | Inglês | Nigam, S., Jawdekar, A. | PR e FM | Nigam, S., Jawdekar, A. | <i>Shri Ram College of Engineering & Management</i> | Índia | 2015 |
| 88 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Inglês | Almatrafi, O., Parack, S., Chavan, B. | Não apresenta | Almatrafi, O., Parack, S., Chavan, B. | <i>George Mason University</i> | EUA | 2015 |
| 92 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Inglês | Kanakaraj, M., Guddeti, R.M.R. | Não apresenta | Kanakaraj, M., Guddeti, R.M.R. | <i>National Institute of Technology</i> | Índia | 2015 |
| 103 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | <i>Movie Reviews</i> | Não | Inglês | Pang et al. | SVM e LR | Vuong, Q.-H., Takasu, A. | <i>National Institute of Informatic</i> | Japão | 2014 |
| 107 | <i>Sentiment e Opinion</i> | Não relaciona | <i>Database</i> | <i>Cornell e Stanford large movie</i> | Não | Inglês | Pang et al. e Maas et al. | C4.5 | Zhi-Hong Deng, Kun-Hu Luo, Hong-Liang Yu | <i>Peking University</i> | China | 2014 |
| 111 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Micro-Wnop</i> | Não | Inglês | Cerini et al. | NB, C45, DT, ME | Li, J., Fong, S., Zhuang, Y., Khoury, R. | <i>University of Macau</i> | Macau | 2014 |
| 114 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>ChnSentiCorp</i> | Não | Chinês | Songbo Tan | SVM | Wang, H., Liu, L., Song, W., Lu, J. | <i>Capital Normal University</i> | China | 2014 |
| 115 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | <i>Movie Reviews</i> | Não | Inglês | Pang et al. | SVM e NB | Francesco C., Massimo De Santo, Luca G. | <i>University of Salerno</i> | Itália | 2014 |
| 116 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Não | Inglês | Li, D., Niu, J., Qiu, M., Liu, M. | REP | Li, D., Niu, J., Qiu, M., Liu, M. | <i>Beihang University</i> | China | 2014 |
| 123 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês | García-Pablos, A., Cuadros, M., Gaines, S., Rigau, G. | SVM e NB | García-Pablos, A., Cuadros, M., Gaines, S., Rigau, G. | <i>Vicomtech-IK4 research centre</i> | Espanha | 2014 |
| 130 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês e Chinês | Chen, X., Tang, W., Xu, H., Hu, X. | Não apresenta | Chen, X., Tang, W., Xu, H., Hu, X. | <i>Shanghai University</i> | China | 2014 |
| 132 | <i>Sentiment e Opinion</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Inglês | Petz, G., Karpowicz, M., Fürschuß, H., (...), Stríteský, V., Holzinger, A. | Não | Petz, G., Karpowicz, M., Fürschuß, H., (...), Stríteský, V., Holzinger, A. | <i>University of Applied Sciences Upper</i> | Áustria | 2014 |
| 139 | <i>Sentiment</i> | Não relaciona | Léxico | Próprio | Não | Árabe | Abdulla, N., Majdalawi, R., Mohammed, S., Al-Ayyoub, M., Al-Kabi, M. | PR e FM | Abdulla, N., Majdalawi, R., Mohammed, S., Al-Ayyoub, M., Al-Kabi, M. | <i>Jordan University of Science and Technology</i> | Jordânia | 2014 |
| 140 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | <i>Cimawa</i> | Não | Inglês | P. Uhr, A. Klahold, M. Fathi | SVM | Uhr, P., Zenkert, J., Fathi, M. | <i>University of Siegen</i> | Alemanha | 2014 |
| 146 | <i>Sentiment</i> | <i>Feldman</i> (2013) | <i>Corpus</i> | Próprio | Não | Espanhol | Del Pilar Salas-Zárate, M., López-López, E., Valencia-García, R., (...), Almela, Á., Alor-Hernández, G. | C4.5, DT e SMO | Del Pilar Salas-Zárate, M., López-López, E., Valencia-García, R., (...), Almela, Á., Alor-Hernández, G. | <i>Universidad de Murcia</i> | Espanha | 2014 |

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|----------------------------|--|---|-------------------------------|-------------------|-------------------|---|----------------------|---|--|--------------------|--------------------|
| 153 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Inglês | Molla, A., Biadgie, Y., Sohn, K.-A. | NodeXL | Molla, A., Biadgie, Y., Sohn, K.A. | <i>Ajou University</i> | Coréia do Sul | 2014 |
| 155 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Sim | Árabe | Akaichi, J. | SVM e NB | Akaichi, J. | <i>ISG-University of Tunis</i> | Tunísia | 2014 |
| 161 | <i>Sentiment</i> | Não relaciona | Léxico | Hu, M., & Liu, B | Não | Inglês | Hu, M., & Liu, B | AQR | Adeborna, E., Siau, K. | <i>Missouri University of Science and Technology</i> | EUA | 2014 |
| 163 | <i>Sentiment e Opinion</i> | Não relaciona | <i>Database Twitter</i> | Próprio | Não | Árabe | Shokry, A. | SVM | Al-Radaideh, Q.A., Twaiq, L.M. | <i>Yarmouk University</i> | Jordânia | 2014 |
| 165 | <i>Sentiment</i> | Não relaciona | <i>Corpus Twitter e Facebook</i> | Próprio | Não | Inglês | Isah, H., Trundle, P., Neagu, D. | SVM | Isah, H., Trundle, P., Neagu, D. | <i>University of Bradford</i> | Reino Unido | 2014 |
| 168 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Wordnet e Sentiwordnet</i> | Sim | Inglês | Fellbaum, C. e Esuli, A., and Sebastiani | Não apresenta | Hasan, K.M.A., Rahman, M., Badiuzzaman | <i>Khulna University of Engineering and Technology</i> | Bangladesh | 2014 |
| 174 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>General Inquirer</i> | Não | Inglês | Stone et al. | SO-PMI | Hassan, A., Abu-Jbara, A., Lu, W., Radev, D. | <i>University of Michigan</i> | EUA | 2014 |
| 176 | <i>Sentiment</i> | <i>Feldman (2013)</i> | <i>Database</i> | <i>Smn e Smt</i> | Não | Português | Neuenschwander, B., Pereira, A., Meira, W., Barbosa, D. | NB e RD | Neuenschwander, B., Pereira, A., Meira, W., Barbosa, D. | <i>UFMG</i> | Brasil | 2014 |
| 189 | <i>Sentiment</i> | Não relaciona | <i>Corpus Twitter</i> | Próprio | Sim | Árabe | Al-Osaimi, S., Badruddin, K.M. | NB | Al-Osaimi, S., Badruddin, K.M. | <i>Imam Muhammad ibn Saud Islamic University</i> | Arábia Saudita | 2014 |
| 191 | <i>Sentiment</i> | Não relaciona | <i>Database Twitter, Facebook e blogs</i> | Próprio | Não | Italiano | Colace, F., Casaburi, L., De Santo, M., Greco, L. | SVM e NB | Colace, F., Casaburi, L., De Santo, M., Greco, L. | <i>University of Salerno</i> | Itália | 2014 |
| 202 | Não | Não relaciona | Léxico | Tenor | Não | Inglês e Espanhol | Mosquera et al. | Não apresenta | Mosquera, A. and P. Moreda | <i>Universitat d'Alicante</i> | Espanha | 2013 |
| 203 | <i>Feeling</i> | Não relaciona | <i>Corpus Facebook</i> | Próprio | Sim | Inglês | Troussas, C., Virvou, M., Espinosa, K.J., Llaguno, K., Caro, J. | NB | Troussas, C., Virvou, M., Espinosa, K.J., Llaguno, K., Caro, J. | <i>University of Piraeus</i> | Grécia | 2013 |
| 206 | <i>Sentiment e emotion</i> | Não relaciona | Léxico <i>Facebook</i> | Próprio | Sim | Inglês | Akaichi, A. | NB e SVM | Akaichi, A. | <i>University of Tunis</i> | Tunísia | 2013 |
| 210 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Sim | Italiano | Gaglio, V e Pilato, A. | NB | Gaglio, V e Pilato, A. | <i>Universit`a di Palermo</i> | Itália | 2013 |
| 214 | <i>Sentiment</i> | <i>Scherer (2007) e Ortony et al. (1990)</i> | Léxico | Próprio | Sim | Inglês | Park, E., Storey, V e Givens, S. | Não apresenta | Park, E., Storey, V e Givens, S. | <i>Georgia State University</i> | EUA | 2013 |

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|-------------------------------|-----------------------|-------------------|------------------------------|-------------------|-------------------|--|----------------------|---|--|--------------------|--------------------|
| 218 | <i>Sentiment</i> | Não relaciona | Léxico | Próprio | Sim | Inglês | Akaichi, J., Z. Dhouioui, and M. J. Lopez-Huertas Perez | NB e SVM | Akaichi, J., Z. Dhouioui, and M. J. Lopez-Huertas Perez | <i>Institut supérieur de gestion</i> | Tunísia | 2013 |
| 221 | <i>Sentiment</i> | Não relaciona | Léxico | Próprio | Não | Chinês | Wen, B., P. Fan, W. Dai, e L. Ding | NB e SVM | Wen, B., P. Fan, W. Dai, e L. Ding | <i>Hubei University of Science</i> | China | 2013 |
| 224 | <i>Personality (Big Five)</i> | <i>Cattell</i> (1943) | Léxico | <i>Senticnet, LIWC e MRC</i> | Sim | Inglês | Cambria, E., Havasi, C. and Hussain, A.; Tausczik, Y. R., & Pennebaker, J. W.; Coltheart, M. | SVM | Poria, S., A. Gelbukh, B. Agarwal, E. Cambria, e N. Howard | <i>Nanyang Technological University</i> | México | 2013 |
| 231 | <i>Sentiment</i> | Não relaciona | Não | Não | Não | Não | Não | SVM | Emma Haddia, Xiaohui Liua, Yong Shib | <i>Brunel University</i> | China | 2013 |
| 234 | <i>Sentiment</i> | Não relaciona | Léxico | <i>SentiStrength</i> | Sim | Sim | Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas | NB e SVM | Bahrainian, S.-A., Dengel, A. | <i>University of Kaiserslautern</i> | EUA | 2013 |
| 237 | <i>Sentiment</i> | Não relaciona | Não | Não | Não | Não | Não | NB e SVM | Puteh, M., Isa, N., Puteh, S., Redzuan, N.A. | <i>UiTM Terengganu</i> | Malásia | 2013 |
| 241 | <i>Sentiment e Feeling</i> | Não relaciona | Léxico | <i>SentiWordNet</i> | Sim | Inglês | Esuli e Sebastiani | Não apresenta | Kaur, A., Gupta, V. | <i>University Institute of Engineering</i> | Índia | 2013 |
| 242 | <i>Sentiment</i> | Não relaciona | Léxico | <i>ANEW</i> | Sim | Inglês | M. M. Bradley and P. J. Lang | ESAF | Zhou, X., Tao, X., Yong, J., Yang, Z. | <i>Queensland University of Technology</i> | Canadá | 2013 |
| 243 | <i>Emotion</i> | <i>Scherer</i> (2001) | Dicionário | Próprio | Sim | Inglês | Huangfu, L., Mao, W., Zeng, D., Wang, L. | FM e PR | Huangfu, L., Mao, W., Zeng, D., Wang, L. | <i>University of Arizona</i> | EUA | 2013 |
| 246 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Não | Inglês | Robaldo, L., Di Caro, L. | SVM | Robaldo, L., Di Caro, L. | <i>University of Turin</i> | Itália | 2013 |
| 250 | <i>Sentiment</i> | Não relaciona | Léxico | <i>WordNet</i> | Sim | Inglês | C. Fellbaum | BOW | Karamibekr, M., Ghorbani, A.A. | <i>University of New Brunswick</i> | EUA | 2013 |
| 251 | <i>Sentiment</i> | Não relaciona | Não | Não | Não | Não | Não | ZMET | Ahn, H.J. | <i>Hongik University</i> | Coréia do Sul | 2013 |
| 252 | <i>Sentiment</i> | Não relaciona | Léxico | <i>Based Approach</i> | Não | Inglês | Hu, M., & Liu, B. | QDA Miner 4.0 | Mostafa, M.M. | <i>University de Lisboa</i> | Portugal | 2013 |
| 253 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | <i>Mobile Phone</i> | Não | Inglês | Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N. | NLP | Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N. | <i>Hellenic University</i> | Grécia | 2013 |
| 256 | <i>Sentiment</i> | Não relaciona | Léxico | <i>MPQA</i> | Não | Inglês | Wiebe, Wilson, and Cardie | CST | Usha, M.S., Indra Devi, M. | <i>KLN College of Information</i> | Índia | 2013 |
| 257 | <i>Sentiment</i> | Não relaciona | Corpora | <i>Cornell</i> | Não | Inglês | Pang, B., & Lee, L. | SVM | Mouthami, K., Devi, K.N., Bhaskaran, V.M. | <i>Kongu Engineering College</i> | Índia | 2013 |
| 277 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>IDF</i> | Não | Inglês | Metzler, D. Generalized inverse document frequency. | LVQ | Jotheeswaran, J., Kumaraswamy, Y.S. | <i>SR technical University</i> | Índia | 2013 |
| 281 | <i>Sentiment</i> | Não relaciona | Léxico | <i>Based Approach</i> | Sim | Inglês | Miao et al. | QDA Miner 4.0 | Mostafa, M.M. | <i>University de Lisboa</i> | Portugal | 2013 |

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ^{27 28} | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|--|--|--------------------|-------------------|-------------------|-------------------|---|----------------------|--|--|--------------------|--------------------|
| 282 | <i>Sentiment</i> | Não relaciona | Léxico | SentiWordNet | Sim | Inglês | Baccianella et al. | C4.5, NB SVM, LR | Bravo-Marquez, F., Mendoza, M., Poblete, B. | <i>University of Chile</i> | EUA | 2013 |
| 284 | <i>Sentiment</i> | Não relaciona | Não | Não | Não | Não | Não | PLSA | Gupta, N., Abhinav, K.R., Annappa, A. | <i>National Institute of Technology</i> | Índia | 2013 |
| 285 | <i>Sentiment</i> | Não relaciona | Léxico | Próprio | Não | Inglês | Satheesh Kumar, R., Vijayan, S. | NB | Satheesh Kumar, R., Vijayan, S. | <i>Hindusthan College of Engineering</i> | Índia | 2013 |
| 288 | <i>Sentiment</i> | Não relaciona | Corpus | Próprio | Não | Inglês | Farhadloo, M., Rolland, E. | BOW | Farhadloo, M., Rolland, E. | <i>University of California</i> | EUA | 2013 |
| 290 | <i>Feeling, Sentiment e Mood</i> | Ortony et al. (1990) | Corpus | Próprio | Não | Inglês | Maria, M., Makoto, N. | Método próprio | Maria, M., Makoto, N. | <i>Webster University Vienna</i> | Holanda | 2013 |
| 302 | <i>Sentiment</i> | Não relaciona | Corpus | MPQA | Não | Inglês | J. Wiebe and C. Cardie | BOW e SVM | Gînsca, A.L. | <i>Alexandru Ioan Cuza University</i> | Reino Unido | 2012 |
| 306 | <i>Sentiment</i> | Não relaciona | Dicionário | Próprio | Não | Chinês e Inglês | Li, C., Ma, J. | NB | Li, C., Ma, J. | <i>Southwest University</i> | China | 2012 |
| 308 | <i>Sentiment</i> | Não relaciona | Corpus | Próprio | Não | Inglês | Martín-Valdivia, M.-T., Montejó-Ráez, A., Ureña-López, A., Saleh, M.R. | SVM, LR e KNN | Martín-Valdivia, M.-T., Montejó-Ráez, A., Ureña-López, A., Saleh, M.R. | <i>University of Jaén</i> | Espanha | 2012 |
| 310 | <i>Sentiment</i> | Não relaciona | Database Twitter | Próprio | Não | Inglês | Bak, J., Kim, S., Oh, A. | SVM | Bak, J., Kim, S., Oh, A. | <i>Korea Advanced Institute</i> | Coréia do Sul | 2012 |
| 317 | <i>Personality (Big Five)</i> | Goldberg (1990) | Léxico | LIWC | Sim | Inglês | J. W. Pennebaker, M. E. Francis, and R. J. Booth, | LR e REP | Wald, R., Khoshgoftaar, T., Sumner, C. | <i>Florida Atlantic University</i> | EUA | 2012 |
| 320 | <i>Feeling e Sentiment</i> | Não relaciona | Léxico readaptável | Cornetto | Sim | Holandês | P. Vossen, I. Maks, R. Segers, H. van der Vliet | Não apresenta | Maks, I., Vossen, P. | <i>VU University</i> | Holanda | 2012 |
| 334 | <i>Sentiment</i> | Não relaciona | Corpus | Próprio | Não | Inglês | Naradhipa, A.R., Purwarianti, A. | NB, SVM e ME | Naradhipa, A.R., Purwarianti, A. | <i>Bandung Institute of Technology Bandung</i> | Indonésia | 2012 |
| 337 | <i>Sentiment</i> | Não relaciona | Database Forums | Próprio | Não | Inglês | Preethi, T., Nirmala Devi, K., Murali Bhaskaran, V. | SVM | Preethi, T., Nirmala Devi, K., Murali Bhaskaran, V. | <i>Kongu Engineering College</i> | Índia | 2012 |
| 342 | <i>Sentiment</i> | Não relaciona | Database Twitter | Próprio | Não | Inglês | Zhang, D., Si, L., Rego, V.J. | SVM | Zhang, D., Si, L., Rego, V.J. | <i>Purdue University</i> | EUA | 2012 |
| 353 | <i>Sentiment</i> | Não relaciona | Database Blog | Próprio | Não | Inglês | Costa, E., Ferreira, R., Brito, P., (...), Machado, A., Marinho, T. | NB | Costa, E., Ferreira, R., Brito, P., (...), Machado, A., Marinho, T. | <i>Universidade Federal de Alagoas</i> | Brasil | 2012 |
| 357 | <i>Sentiment</i> | Não relaciona | Corpus | Próprio | Não | Inglês | Leong, C.K., Lee, Y.H., Mak, W.K. | Não apresenta | Leong, C.K., Lee, Y.H., Mak, W.K. | <i>SIM University</i> | Singapore | 2012 |
| 375 | <i>Sentiment, Personality (Big Five)</i> | Goldberg (1990) e Ortony et al. (1990) | Léxico | Anew e Affinity | Sim e Não | Inglês | Bradley, M.M., Lang, P.J., Cuthbert, B.N.: e Liu, H., Lieberman, H., Selker, T. | SMO | Loureiro, D., Marreiros, G., Neves, J. | <i>Minho University</i> | Portugal | 2011 |

| Nº Artigo | Qp1 ²⁷ | Qp2 ²⁷ | Qp3 ²⁷ | Qp4 ²⁷ | Qp5 ²⁷ | Qp6 ²⁷ | Qp7 ²⁷ | Qp8 ²⁷ | Qp9 ²⁷ | Qp10 ²⁷ | Qp11 ²⁷ | Qp12 ²⁷ |
|-----------|----------------------------|-----------------------|---------------------------|-------------------|-------------------|----------------------------|---|---------------------|--|---|--------------------|--------------------|
| 381 | <i>Sentiment e Mood</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês | Singh, V.K., Mukherjee, M., Mehta, G.K. | SVM | Singh, V.K., Mukherjee, M., Mehta, G.K. | <i>Banaras Hindu University</i> | Índia | 2011 |
| 384 | <i>Sentiment</i> | Não relaciona | <i>Database Amazom</i> | Próprio | Não | Inglês | Delmonte, R., Pallotta, V. | Não apresenta | Delmonte, R., Pallotta, V. | <i>Università Ca Foscari</i> | Itália | 2011 |
| 389 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Muchocine</i> | Não | Espanhol | Cruz, F.L., Troyano, J.A., Enriquez, F., Ortega | SVM, NB, KNN e C4.5 | Martínez-Cámara, E., Martín-Valdivia, M.T., Ureña-López, L.A. | <i>University of Jaén</i> | Espanha | 2011 |
| 406 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Não | Inglês | He, S., Guo, Y., Ghanem, M. | SVM e NB | He, S., Guo, Y., Ghanem, M. | <i>South Kensington Campus</i> | Reino Unido | 2010 |
| 412 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês | Whitehead, M., Yaeger, L. | SVM | Whitehead, M., Yaeger, L. | <i>School of Informatics</i> | EUA | 2010 |
| 423 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Senti</i> | Não | Inglês | Xia, H., Tao, M., Wang, Y. | SVM | Xia, H., Tao, M., Wang, Y. | <i>Wuhan University of Science</i> | China | 2010 |
| 432 | <i>Sentiment</i> | <i>Feldman (2013)</i> | Léxico | <i>LIWC</i> | Sim | Inglês | Tausczik, Y. R., & Pennebaker, J. W. | SVM e NB | Gobron, S., Ahn, J., Paltoglou, G., Thelwall, M., Thalmann, D. | <i>School of Engineering, Neuchatel</i> | Suíça | 2010 |
| 434 | <i>Sentiment</i> | Não relaciona | <i>Database Sentenças</i> | Próprio | Não | Inglês | Shein, K.P.P., Nyunt, T.T.S. | SVM | Shein, K.P.P., Nyunt, T.T.S. | <i>University of Computer Studies</i> | Singapura | 2010 |
| 444 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | Próprio | Sim | Inglês | Cheng, M., Xin, Y., Bao, J., Wang, C., Yang, Y. | SVM | Cheng, M., Xin, Y., Bao, J., Wang, C., Yang, Y. | <i>Beijing University</i> | China | 2009 |
| 449 | <i>Sentiment</i> | Não relaciona | <i>Database sentenças</i> | Próprio | Não | Inglês | Shein, K.P.P. | SVM | Shein, K.P.P. | <i>University of Computer Studies</i> | Singapura | 2009 |
| 452 | <i>Sentiment</i> | Não relaciona | <i>Corpus</i> | <i>Senti</i> | Não | Chinês | Doctor Tan Songbo | SVM, KNN e NB | Xia, Z., Suzhen, W., Mingzhu, X., Yixin, Y. | <i>Hebei University of Economics</i> | China | 2009 |
| 457 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Não | Verma, R. | Não apresenta | Verma, R. | <i>Paris School of Economics</i> | França | 2009 |
| 460 | <i>Sentiment e Emotion</i> | Não relaciona | Léxico | Próprio | Sim | Francês | Verma, R. | Não apresenta | Verma, R. | <i>Paris School of Economics</i> | França | 2009 |
| 461 | <i>Sentiment e Feeling</i> | Não relaciona | Corpora - Blogs e Foruns | Próprio | Não | Inglês, Holandês e Francês | Boiy, E., Moens, M.-F. | SVM, NB e ME | Boiy, E., Moens, M.-F. | <i>Katholieke Universiteit Leuven</i> | Bélgica | 2009 |
| 470 | <i>Sentiment</i> | Não relaciona | <i>Corpus Fórum</i> | Próprio | Não | Inglês | Bautin, M., Vijayarenu, L., Skiena, S. | SVM | Bautin, M., Vijayarenu, L., Skiena, S. | <i>Stony Brook University</i> | EUA | 2008 |
| 477 | <i>Sentiment</i> | Não relaciona | Dicionário | <i>HowNet</i> | Sim e Não | Chinês e Inglês | Zhendong Dong | SVM | Zhang, C., Zeng, D., Xu, Q., (...), Mao, W., Wang, F.-Y. | <i>Jilin University</i> | China | 2008 |
| 500 | <i>Sentiment</i> | Não relaciona | <i>Database</i> | Próprio | Não | Inglês | Matsumoto, S., Takamura, H., Okumura, M. | SVM | Matsumoto, S., Takamura, H., Okumura, M. | <i>Tokyo Institute of Technology</i> | Japão | 2005 |

Em relação Qp1, 99% dos artigos usaram alguma palavra afetiva. A partir disso, 76,4% usaram a palavra “*Sentiment*”, 3,8% usaram as palavras “*Sentiment e Emotion*”, 2,8% “*Sentiment e Mood*”, 5,7% “*Sentiment e Opinion*”, 1,9% “*Sentiment e Feeling*”, 0,9% “*Feeling*”, 1,9% “*Opinion*”, 2,8% “*Personality*”, 1,9% “*Personality e Sentiment*”, e 0,9% “*Sentiment, Mood e Feeling*”. Observa-se que na maioria dos artigos usou-se a palavra “*Sentiment*”. Palavras como “*Feeling*”, “*Opinion*”, “*Emotion*” e “*Personality*” raramente apareceram. Como Munezero *et al.* (2014) afirmou, na área da OM e SA, há falta de coerência no uso da palavra “*Sentiment*” e falta de diferenciação e terminologia entre as palavras afetivas. Essa afirmação foi confirmada, conforme o cenário firmado pela Qp1.

Com relação ao Qp2, 84,1% dos artigos que têm alguma palavra afetiva, não foram relacionadas com qualquer teoria afetiva ou psicológica. Apenas 15,9% desses artigos estão relacionados com alguma teoria afetiva ou psicológica; 8,5% dos artigos usaram a palavra “*Sentiment*” e citaram Feldman (2013), que desenvolveu técnicas e aplicações para análise de sentimento; 2,8% usaram as palavras “*Sentiment e Personality*” e citaram Goldberg (1990), que desenvolveu o Modelo de Personalidade Big Five; 2,8% usaram a palavra “*Personality*” e citaram Cattell (1943), que desenvolveu a *16 Personality Factors Questionnaire*; 0,9% usou a palavra “*Emotion*” e citou Scherer *et al.* (2001), que desenvolveu o Modelo de Processo de componentes; e, finalmente, 0,9% usou as palavras “*Sentiment, Feeling e Mood*” e citou Ortony *et al.* (1990), que desenvolveu o modelo OCC.

Na Qp3, descobriu-se que 27,3% dos artigos usaram léxicos, 25,5% utilizaram *corpus* / corpora, 39,6% empregaram *databases* próprias / específicas, 3,8% usaram dicionários, e 3,8% não utilizaram qualquer banco de dados. Apenas 5,6% dos 106 artigos investigados apresentaram simultaneamente alguma teoria psicológica junto com algum léxico, *corpus*, *databases*, ou dicionário. 21,7% não apresentaram nenhuma teoria psicológica explicitamente ligada à criação do seu próprio léxico. Esses léxicos foram criados pelos seus autores, utilizando técnicas, tais como: *crawler* de *Facebook* e *Twitter*, inferência das palavras de *chats*, fóruns e *blogs*; inferência a partir de resumos de livros e comentários em páginas web (*Amazon* e *Trip Advisor*, por exemplo); e telefonemas gravados (veja mais exemplos na Tabela 6).

Em relação Qp4, alguns exemplos de léxicos são o *SentiWordNet*, *SenticNet*, *WordNet*, *LIWC*, *MRC* e *ANEW*, citados na seção 3.2. Os corpora incluem *Muchocine* (CRUZ *et al.* 2008) e *Senti* (TAN *et al.* 2005). *FactNet* é citado como uma *database*. Há também citações de *large bases of movie reviews e mobile phones* (encontradas nas referências na Tabela 6).

Na Qp5, conclui-se que 73,6% dessas *databases* encontradas nos artigos não apresentaram quaisquer dados afetivos, enquanto que 26,4% apresentaram de fato algum dado afetivo.

Para a Qp6, verificou-se que 83,8% dos bancos de dados foram desenvolvidos para o idioma Inglês, 5,1% para o italiano, 3% para o espanhol e 7,1% para duas ou mais línguas. Deste montante, apenas 1% foi desenvolvido para o idioma Português. (Os autores que desenvolveram as bases de dados, em relação Qp7, foram listados na Tabela 6).

Com relação à Qp8, listaram-se os algoritmos mais usados para mineração afetiva relacionados com a pesquisa. Destes, 36,8% utilizam o *Support Vector Machines* (SVM), 21,7% utilizaram o *Naive Bayes* (NB), e 41,5% outras ferramentas / algoritmos, como, por exemplo, *Linear Regression* (LR), *Max Entropy* (ME), e *Bag of Words* (BOW). O SVM e a NB foram as estratégias computacionais mais utilizadas.

O *Support Vector Machines* (SVM), algoritmo mais utilizado nas análises, é mencionado na literatura como um classificador de alto padrão, é proveniente da teoria de aprendizagem por análise estatística, está inserido no grupo de classificadores não paramétricos, tendo a vantagem de redução de erros empíricos da classificação, além de separar as classes através de uma superfície de decisão que maximiza a margem de separação entre elas. Por isto, esse método de classificação está sendo utilizado em vários trabalhos científicos consequentemente apresentando resultados satisfatórios. As referências sobre todas as ferramentas estão referenciados nos artigos numerados na Tabela 6 e referenciados no Apêndice A.

Já o *Naive Bayes* (NB) é um algoritmo que foi implementado por ferramentas como *MALLET*²⁹, *Apache Mahout*³⁰ e *NLTK*³¹. É um algoritmo de classificação baseado no teorema de *Bayes* para uso de modelagem de previsão. É útil para gerar modelos de mineração rapidamente para descobrir as relações entre as colunas de entrada e as colunas previsíveis. Recomenda-se esse modelo para realizar explorações de dados iniciais, e em seguida, aplicar os resultados para criar modelos de mineração com outros algoritmos.

A Qp9, representada pela Figura 3, aborda que os autores que publicaram mais artigos neste campo são: Akaichi, A. (3 artigos); Li, J. (2 artigos); Mostafa, M.M. (2 artigos); Petz, G. (2 artigos); Shein, K.P.P. (2 artigos); Verma, R. (2 artigos). Os outros autores são apresentados na Tabela 6, publicaram apenas 1 artigo. Considerando como

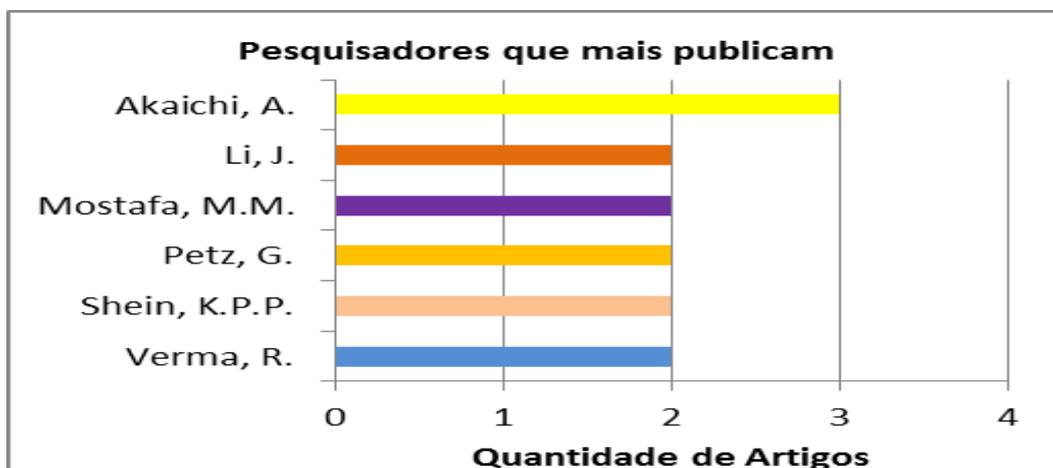
²⁹ <http://mallet.cs.umass.edu>

³⁰ <http://mahout.apache.org>

³¹ <http://nltk.org>

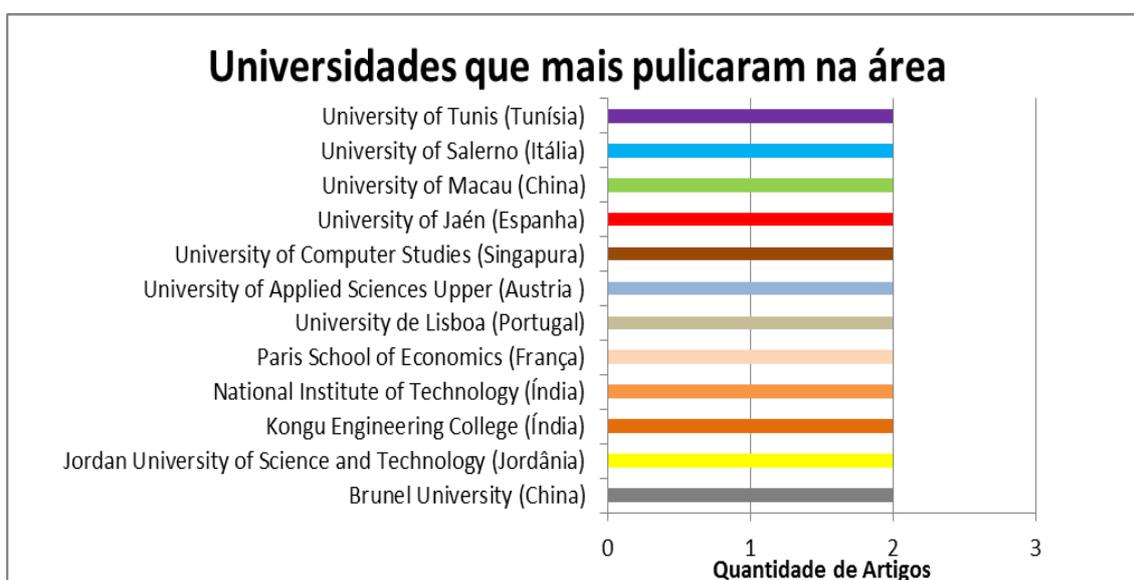
uma área de crescimento, ainda não existe um cientista para ser usado como uma referência.

Figura 3: Pesquisadores que publicaram mais artigos



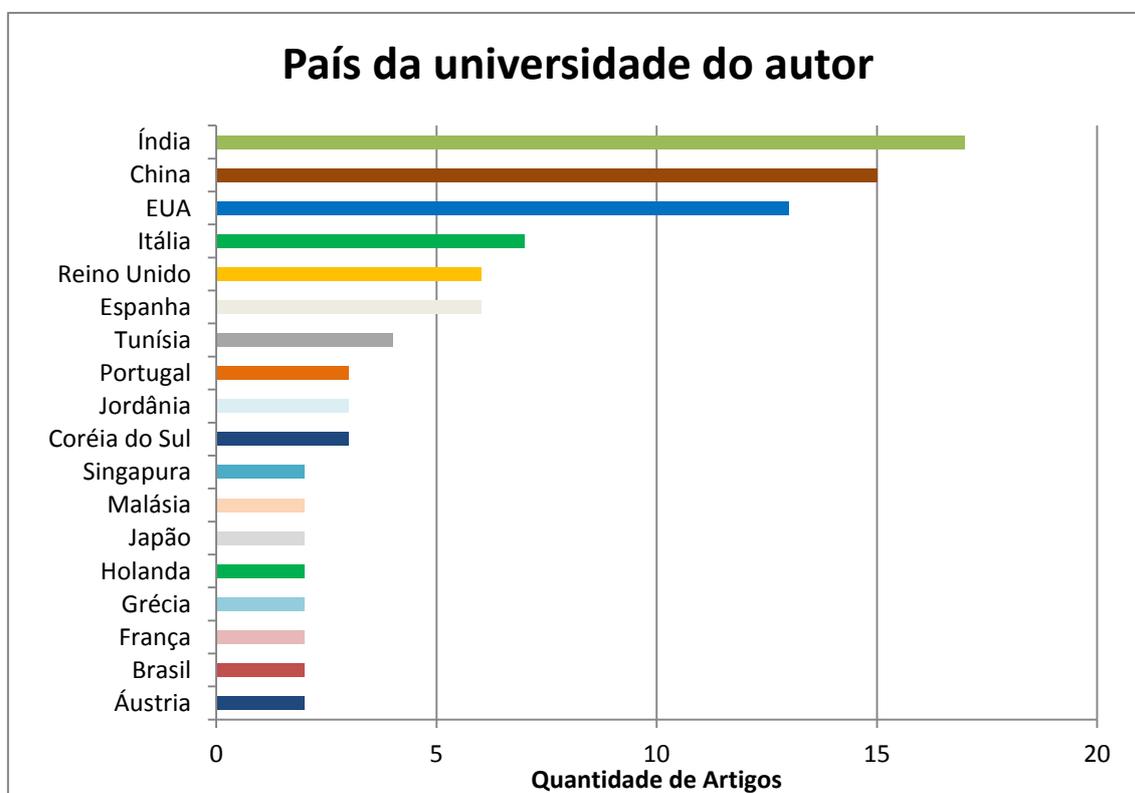
Com a resposta da Qp10, representada pela Figura 4, encontrou-se as universidades onde os autores têm vindo a trabalhar e publicaram mais documentos, considerando nosso escopo de pesquisa. São elas: University of Tunis / Tunísia (2 artigos); University of Salerno / Itália (2 artigos); University of Macau / China (2 artigos); University of Jaén / Espanha (2 artigos); University of Computer Studies / Singapura (2 artigos); University of Applied Sciences Upper / Áustria (2 artigos); University de Lisboa / Portugal (2 artigos); Paris School of Economics / França (2 artigos); National Institute of Technology / Índia (2 artigos); Kongu Engineering College / Índia (2 artigos); Jordan University of Science and Technology / Jordânia (2 artigos); Brunel University / China (2 artigos). As outras universidades, citadas na Tabela 6, publicaram apenas 1 artigo.

Figura 4: Universidades que mais publicaram



Baseado na Qp11, representada na Figura 5, descobriu-se os países dessas universidades. São eles: Índia (17 artigos); China (15 artigos); EUA (13 artigos); Itália (7 artigos); Reino Unido (6 artigos); Espanha (6 artigos); Tunísia (4 artigos); Portugal (3 artigos); Jordânia (3 artigos); Coreia do Sul (3 artigos); Singapura (2 artigos); Malásia (2 artigos); Japão (2 artigos); Holanda (2 artigos); Grécia (2 artigos); França (2 artigos); Brasil (2 artigos); Áustria (2 artigos). Os demais países, citados na Tabela 6, publicaram apenas 1 artigo cada. Um fato a ser observado é que como em um país existem várias universidades, nesta análise foram contabilizados os artigos que foram publicados por todas as universidades investigadas. Por exemplo, no Brasil existem 2 artigos publicados mas, por universidades distintas, por isto, o país aparece nesta estatística e não na Qp10 onde foram computados a quantidade total de artigos por universidade.

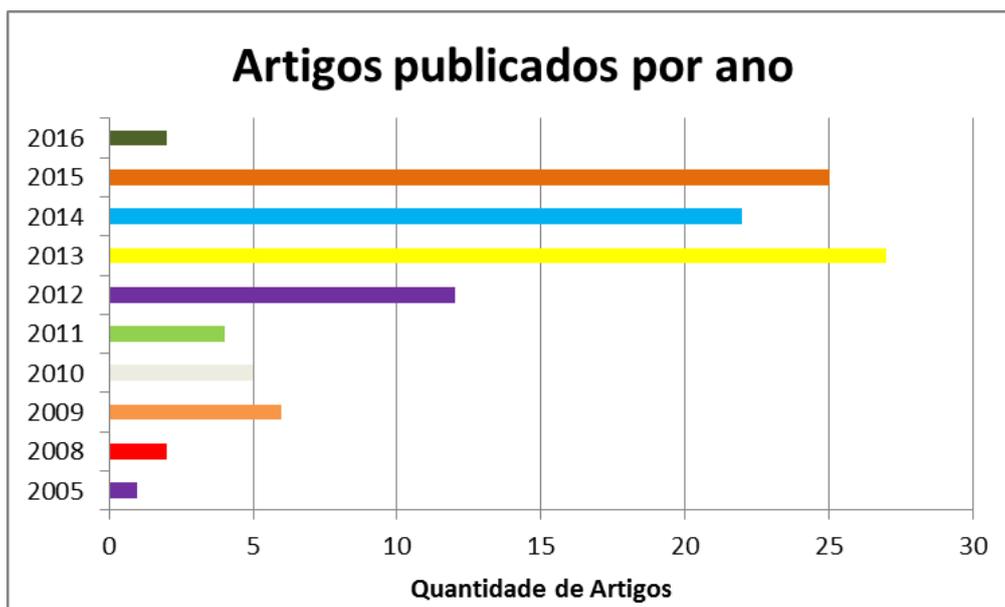
Figura 5: Países da universidade do autor



Finalmente na Qp12, representada pela Figura 6, apresenta-se a quantidade de publicações produzidas por ano. São elas: 2016 (2 artigos); 2015 (25 artigos); 2014 (22 artigos); 2013 (27 artigos); 2012 (12 artigos); 2011 (4 artigos); 2010 (5 artigos); 2009 (6 artigos); 2008 (2 artigos); 2005 (1 artigo).

Considerando-se que muitas universidades e países estão publicando nesse campo e que a publicação está aumentando a cada ano, esse mapeamento mostrou a importância crescente da área aos investimentos em investigações futuras.

Figura 6: Quantidade de artigos publicados por ano.



4.5 Considerações Finais

Esta seção apresentou uma pesquisa com base no método de mapeamento sistemático (PETERSEN *et al.* 2008). Desenvolveu-se a estratégia de pesquisa, considerando as palavras-chave relacionadas com “*affective mining*” e “*personality mining*”. Mineração afetiva inclui noções de *sentiment*, *feeling*, *emotion*, entre outros. A primeira publicação na área surgiu por volta de 2005, e em cada ano, a quantidade de publicações neste campo tem sido crescente. No entanto, encontraram-se algumas evidências de pesquisas feitas para o idioma Português. Apenas um artigo publicado em 2014 descreveu um banco de dados na língua portuguesa, no entanto, não foi baseado em dados afetivos. O desenvolvimento de qualquer banco de dados, léxico, dicionário, ou *corpus* em Português é altamente recomendado se pretende-se inferir dados afetivos ou personalidade a partir de texto, considerando os serviços que geram para o público brasileiro. Caso contrário, a mineração automática de perfis públicos em redes sociais, como *Facebook*, *Twitter* e *LinkedIn*, não será viável. Embora este mapeamento apresente um campo crescente, no Brasil, há uma lacuna importante nesta área de pesquisa. Considera-se que este fato significa uma oportunidade no sentido de inferir informações afetivas para o idioma Português. Concluiu-se também que muitos trabalhos têm vindo a utilizar termos afetivos, sem qualquer teoria subjacente. Talvez essa seja uma explicação sobre o fato de encontrarem-se muito poucos artigos que justamente se conectam a termos afetivos para qualquer teoria psicológica. Este cenário demonstra a confusão sobre este campo de pesquisa, onde o mal-entendido e falta de uma terminologia consistente de termos afetivos na literatura têm sido um problema para os cientistas durante a sua investigação sobre áreas como análises de sentimento e mineração de opinião.

5 CONSTRUÇÃO DO *PERSONALITATEM LEXICON*

A ideia da construção de um léxico afetivo se deu a partir da leitura dos estudos dos autores Pennebaker *et al.* (2001), criadores do *LIWC - Linguistic Inquiry e Word Count*. Originalmente o *LIWC* foi desenvolvido dentro de um contexto sobre a escrita emocional (PENNEBAKER e FRANCIS, 1996) (PENNEBAKER *et al.* 1997). Foi projetado para descobrir experiências negativas de vida dos pacientes e com isso poder prever subsequentes melhorias na saúde. Mais recentemente a utilização de *LIWC* foi expandida para o rastreamento e uso da linguagem natural em fontes de texto abrangendo literatura clássica, narrativas pessoais, conferências de imprensa, e transcrições de conversas diárias (PENNEBAKER e GRAYBEAL, 2001).

O *LIWC* é um programa que consiste em um módulo de processamento de texto principal com um auxílio de um dicionário externo. Esse módulo de processamento de texto executa funções que envolvem o controle e fluxo de processamento de texto e gestão do arquivo do dicionário auxiliar. O arquivo do dicionário é composto por mais de 2000 palavras ou temas de palavras que são atribuídas a uma subcategoria ou escala. Cada subcategoria é composta por grupos de palavras relacionadas que tocam em uma dimensão específica da linguagem, como por exemplo, emoções positivas ou negativas. As palavras de busca foram anteriormente classificadas por juízes independentes em mais de 70 dimensões linguísticas. Estas dimensões incluem categorias de linguagem padrão (por exemplo, artigos, preposições, pronomes - incluindo a primeira pessoa do singular, primeira pessoa do plural, etc.), os processos psicológicos (por exemplo, emotivas com categorias positivas e negativas, processos cognitivos, como o uso de palavras de causalidade, auto discrepâncias), palavras relacionadas com a relatividade (por exemplo, tempo, verbo dimensões de conteúdo tenso, movimento, espaço), e tradicional (por exemplo, o sexo, a morte, em casa, ocupação). As dimensões *LIWC* são organizadas hierarquicamente. Por exemplo, a palavra "chorou" cairia nas categorias "tristeza", "emoção negativa", "afeto global", "e" verbo-passado. O programa é suficientemente flexível para permitir a bem categorias definidas pelo usuário.

Analisando os léxicos pesquisados e apresentados no capítulo anterior, decidiu-se utilizar o *LIWC* que fora construído pelo Pennebaker *et al.*(2001) e traduzido para o português do Brasil pela equipe do NILC, conforme detalhado na seção 3.2.1, para que o mesmo possuía aspectos psicológicos e características de personalidade. Essa construção dará origem a um novo léxico, ou seja, um léxico afetivo, que veio a ser chamado de "*Personalitatem Lexicon*".

Inicialmente foi analisado o *LIWC* com mais acuidade e observou-se que as palavras que continham em seu banco de dados foram classificadas em diversas classes. Foram criadas *labels*³² com as definições para que pudesse ser aplicado a cada palavra

³² Criação de características específicas

correspondente, por exemplo, cada palavra tem uma ou mais *labels* correspondentes. Para detalhes, todas as palavras e classes foram disponibilizadas de forma online³³. Partiu-se então para as *labels* que correspondesse à afetividade. Consequentemente, se fosse possível identificar as palavras que tivessem a *labels* correspondente à afetividade, saber-se-ia que aquela palavra “x” seria uma palavra afetiva.

O *LIWC* original contém 127.159 (cento e vinte sete mil cento e cinquenta e nove) palavras. Esse quantitativo é o total de palavras com todas as *labels* atribuídas. O primeiro trabalho foi a separação das palavras que continham a *labels* “*affect*”, “*posemo*”, “*negemo*”, “*anx*”, “*anger*”, “*sad*” e “*feel*” *labels* que o psicólogo Pennebaker atribuiu para as palavras afetivas. Essas *labels* foram rotuladas com os códigos “125”, “126”, “127”, “128”, “129”, “130” “143” respectivamente. Então, as palavras do *LIWC* que contivessem esses códigos em suas definições, seriam as palavras que ele considerou afetivas, portanto, palavras que interessam a nossa pesquisa. Para detalhes sobre a escolha de palavras, classificação e definições consultar Pennebaker e King (1999).

Separadas essas palavras, obteve-se um retorno de 28.475 (vinte e oito mil quatrocentos e setenta e cinco) palavras. Esse total incluem palavras e suas derivações, por exemplo, a palavra abafar. Essa palavra tem várias derivações, por exemplo, abafara, abafaram, abafaras, e todas com alguma *label* acima especificada. Para reduzir e simplificar ainda mais o quantitativo de palavras, foram eliminadas as palavras derivadas, ficando somente as palavras “chave” ou palavra “principal”. Neste caso a palavra “chave ou principal” é abafar. Portanto, a caracterização do fator(es) do *Big Five* e da(s) faceta(s) do IPIP-NEO que fora atribuída para abafar, consequentemente será atribuída às suas derivações. Com essa nova filtragem, retornaram-se 1.866 (um mil oitocentos e sessenta e seis) palavras. Finalizado esse processo de separação, partiu-se para o processo de classificação.

Como o objetivo é classificar as palavras separadas com fatores do *Big Five* mais precisamente dentro de alguma faceta do IPIP-NEO, o questionário de personalidade IPIP-NEO 120 foi usado para comparar as palavras com as suas características (definidas pelo *Story-based*).

Contudo, inicialmente para classificação, precisaria-se entender o significado de cada palavra. Quando se fala em significado, trata-se do entendimento/compreensão, ou seja, saber o que exatamente aquela palavra quer dizer. Para isso, foram utilizados 3 (três) métodos. O primeiro método foi consultar um dicionário Aurélio (dicionário de mão, manual). No segundo método, foi consultado outro dicionário em versão online³⁴. Caso houvesse divergência ou dúvida entre os significados entre os dicionários

³³ http://www.icmc.usp.br/~sandra/LIWC/LIWC2007_Portugues_win.dic

³⁴ <http://www.dicionarioaurelio.com/>

consultados, um terceiro método foi imposto, ou seja, um terceiro dicionário³⁵ seria consultado. Metodologia consideravelmente parecida com o *LIWC* original que usaram opiniões de avaliadores.

Sabendo o significado das palavras, procurou-se achar relação com os fatores do *Big Five* e com as facetas do IPIP-NEO. Cada fator do *Big Five* (Abertura; Neuroticismo; Extroversão; Socialização e Realização) possuem 6 (seis) facetas conforme explicado e detalhado na seção 2.4 As Tabelas 7 e 8, apresentam a forma de atribuição e distribuição dos códigos, tanto nos fatores do *Big Five* quanto nas facetas do IPIP-NEO, definidos pelo próprio autor. Cada faceta tem características definidas como “alta” e “baixa” ou só “alta” ou só “baixa”, definidas pelo *Story-based*. Para as palavras que obtiveram característica “alta”, foi atribuído o símbolo “+” e as que obtiveram característica “baixa” foi atribuído o símbolo “-”. Essas simbologias estarão ao lado dos códigos dados a cada faceta.

Tabela 7: Códigos atribuídos a cada fator do *Big Five*

| Código | 1000 | 2000 | 3000 | 4000 | 5000 |
|-----------------------|----------|--------------|-------------|--------------|------------|
| Fator Big Five | Abertura | Neuroticismo | Extroversão | Socialização | Realização |

Tabela 8: Códigos atribuídos a cada faceta do IPIP-NEO

| Códigos | Facetas IPIP-NEO | Códigos | Facetas IPIP-NEO | Códigos | Facetas IPIP-NEO | Códigos | Facetas IPIP-NEO | Códigos | Facetas IPIP-NEO |
|---------|-----------------------|---------|--------------------|---------|-----------------------|---------|------------------|---------|------------------|
| 1001+ | Imaginação | 2001+ | Ansiedade | 3001+ | Nível de Atividade | 4001+ | Empenho | 5001+ | Confiança |
| 1002+ | Interesses Artísticos | 2002+ | Falta de Moderação | 3001- | | 4001- | | 5001- | |
| 1002- | | 2002- | | 3002+ | Assertividade | 4002+ | Ordem | 5002- | Moralidade |
| 1003+ | Intelecto | 2003+ | Depressão | 3002- | | | | 4002- | |
| 1003- | | | | 2003- | | 3003+ | Gregarismo | 4003+ | Auto Disciplina |
| 1004+ | Senso Aventureiro | 2004+ | Vulnerabilidade | 3003- | | 4003- | | | 5004- |
| 1004- | | | | 2004- | | 3004+ | Bom Humor | 4004- | Prudência |
| 1005+ | Emotividade | 2005+ | Auto Percepção | 3005+ | Amigabilidade | 4005+ | Auto Eficiência | 5006+ | Compaixão |
| 1005- | | 2005- | | | | 3005- | | 4006+ | |
| 1006+ | Liberalismo | 2006+ | Raiva | 3006+ | Procura por excitação | 4006- | | | |
| | | 2006- | | | | | | | |

Os pesos atribuídos às palavras foram dados da seguinte maneira. Cada faceta, tem as suas características definidas pelo *Story-based*, que por sua vez, propõe 5 respostas, que são: Me identifico completamente; Me identifico parcialmente; Estou neutro; Não me identifico parcialmente; Não me identifico de forma nenhuma. Partindo desse princípio, foram atribuídos pesos baseados nessa lógica. Sendo assim, se a palavra tiver correlação direta com a faceta, ou seja, se o significado da palavra tiver proximidade elevada com a definição atribuída àquela faceta, foi atribuído peso 3 (muito significativo). Se acaso a palavra fosse correlacionada com a faceta através de um sentido figurado, foi atribuído peso 2 (significativo) e caso fosse correlacionado com um sinônimo, ou algo mais distante, mas que tivesse ligação, foi atribuído peso 1 (pouco significado). Caso não fosse encontrado nenhum tipo de correlação com os

³⁵ <http://michaelis.uol.com.br/>

citados acima, foi atribuído peso 0 (nada significativo). Não foi usado um peso “neutro” porque não se tem como mensurar características como tal. Palavras afetivas existem, ou não existem.

5.1 Algoritmo de Correlação Facetas *LIWC*

Para detalhar o processo de readaptação do *LIWC*, foi construído um algoritmo, apresentado abaixo, com o passo-a-passo para se chegar a uma lista de palavras com suas facetas e respectivos pesos.

sub-rotina correlação facetas *LIWC*

parâmetros de entrada: léxico do *LIWC*, dicionários, story-based

parâmetros de saída: listagem de palavras ponderadas para cada faceta

início-procedimento

declare Defs como string

declare listagem_de_palavras como matriz 2x2 (palavras x facetas)

para cada palavra do *LIWC* faça

início-para

se a palavra pertencer a uma das dimensões 125, 126, 127, 128, 129, 130 ou 143, então faça

início-se

recupere o lema da palavra (manualmente)

Defs = para o lema recuperado, busque as definições da palavra nos dicionários Aurélio (impresso) e Michaelis

se Defs forem julgadas (de forma subjetiva) imprecisas, então

Defs = Defs + definição no Aurélio (online) do lema recuperado

para cada faceta faça

início-para

listagem_de_palavras[lema, faceta] =

calcular_sobreposição_de_significado(Defs, faceta no story-based)

fim-para

fim-se

fim-para

retornar listagem_de_palavras

fim-procedimento

sub-rotina calcular sobreposição de significado

parâmetros de entrada: definições de dicionários, faceta no story-based

parâmetros de saída: correlação (peso) da faceta em foco com as definições

início-procedimento

se as definições forem exatamente iguais com as explicações da faceta no story-based, então retorne 3 (correlação muito significativa)

senão se se for feito com palavras sinônimas, então retorne 2 (correlação significativa)

senão se for feito em sentido figurado, então retorne 1 (correlação pouco significativa)

senão se não achar correlação alguma, então retorne 0 (sem correlação)

fim-procedimento

5.2 Exemplos

5.2.1 Exemplo Listagem_de_palavras como matriz

Tabela 9: Exemplo de palavras com facetas e pesos

| Facetas | | |
|------------------|-----------|-----------|
| Palavra | Bom Humor | Compaixão |
| Amor (125 e 126) | 1 | 2 |

| Facetas | | |
|-------------------------|-----------|-----------------|
| Palavra | Ansiedade | Vulnerabilidade |
| Terror (125, 127 e 128) | 3 | 3 |

| Facetas | | |
|---------------------|----------------------|-------------|
| Palavras | Senso Aventureiro | Liberalismo |
| Divirja (125 e 127) | 1 | 1 |

5.2.2 Exemplo de Palavras (calculado_sobreposição_de_significado)

1 - Amor (Dimensões 125 e 126)

Definição Aurélio: **1** Sentimento que induz a aproximar, a proteger ou a conservar a pessoa pela qual se sente afeição ou atração; grande afeição ou afinidade forte por outra pessoa. **2** Sentimento intenso de atração entre duas pessoas. **3** Ligação afetiva com outrem, incluindo geralmente também uma ligação de cariz sexual. **4** Ser que é amado. **5** Disposição dos afetos para querer ou fazer o bem a algo ou alguém.

Definição Michaelis: **1** Sentimento que impele as pessoas para o que se lhes afigura belo, digno ou grandioso. **2** Grande afeição de uma a outra pessoa de sexo contrário. **3** Afeição, grande amizade, ligação espiritual. **4** Objeto dessa afeição. **5** Benevolência, carinho, simpatia.

Sabendo as definições e fazendo as análises com o *Story-based* chegou-se a conclusão que a palavra Amor tem correlação com as facetas, Bom Humor “Alta (+keyed)” “Irradio alegria. Divirto-me bastante. Amo a vida. Olho pelo lado bom da vida” recebendo peso 1 pois a correlação é feita somente em sentido figurado “Amo a vida”. A palavra Amo, está em sentido figurado porque neste contexto significa o mais importante, a razão de ser e a faceta Compaixão “Alta (+keyed)” “Tenho compaixão pelos desabrigados. Sinto compaixão por aqueles menos abastados do que eu”. recebendo peso 2 pois a correlação é feita com os sinônimos fraternidade, afeto, carinho. Para as demais facetas, a palavra não tem correlação alguma, recebendo peso 0.

2 - Terror (Dimensões 125, 127 e 128)

Definição Aurélio: 1 Pavor, pânico, grande medo. 2 Qualidade de terrível. 3 Regime político caracterizado por prisões e morticínios.

Definição Michaelis: 1 Qualidade de terrível. 2 Grave perturbação trazida por perigo imediato, real ou não; medo, pavor. 3 Ameaça que causa grande pavor. 4 Objeto de espanto. 5 Perigo, dificuldade extrema.

Sabendo as definições e fazendo as análises com o *Story-based* chegou-se a conclusão que a palavra Terror tem correlação com as facetas, Ansiedade “Alta (+keyed)” “Me preocupo com as coisas. Temo o pior. Tenho medo de muitas coisas. Me estresso facilmente”. Exemplo: Pânico, medo de alguma coisa, preocupação excessiva, tem correlação muito significativa com ansiedade e a faceta Vulnerabilidade “Alta (+keyed)” “Entro em pânico com facilidade. Muitas vezes me sinto sobrecarregado. Sinto que sou incapaz de lidar com as situações”. Exemplo: Pessoa que se perturba com algum perigo imediato, real ou não, também é característica muito significativa com vulnerabilidade. Neste caso, ambas as facetas receberam peso 3, muito significativo.

3 - Divirja (Dimensões 125 e 127)

Para esta palavra foi necessário o auxílio do 3º dicionário, pois as definições dos dois primeiros não foram suficientes para as análises.

Definição Aurélio: 1 Aparte; difira; separe. 2 Afastar(-se), de maneira progressiva, uma coisa de outra coisa. Afastar-se cada vez mais do ponto de partida - separar-se ou desviar-se.

Definição Michaelis: 1 Mover-se ou estender-se em direções diferentes a partir de um ponto comum; afastar-se progressivamente um do outro a partir de um ponto de partida comum. 2 Não se combinar; discordar: Divirjo dessa opinião. Ele e eu divergimos no temperamento. Muitos divergiram, e isso prejudicou a louvável tentativa.

3 Diferir na forma, caráter ou opinião: Quanto ao motivo da contenda, as opiniões divergem. Antônimo (acepção 1): convergir.

Definição Aurélio (Online): **1** Ter ou sofrer divergência. **2** Desviar-se; afastar-se cada vez mais. **3** Ser de opinião diferente; não concordar.

Sabendo as definições e fazendo as análises com o *Story-based* chegou-se a conclusão que a palavra Divirja tem correlação com as facetas, Senso Aventureiro “Baixo (-keyed)” “Prefiro lidar com coisas que já conheço. Não gosto de mudanças. Estou preso(a) a tradições”. Exemplo: Divirjo dessa opinião, a palavra divirjo está em sentido figurado (Discordar) e a faceta Liberalismo “Alta (+keyed)” “Acredito que valores mudam com o tempo. Acredito que não existe verdade absoluta”. Exemplo: Antonio acredita que as pessoas mudam com o tempo, já maria tem diverge dessa opinião, a palavra diverge, está em sentido figurado (opiniões diferentes). Neste caso, ambas as facetas receberam peso 1, pois a correlação é feita somente em sentido figurado.

Um detalhe a se chamar atenção é que da base inteira de palavras, existem noventa e duas que foram marcadas como afetivas pelo *LIWC*, mas que não possui correlação alguma com nenhuma faceta. Também não foi feito nenhum tipo de cálculo matemático para chegar aos pesos atribuídos, apenas foi analisado e atribuído os passos acima conforme descrito.

5.3 Considerações Finais

Essa seção apresentou a metodologia usada na construção do *Personalitatem Lexicon*. Foi adotado o léxico *LIWC* por observar que as palavras utilizadas foram classificadas em diversas classes. Outro motivo foi que o léxico foi traduzido pela equipe do NILC, conforme detalhado na seção 3.2.1.

Foram analisadas todas as palavras do *LIWC* e separadas aquelas que continham características afetivas. O psicólogo Pennebaker atribuiu “*affect*”, “*posemo*”, “*negemo*”, “*anx*”, “*anger*”, “*sad*” e “*feel*” como tal característica.

Para as classificações foi necessário o conhecimento do significado das palavras para poder encontrar as relações com os fatores do *Big Five* e com as facetas do IPIP-NEO. As palavras foram organizadas a partir de códigos que foram atribuídos para cada fator do *Big Five* e para cada faceta do IPIP-NEO.

Após as palavras serem separadas e classificadas foi definido um algoritmo de correlação facetas_*LIWC* para que se fosse possível chegar a uma lista com essas palavras com suas facetas e seus respectivos pesos.

6 EXPERIMENTOS E ANÁLISES DE RESULTADOS

Neste capítulo, serão apresentados os experimentos realizados e os resultados obtidos nas avaliações. Os experimentos foram pensados a partir da hipótese da dissertação, onde acredita-se na possível inferência de personalidade a partir de textos de rede social.

Inicialmente foram necessárias pessoas que tinham cadastro e faziam uso contínuo da rede social *Facebook*. A partir daí, essas pessoas foram convidadas a se cadastrar no *Personalitatem Inventory*³⁶ e responderem os questionários propostos. De posse das respostas dos questionários, foram feitas correlações com as postagens divulgadas no *Facebook* e conseqüentemente chegando-se aos resultados.

O objetivo dos experimentos é validar a base *Personalitatem Lexicon* como um recurso lexical, quanto ao seu conteúdo. Sendo assim, o método definitivo da pesquisa foi correlacionar as respostas dos questionários com as postagens feitas no *Facebook*. Nas próximas seções, serão explicados com mais acuidade, o passo a passo de cada etapa e os métodos avaliativos para se chegar aos resultados e as conclusões.

6.1 Participantes

Os participantes do experimento são estudantes (técnicos, graduação e pós-graduação) e professores de diferentes instituições de ensino, a exemplo da UFS³⁷, IFS³⁸ e UNIT³⁹. Eles foram convidados a participar do experimento de forma voluntária a fim de contribuir com a dissertação de mestrado. Foram escolhidas essas instituições de ensino porque a maioria dos alunos e professores utilizam redes sociais para comunicação.

6.2 *Personalitatem Inventory*

O *Personalitatem Inventory* disponibiliza, de forma online, os questionários, que os participantes precisam responder para poder participar da pesquisa. Na sua tela principal, assim que o participante acessasse, pela primeira vez, ele era conduzido a abrir uma conta, ou seja, realizar um cadastro. Neste cadastro é solicitado o seu nome completo, um nome de usuário (para ser usado em próximos acessos), o e-mail, uma senha pessoal, a data de nascimento e o sexo. A tela de cadastro/acesso é ilustrada pela Figura 7. Após a inserção desses dados, o participante deveria ler os termos de condições de uso e aceitar caso estivesse de acordo. Este termo está presente no

³⁶ <http://personalitatem.ufs.br/inventory>

³⁷ Universidade Federal de Sergipe - UFS

³⁸ Instituto Federal de Sergipe - IFS

³⁹ Universidade Tiradentes - UNIT

Apêndice B. Após o cadastro e acesso ao site, é solicitado o nome de usuário de sua rede social *Facebook*, procedimento essencial para as análises e comparações dos resultados. Caso o participante não lembrasse ou não quisesse cadastrar seu usuário naquele momento, ele poderia cadastrar posteriormente.

Figura 7: Tela de cadastro/acesso

Na página inicial do portal, estão disponíveis os questionários propostos, desenvolvidos e cedidos gentilmente pelo psicólogo Johnson (2000b), na língua inglesa. A tradução para o português do Brasil foi feita por Nunes (2013). O site oferece quatro tipos de questionários, sendo eles: IPIP-NEO 120 (JOHNSON, 2014); TIPI (GOSLING *et al.* 2003); IPIP-NEO 300 (JOHNSON, 2014) Os questionários contêm 120, 10 e 300 questões respectivamente. Para a pesquisa e análise da presente dissertação, foram consideradas as respostas dos questionários IPIP-NEO 120 e TIPI. A Tabela 10 indica o quantitativo de homens e mulheres que participaram da pesquisa respondendo tanto o IPIP-NEO 120 quanto ao TIPI.

Tabela 10: Quantidade de participantes (homens e mulheres)

| Estatísticas dos selecionados | | | |
|--------------------------------------|---------------|-----------------|--------------|
| Quantidade | Homens | Mulheres | Total |
| | 20 | 23 | 43 |
| Média Idade | 25 | 24 | |

Uma observação importante a fazer é que o site disponibiliza as duas versões dos questionários, a inglesa e a portuguesa. Por padrão, assim que o participante se cadastra, ele será direcionado para a versão portuguesa, mas caso o usuário deseje também responder a versão inglesa, basta ele trocar o idioma dos questionários para o inglês (Estados Unidos). Para isto, basta o usuário acessar a aba preferências e depois região e idioma.

Foram escolhidos estes questionários por utilizarem o modelo do *Big Five*, apresentado na seção 2.4, por estarem sendo amplamente utilizados por pesquisadores em inferência de personalidade por meio de texto e por identificar e descrever de forma simples, clara e objetiva os traços de personalidade. A aplicação dos questionários tem o intuito de inferir a personalidade do participante para posteriormente comparar com as postagens publicadas em sua rede social *Facebook*.

O questionário IPIP-NEO 120 possui uma quantidade aceitável de questões para preenchimento de forma *online*. O fator de ter sido considerado o questionário IPIP-NEO 120 e não o IPIP-NEO 300 foi a diferença do quantitativo de questões. Um questionário com muitas questões, às vezes se torna irrealizável por requerer muito tempo do participante, o que pode ocasionar uma não conclusão do questionário, deixando de obter o resultado de sua forma completa.

Contudo, como estão disponíveis para respostas, os participantes podiam responder os três questionários se assim desejassem. Ao clicar para iniciar qualquer questionário, são apresentadas as instruções, Apêndice C, que os participantes necessitam saber para um bom entendimento. A importância da honestidade nas respostas resultará diretamente no seu resultado. Foram mantidos o número de questões, as características e o formato de respostas desenvolvidas inicialmente pelo psicólogo Johnson (2000b).

Os participantes responderam cada questão selecionando uma das cinco alternativas que foram lhes fornecidas. Cada alternativa corresponde a um nível de concordância relacionada à pergunta. Para isso foi utilizada a escala tipo Likert (LIKERT, 1932), conforme demonstrado na primeira pergunta do IPIP-NEO 120 apresentado pela Figura 8. Todas as questões referentes ao IPIP-NEO 120 são apresentadas no Apêndice D, incluindo relação das questões com os fatores do *Big Five* com as Facetas, ver seção 2.4, desenvolvidas também pelo psicólogo Johnson (2000b). As respostas referentes a essas perguntas são: Discordo plenamente; Discordo um pouco; Nem discordo nem concordo; Concordo um pouco e Concordo plenamente.

Ao final do preenchimento das 120 questões são analisadas as respostas dadas pelos participantes para iniciar o processo de contabilização do resultado. Para o cálculo do resultado, são atribuídos valores que correspondem de 1 a 99 para cada um dos fatores do *Big Five*. O valor 1, por exemplo, corresponde ao valor mais baixo dado aquele determinado fator, conseqüentemente, o valor 99 é atribuído ao valor mais alto. Na Figura 9, será apresentado um resultado de um dos participantes. Observa-se que o participante tem valor superior a 90 para extroversão, que segundo o psicólogo Johnson, o participante se mostra um indivíduo têm tendência a procurar estimulações, companhia dos outros e se envolve positivamente com suas atribuições.

Figura 8: Questão 1 do questionário IPIP-NEO 120 (JOHNSON, 2000b)

F

IPIP-NEO 120

Item 1 de 120. 0% concluído

Preocupo-me com as coisas.

Discordo plenamente
Discordo um pouco
Nem discordo nem concordo
Concordo um pouco
Concordo plenamente

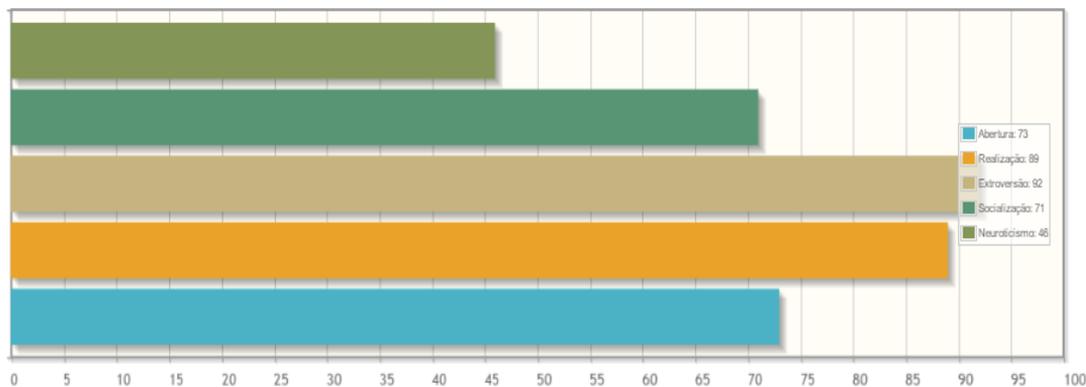
Figura 9: Resumo do resultado do questionário IPIP-NEO 120 de um participante.

Veja um resumo do resultado do seu inventário:

120 questões | Português (Brasil) | Tipo de inventário: International Personality Item Pool (Associação Internacional de Itens de Personalidade)

Estadísticas

Iniciado em: 01/11/2015 às 14:21
Concluído em: 01/11/2015 às 14:33



| Item ↕ | Escore ↕ |
|--------------|----------|
| Abertura | 73 |
| Realização | 89 |
| Extroversão | 92 |
| Socialização | 71 |
| Neuroticismo | 46 |

O outro questionário que foi considerado para as análises foi o TIPI (Apêndice E). Este questionário segue os mesmos padrões de interação com o participante que o IPIP-NEO 120, tendo as mesmas instruções, expostas no Apêndice C. O diferencial para o IPIP-NEO 120 é o quantitativo de questões. O TIPI aborda 10 questões, originalmente desenvolvidas na língua inglesa e traduzidas para o português por Nunes (2013), conforme apresentado na Figura 10.

Figura 10: Questão 1 do questionário TIPI (JOHNSON, 2000b)

TIPI

Item 1 de 10. 0% concluído

Eu me vejo como um homem...

Extrovertido, entusiasta.

Discordo plenamente Discordo bastante Discordo um pouco Nem discordo nem concordo Concordo um pouco

Concordo bastante Concordo plenamente

Observa-se que este questionário apresenta diferenças relacionadas nas alternativas para respostas. Enquanto o IPIP-NEO 120 disponibilizam 5 (cinco) alternativas para o participante, o TIPI disponibilizaram 7 (sete) alternativas. São elas: Discordo plenamente; Discordo bastante; Discordo um pouco; Nem discordo nem concordo; Concordo um pouco; Concordo bastante e Concordo plenamente. Para as análises dos resultados, também foi utilizado à escala Likert (LIKERT, 1932). O cálculo dos resultados também foi mantido, seguindo a mesma metodologia utilizada para o questionário IPIP-NEO 120. A Figura 11 apresenta o resultado do mesmo participante que foi apresentado acima.

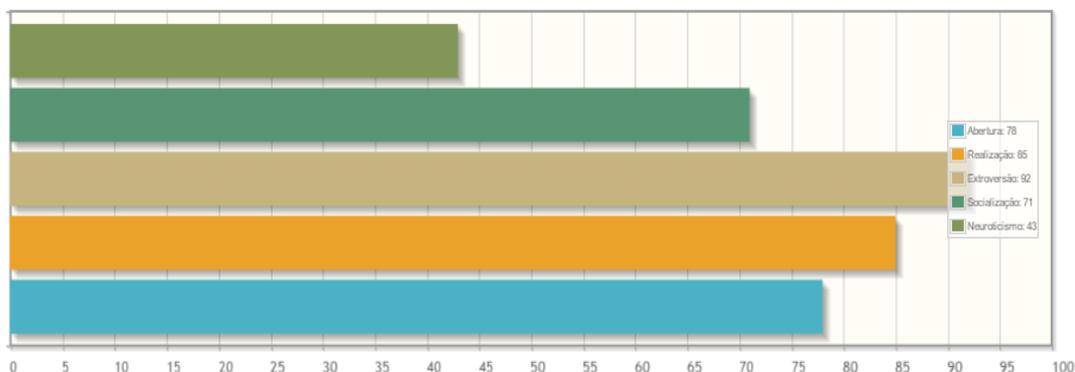
Figura 11: Resumo do resultado do questionário TIPI de um participante.

Veja um resumo do resultado do seu inventário:

10 questões | Português (Brasil) | Tipo de inventário: Ten Item Personality Inventory (Inventário de Personalidade de Dez Itens)

Estatísticas

Iniciado em: 01/11/2015 às 15:15
Concluído em: 10/11/2015 às 17:14



| Item ↕ | Escore ↕ |
|--------------|----------|
| Abertura | 78 |
| Realização | 85 |
| Extroversão | 92 |
| Socialização | 71 |
| Neuroticismo | 43 |

Observa-se que por ser o mesmo participante, os resultados apresentaram uma proximidade muito alta, no caso até do traço extroversão e socialização ter dado o mesmo resultado (92 e 71, respectivamente) o que prova que este participante respondeu os questionários utilizando uma mesma lógica.

6.3 Análises de Resultados

Para a comprovação da hipótese da dissertação, foram implementados e adaptados três questionários de personalidade para a língua portuguesa do Brasil. Cada questionário com a sua característica a fim de gerar os resultados esperados. Além dos questionários foram coletados e analisados *posts* do *Facebook* dos usuários para inferir personalidade através do *Personalitatem Lexicon*, conforme explicado na seção 5. As coletas dos *posts* foram feitas de forma manual e somente dos usuários que disponibilizaram seus *logins* do *Facebook* na página de cadastro do *Personalitatem Inventory*.

Para satisfazer a pesquisa supracitada, foram levados em consideração o IPIP-NEO 120 e o TIPI por razões do tamanho do questionário e o tempo utilizado para resposta. O método de avaliação dos resultados da inferência de personalidade irá analisar a correlação de Pearson (BUSSAB e MORETTIN, 1986) para os fatores do *Big Five* e facetas do IPIP-NEO, determinando quais os índices de correlações entre eles.

O coeficiente de correlação de Pearson é uma das medidas mais utilizadas para calcular o grau de associação linear entre duas variáveis quantitativas. O coeficiente de correlação de Pearson (r), apresentado na Equação 1, calcula a semelhança existente entre dois vetores de valores. Esse coeficiente varia sempre de -1 a 1. Uma relação diretamente proporcional exata ocorre quando o coeficiente é igual a 1 e inversa exata ocorre quando o coeficiente é igual a -1. Quando o coeficiente é igual a 0, significa que não existe relação linear entre os valores (MUKAKA, 2012).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Nesta equação, r representa o coeficiente de correlação, n o número de elementos no vetor, x o vetor de valores dos dados reais, y o vetor dos valores dos dados obtidos, i representa o i -ésimo elemento do vetor, \bar{x} a média dos valores do vetor x e \bar{y} a média dos vetores de y .

A interpretação dos valores de referência do coeficiente de correlação de Pearson pode ser interpretado e avaliado qualitativamente (APPOLINÁRIO, 2006) como apresentado na Tabela 11.

Tabela 11 - Valores de referência para a interpretação do coeficiente de correlação Pearson.

| Valores de Correlação | Interpretação |
|-----------------------|---------------|
| 0,00 | Nula |
| 0,01 até 0,10 | Muito Fraca |
| 0,11 até 0,30 | Fraca |
| 0,31 até 0,59 | Moderada |
| 0,60 até 0,80 | Forte |
| 0,81 até 0,99 | Muito Forte |
| 1,00 | Absoluta |

O experimento se baseou na avaliação do julgamento das palavras, do que era sentimento ou não, realizada somente por uma pessoa (pesquisador), de lexemas de conotação afetiva a partir da leitura de mensagens disponibilizadas em *posts* do *Facebook*, extraídos com o consentimento dos participantes. Faz necessário informar que a análise se deu pela observação das palavras no *Personalitatem Lexicon*. Outro fator importante foi a dificuldade de identificar o sentido das palavras, notou-se ambiguidade na análise dentre algumas dessas, o que dificultou na afirmação se havia ou não conotação afetiva.

Assim, todas as palavras existentes nos *posts*, receberam uma marcação de afetiva ou não afetiva. Para as palavras marcadas como afetivas (w), foi verificado a qual grande traço de personalidade e faceta cada uma pertencia, ou estaria associada. A ela foi aplicada uma escala de significação da palavra (p) conforme a relação da palavra com o traço de personalidade (não significativa, pouco significativa, significativa, muito significativa). A partir disso, foi aplicado um cálculo que indica o possível traço de personalidade e faceta (TPFs) demonstrado na situação avaliada seria um somatório das palavras afetivas multiplicado pelo peso atribuído (não significativa, pouco significativa, significativa, muito significativa) dividido pela quantidade de palavras afetivas (qw) encontradas no texto. O cálculo é demonstrado pela Equação 2.

$$TPF_s = \frac{\sum w * p}{qw} \quad (2)$$

Para calcular e analisar os *posts*, foi desenvolvida uma ferramenta chamada de *Lexicon Miner*. Essa ferramenta faz a leitura dos *posts* e determina o resultado a partir da equação atribuída. A contribuição da dissertação está em correlacionar os resultados dos questionários com os *posts* dos participantes. O resultado das correlações, apresentados na Tabela 12, mostrará o grau de intensidade com cada fator do Big Five e faceta do IPIP-NEO.

Tabela 12: Correlações entre IPIP-NEO 120, TIPI e *posts* do *Facebook*.

| IPIP-NEO 120 e TIPI | | | | |
|------------------------------|-----------------|-----------------|----------|----------|
| Índices de Correlação | | | | |
| A | R | E | S | N |
| 0,495471 | 0,730599 | 0,794991 | 0,450947 | 0,540561 |

| IPIP-NEO 120 e <i>posts</i> do <i>Facebook</i> | | | | |
|---|----------|-----------------|----------|-----------------|
| Índices de Correlação | | | | |
| A | R | E | S | N |
| 0,111841 | 0,230535 | 0,481998 | 0,113808 | 0,610616 |

| IPIP-NEO 120 e <i>posts</i> do <i>Facebook</i> | | | | |
|---|----------|-----------------|----------|-----------------|
| Índices de Correlação | | | | |
| A | R | E | S | N |
| 0,115074 | 0,128952 | 0,493753 | 0,134556 | 0,605224 |

Legenda: **A** - Abertura; **R** - Realização; **E** - Extroversão; **S** - Socialização; **N** - Neuroticismo.

Os resultados obtidos mostram que a correlação entre IPIP-NEO 120 e TIPI indicam que os fatores Realização (0,730599) e Extroversão (0,794991) obtiveram correlação forte. Já os fatores Neuroticismo (0,540561), Abertura (0,495471) e Socialização (0,450947) obtiveram correlação moderada.

Já para as correlações entre o questionário IPIP-NEO 120 e *posts* do *Facebook* e o questionário TIPI e *posts* do *Facebook* os fatores Extroversão (0,481998 e 0,493753) obtiveram correlação moderada, porém Neuroticismo (0,610616 e 0,605224) apresentou correlação forte. Porém os demais fatores, Abertura, Realização e Socialização, obtiveram correlação fraca. O fato dessas correlações, estarem ou serem fracas, quer dizer que no momento das extrações dos *posts*, foram encontradas poucas características que evidenciassem esses fatores de personalidade.

Devem-se levar em consideração que o IPIP-NEO 120 e o TIPI apresentaram melhores resultados por serem questionários completos em sua inferência, onde é possível a abordagem e identificação de todos os fatores. Já para os resultados das análises de *posts* do *Facebook* considera-se satisfatório por tratar de *posts* que foram publicados entre amigos ou publicamente. Evidenciando sempre que esses resultados dos *posts* são análises preliminares que retratam o sentimento do sujeito no momento.

Tabela 13: Correlação entre Facetas do IPIP-NEO 120 e *posts* do Facebook

| Índices de correlação | | | | |
|-----------------------|-----------------|------------------|------------------|------------------|
| A1 | R1 | E1 | S1 | N1 |
| -0,097067 | 0,035665 | -0,259340 | -0,003453 | -0,184351 |
| A2 | R2 | E2 | S2 | N2 |
| -0,024348 | 0,112410 | -0,663698 | -0,005866 | 0,631622 |
| A3 | R3 | E3 | S3 | N3 |
| -0,112547 | -0,166373 | -0,127293 | -0,006289 | -0,164185 |
| A4 | R4 | E4 | S4 | N4 |
| 0,255202 | 0,061453 | 0,019087 | -0,007939 | -0,372612 |
| A5 | R5 | E5 | S5 | N5 |
| -0,040894 | 0,052399 | 0,165767 | 0,357868 | 0,176668 |
| A6 | R6 | E6 | S6 | N6 |
| 0,130461 | 0,346550 | -0,055985 | -0,405799 | 0,211336 |

Legenda:

A1 - Imaginação; **A2** - Interesses Artísticos; **A3** - Emotividade; **A4** - Senso Aventureiro; **A5** - Intelecto; **A6** - Liberalismo; **R1** - Auto Eficácia; **R2** - Ordem; **R3** - Senso de dever; **R4** - Empenho; **R5** - Autodisciplina; **R6** - Prudência; **E1** - Amigabilidade; **E2** - Gregarismo; **E3** - Assertividade; **E4** - Nível de Atividade; **E5** - Procura por excitação; **E6** - Bom humor; **S1** - Confiança; **S2** - Moralidade; **S3** - Altruísmo; **S4** - Cooperação; **S5** - Modéstia; **S6** - Compaixão; **N1** - Ansiedade; **N2** - Raiva; **N3** - Depressão; **N4** - Auto Percepção; **N5** - Falta de moderação; **N6** - Vulnerabilidade

Os resultados apresentados na Tabela 13 mostram as correlações entre as facetas atribuídas pelo IPIP-NEO 120 com as facetas atribuídas pelos *posts* do Facebook. Observa-se que as facetas E2 - Gregarismo (-0,663698) e N2 - Raiva (0,631622) obtiveram correlações fortes. Já as facetas R6 - Prudência (0,346550), S5 - Modéstia (0,357868), S6 - Compaixão (-0,405799) e N4 - AutoPercepção (-0,372612) obtiveram correlações moderadas. As facetas A1 - Imaginação, A2 - Interesses Artísticos, A4 - Senso Aventureiro, R1 - Auto Eficácia, R5 - Autodisciplina, E4 - Nível de Atividade, S1 - Confiança, S2 - Moralidade, S3 - Altruísmo e S4 - Cooperação obtiveram correlações muito fraca, enquanto as facetas A3 - Emotividade, A6 - Liberalismo, R2 - Ordem, R3 - Senso de dever, E3 - Assertividade, E5 - Procura por excitação, N1 - Ansiedade, N3 - Depressão, N5 - Falta de moderação e N6 - Vulnerabilidade obtiveram correlações fracas.

Nesta análise observa-se que muitas facetas obtiveram correlações fracas e muito fracas. Salienta-se que quando uma pessoa se dispõe a participar da pesquisa, primeiramente ela é orientada a responder os questionários propostos. Nestes questionários existem tipos de perguntas que através de suas respostas é possível inferir personalidade, em outras palavras, pelos questionários é possível a identificação de todos os fatores do *Big Five* e todas as facetas do IPIP-NEO. Sabe-se que os *posts* são ditados pelos participantes e nem sempre determinadas situações vividas por eles são postados, por isso a dificuldade em encontrar maiores correlações. Essa foi uma das dificuldades encontradas no momento das análises.

Outro fator que chamou muito a atenção foi que durante as análises dos *posts* do *Facebook* foram encontradas dificuldades em identificar o verdadeiro sentido da palavra. Algumas palavras, em português, podem ter sentidos diferentes dependendo do contexto a qual foram atribuídas. Sendo assim, inferir personalidade por *posts* é bem mais suscetível a ruídos do que a inferência por questionários. Nos questionários o índice de ruídos é quase zero, sendo o único e possível erro é se o usuário responder uma pergunta de forma errada ou equivocada (na sua visão de respostas), fora esse detalhe nada mais compromete o resultado. Já em *posts*, os ruídos são muitos, por exemplo, o usuário pode “ironizar” ou “postar frases com duplos sentidos” ou simplesmente por estar em um “dia feliz”, postar textos felizes, como também pode estar em um “dia triste” e postar textos tristes.

6.5 Considerações Finais

Neste capítulo foram apresentados experimentos a fim de inferir personalidade através dos questionários IPIP-NEO 120 e TIPI e *posts* da rede social *Facebook*. Primeiramente foram convidadas pessoas e solicitado que realizassem um cadastro no *Personalitatem Inventory* e que respondessem os questionários propostos. Após responderem os questionários, os participantes que cadastraram a sua rede social *Facebook* tiveram alguns *posts* extraídos para análises e correlacionar com as respostas dadas nos questionários.

Os resultados das correlações permitem afirmar que as respostas dadas nos questionários IPIP-NEO 120 e TIPI obtiveram correlações fortes e moderadas. Já as análises dos *posts* extraídos do *Facebook* resultam em correlações moderadas e fracas, conforme apresentado e justificado no decorrer do capítulo.

O próximo capítulo apresentará as considerações finais, contribuições, limitações, assim como trabalhos futuros.

7 CONSIDERAÇÕES FINAIS

O estudo descrito no presente documento propõe inferir personalidade a partir de textos de rede social utilizando um léxico afetivo em português brasileiro, diminuindo a lacuna encontrada no estado da arte para tal idioma. O objetivo foi quantificar os fatores do *Big Five* que descrevem a personalidade de um indivíduo.

Para concretizar a proposta, foram realizadas pesquisas, em diversas bases de busca, acerca das inferências de personalidade por meio de textos, na língua inglesa, referência na área, e na língua portuguesa, visando à aquisição de bom embasamento teórico, essencial para a construção do método proposto.

A partir das pesquisas realizadas, notou-se a falta de métodos para inferir personalidade na língua portuguesa. Sendo assim, foi construído um léxico afetivo chamado de *Personalitatem Lexicon* a partir do léxico *LIWC*. Sua origem se deu utilizando um algoritmo de correlação, detalhado na seção 5.1, para a geração de métodos de inferência de personalidade.

Para confirmar a hipótese, foram aplicados aos participantes da pesquisa os questionários IPIP-NEO 120, que permite quantificar os fatores do *Big Five* e facetas que descrevem a personalidade de um indivíduo e o TIPI, que quantifica somente os fatores do *Big Five*. Dos mesmos participantes, foram solicitados seus nomes de usuário da rede social *Facebook* para que seus *posts* fossem coletados. As análises destes *posts* se deu por uma ferramenta chamada *Lexicon Miner*.

Concluída a fase de implementações, iniciou-se a etapa de testes e experimentos para validação do trabalho. Neste contexto, realizaram-se as análises dos *posts* coletados dos participantes e a partir daí correlacionaram-se com as respostas dadas nos questionários propostos. Foi comprovado que as correlações entre o questionário IPIP-NEO 120 e *posts* do *Facebook* e o questionário TIPI e *posts* do *Facebook* obtiveram correlações fortes para Neuroticismo e moderadas para Extroversão. Já relacionados às facetas, E2 - Gregarismo e N2 - Raiva obtiveram correlações fortes, porém R6 - Prudência, S5 - Modéstia, S6 - Compaixão e N4 - AutoPercepção obtiveram correlações moderadas. Estes testes permitem afirmar, que o método proposto é possível inferir personalidade.

7.1 Contribuições

São várias as contribuições obtidas nesta pesquisa. Além de ser a primeira investigação em inferência de personalidade baseado no *Big Five* e facetas para a língua portuguesa, este trabalho permitiu a ampliação teórica e prática da referida linha de pesquisa. Das principais contribuições alcançadas nesta dissertação de mestrado, pode-se citar:

- Criação de um léxico afetivo chamado de *Personalitatem Lexicon*, em português brasileiro, com finalidade de inferência de personalidade por meio de texto;
- Construção de um algoritmo de correlação facetas *LIWC*;
- Correlação dos questionários IPIP-NEO 120 e TIPI com os *posts* do *Facebook*.
- Os resultados das correlações mostram que existem correlações fracas e moderadas através dos *posts* extraídos.

7.2 Limitações

Mesmo com as contribuições feitas nesta dissertação, foram identificadas também algumas limitações. Uma das principais limitações nesta pesquisa está relacionada ao léxico *LIWC*. Este léxico foi utilizado por suas palavras estarem disponíveis na língua portuguesa do Brasil, sendo que sua tradução foi feita pela equipe do NILC. No léxico não existe evidência de palavras afetivas. Entretanto, é apresentado codificações (*labels*) que a partir delas foi possível identificar as palavras que o autor considerou como afetivas. A construção do *Personalitatem Lexicon* foi feita somente com as palavras contidas no léxico *LIWC*, nenhuma outra palavras que não contivesse no léxico original foi considerada.

Na atual situação, a inferência de personalidade é feita analisando as palavras pelo sentido real. Quando é utilizado um sentido figurado ou houver uma mudança de sentido real da palavra, a frase “ganha” outro sentido e isso não foram considerados.

Outra limitação desta pesquisa foi à falta de outros tipos de textos, a exemplo de chats e fóruns, para as análises. Isso comprovaria com mais robustez a hipótese principal que é a inferência de personalidade. Até o momento, não se tem conhecimento de pesquisas que abordem tais características.

7.3 Trabalhos Futuros

A partir deste trabalho de mestrado, foi possível identificar outros cenários a serem investigados no futuro. Entre os principais trabalhos futuros que surgem a partir desta dissertação, cita-se:

- Coletar textos de outras redes sociais, como por exemplo, o *Twitter*, podendo os resultados serem comparados com o *Facebook*.
- Adaptar o método desenvolvido para outras línguas, por exemplo, o inglês.
- Construir um algoritmo que seja capaz de interpretar o sentido das palavras nas frases.
- Aumentar o quantitativo de palavras do *Personalitatem Lexicon*. Atualmente o léxico é composto somente com palavras do *LIWC*.

Por fim, não existe uma ferramenta, em português do Brasil, que infere personalidade, por exemplo, para recrutamento de pessoas ou sistemas de recomendação, sendo viável o aprimoramento e o desenvolvimento de tal ferramenta para fins comerciais.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFONSO, S., BICK, E., HABER, R. and SANTOS, D. **Floresta Sintá (c) tica: A treebank for Portuguese**. In *LREC*, 2002.
- ALLPORT F. H. and ALLPORT G. W. **Personality traits: Their classification and measurement**. *Journal of Abnormal and Social Psychology*, (16):6–40, 1921.
- ALLPORT, G. W. **Concepts of trait and personality**. *Psychological Bulletin*, (24):284–293, 1927.
- ALLPORT, G. W. **Personality and social encounter**. Boston, MA: Beacon. 1960.
- ALLPORT, G. W. **Personality: A psychological interpretation**. New York: H. Holt and Company, 1937.
- ALLPORT, G. W. and ODBERT, H. S. **Trait-names: A psycho-lexical study**. *Psycho-ology Monographs*, 47 (1, Whole No. 211), 1938.
- ALUÍSIO, S., PINHEIRO, G. M., MANFRIM, A. M. P, OLIVEIRA, L. H. M. de, GENOVES JR., L. C. and TAGNIN, S. E. O. **The Lácio-Web: Corpora and Tools to advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools**. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*. Lisboa, Portugal, pp. 1779-1782, 2004.
- AMES, C. **Motivation: What Teachers Need to Know**. *Teachers College Record*, [s.l.], v. 91, n. 3, p. 409-421, 1990.
- ANDRÉ, E., KLESEN, M., GEBHARD, P., ALLEN, S. and RIST, T. **Integrating models of personality and emotions into lifelike characters**. In *Affective interactions* (pp. 150-165). Springer Berlin Heidelberg, 2000.
- APPOLINÁRIO, F. **Metodologia da Ciência: filosofia e prática da pesquisa**. 1 ed. São Paulo: Editora Thomson, 2006.
- ASTON, G. and BURNARD, L. **The BNC Handbook: Exploring the British National Corpus with SARA**, Edinburgh University Press, Edinburgh, 1998.
- AVERILL, J.R. **A Semantic Atlas of Emotion Concepts**. *JSAS Catalog of Selected Documentus in Psychology*, 5, 330. (Manuscript # 421), 1975.
- AZEREDO, J. C. **Gramática Houaiss da Língua Portuguesa**. São Paulo: Publifolha. 2008.
- AZIZ, W. e SPECIA, L. **Fully Automatic Compilation of Portuguese-English and Portuguese-Spanish Parallel Corpora**. In *Proceedings of the The 8th Brazilian Symposium in Information and Human Language Technology (STIL 2009)*. Cuiabá, MT, Brazil, 2011.

BALAGE FILHO, P.P.; ALUÍSIO, S.M.; PARDO, T.A.S. **An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis**. In the Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology – STIL, pp. 215-219. October 21-23, Fortaleza/Brazil. 2013.

BARBARANELLI, C. e CAPRARA, G. V. **Studies of the Big Five Questionnaire**. In B. De Raad and M. Perugini, editors, *Big Five Assessment*, chapter 5, pages 109–128. Hogrefe Huber, Germany, first edition, 2002.

BARKHUSS, L. e CSANK P. **Allport's theory of traits** - A critical review of the theory and two studies. Technical report, Concordia University. 1999.

BATSON, C., SHAW, L. e OLESON, K. **Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions**,” *Rev. Personality Soc. Psychology*, vol. 13, pp. 294–326, 1992.

BERCHT, M. **Computação afetiva: Vínculos com a psicologia e aplicações na educação**. *Psicologia & Informática. Produções do III PSICOINFO e II Jornada do NPPI / Oliver Zancul Prado, Ivelise Fortim e Leonardo Consetino (Organizadores) – São Paulo: Conselho Regional de Psicologia de São Paulo: CRP/SP, 1ª edição, 2006.*

BICK, E. **The Parsing System “Palavras”**: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus: Aarhus University Press, 2000.

BRADLEY, M. M., e LANG, P.J. **Affective norms for English words (ANEW)**: Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

BRUCKSCHEN, M., MUNIZ, F., SOUZA, J. G. C., FUCHS, J. T., INFANTE, K., MUNIZ, M., GONÇALVES, P. N., VIEIRA, R. e ALUÍSIO, S. M. **Anotação Lingüística em XML do Corpus PLN-BR**. Série de Relatórios do NILC. NILC-TR-09-08, 39 p., 2008.

BURGER, J. M. **Personality**. Wadsworth, fifth edition, 2000.

BUSS, A. H. e FINN, S. E. **Classification of Personality Traits**. *Journal of Personality and Social Psychology*, Vol 52(2), 432-444, 1987.

BUSSAB, W. O and MORETTIN, A. **Estatística Básica**. 3 ed. São Paulo: Atual, 1986.

BUTCHER, J. N. **Minnesota multiphasic personality inventory**. *Corsini Encyclopedia of Psychology*, 1989.

CAMBRIA, E.; OLSHER, D.; RAJAGOPAL, D. **SenticNet 3**: A common and common-sense knowledge base for cognition-driven sentiment analysis. In: *AAAI*, pp. 1515-1521, Quebec City, 2014.

CAMBRIA, E.; SPEER, R.; HAVASI, C. e HUSSAIN, A. **SenticNet: A Publicly Available Semantic Resource for Opinion Mining.** In: AAAI fall symposium: commonsense knowledge, v. 10, p. 02, 2010.

CATTELL, R. B. **The description of personality: basic traits resolved into clusters.** *The Journal of Abnormal and Social Psychology*, 38(4), 476–506, 1943.

CATTELL, R. B. **The description of personality: Principles and findings in a factor analysis.** *The American Journal of Psychology*, 58(1):69–90, 1945.

CATTELL, R. **Sentiment or attitude? The core of a terminology problem in personality research,**” *J. Personality*, vol. 9, pp. 6–17, 2006.

COLTHEART, M. **The Mrc Psycholinguistic Database.** *Quarterly J. of Experimental Psychology*, 33A:497–505, 1981.

COMREY, A. L. **Factor-analytic methods of scale development in personality and clinical psychology.** *Journal of Consulting and Clinical Psychology*, Vol 56(5), 754–761, 1988.

COOK, M. **Levels of Personality.** New York: Praeger, 1984.

COSTA, P. T. e MCCRAE R. R. **Revised neo personality inventory (neo-pi-r) and neo five-factor inventory (neo-ffi):** Professional manual., 1992.

COSTA, P. T. e MCCRAE, R. R. **From catalog to classification: Murray’s needs and the Five- Factor Model.** *Journal of Personality and Social Psychology*, 55 (2), 255-265, 1988.

COSTA Jr, P. T. e WIDIGER, T. A. **Personality disorders and the Five-Factor Model of Personality** American Psychological Association, 1993.

CRUZ, F. L., TROYANO, J. A., ENRIQUEZ, F., and ORTEGA, J. **Clasificación de documentos basada en la opinión: experimentos con un corpus de criticas de cine en espanol.** *Procesamiento de Lenguaje Natural*, 41, 2008.

DAMÁSIO, A. **Em busca de Espinosa: prazer e dor na ciência dos sentimentos.** MOTTA, L. T. (Trad.). São Paulo: Cia das Letras, 2004.

DAMÁSIO, A. R. **Descartes’ Error: Emotion, Reason, And The Human Brain.** Quill, New York, 1994.

DAMÁSIO, A. R. **O Erro de Descartes.** Emoção, Razão e o Cérebro Humano. São Paulo: Companhia das Letras, 1996.

DAMASIO, A. R. **The Feeling of What Happens.** Harcourt, Orlando, Florida, 1999.

DE RAAD, B.; PERUGINI, M.; HREBÍCKOVÁ, M. e SZAROTA, P. **Lingua franca personality: Taxonomies and structures based on the psycholexical approach.** *Journal of Cross-Cultural Psychology*, 29 (1): 212-232, 1998.

DEMOS, V. E. **An Affect Revolution**: Silvan Tompkin's Affect Theory. Exploring Affect: The Selected Writings of Silvan S. Tompkins. Ed. Virginia E. Demos. New York: Press Syndicate of the U. of Cambridge, 17-26, 1995.

DIAS-DA-SILVA, B. C.; DI FELIPPO, A.; NUNES, M. G. V. **The automatic mapping of Princeton WordNet lexical conceptual relations onto the Brazilian Portuguese WordNet database**. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrocos, 2008.

DOLAN, R. **Emotion, cognition and behavior**. Science, vol. 298, no. 5596, pp. 1191–1194, 2002.

DORSCH, F.; HÄCKER, H. e STAPF, K.H. **Dicionário de psicologia Dorsch**. LEÃO, E.C. (Trad.). Petrópolis, RJ: Vozes, 2008.

ELLIOT, C. D. **The Affective Reasoner**: a process model of emotions in a multi-agent system. Ph.D. thesis, Northwestern University, Evanston, Illinois, 1992.

ESULI, A. e SEBASTIANI F. **Determining term subjectivity and term orientation for opinion mining**. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming, 2006a.

ESULI, A., e SEBASTIANI F. **Determining the Semantic orientation of terms through gloss analysis**. In Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management, pages 617–624, Bremen, DE, 2005.

ESULI, A. e SEBASTIANI F. **Sentiwordnet**: A publicly available lexical resource for opinion mining. Proceedings of LREC. Vol. 6. 2006b.

EVANS, R. e KILGARRIFF, A. **Mrds, standards and how to do lexical engineering**. Technical Report ITRI- 95-19, University of Brighton, 1995.

EYSENCK, H. J. **Uses and abuses of psychology**. London: Penguin Books, 1953.

FELDMAN R. Techniques and applications for sentiment analysis. *Commun ACM*, 56:82–89, 2013.

FELDMAN, R. e FEINMAN, J. **Who are you**: Personality and its development. New York: Franklin Watts, 1992.

FRANCIS, G. e HUNSTON, S. **Grammar Patterns 2**: Nouns and Adjectives. London: HarperCollins, COBUILD, 1998.

FRANCIS, G. e HUNSTON, S. **Grammar Patterns 1**: Verbs. London: HarperCollins, COBUILD, 1996.

FREITAS, C. **Sobre a construção de um léxico da afetividade para o processamento computacional do português**. Revista Brasileira de Linguística Aplicada, 13(4), 1013-1059, 2013.

FREITAS, C.; MOTTA, E.; MILIDIÚ, R.; CESAR, J. **Vampiro que brilha... rá!** Desafios na anotação de opinião em um corpus de resenhas de livros. In: XI Encontro de Linguística de Corpus (ELC 2012), São Paulo, Brasil, 2012.

FRIEDENBERG, J. e SILVERMAN, G. **Cognitive Science: An Introduction to the Study of Mind**. 1st ed. Newbury Park, CA, USA: SAGE, 2005.

FUNDER, D. **The Personality Puzzle**. Norton, 2001.

GARCIA, D. e SCHWEITZER, F. **Emotions in Product Reviews--Empirics and Models**. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), p. 483-488, 2011.

GARCIA, O. M. **Comunicação em prosa moderna: aprenda a escrever, aprendendo a pensar**. Rio de Janeiro: FGV, 1982.

GOLDBERG, L. R. **An alternative "description of personality": The Big-Five factor structure**. *Journal of Personality and Social Psychology*, Vol 59(6), 1216-1229, 1990.

GOLDBERG, L. R. **Language and individual differences: The search for universals in personality lexicons**. Em L. Wheeler (Org.), *Review of personality and social psychology* (pp. 141-165). Beverly Hills, CA: Sage, 1981.

GOLDBERG, L. R., JOHNSON, J. A., EBER, H. W., HOGAN, R., ASHTON, M. C., CLONINGER, C. R., & GOUGH, H. G. **The International Personality Item Pool And The Future Of Public-Domain Personality Measures**. *Journal Of Research In Personality*, 40, 84–96, 2006.

GOLDBERG, L., R. **The Development of Markers for the Big Five factor Structure**. In *Psychological Assessment*, 4(1). pp. 26–42. 1992.

GOLEMAN, D. **Emotional Intelligence - Why it can matter more than IQ?** Bloomsbury, London, first edition, 1995.

GORDON, S. L. **The sociology of sentiments and emotion**. in *Social Psychology: Sociological Perspectives*, M. Rosenberg and R. H. Turner, eds., New York, NY, USA: Basic Books, pp. 562–592, 1981.

GOSLING, S. D., RENTFROW, P. J., & SWANN, W. B., JR. **A Very Brief Measure of the Big Five Personality Domains**. *Journal of Research in Personality*, 37, 504-528, 2003.

GREGORY M. L.; CHINCHOR N.; WHITNEY P.; CARTER R.; HETZLER E.; e TURNER A. **User-directed sentiment analysis: Visualizing the affective content of documents**. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, Association for Computational Linguistics, pages 23–30, Sydney, Australia, July 2006.

HASSIN, M. H. M.; AZIZ, A. A.; NORWAWI, N. M. **Affective computing**: knowing how you feel. IN: The National Seminar of Science Technology and Social Science (STSS '04), UiTM Pahang, 2004.

HENDRINKS A. A. J., HOFSTEE W. K. B., e RAAD, B. **The Five-Factor Personality Inventory**: Assessing the Big Five by means of brief and concrete statements. In B. De Raad and M. Perugini, editors, *Big Five Assessment*, chapter 4, pages 79–108. Hogrefe Huber, Germany, first edition, 2002.

IPIP. **The international Personality item pool**, 2006. (Available at <http://ipip.ori.org/ipip/>)

IZARD, C.E. **Emotion-cognition relationships and human development**. In C.E. Izard, J. Kagan, & R.B. Zajonc (eds.). *Emotions, cognition, and behavior*. New York: Cambridge University Press, 1984.

JOHN, O. P. e SRIVASTAVA S. **The big five trait taxonomy**: History, measurement, and theoretical perspectives. In Lawrence A. Pervin and Oliver P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. The Guilford Press, New York, second edition, 1999.

JOHNSON, J. A. Ascertaining the validity of individual protocols from webbased personality inventories. *Journal of research in Personality*, 39(1):103–129, 2005.

JOHNSON, J. A. **Computer narrative interpretations of individual profiles**. R. Hogan, J. Johnson and S. Briggs, (chapter withdrawn due to space limitations. Available from the author at Penn State DuBois, DuBois, PA 15801.) 1994.

JOHNSON, J. A. **Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory**: Development of the IPIP-NEO-120, In: *Journal of Research in Personality*, p. 78-89, 2014.

JOHNSON, J. A. **Predicting Observers Ratings Of The Big Five From The Cpi, Hpi, And Neo-Pi-R**: A Comparative Validity Study. *European Journal Of Personality*, 14, 1–19, 2000a.

JOHNSON, J. A. **Web-based personality assessment**. In 71st Annual Meeting of the Eastern Psychological Association, Baltimore, MD, 2000b.

JURAFSKY, D. e MARTIN, J. **Speech and Language Processing**: Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Upper Saddle. River, New Jersey, Prentice Hall. (pp. 1-18). 2000.

KAMPS, J., MARX, M., MOKKEN, R. J., e DE RIJKE, M. **Using WordNet to measure semantic orientation of adjectives**. In *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, pages 1115–1118, Lisbon, PT, 2004.

KANFER, A.; TANAKA, J.S. **Unraveling the Web of Personality Judgments: The Influence of Social Networks on Personality Assessment.** *Journal of Personality*, 61(4) pp. 711–738. 1993.

KASSIN, S. **Psychology.** Prentice-Hall, USA, 2003.

KIPPER-SCHULER, K. **VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon.** PhD Thesis, University of Pennsylvania, 2005.

KIROUAC, G. **Les émotions.** In: Richele, M. et alii. *Traité de Psychologie Experimentale.* Paris, PUF, 1994.

KLEINGINNA, P. R. e KLEINGINNA, A. **A categorized list of emotion definitions, with suggestions for a consensual definition.** *Motivation Emotion*, vol. 5, no. 4, pp. 345–379, 1981.

KLEINMUNTZ, B. **Concepts and the Structure of Memory.** Department of Psychology Carnegie Institute of Technology Pittsburgh, Pennsylvania, 1967.

KRISTENSEN, C. H., GOMES, C. F. D. A., JUSTO, A. R., & VIEIRA, K. **Brazilian norms for the affective norms for English words.** *Trends in psychiatry and psychotherapy*, 33(3), 135-146, 2011.

LEVIN, B. **English Verb Classes and Alternation, A Preliminary Investigation.** In: The University of Chicago Press, 1993.

LIKERT, R. **A technique for the measurement of attitudes.** *Archives of psychology*, 1932.

LISSETTI, C. L. **Personality, affect and emotion taxonomy for socially intelligent agents.** In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, pages 397–401. AAAI Press, 2002.

LIU, H., LIEBERMAN H. e SELKER T. **A model of textual affect sensing using real-world knowledge.** *Proceedings of the 8th international conference on Intelligent user interfaces*, 12-15, Miami, Florida, USA. 2003.

LONGHI, M. T. **Mapeamento de aspectos afetivos em um ambiente virtual de aprendizagem.** PPGIE/UFRGS. Tese de Doutorado. Porto Alegre, 2011.

LONGHI, M. T.; BEHAR, P. A. e BERCHT, M. **In Search of the Affective Subject Interacting in the ROODA Virtual Learning Environment.** In: KCKS'2010, KEY COMPETENCIES IN THE KNOWLEDGE SOCIETY, Brisbane, Australia. IFIP Advances in Information and Communication Technology, 2010a, v.324, pp.234-245, 2010.

MA, C.; PRENDINGER, H. e ISHIZUKA, M. **Emotion Estimation and Reasoning Based on Affective Textual Interaction, in Affective Computing and Intelligent Interaction.** (First Int'l Conf. ACII. Beijing, China, p.622-628. 2005.

MADDI, S. R. **Personality Theories: A comparative analysis.** Illinois: Dorsey Press, 1980.

MASSUMI, B. **Parables for the Virtual.** Durham: Duke UP, 2002.

MCCRAE, R. R. e JOHN, O. P. **An introduction to the five-factor model and its applications.** *Journal of Personality*, 60(2):175–216, June 1992.

MCCRAE, R. R. **Openness to experience as a basic dimension of personality.** *Imagination, Cognition and Personality*, 13, 39-55, 1993.

MIHALCEA, R., BANEA, C., e WIEBE, J. **Learning multilingual subjective language via cross-lingual projections.** In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 976–983, Prague, CZ, 2007.

MUKAKA, M. M. **A guide to appropriate use of Correlation coefficient in medical research.** *Malawi Medical Journal* 24, no. 3: 69-71, 2012.

MUNEZERO, M., MONTERO, C. S., SUTINEN, E., e PAJUNEN, J. **Are they different?** affect, feeling, emotion, sentiment, and opinion detection in text. *Affective Computing, IEEE Transactions on*, 5(2), 101-111, 2014.

MUNIZ, M. C. M. e NUNES, M. G. V. **A construção de recursos linguístico-computacionais para o português do Brasil: o projeto de Unitex-PB.** São Carlos, 2004.

MURRAY, H. e MORGAN, C. **A clinical study of sentiments i.** *Genetic Psychological Monograph*, vol. 32, pp. 153–311, 1945.

NUNES, C. H. S., HUTZ, C. S. e GIACOMONI, C. H. **Associação entre bem estar subjetivo e personalidade no modelo dos cinco grandes fatores.** *Avaliação Psicológica*, 8(1), 99-108, 2009.

NUNES, M. A. S. N.; TELES, F. R. e DE SOUZA, J. G. **Inferindo personalidade via tweets.** In: *GEINTEC-Gestão, Inovação e Tecnologias* 3.3, p. 045-057. 2013.

NUNES, M. A. S. N. **Recommender System based on Personality Traits.** (Tese De Doutorado). Université Montpellier 2-Lirimm- Franca, 2008.

NUNES, M. A. S. N. e HU, R. **Personality-based Recommender Systems: an Overview.** In: *ACM Conf. on Recommender System.* Dublin. Proceedings of (RecSys '12). New York: ACM, p. 5-7, 2012.

NUNES, M. A. S. N.; CAZELLA, S.C. **O que sua Personalidade revela?** Fidelizando clientes web através de Sistemas de Recomendação e Traços de Personalidade. In: (Webmedia 2011 Minicursos): Patricia Vilain e Valter Roesler. (Org.). Tópicos em Banco de Dados e Multimídia e Web. Porto Alegre: SBC, v. 1, p. 91-122, 2011.

ORTONY, A., CLORE, G. L., e FOSS, M. A. **The Referential Structure of the Affective Lexicon**. *Cognitive science*, 11(3), 341-364, 1987.

ORTONY, A.; CLORE, G. e COLLINS, A. **The cognitive structure of emotions**. Cambridge: Cambridge University Press, 1990.

PAIVA, A. **Affective interactions**: towards a new generation of computer interfaces. pages 1–8, 2000.

PALMER, M.; GILDEA, D.; XUE, N. **Semantic Role Labeling**. Synthesis Lectures on Human Language Technology Series, Ed. Graeme Hirst, Morgan & Claypoole Publishers, 2010.

PANG, B. e LEE, L. **Opinion Mining and Sentiment Analysis**. New York, NY, USA: Cambridge Univ. Press, 2008.

PASQUALI, L. **Instrumentos psicológicos**: Manual prático de elaboração. Brasília: LabPAM / IBAPP. 1999.

PASQUALI, L. **Técnicas de exame psicológico**: T.E.P manual. Vol.1. Fundamentos das Técnicas Psicológicas. São Paulo: Casa do Psicólogo; Conselho Federal de Psicologia. 2001.

PASQUALOTTI, P. R. e VIEIRA, R. **WordnetAffectBR**: uma base lexical de palavras de emoções para a língua portuguesa, 2008.

PENNEBAKER, J. W. e FRANCIS, M. E. **Cognitive, emotional, and language processes in disclosure**. *Cognition & Emotion*, 10(6), 601-626, 1996.

PENNEBAKER, J. W. e GRAYBEAL, A. **Patterns of natural language use**: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10(3), 90-93, 2001.

PENNEBAKER, J. W. e KING, L. A. **Linguistic styles**: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312. 1999.

PENNEBAKER, J. W., FRANCIS, M. E., e BOOTH, R. J. **Linguistic Inquiry and Word Count**: LIWC. Mahwah, NJ: Erlbaum Publishers (www.erlbaum.com). 2001.

PENNEBAKER, J. W. e KING, L. A. **Linguistic Inquiry and Word Count**: LIWC. Erlbaum Publishers, 1999.

PENNEBAKER, J. W., MAYNE, T. J. e FRANCIS, M. E. **Linguistic predictors of adaptive bereavement**. *Journal of personality and social psychology*, 72(4), 863, 1997.

PETERSEN, K., FELDT, R., MUJTABA, S. e MATTSSON, M. **Systematic mapping studies in software engineering**. In 12th international conference on evaluation and assessment in software engineering (Vol. 17, No. 1). Sn, 2008.

PICARD, R. W. **Affective Computing**. MIT Press, Cambridge, MA, USA, 1997.

PICARD, R. W. **An interview with rosalind picard**, author of *Affective computing* book. pages 219–228, 2000.

PICARD, R. W. **What does it mean for a computer to ‘have’ Emotions?** In R. Trapp, P. Petta, and S. Payr, editors, *Emotions in humans and artefacts*, chapter 7, pages 213–235. A Bradford Book - MIT Press, Cambridge, Massachusetts, 2002.

PORIA, S., A. GELBUKH, B. AGARWAL, E. CAMBRIA, AND N. HOWARD. **Common sense knowledge based personality recognition from text**. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 8266 LNAI, 2013.

RAHIM M.A., **A measure of styles of handling interpersonal conflict**. *Academy of Management journal*, vol. 26, no. 2, pp. 368–376, 1983.

RAMMSTEDT B. e JOHN, O. P. **Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german**, *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

REEVES, B. e NASS, C. **The media equation: how people treat computers, television, and new media like real people and places**. Cambridge University Press, New York, NY, USA, 1996.

ROSEMAN, I. J.; SPINDEL, M. S. e JOSE, P. **Appraisals of emotion-eliciting events: Testing a theory of discrete emotions**. *Journal Personality and Social Psychology*, v. 5, n. 59, p. 899– 915. 1990.

ROSEMAN, IJ e SMITH, C. A. **Appraisal Theory: overview, assumptions, varieties, controversies**. In: SCHERER, K.R.; SCHORR, A; JOHNSTONE, T. (Eds) *Appraisal Processes in Emotion: Theory, Methods, Research*, New York and Oxford: Oxford University Press. p. 3-19, 2001.

ROUSSEAU, D. e HAYES-ROTH, B. **Personality in synthetic agents**. Technical Report KSL-96-21, July 1996.

RYFF, C. D. e KEYES, C. L. M. **The structure of psychological well-being revisited**. *Journal of Personality and Social Psychology*, 69, 719-727, 1995.

SARDINHA, A. B. **Corpus linguistics: history and problematization**. *Delta* v. 16, n. 2, p. 323-367, 2000.

SCARTON, C. E. e ALUISIO, S. **Towards a cross-linguistic VerbNet-style lexicon for Brazilian Portuguese**. In: *Workshop on Creating Cross-language Resources for Disconnected Languages and Styles Workshop Programme*, p.11, 2012.

SCHERER, K. R. **Psychological models of emotion**. in *The Neuropsychology of Emotion*, J. Borod, ed., Oxford, U.K.: Oxford Univ. Press, pp. 137–167, 2000b.

SCHERER, K. R. **Studying the emotion-antecedent appraisal process: an expert system approach**. *Cognition and Emotion*, [S.l.], n.7, p. 1-141, 325-355, 2000a.

SCHERER, K. R., SCHORR, A. E JOHNSTONE, T. **Appraisal processes in emotion: Theory, methods, research.** Oxford University Press, 2001.

SCHIMIT, M.; J. KIHM, J. A. e ROBIE, C. **The Global Personality Inventory (GPI).** In B. De Raad and M. Perugini, editors, Big Five Assessment, chapter 9, pages 195–236. Hogrefe Huber, Germany, first edition, 2002.

SCARTON, C. E. **VerbNet.Br:** construção semiautomática de um léxico verbal online e independente de domínio para o português do Brasil. Universidade de São Paulo, 2013.

SCHULTZ, D. **Theories of Personality.** Brooks/Cole, forth edition, 1990.

SENTISTRENGTH. **SentiStrength.** Disponível em: < <http://sentistrength.wlv.ac.uk/>>. Acesso em: 17 nov. 2015.

SHOUSE, E. **Feeling, emotion, affect.** Media Culture J., vol. 8, no. 6, p. 1, 2005.

SILVA, I. B., e NAKANO, T. C. **Modelo dos Cinco Grandes Fatores da personalidade:** análise de pesquisas. Avaliação Psicológica, 10(1), 51-62, 2011.

SIMON, H. A. **Reason in Human Affairs.** Stanford University Press, California, 1983.

SINCLAIR, J. **Eagles Preliminary Recommendations On Corpus Typology EAG-TCWG-CTYP/P.** Version of May, ILC-CNR, Pisa, 1996.

SINGER, R. N. **Psicologia dos esportes:** mitos e verdades. São Paulo: Harba, 1986.

SOLDZ, S. e VAILLANT G. E. **The big five personality traits and the life course:** A 45 years longitudinal study. Journal of Research in Personality, 33:208–232, 1998.

SOUZA, M., VIEIRA, R., Busetti, D., CHISHMAN, R., e ALVES, I. M. **Construction of a portuguese opinion lexicon from multiple resources.** STIL, 2011.

SPECIA, L. e NUNES, M. **Desambiguação Lexical Automática de Sentido:** Um Panorama. Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional. NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP, Brasil, 2004.

TAN, S., CHENG, X., GHANEM, M.M., WANG, B. and XU, H. **A novel refinement approach for text categorization.** In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 469-476). ACM, 2005.

TAUSCZIK, Y., and PENNEBAKER, J.W. **The psychological meaning of words:** LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29, 24-54. 2010.

THAGARD, P. **Hot Thought:** Mechanisms and Applications of Emotional Cognition. A Bradford Book - Mit Press, Cambridge, Ma, Usa, 2006.

THELWALL, M., BUCKLEY, K., PALTOGLOU, G., CAI, D., and KAPPAS, A. (2010). **Sentiment strength detection in short informal text**. In: Journal of the American Society for Information Science and Technology, p. 2544-2558, 2010.

THOITS, P. A. **The sociology of émotions**. Annu. Rev. Sociology, vol. 15, pp. 317–342, 1989.

TRAPPL, R.; PAYR, S. e PETTA, P. **Emotions In Humans And Artifacts**. Mit Press, Cambridge, Ma, Usa, 2003.

TRASK, R. L. **Dicionário de Linguagem e Linguística**. ILARI, R. (Trad.), São Paulo: Contexto. 2008.

TURNEY, P. D. **Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews**. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, pages 417–424, Morristown,US, 2002.

URQUIJO, S. **Modelos circunplexos da personalidade**. Em F. F. Sisto, E. T. B. Sbardelini & R. Primi. (Orgs.), Contextos e questões da avaliação psicológica. São Paulo: Casa do Psicólogo, pp. 31-49, 2001.

VALITUTTI, A., STRAPPARAVA, C. e, STTOCK, O. **Developing Affective Lexical Resources**. PsychNology Journal, Volume 2, Number 1, 61 – 83, 2004.

VAN ZALK, N., VAN ZALK, M., KERR, M. e STATTIN, H. **Social Anxiety as a Basis for Friendship Selection and Socialization in Adolescents' Social Networks**. Journal of Personality, 79: pp. 499–526. 2011.

WANG H.; HELMUT P.; MITSURU I.; e TAKEO I. **Affective Communication in Online Chat Using Physiological Sensors and Animated Text**, Trans. Human Interface Society, Vol.7, No.1, pp.39-45. 2005.

WECHSLER, S. M. **Guia de procedimentos éticos para a avaliação psicológica**. Em S. M. Wechsler & R. S. L. Guzzo (Orgs.), Avaliação psicológica: Perspectiva internacional (pp. 133-141). São Paulo: Casa do Psicólogo. 1999.

WECHSLER, S. M. **Princípios éticos e deontológicos na avaliação psicológica**. Em L. Pasquali (Org.), Técnicas de exame psicológico: T.E.P manual. Fundamentos das técnicas psicológicas (Vol. 1, pp. 171-193). São Paulo: Casa do Psicólogo; Conselho Federal de Psicologia. 2001.

WIEBE J. **Identifying subjective characters in narrative**. in Proc. 13th Int. Conf. Comput. Linguistics, vol. 2, pp. 401–406, 1990.

WIERZBICKA, A. **Emotions Across Languages and Cultures: Diversity and Universals**. Cambridge, U.K.: Cambridge Univ. Press, 1999.

WILSON, M. D. **The MRC Psycholinguistic Database**: Machine Readable Dictionary, Version 2. In: Behavioral Research Methods, Instruments and Computers, p. 6–11, 1988.

ZANETTE, A. **Aquisição de Subcategorization Frames para Verbos da Língua Portuguesa**. 53f. Trabalho de Conclusão de Curso (Bacharel em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

ZAVAGLIA, C. **O Papel do Léxico na Elaboração de Ontologias Computacionais: do seu resgate à sua disponibilização**. *Linguística IN FOCUS - Léxico e morfofonologia: perspectivas e análises*, EDUFU, Uberlândia, v. 4, n. 1, pp. 233-274, 2006.

Apêndice A - Referências Bibliográficas do Mapeamento Sistemático

| | |
|----|---|
| 2 | Khan, F. H., U. Qamar, and S. Bashir. 2016. Multi-objective model selection (MOMS)-based semi-supervised framework for sentiment analysis. <i>Cognitive Computation</i> : 1-15. |
| 10 | Nasseri, A. A., A. Tucker, and S. De Cesare. 2015. Quantifying StockTwits semantic terms' trading behavior in financial markets: An effective application of decision tree algorithms. <i>Expert Systems with Applications</i> 42, no. 23: 9192-9210. |
| 12 | Alahmadi, D. H. and X. -J Zeng. 2015. ISTS: Implicit social trust and sentiment based approach to recommender systems. <i>Expert Systems with Applications</i> 42, no. 22: 8840-8849. |
| 13 | Musto, C., G. Semeraro, P. Lops, and M. D. Gemmis. 2015. CrowdPulse: A framework for real-time semantic analysis of social streams. <i>Information Systems</i> 54, : 127-146. |
| 15 | Noferesti, S. and M. Shamsfard. 2015. Using linked data for polarity classification of patients' experiences. <i>Journal of Biomedical Informatics</i> 57, : 6-19. |
| 16 | Ajitha, P. and G. Gunasekaran. 2015. Sentiment prediction based on valence and arousal using concept search engine. |
| 17 | Kazi, F., S. Joshi, and S. Machchhar. 2015. A survey on online forum hotspot detection. |
| 19 | Li, J., S. Fong, Y. Zhuang, and R. Khoury. 2015. Hierarchical classification in text mining for sentiment analysis of online news. <i>Soft Computing</i> . |
| 21 | Petz, G., M. Karpowicz, H. Fürschuß, A. Auinger, V. Střiteský, and A. Holzinger. 2015. Reprint of: Computational approaches for mining user's opinions on the web 2.0. <i>Information Processing and Management</i> 51, no. 4: 510-519. |
| 31 | Vinodhini, G. and R. M. Chandrasekaran. 2015. Sentiment classification using principal component analysis based neural network model. |
| 32 | Zhang, P. and Z. He. 2015. Using data-driven feature enrichment of text representation and ensemble technique for sentence-level polarity classification. <i>Journal of Information Science</i> 41, no. 4: 531-549. |
| 35 | Ravi, K., V. Ravi, and C. Gautam. 2015. Online and semi-online sentiment classification. |
| 43 | Hegde, Y. and S. K. Padma. 2015. Sentiment analysis for kannada using mobile product reviews: A case study. |
| 54 | Serrano-Guerrero, J., J. A. Olivas, F. P. Romero, and E. Herrera-Viedma. 2015. Sentiment analysis: A review and comparative analysis of web services. <i>Information Sciences</i> 311, : 18-38, |
| 59 | Candelieri, A. and F. Archetti. 2015. Detecting events and sentiment on twitter for improving urban mobility. |
| 61 | Khasawneh, R. T., H. A. Wahsheh, I. M. Alsmadi, and M. N. Ai-Kabi. 2015. Arabic sentiment polarity identification using a hybrid approach. |
| 64 | Polpinij, J. 2015. Multilingual sentiment classification on large textual data. |
| 65 | Walha, A., F. Ghazzi, and F. Gargouri. 2015. ETL transformation algorithm for facebook opinion data. |
| 66 | Muhammad, A., N. Wiratunga, and R. Lothian. 2015. A hybrid sentiment |

| | |
|-----|---|
| | lexicon for social media mining. |
| 68 | Poddar, S., V. Kattagoni, and N. Singh. 2015. Personality mining from biographical data with the "adjectival marker" technique. |
| 73 | Cherif, W., A. Madani, and M. Kissi. 2015. A new modeling approach for arabic opinion mining recognition. |
| 75 | Alsaffar, A. and N. Omar. 2015. Integrating a lexicon based approach and K nearest neighbour for malay sentiment analysis. <i>Journal of Computer Science</i> 11, no. 4: 639-644. |
| 77 | Tang, D., B. Qin, and T. Liu. 2015. Deep learning for sentiment analysis: Successful approaches and future challenges. <i>Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery</i> 5, no. 6: 292-303. |
| 78 | Thakur, D. and J. Singh. 2015. The SAFE miner: A fine grained aspect level approach for resolving the sentiment. |
| 87 | Nigam, S. and A. Jawdekar. 2015. An efficient person name bipolarization using KPCA. |
| 88 | Almatrafi, O., S. Parack, and B. Chavan. 2015. Application of location-based sentiment analysis using twitter for identifying trends towards indian general elections 2014. |
| 92 | Kanakaraj, M. and R. M. R. Guddeti. 2015. Performance analysis of ensemble methods on twitter sentiment analysis using NLP techniques. |
| 103 | Vuong, Q. -H and A. Takasu. 2014. Transfer learning for emotional polarity classification. |
| 107 | Deng, Z. -H, K. -H Luo, and H. -L Yu. 2014. A study of supervised term weighting scheme for sentiment analysis. <i>Expert Systems with Applications</i> 41, no. 7: 3506-3513. |
| 111 | Li, J., S. Fong, Y. Zhuang, and R. Khoury. 2014. Hierarchical classification in text mining for sentiment analysis. |
| 114 | Wang, H., L. Liu, W. Song, and J. Lu. 2014. Feature-based sentiment analysis approach for product reviews. <i>Journal of Software</i> 9, no. 2: 274-279. |
| 115 | Colace, F., M. D. Santo, and L. Greco. 2014. Sentiment mining through mixed graph of terms. |
| 116 | Li, D., J. Niu, M. Qiu, and M. Liu. 2014. Sentiment analysis on weibo data. |
| 123 | García-Pablos, A., M. Cuadros, S. Gaines, and G. Rigau. 2014. Unsupervised acquisition of domain aspect terms for aspect based opinion mining. <i>Procesamiento De Lenguaje Natural</i> 53, : 121-128, |
| 130 | Chen, X., W. Tang, H. Xu, and X. Hu. 2014. Double LDA: A sentiment analysis model based on topic model. |
| 132 | Petz, G., M. Karpowicz, H. Fürschuß, A. Auinger, V. Stříteský, and A. Holzinger. 2014. Computational approaches for mining user's opinions on the web 2.0. <i>Information Processing and Management</i> 50, no. 6: 899-908. |
| 139 | Abdulla, N., R. Majdalawi, S. Mohammed, M. Al-Ayyoub, and M. Al-Kabi. 2014. Automatic lexicon construction for arabic sentiment analysis. |
| 140 | Uhr, P., J. Zenkert, and M. Fathi. 2014. Sentiment analysis in financial markets: A framework to utilize the human ability of word association for analyzing stock market news reports. |
| 146 | Del Pilar Salas-Zárate, M., E. López-López, R. Valencia-García, N. Aussenac-Gilles, Á. Almela, and G. Alor-Hernández. 2014. A study on LIWC categories for opinion mining in spanish reviews. <i>Journal of Information Science</i> 40, no. 6: 749-760. |

| | |
|-----|---|
| 153 | Molla, A., Y. Biadgie, and K. -A Sohn. 2014. Network-based visualization of opinion mining and sentiment analysis on twitter. |
| 155 | Akaichi, J. 2014. Sentiment classification at the time of the tunisian uprising: Machine learning techniques applied to a new corpus for arabic language. |
| 161 | Adeborna, E. and Siau, K., 2014. An Approach to Sentiment Analysis-the Case of Airline Quality Rating. In <i>PACIS</i> (p. 363). |
| 163 | Al-Radaideh, Q. A. and L. M. Twaiq. 2014. Rough set theory for arabic sentiment classification. |
| 165 | Isah, H., P. Trundle, and D. Neagu. 2014. Social media analysis for product safety using text mining and sentiment analysis. |
| 168 | Hasan, K. M. A., M. Rahman, and Badiuzzaman. 2014. Sentiment detection from bangla text using contextual valency analysis. |
| 174 | Hassan, A., A. Abu-Jbara, W. Lu, and D. Radev. 2014. A random walk-based model for identifying semantic orientation. <i>Computational Linguistics</i> 40, no. 3: 539-562. |
| 176 | Neuenschwander, B., A. Pereira, W. Meira, and D. Barbosa. 2014. Sentiment analysis for streams of web data: A case study of brazilian financial markets. |
| 189 | Al-Osaimi, S. and K. M. Badruddin. 2014. Role of emotion icons in sentiment classification of arabic tweets. |
| 191 | Colace, F., Casaburi, L., De Santo, M. and Greco, L., 2015. Sentiment detection in social networks and in collaborative learning environments. <i>Computers in Human Behavior</i> , 51, pp.1061-1067. |
| 202 | Mosquera, A. and P. Moreda. 2013. Improving Web 2.0 opinion mining systems using text normalisation techniques. |
| 203 | Troussas, C., M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro. 2013. Sentiment analysis of facebook statuses using naive bayes classifier for language learning. |
| 206 | Akaichi, J. 2013. Social networks "facebook" statuses updates mining for sentiment classification. |
| 210 | Mazzonello, V., S. Gaglio, A. Augello, and G. Pilato. 2013. A study on classification methods applied to sentiment analysis. |
| 214 | Park, E. H., V. C. Storey, and S. Givens. 2013. An ontology artifact for information systems sentiment analysis. |
| 218 | Akaichi, J., Z. Dhouioui, and M. J. Lopez-Huertas Perez. 2013. Text mining facebook status updates for sentiment classification. |
| 221 | Wen, B., P. Fan, W. Dai, and L. Ding. 2013. Research on analyzing sentiment of texts based on semantic comprehension. |
| 224 | Poria, S., A. Gelbukh, B. Agarwal, E. Cambria, and N. Howard. 2013. <i>Common sense knowledge based personality recognition from text</i> . Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 8266 LNAI. |
| 231 | Haddi, E., X. Liu, and Y. Shi. 2013. The role of text pre-processing in sentiment analysis. |
| 234 | Bahrainian, S. -A and A. Dengel. 2013. Sentiment analysis using sentiment features. |
| 237 | Puteh, M., N. Isa, S. Puteh, and N. A. Redzuan. 2013. Sentiment mining of malay newspaper (SAMNews) using artificial immune system. |
| 241 | Kaur, A. and V. Gupta. 2013. A survey on sentiment analysis and opinion mining techniques. <i>Journal of Emerging Technologies in Web Intelligence</i> 5, |

| | |
|-----|---|
| | no. 4: 367-371, |
| 242 | Zhou, X., X. Tao, J. Yong, and Z. Yang. 2013. Sentiment analysis on tweets for social events. |
| 243 | Huangfu, L., Mao, W., Zeng, D. and Wang, L., 2013, June. OCC model-based emotion extraction from online reviews. In <i>Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on</i> (pp. 116-121). IEEE. |
| 246 | Robaldo, L. and L. Di Caro. 2013. Opinion Mining-ML. <i>Computer Standards and Interfaces</i> 35, no. 5: 454-469. |
| 250 | Karamibekr, M. and A. A. Ghorbani. 2013. Sentiment analysis of social issues. |
| 251 | Ahn, H. J. 2013. Mining texts to understand customers' image of brands. <i>International Journal of Electronic Commerce Studies</i> 4, no. 1: 131-134. |
| 252 | Mostafa, M. M. 2013. More than words: Social networks' text mining for consumer brand sentiments. <i>Expert Systems with Applications</i> 40, no. 10: 4241-4251. |
| 253 | Nithish, R., Sabarish, S., Kishen, M.N., Abirami, A.M. and Askarunisa, A., 2013, December. An Ontology based Sentiment Analysis for mobile products using tweets. In <i>Advanced Computing (ICoAC), 2013 Fifth International Conference on</i> (pp. 342-347). IEEE. |
| 256 | Usha, M. S. and M. Indra Devi. 2013. Analysis of sentiments using unsupervised learning techniques. |
| 257 | Mouthami, K., K. N. Devi, and V. M. Bhaskaran. 2013. Sentiment analysis and classification based on textual reviews. |
| 277 | Jotheeswaran, J. and Y. S. Kumaraswamy. 2013. Opinion mining using decision tree based feature selection through manhattan hierarchical cluster measure. <i>Journal of Theoretical and Applied Information Technology</i> 58, no. 1: 72-80. |
| 281 | Mostafa, M. M. 2013. An emotional polarity analysis of consumers' airline service tweets. <i>Social Network Analysis and Mining</i> 3, no. 3: 635-649. |
| 282 | Bravo-Marquez, F., Mendoza, M. and Poblete, B., 2013, August. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In <i>Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining</i> (p. 2). ACM. |
| 284 | Gupta, N., K. R. Abhinav, and A. Annappa. 2013. Fuzzy sentiment analysis on microblogs for movie revenue prediction. |
| 285 | KUMAR, R.S. and VIJAYAN, S., 2013. MINING MOVIE REVIEWS--AN EVALUATION. <i>Journal of Theoretical & Applied Information Technology</i> , 56(2). |
| 288 | Farhadloo, M. and E. Rolland. 2013. Multi-class sentiment analysis with clustering and score representation. |
| 290 | Maria, M. and N. Makoto. 2013. On top of the world, down in the dumps: Text mining the emotionality of online consumer reviews. |
| 302 | Ginsca, A. L. 2012. Fine-grained opinion mining as a relation classification problem. |
| 306 | Li, C. and J. Ma. 2012. Research on online education teacher evaluation model based on opinion mining. |
| 308 | Martín-Valdivia, M. -T, A. Montejo-Ráez, A. Ureña-López, and M. R. Saleh. 2012. Learning to classify neutral examples from positive and negative |

| | |
|-----|---|
| | opinions. <i>Journal of Universal Computer Science</i> 18, no. 16: 2319-2333. |
| 310 | Bak, J., S. Kim, and A. Oh. 2012. Self-disclosure and relationship strength in twitter conversations. |
| 317 | Wald, R., T. Khoshgoftaar, and C. Sumner. 2012. Machine prediction of personality from facebook profiles. |
| 320 | Maks, I. and P. Vossen. 2012. A lexicon model for deep sentiment analysis and opinion mining applications. <i>Decision Support Systems</i> 53, no. 4: 680-688. |
| 334 | Naradhipa, A. R. and A. Purwarianti. 2012. Sentiment classification for indonesian message in social media. |
| 337 | Preethi, T., K. Nirmala Devi, and V. Murali Bhaskaran. 2012. A semantic enhanced approach for online hotspot forums detection. |
| 342 | Zhang, D., L. Si, and V. J. Rego. 2012. Sentiment detection with auxiliary data. <i>Information Retrieval</i> 15, no. 3-4: 373-390. |
| 353 | Costa, E., R. Ferreira, P. Brito, I. I. Bittencourt, O. Holanda, A. Machado, and T. Marinho. 2012. A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis. <i>Expert Systems with Applications</i> 39, no. 5: 4813-4834. |
| 357 | Leong, C. K., Y. H. Lee, and W. K. Mak. 2012. Mining sentiments in SMS texts for teaching evaluation. <i>Expert Systems with Applications</i> 39, no. 3: 2584-2589. |
| 375 | Loureiro, D., G. Marreiros, and J. Neves. 2011. <i>Sentiment analysis of news titles: The role of entities and a new affective lexicon</i> . Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 7026 LNAI. |
| 381 | Singh, V. K., M. Mukherjee, and G. K. Mehta. 2011. <i>Sentiment and mood analysis of weblogs using POS tagging based approach</i> . Communications in computer and information science. Vol. 168 CCIS. |
| 384 | Delmonte, R. and V. Pallotta. 2011. <i>Opinion mining and sentiment analysis need text understanding</i> . Studies in computational intelligence. Vol. 361. |
| 389 | Martínez-Cámara, E., M. T. Martín-Valdivia, and L. A. Ureña-López. 2011. <i>Opinion classification techniques applied to a spanish corpus</i> . Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 6716 LNCS. |
| 406 | Manuel, K., K. V. Indukuri, and P. R. Krishna. 2010. Analyzing internet slang for sentiment mining. |
| 412 | Whitehead, M. and L. Yaeger. 2010. Sentiment mining using ensemble classification models. |
| 423 | Xia, H., M. Tao, and Y. Wang. 2010. Sentiment text classification of customers reviews on the web based on SVM. |
| 432 | Gobron, S., J. Ahn, G. Paltoglou, M. Thelwall, and D. Thalmann. 2010. From sentence to emotion: A real-time three-dimensional graphics metaphor of emotions extracted from text. <i>Visual Computer</i> 26, no. 6-8: 505-519, |
| 434 | Shein, K. P. P. and T. T. S. Nyunt. 2010. Sentiment classification based on ontology and SVM classifier. |
| 444 | Cheng, M., Y. Xin, J. Bao, C. Wang, and Y. Yang. 2009. A random walk method for sentiment classification. |
| 449 | Shein, K. P. P. 2009. Ontology based combined approach for sentiment classification. |

| | |
|-----|--|
| 452 | Xia, Z., W. Suzhen, X. Mingzhu, and Y. Yixin. 2009. Chinese text sentiment classification based on granule network. |
| 457 | Verma, R. 2009. <i>Extraction and classification of emotions for business research</i> . Communications in computer and information science. Vol. 31. |
| 460 | Verma, R. 2009. <i>Data quality issues and dual purpose lexicon construction for mining emotions</i> . Lecture notes in business information processing. Vol. 37 LNBIP. |
| 461 | Boiy, E. and M. -F Moens. 2009. A machine learning approach to sentiment analysis in multilingual web texts. <i>Information Retrieval</i> 12, no. 5: 526-558, |
| 470 | Funk, A., Y. Li, H. Saggion, K. Bontcheva, and C. Leibold. 2008. Opinion analysis for business intelligence applications. |
| 477 | Zhang, C., D. Zeng, Q. Xu, X. Xin, W. Mao, and F. -Y Wang. 2008. <i>Polarity classification of public health opinions in chinese</i> . Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 5075 LNCS. |
| 500 | Matsumoto, S., H. Takamura, and M. Okumura. 2005. <i>Sentiment classification using word sub-sequences and dependency sub-trees</i> . Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 3518 LNAI. |

Apêndice B – Termos de Consentimento Livre, Esclarecido e Condições de Uso

SOBRE ASPECTOS CIENTÍFICOS

SUJEITO DA PESQUISA: CONFIDENCIALIDADE DE PARTICIPAÇÃO

Ao se cadastrar neste software você estará participando, como voluntário e de forma confidencial e implícita, de pesquisas científicas que envolvem a plataforma *Personalitatem*. As pesquisas serão desenvolvidas pela Universidade Federal de Sergipe e poderão contar com a colaboração de outras instituições. A sua participação nas pesquisas ocorrerá mediante o uso dos seus dados e atividades no software, cf. Resolução 466/12, III. 2.i. “As pesquisas preveem procedimentos que assegurem a confidencialidade e a privacidade, a proteção da imagem e a não estigmatização dos participantes da pesquisa, garantindo a não utilização das informações em prejuízo das pessoas e/ou das comunidades, inclusive em termos de autoestima, de prestígio e/ou de aspectos econômico-financeiros” (Resolução, 466/12, III. 2.i). As pesquisas conduzidas poderão ser de qualquer natureza, inclusive validação de questionários de personalidade e inferência das características pessoais por meio de questionários e mineração de texto usando dicionários afetivos propostos pela equipe ou por parceiros, sem que você seja notificado. Salienta-se que tais pesquisas trarão contribuições significativas nas áreas de Ciência da Computação, tais como Computação Afetiva e Linguística Computacional.

GARANTIA: Os dados utilizados na pesquisa não serão vinculados a nomes e/ou identidades pessoais e serão usados somente e exclusivamente para fins de pesquisa.

SOBRE ASPECTOS DO APLICATIVO

DADOS

A equipe responsável por este software e/ou parceiros não são responsáveis pelos dados inseridos pelos usuários, nem pela sua veracidade, incluindo informações pessoais de terceiros ou conteúdo protegido por direitos autorais.

Ao se cadastrar, você declara que é o responsável por todas as informações que insere no software, incluindo informações pessoais, e-mail, foto, textos, informações do perfil, e respostas às questões dos questionários e garante a sua veracidade.

Ao se cadastrar, você declara que não irá inserir indevidamente, no software conteúdo protegido por direitos autorais de qualquer natureza.

Ao se cadastrar, você declara em não usar informações de terceiros nem se passar por terceiro durante a inserção de dados no software.

PRIVACIDADE

A equipe responsável por este software e/ou parceiros dão “garantia de manutenção do sigilo e da privacidade dos participantes da pesquisa durante todas as fases da pesquisa” (Resolução 466/12 IV.3.e). Por esta razão, este software possui algumas ferramentas que auxiliam no controle de privacidade. A equipe responsável por este software e/ou parceiros não são responsáveis pela forma como os usuários fazem uso das ferramentas de privacidade, não se responsabilizando em caso de mal-uso, ou acesso indevido aos dados em consequência de uso inadequado das ferramentas.

Para fins de entendimento, seguem os significados de alguns termos presentes nesta seção:

Perfil invisível: Perfil que, por padrão, pode ser visualizado apenas pelo proprietário. Todos os usuários cadastrados no software poderão encontrar o perfil, desde que saibam o nome de usuário do proprietário, mas não poderão visualizar os dados.

Perfil visível: Perfil que pode ser visualizado por todos os usuários cadastrados no software, desde que saibam o nome de usuário do proprietário.

Código secreto: Código que permite o acesso ao perfil de um usuário, mesmo que esteja invisível. Não é necessário saber o nome de usuário para acessar um perfil através do código secreto.

Através de todos estes recursos, é necessário possuir uma conta e entrar no software para poder visualizar o perfil de outro usuário.

Todos os usuários cadastrados neste software possuem um perfil definido como “invisível”, por padrão. A partir do momento em que o seu perfil passa a ser “visível”, você declara que é o responsável por esta alteração. A equipe responsável por este software e/ou parceiros não são responsáveis pela divulgação de nomes de usuário dos usuários deste software.

Todos os usuários cadastrados neste software podem criar um código secreto para definir acesso seletivo ao perfil. Apenas um único código é válido por vez para um único usuário. Não é possível que dois usuários diferentes possuam um mesmo código secreto, e não é possível que um mesmo código seja gerado mais de uma vez, ou seja, cada código gerado é único. Um código pode ser gerado ou removido a qualquer momento através das configurações de privacidade. Nenhum código secreto é criado automaticamente para nenhum usuário cadastrado neste software. A partir do momento que um código secreto é criado na sua conta, você declara que é o responsável por esta alteração. A equipe responsável por este software e/ou parceiros não são responsáveis pela divulgação de códigos secretos dos usuários deste software.

Para sua segurança e comodidade, você poderá receber e-mails sobre as suas atividades neste software.

SEGURANÇA

O acesso a este software é controlado através de e-mail/nome de usuário e senha. A equipe responsável por este software e/ou parceiros não são responsáveis pela forma como os usuários fazem uso do e-mail e/ou nome de usuário e/ou senha. A equipe responsável por este software e/ou parceiros não são responsáveis pela divulgação de e-mails e/ou nome de usuário e/ou senhas dos usuários deste software.

CONSIDERAÇÕES

Ao se cadastrar, você declara que leu e aceita todos os termos citados neste documento. Você tem o direito de retirar seu consentimento depois de cadastrado, mas deve demonstrar isso mediante a exclusão da sua conta através do próprio software. Enquanto estiver cadastrado, você declara que aceita todos os termos citados neste documento, mesmo que o uso do software esteja interrompido.

Para mais informações sobre a equipe, acesse personalitresearch.ufs.br, personalitresearch.com.br ou <http://200.17.141.213/~gutanunes>.

Apêndice C - Instruções para responder os questionários de pesquisa

Serão apresentadas frases que descrevem o comportamento das pessoas. Utilize a escala de classificação abaixo de cada frase para descrever a precisão com que cada frase descreve você.

Descreva-se como você geralmente se sente agora, não como você gostaria de ser no futuro. Descreva-se como você honestamente se vê, em relação a outras pessoas que você conhece do mesmo sexo que você e mais ou menos de sua mesma idade.

Para que você possa descrever a si mesmo de uma forma honesta, suas respostas serão mantidas em absoluto sigilo. Por favor, leia cuidadosamente cada item, e, em seguida, clique na opção que corresponde exatamente a como você é.

Responda a cada item. Você não poderá visualizar um novo item antes de responder o anterior. Note que as opções de respostas aparecem diretamente abaixo de cada pergunta. Por favor, certifique-se de que a opção que você está escolhendo corresponde à sua resposta para o item que você está considerando. Você não poderá mudar sua resposta depois de clicar em uma das opções.

Para responder um item, basta escolher uma das opções que serão apresentadas em seguida, e um novo item será carregado automaticamente até que todos os itens sejam respondidos.

Você pode responder apenas UMA VEZ cada item.

Todas as respostas a este questionário de todos os entrevistados são totalmente confidenciais e não serão associados a você como um indivíduo. As respostas entrarão automaticamente em um banco de dados a fim de melhorar as normas, por idade e sexo e para avaliar as propriedades estatísticas de respostas de itens para grupos de pesquisados.

Apêndice D - Questões IPIP-NEO 120

1. "Preocupo-me com as coisas."
2. "Faço amigos facilmente."
3. "Tenho uma imaginação vívida."
4. "Confio nos outros."
5. "Completo tarefas com sucesso."
6. "Fico com raiva facilmente."
7. "Adoro festas com muitas pessoas."
8. "Acredito na importância da arte."
9. "Nunca sonitaria impostos."
10. "Gosto de ordem"
11. "Frequentemente me sinto triste."
12. "Assumo o comando das situações"
13. "Vivo minhas emoções intensamente."
14. "Faço as pessoas se sentirem bem vindas"
15. "Tento obedecer as regras"
16. "Sou intimidado facilmente."
17. "Estou sempre ocupado."
18. "Prefiro variedade à rotina."
19. "Sou fácil de satisfazer."
20. "Vou direto ao objetivo"
21. "Frequentemente como demasiadamente."
22. "Adoro adrenalina."
23. "Gosto de solucionar problemas complexos."
24. "Detesto ser o centro das atenções."
25. "Faço minhas tarefas o mais rápido possível"
26. "Entro em pânico com facilidade."
27. "Irradio alegria."
28. "Tendo a votar em políticos de esquerda."
29. "Tenho compaixão pelos desabrigados."
30. "Evito cometer erros."
31. "Tenho medo do pior"
32. "Aproximo-me das pessoas com facilidade"
33. "Curto altos vãos na minha imaginação."
34. "Acredito que os outros têm boas intenções."
35. "Sobressaio nas coisas que faço."
36. "Irrito-me facilmente."
37. "Converso com diversas pessoas em festas."
38. "Gosto de música."
39. "Sigo as regras."
40. "Gosto de arrumar as coisas."
41. "Não gosto de mim mesma."
41. "Não gosto de mim mesmo."
42. "Tento liderar os outros."
43. "Sinto as emoções dos outros."
44. "Antecipo as necessidades dos outros."
45. "Matenho as minhas promessas."
46. "Tenho medo de fazer a coisa errada."

47. "Estou sempre ativa."
47. "Estou sempre ativo."
48. "Gosto de conhecer lugares novos."
49. "Não suporto confrontos."
50. "Trabalho duro."
51. "Não sei porque faço algumas das coisas que faço."
52. "Busco aventura."
53. "Adoro ler coisas que me desafiam."
54. "Não gosto de falar sobre mim mesma."
54. "Não gosto de falar sobre mim mesmo."
55. "Estou sempre preparada."
55. "Estou sempre preparado."
56. "Muitas vezes me sinto sobrecarregada."
56. "Muitas vezes me sinto sobrecarregado."
57. "Divirto-me bastante."
58. "Acredito que não existe verdade absoluta."
59. "Sinto compaixão por aqueles menos abastados que eu."
60. "Escolho minhas palavras com cuidado."
61. "Tenho medo de muitas coisas."
62. "Sinto-me à vontade perto das pessoas."
63. "Amo sonhar acordada."
63. "Amo sonhar acordado."
64. "Confio no que as pessoas falam."
65. "Lido com minhas tarefas tranquilamente."
66. "Aborreço-me facilmente."
67. "Gosto de fazer parte de um grupo."
68. "Vejo beleza em coisas que outros podem não notar."
69. "Uso de bajulação para avançar."
70. "Quero que tudo esteja perfeito."
71. "Frequentemente me sinto um lixo."
72. "Convenço pessoas a agirem."
73. "Sou apaixonada por causas."
73. "Sou apaixonado por causas."
74. "Adoro ajudar o próximo."
75. "Pago minhas contas em dia."
76. "Tenho dificuldade de me aproximar das pessoas."
77. "Faço diversas coisas no meu tempo livre."
78. "Interesso-me por muitas coisas"
79. "Odeio parecer muito controlador ou exigente."
80. "Transformo planos em ações."
81. "Faço coisas de que me arrependo posteriormente."
82. "Adoro ação."
83. "Tenho um vocabulário rico."
84. "Considero-me uma pessoa comum."
85. "Inicio meus trabalhos o mais rápido possível."
86. "Sinto que sou incapaz de lidar com as situações."
87. "Expresso alegria como uma criança."
88. "Acredito que criminosos deveriam receber ajuda ao invés de punição."
89. "Valorizo mais cooperação do que competição."
90. "Sigo no caminho que escolho."

91. "Estresso-me facilmente."
92. "Ajo confortavelmente perto de outras pessoas."
93. "Gosto de me perder dos meus pensamentos."
94. "Acredito que as pessoas são essencialmente boas."
95. "Sei da minha capacidade."
96. "Estou frequentemente de mau humor."
97. "Envolver outras pessoas no que estou fazendo."
98. "Amo flores."
99. "Uso outras pessoas para conseguir meus objetivos."
100. "Gosto de ordem e harmonia."
101. "Tenho uma opinião ruim sobre mim mesma."
101. "Tenho uma opinião ruim sobre mim mesmo."
102. "Procuo influenciar outros."
103. "Gosto de analisar a mim mesma e minha vida."
103. "Gosto de analisar a mim mesmo e minha vida."
104. "Preocupo-me com os outros."
105. "Falo a verdade."
106. "Tenho medo de chamar atenção."
107. "Consigo fazer muitas coisas ao mesmo tempo."
108. "Gosto de iniciar coisas novas."
109. "Tenho uma língua afiada."
110. "Mergulho de coração nas minhas tarefas."
111. "Gosto de farras."
112. "Gosto de fazer parte de multidões barulhentas."
113. "Consigo lidar com muitas informações."
114. "Raramente conto vantagem."
115. "Começo logo a trabalhar"
116. "Não consigo me decidir."
117. "Estou sempre de bem com a vida."
118. "Acredito numa única religião verdadeira."
119. "Sofro com as perdas dos outros."
120. "Faço coisas sem pensar."

Apêndice E - Questões TIPI

"Eu me vejo como um homem" para o sexo masculino

ou

"Eu me vejo como uma mulher" para o sexo feminino

1. "Extrovertida, entusiasta."
1. "Extrovertido, entusiasta."
2. "Crítica, conflituosa."
2. "Crítico, conflituoso."
3. "De confiança, com auto-disciplina."
4. "Ansiosa, que se chateia/aborreça facilmente."
4. "Ansioso, que se chateia/aborreça facilmente."
5. "Aberta a experiências novas, complexa/difícil/complicada."
5. "Aberto a experiências novas, complexo/difícil/complicado."
6. "Reservada, calada."
6. "Reservado, calado."
7. "Compreensiva/solidária, afetuosa."
7. "Compreensivo/solidário, afetuoso."
8. "Desorganizada, descuidada."
8. "Desorganizado, descuidado."
9. "Calma, emocionalmente estável."
9. "Calmo, emocionalmente estável."
10. "Convencional, pouco criativa."
10. "Convencional, pouco criativo."