

**XII** Congresso  
Fluminense  
de Iniciação Científica  
e Tecnológica



**V** Congresso  
Fluminense  
de Pós-Graduação

**Ciência para o Desenvolvimento Sustentável**

## **Transformação de fontes de dados heterogêneas desconexas em um grafo de conhecimento semântico para Data Science**

*Renato Gomes de Souza Vale Júnior, Mark Douglas de Azevedo Jacyntho*

O mercado empresarial está cada vez mais competitivo, impulsionando as empresas a investirem em Tecnologia de Informação (TI) e em sistemas robustos, capazes de manipular informações de forma mais ágil e precisa. Trata-se de vultosa quantidade de informações disponibilizadas por diversas fontes de dados heterogêneas, a saber: banco de dados relacionais, arquivos de TXT, planilhas eletrônicas, arquivos CSV, arquivos DOC, arquivos PDF, dentre outras. As tecnologias e padrões da Web Semântica, propostos por Tim Berners-Lee, em 2001, viabilizam a criação de (meta)dados estruturados com semântica formalmente explícita, ou seja, dados inteligíveis por máquinas, que, portanto, podem nos auxiliar na busca, integração, compartilhamento, reuso e interpretação desses dados. Para tal, as fontes de dados devem compartilhar e reusar um vocabulário semântico (ontologia) comum, voltado para uma área de conhecimento, classificando e descrevendo as entidades de dados presentes. O objetivo desse trabalho é desenvolver uma base de conhecimento semântica, utilizando a plataforma Ontotext, de modo a integrar fontes de dados heterogêneas, transformando-as em um grafo de conhecimento semântico, descrito com propriedades/classes de ontologias cuidadosamente selecionadas e/ou estendidas. Uma vez criado o grafo de conhecimento, funções de análise de Data Science podem facilmente ser aplicadas sobre a visão formal integrada provida pelo mesmo, a fim de se obter melhor discernimento nos processos de análise de dados e tomada de decisão. Para corroborar o resultado satisfatório, será realizado um estudo de caso com uma empresa atacadista do ramo alimentício da região Norte Fluminense, desenvolvendo para tal um modelo ontológico específico, que permita que a máquina realize inferências referentes a vendas, clientes e produtos, gerando mais insumos para tomadas de decisão mais consistentes.