*“O significado das coisas não está nas coisas em si,*

*mas em nossa atitude em relação à elas.”*

Antoine de Saint-Exupéry

# Agradecimentos

Com o desenrolar de dois anos de trabalho, gostaria de agradecer, inicialmente a professora Zulma Medeiros, por ter me concedido a oportunidade de conhecer o meu orientador Fred Freitas, ao qual sou igualmente grato, pela confiança, ensinamentos, empenho e disponibilidade.

Nesse mesmo meio, não posso esquecer as contribuições e a amizade do Prof. Stefan Schulz, da Roberta Fernandes e do Daniel Schober. Trabalhando com todas essas pessoas, e por causa deles, tive a oportunidade de expandir meus conhecimentos e conhecer lugares e outras pessoas fantásticas.

Gostaria, também, de agradecer ao meu pai, Silvio Santana, e a minha irmã, Mariana Santana, pelo suporte, troca de idéias, conselhos e palavras de conforto. E também, a Eliane Leal e Família, por ter sempre auxiliado nos momentos de dificuldade com muito zelo e paciência. Antes de tudo, tenho que pedir desculpa aos meus demais familiares pela ausência freqüente.

Ainda, devo também agradecimentos as Profas. Maria Tereza Cartaxo e Maria do Socorro Cavalcanti, do Instituto de Ciências Biológicas da Universidade de Pernambuco, por terem me introduzido na vida acadêmica e terem contribuído substancialmente para a minha formação, tanto como pesquisador quanto como pessoa.

Por fim, gostaria de lembrar aqueles que já se foram, que marcam, e irão marcar a minha vida eternamente, em caráter e pensamentos principalmente. Sou grato a elas: minha falecida mãe Lucienne Santana; e a minha falecida tia-avó Edite Nunes, por não saírem da minha mente em momento algum e por ter na lembrança delas o meu alicerce, minha admiração e grande parte da minha motivação.

Sem mais, sou eternamente grato a todos.

# Resumo

Muitas aplicações não conseguem tratar a ambiguidade presente em fontes de dados e informação. Tal fato ganhou maior notoriedade a partir do desenvolvimento de tecnologias relacionadas à web semântica, principalmente com as ontologias. O estudo de modelos com certo grau de complexidade representacional relacionados às doenças infecciosas, especificamente as Doenças Tropicais Negligenciáveis (DTNs), vem gradualmente ganhando interesse por parte dos pesquisadores. O presente estudo visa representar um conjunto de conhecimento complexo sobre a transmissão de Doenças Tropicais Negligenciáveis e os possíveis processos que ocorrem a partir do desenvolvimento destas, como o falecimento de indivíduos, em uma ontologia: a *NTDO* (*Neglected Tropical Disease Ontology*). A partir do modelo básico de transmissão de doenças, incluindo vetores artrópodes, e do conteúdo tabular com a representação de vetores, patógenos, hospedeiros, locais de ocorrência e doenças causadas, foi possível descrever um Padrão de Projeto Ontológico (PPO) para a representação de tais processos, refinados e testados segundo consultas em Lógica de Descrições. Outros resultados foram encontrados a partir da representação de processos complexos relacionados ao falecimento de indivíduos por causas específicas. No presente estudo, conhecimento acerca das DTNs foi descrito a partir de informações legadas presentes em tabelas, e puderam ser expressas em uma ontologia formal. A NTDO evidencia eventos complexos com marcações temporais e sequência de processos, desde a transmissão de um patógeno ao falecimento de um indivíduo por uma doença. Assim, a NTDO pode permitir a construção de consultas inteligentes em bancos de dados de Morbidade e Mortalidade. Ainda, pode permitir uma inovação no que concerne a vigilância de casos de doenças relacionados a infecções por doenças, principalmente negligenciadas, por possibilitar o estudo de um amplo conjunto de variáveis, inerentes aos registros de morbidade e mortalidade, e a conseqüente construção de novo conhecimento sobre os dados de saúde.

**Palavras-chave***:* Ontologia, Doenças Tropicais Negligenciáveis, PPO, Representação Tabular, Mortalidade.

# Abstract

Many applications are not able to handle the ambiguity present in data and information sources. These facts gained notoriety due to the development of semantic web related technologies, mainly due to ontologies, that appear to be a solution for those problems. The study of complex representational models related to infectious diseases, specifically Neglected Tropical Diseases (NTDs), interests many researchers. The current study aims at transforming a complex knowledge body about NTD transmission and the possible processes which occurs due to their development, linking it to mortality information, as an ontology: The Neglected Tropical Disease Ontology (NTDO). From the basic disease transmission model, including arthropod vectors, pathogens, hosts, geographic locations and caused diseases, it was possible to describe an Ontology Design Pattern (ODP) to represent such processes, refined and tested by means of Description Logics (DL) queries. Other result was about the representation of complex processes related to passing from life to death, by specific causes. In the current study, the knowledge about NTD was described as legacy data available in tables, and was expressed as a formal ontology. The ontology highlights complex events with temporal marks and process sequences, from a pathogen transmission to the death of a given person due to a disease. As for usefulness, NTDO can support be used for querying Morbidity and Mortality databases and enabling the study of multiple variables, inherent to morbidity and mortality registries, and creating new knowledge using health related data.

**Keywords**: Ontology, Neglected Tropical Diseases, Ontology Design Pattern, Tabular Representation, Mortality.

**SUMÁRIO**

|  |  |
| --- | --- |
| 1. Introdução | 12 |
| 1.1.   Objetivos | 16 |
| 1.1.1. Objetivo Geral | 16 |
| 1.1.2. Objetivos Específicos | 16 |
| 1.2. Justificativa | 18 |
| 2. Estado da Arte | 21 |
| 2.1. Ontologias | 21 |
| 2.2. Ontologias Biológicas | 27 |
| 2.3. Engenharia de Ontologias | 31 |
| 2.4. Lógica de Descrições - DL | 35 |
| 2.5. Web Ontology Language - OWL | 42 |
| 2.6. Padrões de Projetos Ontológicos– PPO | 45 |
| 3. Doenças Tropicais Negligenciáveis – DTN | 48 |
| 3.1. O que são as doenças tropicais negligenciáveis? | 48 |
| 3.2. Quais são e o no que resultam? | 49 |
| 3.3. Regiões de Ocorrência | 51 |
| 3.4. Transmissão de DTN por Vetores Artrópodes | 53 |
| 3.4.1. Processo de Transmissão | 53 |
| 4. Aspectos Metodológicos | 59 |
| 4.1. Modelagem Ontológica | 59 |
| 4.1.1. Propósito | 59 |
| 4.1.2. Escopo | 60 |
| 4.1.3. Fontes de Conhecimento | 61 |
| 4.1.4. Transferência de Conhecimento Tabular para a NTDO | 62 |
| 4.1.5. Integração com outras Ontologias | 64 |
| 4.1.6. Desenvolvimento da NTDO | 65 |
| 4.1.7. Requisitos, Normalização e Documentação | 66 |
| 4.1.8. Manutenção | 66 |
| 5. Uma Ontologia para as Doenças Tropicais Negligenciáveis | 68 |
| 5.1. Representações Tabulares e DTN | 69 |
| 5.1.1. Formalização do PPO para representação da Transmissão | 71 |
| 5.1.2. Entidades Geográficas | 72 |
| 5.1.3. Transmissão do Patógeno | 73 |
| 5.1.4. Disposições e Manifestações | 75 |
| 5.1.5. Dependência das Doenças nos Patógeno | 76 |
| 5.1.6. Axiomas de Inclusão Geral | 77 |
| 5.1.7. Axiomas de Disjunção | 79 |
| 5.1.8. Geração de Ontologia baseada em script | 79 |
| 5.1.9. Pós-processamento da Ontologia | 79 |
| 5.1.10. Avaliação da Ontologia | 80 |
| 5.1.11. Questões relacionadas ao desempenho | 87 |
| 5.2. Modelo de Injúrias e Mortalidade por Causas Específicas | 89 |
| 5.2.1. Representando Injurias e Mortalidade | 89 |
| 5.2.2. Desafios Representacionais do Modelo de Mortalidade | 92 |
| 5.3. Estrutura de Importação e Hierarquia Completa | 104 |
| 6. Discussão | 108 |
| 6.1. Considerações sobre o PPO | 108 |
| 6.2. Considerações sobre o Modelo de Mortalidade | 110 |
| 6.3. Trabalhos Relacionados | 110 |
| 6.4. Casos de Uso | 112 |
| 7. Conclusão | 113 |
| 8. Referências Bibliográficas | 117 |

# Lista de Abreviaturas

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | A.C. | Antes de Cristo | | AIDS | Sindrome da Imunodeficiência Adquirida (*Acquired Immunodeficiency Syndrome*) | | BFO | Ontologia Básica Formal (*Basic Formal Ontology*) | | BFO-RO | Ontologia Básica Formal - Ontologia de Relações (*Basic Formal Ontology - Relation Ontology*) | | BioTop | *Top-domain ontology for the life Sciences* | | CQ | Questões de Competência (*Competency Questions*) | | DAML | Linguagem de Marcação para Agentes DARPA (*Darpa Agent Markup Language*) | | DL | Lógica de Descrições (Description Logics ) | | DO | Ontologia de Doenças (*Disease Ontology*) | | DOLCE | Ontologia Descritiva para a Linguística e Engenharia Cognitiva (*Descriptive Ontology for Linguistic and Cognitive Engineering*) | | DTN | Doença Tropical Negligenciável | | GCI | Classe de Inclusão Geral (General Class Inclusion) | | GFO | Ontologia Geral Formal (*General Formal Ontology*) | | GO | Ontologia dos Genes (*Gene Ontology*) | | HIV | Vírus da Imunodeficiência Humana (*Human Immunodeficiency Virus*) | | IDH | Índice de Desenvolvimento Humano | | IDO | Ontologia para as Doenças Infecciosas (Infectious Disease Ontology) | | IDOMAL | Ontologia para as Doenças Infecciosas - Malária (Infectious Disease Ontology - Malaria) | | KB | Base de Conhecimento (Knowledge Base) | | LC | Leishmaniose Cutânea | | LCD | Leishmaniose Cutânea Difusa | | LMC | Leishmaniose Mucocutânea | | LV | Leishmaniose Visceral | | MSF | Médicos Sem Fronteiras (*Medicins Sans Frontieres*) | | NTDO | *Ontologias para as Doenças Tropicais Negligenciáveis (Neglected Tropical Disease Ontology* ) | | OBI | Ontologia para Investigações Biomédicas (*Ontology for Biomedical Investigations*) | | OBO | Ontologias Biológicas Livres (*Open Biomedical Ontologies*) | | ODL | *Ontology Design Pattern* | | OIL | *Linguagem de Troca de Ontologia* (*Ontology Interchange Language*) | | OMS | Organização Mundial de Saúde | | ONU | Organização das Nações Unidas | | OWL | Linguagem de Ontologias para a Web(*Ontology Web Language*) | | OWL2 | Linguagem de Ontologias para a Web v.2 (*Ontology Web Language* v.2) | | OWL2 EL | Linguagem de Ontologias para a Web v.2 - Linguagem EL (*Ontology Web Language v.2 - EL Language*) | | OWL2 QL | Linguagem de Ontologias para a Web v.2 - Linguagem de Consulta (*Ontology Web Language v.2 - Query Language*) | | OWL2 RL | Linguagem de Ontologias para a Web v.2 - Linguagem de Regras (*Ontology Web Language v.2 - Rule Language*) | | PPO | Padrão de Projeto Ontológico | | RDF | Estrutura de Descrição de Recursos (*Resource Description Framework*) | | RDF *Schema* | Esquema de Estrutura para Descrição de Recursos (*Resource Description Framework Schema*) | | RI | Inclusão de Papel (Role Inclusion) | | SGBD | sistemas de gerenciamento de bancos de dados | | SIM | Sistema de Informação de Mortalidade | | SINAN | Sistema de Informação de Agravos de Notificação | | SNOMED RT | Nomenclatura Sistematizada da Medicina - Terminologia de Referência (Systematized Nomenclature of Medicine - Reference Terminology) | | SNOMED-CT | Nomenclatura Sistematizada da Medicina - Termos Clínicos (Systematized Nomenclature of Medicine-Clinical Terms) | | SPARQL | Linguagem SPARQL para consulta em RDF (*SPARQL Query Language for RDF*) | | SVN | Subversão (*Subversion*) | | UMLS | Sistema de Linguagem Médica Unificada (*Unified Medical Language System*) | | W3C | Consórcio *World Wide Web* (*World Wide Web Consortium*) | | WS | Web Semântica | | XML | Linguagem de Marcação Extensível (*eXtensible Markup Language*) | |  |

# Lista de Figuras

|  |  |
| --- | --- |
| FIGURA 1: Hierarquia de Bibliotecas de Ontologias segundo (GUARINO, 1998) | 25 |
| Figura 2: Hierarquia de Bibliotecas de Ontologias, segundo Freitas (FREITAS, 2003). | 25 |
| FIGURA 3: Atividades e Estados para a Engenharia de Ontologias (FERNANDEZ *et al.*, 1997). | 32 |
| FIGURA 4: Relacionamento entre a *Tbox* e a *Abox*. Adaptado de (BAADER *et al.*, 2003). | 36 |
| FIGURA 5: Tríade Epidemiológica com a Inclusão do Vetor. | 54 |
| FIGURA 6: Processo de Transferência de Conteúdo Tabular para uma Ontologia. | 64 |
| FIGURA 7: Estrutura da NTDO. | 65 |
| FIGURA 8: Ciclo de vida de um organismo. | 89 |
| FIGURA 9: O intervalo de tempo descrito para um gfo:*Chronoid* e seus limites (gfo:*LeftTimeBoundary* e gfo:*RightTimeBoundary*) no eixo temporal. | 95 |
| FIGURA 10: Modelo Gráfico de um evento de um *DeathProcess*, *InjuryProcess* e *BiologicalDeathProcess* e suas relações de participação (**hasDeathPatient**, **hasInjuredPatient** e **hasConvalescentPatient**), de acordo com os participantes dos processos (*e.g.* *Human*). | 102 |
| FIGURA 11: Conexões entre a NTDO, BioTop e GFO. | 103 |
| FIGURA 12: Modelo de Exportação da NTDO. | 104 |
| FIGURA 13: Processos da NTDO | 105 |
| FIGURA 14: Objetos Materiais e Imateriais na NTDO | 106 |
| FIGURA 15: Disposições e Tempo | 107 |

# Lista de Tabelas

|  |  |
| --- | --- |
| TABELA 1: Espécies patogênicas (aos humanos) de *Leishmania*, seus vetores no Novo Mundo, manifestações clínicas e suas distribuições geográficas, baseado em Sharma & Singh (SHARMA; SINGH, 2008)(SHARMA; SINGH, 2008). | 56 |
| TABELA 2: Tabela de DTN transmitidas por vetores, por localização geográfica, patógeno e manifestação clínica. | 62 |
| TABELA 3: Padrão geral de uma matriz para DTN transmitidas por vetores. | 71 |
| TABELA 4: Ontologia simples para realização de testes de raciocínio. | 81 |
| TABELA 5: Mapeamentos entre as entidades presentes na BioTop e GFO. | 96 |