

## IV Workshop de Dissertações de Mestrado do PPgSI (2017)

### FICHA DA PESQUISA<sup>1</sup>

DADOS GERAIS				
<b>Título do projeto de pesquisa</b>	Determinação de conteúdo para geração de língua natural baseada em personalidade			
<b>Orientando</b>	Ricelli Moreira Silva Ramos			
<b>Orientador(es)</b>	Prof. Dr. Ivandré Paraboni			
<b>Semestre no curso, na data do workshop</b>	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input checked="" type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre
<b>Qualificação</b>	<input checked="" type="checkbox"/> Qualificação já realizada em: 05/07/2017 <input type="checkbox"/> Realização da qualificação planejada para: dd/mm/aaaa			
<b>Defesa</b>	Prazo máximo para depósito: 29/08/2018 Realização da defesa planejada para: 03/2018			
<b>Linha e Área de pesquisa</b>	Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software <input type="checkbox"/> IHC		Inteligência de Sistemas: <input checked="" type="checkbox"/> IA  <input type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico	
<b>Área de aplicação</b>	<input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Economia	<input type="checkbox"/> Educação <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Internet <input type="checkbox"/> Jogos	<input checked="" type="checkbox"/> Linguagem Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Química	<input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input type="checkbox"/> [outro – escrever]
<b>Publicações associadas ao projeto de mestrado</b>	Artigo submetido: O <i>corpus</i> b5 de textos e inventários de personalidade humana Autores: Ricelli Ramos, Georges Basile Stavracas Neto, Barbara Barbosa C. Da Silva, Danielle Monteiro, Fabio Lopes, Vitor Santos, Ivandré Paraboni Data de submissão: 25/mai/2017 Título da conferência: STIL – Symposium in Information and Human Language Technology Trilha da conferência: Principal H-Index Google Scholar (h5): Qualis b3			

DESCRIÇÃO DO PROJETO DE PESQUISA	
<b>Contextualização / motivação</b>	Observou-se que a literatura em Geração de Língua Natural (GLN) de modo geral não trata da questão de como a personalidade humana é refletida na determinação de conteúdo.
<b>Problema de pesquisa</b>	O projeto busca utilizar o modelo dos Cinco Grandes Fatores (CGF) de personalidade para que a geração de texto seja feita de forma mais próxima ao desempenho humano, na etapa de determinação de conteúdo de um sistema GLN.
<b>Objetivo geral da pesquisa</b>	O objetivo da pesquisa proposta é desenvolver modelos computacionais de determinação de conteúdo baseados nos fatores de personalidade CGF, de tal modo que esta determinação seja o mais próxima possível das escolhas humanas na tarefa de descrição de imagens.
<b>Trabalhos relacionados</b>	MAIRESSE, F. Learning to adapt in dialogue systems: Data-driven models for personality recognition and generation. Tese (Doutorado) — University of Sheffield, United Kingdom, 2008. Esse é o principal trabalho relacionado, pois apresenta o sistema PERSONAGE por completo, um sistema de GLN que se utiliza do modelo CGF. Os demais trabalhos relacionados são de variações e/ou evolução do sistema.
<b>Justificativa e relevância</b>	Além do pouco conhecimento disponível sobre a determinação de conteúdo baseada em personalidade, observa-se que, no caso específico do Português brasileiro, estudos desta natureza são ainda mais escassos. Diante deste cenário, evidencia-se assim a oportunidade para um estudo de como os fatores do modelo CGF influenciam a determinação de conteúdo para o desenvolvimento de sistemas de GLN baseados em personalidade.
<b>Proposta para Solução</b>	Foi feita uma coleta de uma base de textos e inventário de personalidade de seus autores, e o tratamento destes dados. Estes dados serão utilizados para o treinamento de modelos supervisionados de Aprendizado de Máquina (AM) com o objetivo de aprender decisões de determinação de conteúdo com informações de personalidade.
<b>Dados</b>	A coleta de dados consiste em um levantamento de descrição de imagens acompanhadas das informações de personalidade de seus respectivos autores, obtidas via preenchimento de inventário de autoavaliação.
<b>Validação</b>	Teste dos modelos desenvolvidos com uso de métodos de AM supervisionada, de modo a estimar quais objetos discursivos incluir, e sua ordem, e comparação com o resultado obtidos utilizando sistemas de <i>baseline</i> pertinentes, ou seja, sistemas de determinação de conteúdo sem a informação de personalidade.
<b>Limitações</b>	A pesquisa proposta não objetiva o desenvolvimento de artefatos completos de GLN, tratará apenas da fase inicial



! Esta ficha é uma adaptação da usada no “VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI 2015)” realizado como parte do “XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)”