

INFERINDO PERSONALIDADE VIA TWEETS

EXTRACTING PERSONALITY FROM TWEETS

Maria Augusta Silveira Netto Nunes¹; Fernando Rodrigues Teles²; Janisson Góis de Souza³;

¹Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil

gutanunes@gmail.com

²Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil

janissongois@gmail.com

³Universidade Federal de Sergipe – UFS – São Cristóvão/SE – Brasil

nanoteles@gmail.com

RESUMO

Esse artigo propõe uma alternativa de direcionamento das pesquisas na área de Computação Afetiva no que tange a inferência de personalidade humana de forma não intrusiva e implícita. O uso de personalidade é importante para auxiliar a tomada de decisão computacional visando um melhor desempenho na personalização de informações, produtos e serviços aos clientes na *World Wide Web*, por exemplo. Nesse contexto, o artigo objetiva apresentar os resultados de um experimento que busca algum indício que possibilite provar a correlação existente entre ferramentas tradicionais, na língua portuguesa, que inferem personalidade explicitamente por meio de inventários e uma ferramenta baseada em *text-mining*, que infere personalidade de maneira implícita via texto. Os resultados encontrados no experimento foram promissores e incitam outras pesquisas na área.

Palavras-Chave: Inferência, Personalidade, *text-mining*, traços, *PersonalityML*

ABSTRACT

This paper researches an alternative on Affective Computing area towards of how to extract the human personality in an implicit and non-intrusive way. The human personality is important in order to improve the computational decision making during the personalization process on the World Wide Web (by means of information, products or services). In addition, it presents the preliminary results of an experiment that shows the correlation between Personality-based tools in

Portuguese that infer Personality explicitly by using inventories and implicitly by using text-mining. The results of the experiment were promising and encourage further research in this field.

Keywords: Inference, Personality, *text-mining*, *PersonalityML*

1. INTRODUÇÃO

Vários estudos tem comprovado a importância de informações psico-afetivas e sutis, tais como personalidade e emoção, na tomada de decisão humana e, conseqüentemente, computacional, para uma melhor desempenho na personalização de informações, produtos e serviços aos clientes na *World Wide Web*. Para personalizar a oferta e recomendar bons produtos é necessário que o computador conheça quem é o cliente, em uma maior amplitude possível, incluindo aspectos psico-afetivos. Um problema enfrentado pelos cientistas de Computação Afetiva é a dificuldade de inferir a personalidade humana de forma não intrusiva e não implícita (NUNES e HU, 2012; NUNES e CAZELLA, 2011). Alguns trabalhos com uma abordagem não intrusiva tem sido propostos por Mairesse et al. (2007) e Celli (2012), entre outros, que usaram perspectivas baseadas em text-mining. Já com outros avanços significativos em final de 2012 e 2013, Quercia *et al.* (2011; 2012), entre outros, tem proposto estudos usando informações de bases de dados proveniente de redes sociais (*Twitter* e *Facebook*) tentando replicar e expandir resultados já existentes no estado da arte (usando a língua inglesa como base). Celli et al. (2013) têm demandado à comunidade científica avanços visando elevar a desempenho algorítmico desses experimentos baseados em personalidade, propondo, assim, o uso de um *benchmark* comum. Apesar das diversas propostas evoluírem o estado da arte e técnica da área no mundo, as mesmas apresentam limitações, principalmente, ligadas à língua base onde a abordagem é aplicada, que nesse caso é a língua inglesa. No caso da língua portuguesa, não existem indícios de estudos computacionais com o objetivo de inferir a personalidade humana, limitando, consideravelmente, e conseqüentemente seu uso e aplicação para o público brasileiro.

Esse artigo propõe uma alternativa de direcionamento das pesquisas nessa área apresentado alguns resultados preliminares na tentativa de diminuir a lacuna encontrada no estado da arte e técnica existente. Dessa forma, esse artigo tem como objetivo apresentar os resultados de um experimento que busca algum indicio que possibilite provar a correlação existente entre

ferramentas tradicionais, na língua portuguesa, que inferem personalidade explicitamente por meio de inventários e uma ferramenta baseada em *text-mining*, que infere personalidade de maneira implícita.

O artigo está organizado como segue: na seção 2 apresenta-se uma breve introdução sobre personalidade, seguido, na seção 3 da metodologia usada no artigo. Na seção 4, descreve-se as etapas do experimento apresentando os protótipos computacionais criados. Na seção 5, apresenta-se a análise dos resultados, seguido pelas conclusões, agradecimentos e referências.

2. PERSONALIDADE

Na Psicologia não existe um consenso para a definição de personalidade. Schultz (1990) define personalidade como “um conjunto permanente e exclusivo de características identificáveis nas ações/interações do indivíduo em diferentes situações”. A personalidade humana pode ser definida segundo diversas abordagens. Uma abordagem bastante interessante é a abordagem de traços de personalidade que permite diferenciar psicologicamente pessoas usando traços mensuráveis, conceituáveis e implementável em computadores (NUNES e HU, 2012; NUNES e CAZELLA, 2011). Os traços de personalidade foram historicamente definidos por Allport, que criou 17.953 traços para descrever a personalidade de um indivíduo. Posteriormente outros psicólogos reduziram mais de 99% desses traços (Schultz, 1990). Dessa redução acabaram restando cinco fatores que se replicaram na maioria dos estudos empíricos desses psicólogos, como resultado, o modelo *Big Five* foi criado (JOHN e SRIVASTAVA, 1999).

2.1 Extração de personalidade

Considerando a abordagem de traços, escolhida por ser a melhor forma para representação de personalidade em computadores, psicólogos geralmente usam questionários intitulados de inventários de personalidade. Esses inventários podem ter uma pequena ou grande quantidade de questões. O número de questões é diretamente proporcional a granularidade e precisão dos traços de personalidade extraídos do usuário. Segundo Gosling *et al.* (2003), os longos e mais precisos inventários tomam um tempo bastante considerável do usuário e muitas vezes torna-se inviável sua aplicação. Ainda, via *World Wide Web*, em sites de apelo comercial, dificilmente usuários tem disponibilidade e paciência para responder explicitamente à qualquer questionário/inventário, dessa

forma o uso de outra abordagem diferente dessa seria preferível. É essa abordagem que tem motivado esse experimento apresentado a seguir.

3. METODOLOGIA

3.1 Materiais

A metodologia usada nesse experimento foi fragmentada em duas etapas: (1) A primeira etapa consistiu em coletar textos/*tweets* públicos de usuários do *Twitter* (participantes do experimento) e em paralelo construir uma ferramenta em português que possibilitasse a mineração desses textos inferindo traços de personalidade dos participantes por meio das pistas deixadas no texto; (2) A segunda etapa consistiu em coletar os traços de personalidade desses mesmos participantes usando dois inventários explícitos, tradicionais e reputados para a extração de personalidade (a versão dos inventários utilizada foi em língua portuguesa).

3.2 Procedimento

Para inferir a personalidade de cada participante foram implementadas 2 diferentes plataformas experimentais: (1) uma a ser usada *online*, para inferência implícita via *text-mining* (que foi usada pelos proponentes do experimento após coleta do conjunto de *tweets* de cada usuário. Poderia também ser utilizada pelos participantes, caso os mesmos quisessem obter o resultado específico relacionado a um extrato de texto somente. Visando facilitar e agilizar o experimento essa etapa foi realizada pelos proponentes); (2) outra *off-line* – *android-based* para inferência explícita via questionário/inventário (foi utilizada e preenchida pelo próprio usuário). Portanto, os participantes podiam facilmente participar desse estudo, em qualquer lugar e horário que lhes fosse adequado e disponível. As duas plataformas disponibilizadas continham instruções de uso, onde os participantes facilmente puderam seguir as instruções passo-a passo fornecidas. Os participantes podiam finalizar o experimento a qualquer momento e retornar posteriormente a fim de concluí-lo. Os dois experimentos permitiam a geração de uma memória *XML-based*.

3.3 Participantes

Foram convidadas aproximadamente cem de pessoas (colegas de trabalho, de universidade, de curso, amigos, familiares). Os meios utilizados para divulgação do experimento foram: lista de discussão de alunos, contatos pessoais e redes sociais. Todos os participantes forneceram seu

acordo e seu *username* do *Twitter*. Participaram do estudo, aproximadamente, 3 dezenas de pessoas. Destas pessoas, foram filtrados e selecionados 28 participantes. Esses 28 participantes cumpriram todas as etapas relacionadas à coleta explícita de dados. Os 28 participantes foram aprovados na etapa de filtragem de dados. Os dados coletados explicitamente sofreram validação psicométrica segundo estudos de Gosling e Johnson (2010). Foi constatado também que os 28 participantes tinham contas ativas e extratos de dados textuais/*tweets* públicos que permitiam a extração e análise textual de conteúdo. A amostra de participantes apresentou participantes do sexo feminino e masculino, com idades variadas. Os participantes eram todos brasileiros.

3.4 Hipótese

Os autores acreditam que algumas características de personalidade podem ser inferidas a partir de *tweets* públicos provenientes de perfis de usuários em redes sociais no contexto Brasileiro e na língua portuguesa.

4. EXPERIMENTO

4.1 Etapa 1 – Coleta implícita via Mineração de Texto

4.1.1 DataSet

O *dataset* foi composto por uma amostra dos *posts* públicos dos 28 participantes. Foi coletado, de cada voluntário um conjunto de *posts*. As amostras foram coletadas no mês dezembro de 2012. Os *posts* públicos (*tweets*) foram coletados no *Twitter* de forma explícita pelos pesquisadores (mesmo sendo públicos, a coleta e uso foi autorizada pelos participantes). A seguir os pesquisadores procederam a análise individual do conjunto de *posts* de cada participante aplicando o *PersonlityMiner*. Após a análise da ferramenta foi gerado o padrão de personalidade de cada participante considerando os 5 traços de personalidade relacionados ao *Big Five* (JOHN e SRIVASTAVA, 1999) e armazenados no *XML-based (PersonalityML)* (NUNES *et al.*, 2012). Os resultados foram mantidos em sigilo e a identidade do participante não foi divulgada em nenhuma hipótese.

4.1.2 Ferramenta de Inferência –*PersonalityMiner*

A mineração de texto (*text-mining*) é um processo de descoberta de conhecimento a partir de bases não estruturadas (FELDMAN e DAGAN, 1995), ou seja, difere da mineração de dados por ter uma etapa de pré-processamento, pois é nela que se define a estrutura a ser utilizada pelos algoritmos de mineração. As etapas da mineração de textos são: seleção, pré-processamento, mineração e interpretação.

Na etapa de seleção são escolhidos os textos dos usuários que serão analisados pelo minerador e posteriormente submetidos à etapa de pré-processamento. Nesta etapa, os textos são transformados em vetores de palavras e suas frequências são sumarizadas seguindo algumas das técnicas de redução de dimensionalidade como, por exemplo, o *stemming*, para obter melhor desempenho (ORENGO e HUYCK, 2001).

O *stemming* é um processo que consiste em reduzir a palavra ao seu radical. Essa redução ajuda a diminuir a variação gramatical das palavras, como gênero e número, principalmente em línguas flexíveis como a língua portuguesa (ORENGO e HUYCK, 2001). Para a língua portuguesa destacam-se dois algoritmos de *Stemmer*: Porter (1980) e Orenge e Huyck (2001). O primeiro apenas retira os sufixos das palavras, enquanto o segundo possui uma base de exceções para reduzir os casos conhecidos como o *understemming* que ocorre quando é removido parte do radical, por exemplo, o termo “gramática” após ser processado, resulta no radical “gramá-“, ao invés de “gramát-“. Existe também o *overstemming* que consiste na não remoção correta do sufixo, por exemplo, o termo “referência” após ser processado, resulta no radical “referên-“ ao invés de “refer-“. Em seguida são sumarizadas as frequências dos termos, separando-os em categorias orientadas pelo dicionário descrito a seguir.

4.1.2.1 Dicionário

Para auxiliar na sumarização dos termos durante o pré-processamento, é recomendado a utilização de um dicionário. Nele é possível definir as categorias utilizadas para rápida classificação e sumarização dos termos. No *PersonalityMiner* foi utilizado como dicionário base o TeP 2.0 (*Thesaurus* Eletrônico para o Português do Brasil) (MAZIERO *et al.* 2008), um dicionário criado e mantido pelo grupo de pesquisa do NILC (Núcleo Interinstitucional de Linguística Computacional) que propõe uma análise sintática das palavras. No entanto, o dicionário foi adaptado para o modelo LIWC, um modelo proposto por Pennebaker e King (1999) para contagem de palavras e, atualmente, está disponível em diversos idiomas, porém não há versão para o

português. Nele são separadas as categorias das palavras para diversos tipos de análises. Esta abordagem permite, facilmente, adicionar ou remover categorias de palavras, tornando, assim, o dicionário mais flexível. A Tabela 1 mostra um exemplo de alguns termos do dicionário:

Tabela 1. Exemplo de termos do dicionário

Dimensões	Exemplos de Palavras
Total de pronomes	Eu, nosso, eles, vocês
1ª pessoa singular	Eu, meu, mim
Negação	Não, nunca, nenhum
Preposições	Em, para, de
Números	Um, trinta, milhão
Processo Afetivo ou Emocional	Feliz, feio, amargo
Emoções Positivas	Feliz, bastante, bom
Sentimentos Positivos	Feliz, alegria, amor

Fonte: produzido pelos autores baseado em (PENNEBAKER e KING, 1999)

Após a realização da etapa de pré-processamento, segue-se para a etapa de mineração, na qual são realizados os cálculos baseados no trabalho do Mairesse et al. (2007), conforme descrito abaixo.

4.1.2.2 Mapeamento para o *Big Five*

Mairesse et al. (2007), realizou um estudo no qual tentou mapear a personalidade de alguns voluntários a partir de seus textos (em inglês) procurando padrões em correlação com as respostas de seus respectivos questionários. Nesse estudo, foram utilizados diversos tipos de ferramentas de análise textual como: LIWC e MRC. Mairesse et al. (2007) ainda apresenta, explicitamente, a expressão de mapeamento apenas do fator Extroversão (E) como sendo a soma das frequências dos termos da seguinte forma:

$$E=(\text{substantivo}+\text{adjetivo}+\text{preposição}+\text{artigo-pronomes-verbos-adverbios-interjeição}+100)/2$$

Seguindo uma lógica semelhante e baseando-se em padrões apontados por Mairesse propõe-se as seguintes expressões para cada um dos fatores: Amabilidade (A), Conscientização (C), Neurotismo (N), Abertura (O).

$$A=(\text{palavras positivas-palavras negativas-artigos}+100)/2$$

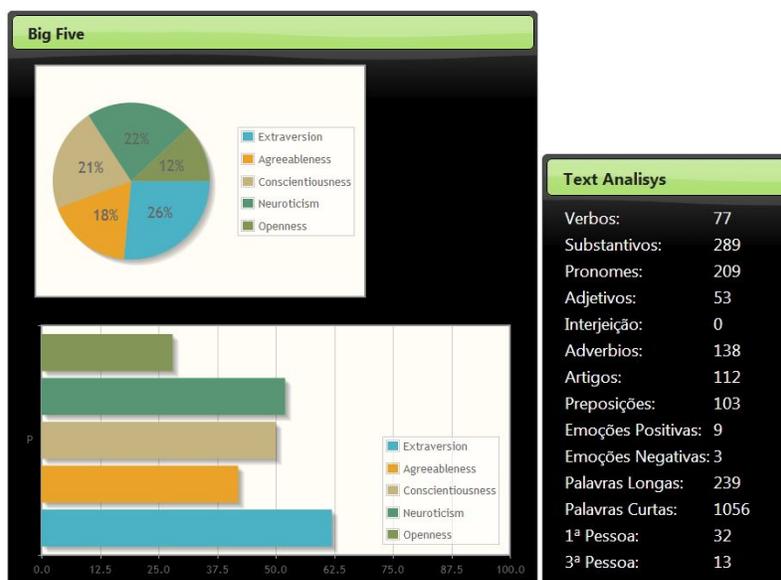
$$C=(\text{palavras positivas-negações-palavras negativas}+100)/2$$

$$N=(\text{palavras negativa-palavras positivas+primeira pessoa-terceira pessoa}+100)/2$$

$$O=(\text{palavras longas-palavras curtas}/2+\text{terceira pessoa-primeira pessoa}+100)/2$$

(E, A, C, N, O são os 5 fatores do *Big Five*). Na Figura 1, observe um exemplo da aplicação de textos através do *PersonalityMiner* PV1.0 (*Portuguese Version*).

Figura 1. Resultado do *PersonalityMiner*



Fonte: produzido pelos autores

Finalmente, os resultados obtidos pelo *PersonalityMiner* e gravados no *PersonalityML* (PML) (NUNES *et al.*, 2012) poderão ser analisados e interpretados por um analista do conhecimento ou submetidos a alguma outra ferramenta que manipule os dados gerados. No caso desse artigo será a busca em demonstrar a existência de correlação entre dois inventários que inferem personalidade de forma explícita e um protótipo experimental que infere personalidade de forma implícita, todos em língua portuguesa.

4.2 Etapa 2 – Coleta explícita via Inventários

4.2.1 DataSet

O *dataset* foi composto pelo conjunto de respostas dos 28 participantes (os mesmos que participaram da etapa 1 desse experimento) que responderam de forma voluntária aos inventários

de personalidade *TIPI* e *NEO-IPIP*, disponíveis na língua portuguesa no *Personality Inventory Mobile*¹. As amostras foram coletadas no mês de dezembro de 2012. O acesso ao *Personality Inventory Mobile* é realizada de forma gratuita mediante um cadastro. Durante o cadastro o participante aceita um termo de confidencialidade no qual a finalidade da pesquisa é exposta e garantida a manutenção das informações em sigilo absoluto.

4.2.2 Ferramenta de Inferência –*PersonalityMobile*

O *Personality Inventory Mobile (PV1.0 Portuguese Version)*¹ é um software que tem como objetivo centralizar os diversos inventários de personalidade existentes. Inicialmente foram escolhidos dois inventários para compor o software: *NEO-International Personality Item Pool (NEO-IPIP)* (JOHNSON, 2005) e o *Ten-Item Personality Inventory (TIPI)* (GOSLING *et al.*, 2003).

O *Personality Inventory Mobile* é uma evolução do software em versão *World Wide Web* desenvolvido em 2010. Ele foi desenvolvido na plataforma *android*, e está disponível nos idiomas Inglês e Português. Além disso, o *Personality Inventory Mobile* armazena as informações do indivíduo, tal como *username* da rede social *Twitter* (que permite-se obter e guardar posteriormente seus *tweets*, informações estas que foram usadas na etapa anterior apresentada). Após a realização dos inventários, os usuários tem acesso aos prognósticos do *Big Five*, descrevendo seu perfil de personalidade.

Outra funcionalidade importante é a exportação dos resultados em *PersonalityML (PML)*. O *PersonalityML* (NUNES *et al.*, 2012) é uma linguagem de marcação baseada em *XML* que foi desenvolvida para padronizar os resultados obtidos em inventários de personalidade possibilitando a geração de *cookies*, usados como *recommender inputs* em Sistemas de Recomendação, por exemplo. (Informações adicionais e *screenshots* do *Personality Inventory Mobile* e *PersonalityML* são encontrados no portal do projeto¹).

4.3 Resultados

Objetivando validar a hipótese buscou-se provar que algumas características de personalidade poderiam ser inferidas a partir de *tweets* públicos provenientes de perfis de usuários em redes sociais; após a etapa 1 e 2 do experimento os dados foram analisados buscando a

¹ <http://www.personalityresearch.com.br/>

correlação entre os padrões de personalidade dos participantes obtidos via *PersonalityMiner* em relação aos padrões obtidos via *PersonalityMobile* nos inventários *NEO-IPIP* e *TIPI*.

Os resultados obtidos foram bastante promissores, apresentado uma razoável correlação da etapa 1 (*text-mining* via *PersonalityMiner*) em relação aos resultados extraídos na etapa 2 (inventários via *Personality Inventory Mobile*). Observe, na Tabela 2, a correlação de *Pearson*² obtida em relação ao *NEO-IPIP*.

Tabela 2. Correlação entre *NEO-IPIP* & *PersonalityMiner*

Personality Miner/ NEO-IPIP	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Open-to-Experience
Extraversion	0,1267	0,2441	0,1464	0,0144	0,1055
Agreeableness	-0,0282	-0,1611	0,0706	-0,1274	-0,0382
Conscientiousness	-0,2260	-0,4360	-0,1417	0,4122	0,2076
Neuroticism	0,2056	0,4176	0,1723	-0,0314	0,1982
Open-to-Experience	-0,0734	0,0493	-0,0282	-0,1445	-0,0271

Fonte: produzido pelos autores

Observe, na Tabela 3, a correlação de *Pearson* obtida em relação ao *TIPI*.

Tabela 3. Correlação entre *TIPI* & *PersonalityMiner*

Personality Miner/ TIPI	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Open-to-Experience
Extraversion	0,2253	0,3535	0,1389	0,2593	0,2705
Agreeableness	-0,1545	-0,2559	-0,0203	-0,0555	-0,1806
Conscientiousness	-0,2431	-0,0613	-0,2262	0,0384	-0,3753
Neuroticism	0,4161	0,2849	0,2758	0,1722	0,2764
Open-to-Experience	0,0842	-0,2132	0,0397	-0,2206	-0,0627

Fonte: produzido pelos autores

Para medir a qualidade dos resultados obtidos, foi considerado nesse trabalho uma correlação positiva forte: os intervalos entre 0,70 e 1; moderada: entre 0,3 e 0,7; e fraca: entre 0 e 0,3. Enquanto a correlação negativa forte: entre -0,7 e - 1; moderada: -0,3 e - 0,7; e fraca: entre 0 e - 0,3. Alguns traços de personalidade obtiveram uma correlação cruzada superior a 0,4 chegando a

² http://pt.wikipedia.org/wiki/Coeficiente_de_correla%C3%A7%C3%A3o_de_Pearson

uma escala moderada, demonstrando uma possibilidade de boa correlação dos termos do dicionário com alguns traços do *Big Five* e/ou cruzamento de traços.

5. ANÁLISES E CONCLUSÕES

O artigo apresentou um experimento realizado em duas etapas que objetivou provar a correlação existente entre ferramentas tradicionais, na língua portuguesa, que inferem personalidade explicitamente por meio de inventários (*NEO-IPIP* e *TIPI*) e uma ferramenta baseada em *text-mining* (*PersonalityMiner*), que infere personalidade de maneira implícita.

O experimento permitiu demonstrar, para essa amostra, uma existência de correlação fraca entre o *PersonalityMiner* e *TIPI* no fator Extroversão e Neuroticismo; e a correlação fraca entre o *PersonalityMiner* e o *NEO-IPIP* no fator Extroversão. Além de algumas outras correlações cruzadas a serem estudadas e compreendidas. A correlação com a Extroversão já foi comprovada por Mairesse em experimento realizado na língua inglesa (2007).

Os resultados obtidos foram promissores pois mostraram algum tipo de correlação. Mesmo que a correlação encontrada tenha sido fraca a existência da mesma motiva os pesquisadores a ampliarem a amostra do experimento visando buscar melhores resultados e garantir uma escalonagem dos mesmos num futuro próximo. Visualizando sob esse aspecto, a hipótese proposta pelo artigo foi parcialmente alcançada, porém sem base estatística suficiente, não sendo possível replicar o caso. Portanto, o resultado parece promissor e os autores acreditam que em experimentos futuros com uma amostra adequada e com um dicionário melhorado algumas características de personalidade poderão ser inferidas a partir de *tweets* públicos provenientes de perfis de usuários em redes sociais. Outros experimentos com uma amostra maior e dicionário mais completo deve ser conduzido em breve.

6. AGRADECIMENTOS

Ao CNPq, à FAPITEC/SE e à Universidade Federal de Sergipe pelos fomentos durante as etapas dessa pesquisa experimental. Ao J. Johnson e S. Gosling por cederem as versões em inglês de seus inventários para a adaptação e validação em português.

7. REFERENCES

CELLI, F., PIANESI, F., STILLWELL, D.S., KOSINSKI, M. Workshop on Computational Personality Recognition. The 7th International AAAI Conference On Weblogs And Social Media. Boston. 2013.

CELLI, F., Unsupervised Personality Recognition for Social Network Sites. In: Proceedings of ICDS, pp. 59–62. 2012.

FELDMAN, R., AND DAGAN, I. Knowledge discovery in textual databases. In: Proceedings of KDD 1995, 112–117. AAAI Press. 1995.

GOSLING, S. D.; RENTFROW, P, J.; SWANN J. R., W. B. A very brief measure of the big-five personality domains. *Journal of Research in Personality*. Elsevier, (37):504–528, 2003.

GOSLING, S. D.; JOHNSON, J. A. Advanced methods for conducting online behavioral research. 1^a. ed. Washington, DC: American Psychological Association, 2010.

JOHN, OLIVER P. E SRIVASTAVA, SANJAY. (1999) The Big Five Trait Taxonomy: History, Measurement, And Theoretical Perspectives. In: *Handbook Of Personality: Theory And Research*. New York. p. 102–138. 1999.

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

MAIRESSE, F.; WALKER, M.; MEHL, M.; MOORE, R. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 1, 457-500. 2007.

MAZIERO, E. G.; PARDO, T. A. S.; FELIPPO, A. DI, AND DIAS-DA-SILVA, B. C.. A base de dados lexical e a interface web do TeP 2.0: thesaurus eletrônico para o Português do Brasil. In *Proceedings of (WebMedia '08)*. ACM, New York, NY, USA, 390-392. 2008.

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

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

NUNES, M.A.S.N. ; BEZERRA, J. S. ; OLIVEIRA, A. A. . PersonalityML: A Markup Language To Standardize The User Personality In Recommender Systems. *GEINTEC - Gestão, Inovação e Tecnologias*, v. 2, p. 255-273, 2012.

ORENGO, V. AND HUYCK, C.A. Stemming Algorithm for The Portuguese Language. In Proceedings of 8th Symposium on String Processing and Information Retrieval (SPIRE 2001),p.186-193, Chile, 2001.

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

PORTER, M. An Algorithm for Suffix Stripping. *Program*, 14(3), 130-137, 1980.

QUERCIA, D.; KOSINSKI, M.; STILLWELL, D.; CROWCROFT, J. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter.. In: *SocialCom/PASSAT : IEEE*, S. 180-185.2011

QUERCIA, D., LAMBIOTTEZ, R., STILLWELL, D., KOSINSKIY, M., CROWCROFT, J. The Personality of Popular Facebook Users. In *Proceedings of ACM CSCW 2012*. pp 1–10. 2012.

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