

ONTOLOGIAS PARA DESCRIÇÃO DE FONTES DE DADOS (APOIO UNIP)

Aluno: Henrique Guerra França

Orientadora: Profa. Eliane Oliveira Santiago

Curso: Ciência da Computação

Campus: Marquês

Nascido da Filosofia, o termo Ontologia indica o campo de estudo da metafísica que se preocupa com a existência do ser enquanto ser. Na ciência da computação, ontologia é o nome dado a um conjunto de técnicas cuja finalidade é nomear e definir tipos, propriedades, atributos e relacionamentos de entidades dentro de um domínio, permitindo então que dados contidos em sua estrutura sejam analisados e transformados em informação – e, futuramente, em conhecimento – para a produção de *insights* e outras amostragens mais condizentes com a realidade. Em outras palavras, trata-se de um campo científico em Teoria da Informação que busca significar e contextualizar dados ao desenvolver uma semântica para a compreensão dos mesmos. No contexto desta Iniciação Científica, a fonte dados utilizada é o DataSUS – *data lake* do Sistema Único de Saúde – que conta com diversas estruturas de dados, sendo nove delas – CIH, CIHA, CMD, CNES, SIASUS, SIHSUS, SIM, SINAC e Base Territorial – usadas no processo de criação das ontologias. Para a criação das ontologias, é utilizado o processo de conversão dos dados estruturados em diagramas entidade-relacionamento (DER) que são, então, analisados para o estabelecimento de relacionamentos e a emergência de abstrações. Finalmente, é feita a transposição do DER para a ontologia usando *Ontology Web Language* (OWL), uma linguagem de marcação que adiciona dimensão semântica a dados, tornando, por exemplo, entidades em conceitos. Nesse momento podem aparecer relacionamentos não triviais entre N conceitos ou atributos, o que torna possível a realização

das outras etapas do ciclo de vida de uma ontologia: alinhamento, refinamento e evolução.