UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

ONTOLOGIAS E AS DOENÇAS TROPICAIS NEGLIGENCIÁVEIS

RECIFE/2011

UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

ONTOLOGIAS E AS DOENÇAS TROPICAIS NEGLIGENCIÁVEIS

MESTRANDO: FILIPE SANTANA DA SILVA

ORIENTADOR: FREDERICO LUÍS GONÇALVES DE FREITAS

RECIFE/2011

SUMÁRIO

[1. INTRODUÇÃO 4](#_Toc288466146)

[2. REVISÃO DA LITERATURA 6](#_Toc288466147)

[2.1. WEB SEMÂNTICA E ONTOLOGIAS 6](#_Toc288466148)

[2.3. LÓGICA DE DESCRIÇÕES – DL 8](#_Toc288466149)

[2.4. DOENÇAS TROPICAIS NEGLIGENCIÁVEIS 10](#_Toc288466150)

[3. JUSTIFICATIVA 12](#_Toc288466151)

[4. OBJETIVOS 14](#_Toc288466152)

[4.1. OBJETIVO GERAL 14](#_Toc288466153)

[4.2. OBJETIVOS ESPECÍFICOS 14](#_Toc288466154)

[5. MÉTODOS 15](#_Toc288466155)

[5.1. GRUPO DE ESTUDO 16](#_Toc288466156)

[5.2. ATIVIDADES E CRONOGRAMA 17](#_Toc288466157)

[6. RESULTADOS PARCIAIS 18](#_Toc288466158)

[7. REFERÊNCIAS BIBLIOGRÁFICAS 19](#_Toc288466159)

# INTRODUÇÃO

Com a explosão de informações criada nos últimos anos pela criação de novos campos de pesquisa na biologia e pelo avanço tecnológico, tratar uma enorme quantidade de conhecimento tornou-se uma tarefa difícil e complexa. Para tentar resolver tal problema, pesquisadores se voltaram para o estudo das ontologias.

Grande parte do conteúdo das informações da área biomédica está disposta em sites, artigos científicos, livros, entre outras formas de publicações, além de bases de dados. A maneira mais comum de identificação dessas fontes de informações se dá pela utilização de palavras chaves, ou termos. Estes não são resguardados com uma base semântica, implicando na dispersão do conhecimento.

Para tratar as informações de forma que os computadores possam entender e, assim, “raciocinar” sobre o conteúdo, Tim Berners-Lee e colaboradores, propuseram a criação da web semântica, a qual apresenta-se como uma extensão da web atual. Concomitantemente, nesse novo modelo proposto a informação é representada com seu sentido bem definido, possibilitando a cooperação de trabalho humano-máquina.

Um dos componentes da web semântica, a qual é utilizada para representar conhecimento de um domínio, são as ontologias. Estas são artefatos computacionais as quais se prestam para a representação de entidades presentes na natureza, como elas realmente se apresentam e se relacionam com outras entidades.

Para a ciência computacional, as ontologias são utilizadas para descrever uma dada conceitualização, provendo aos computadores a habilidade de realizar inferências lógicas. Com efeito, isso proporciona a descoberta de eventos dentro das informações já registradas, resultando na produção de novo conhecimento.

Sendo assim, a *World Wide Web Consortium* (W3C) criou, a partir de experiências relacionadas pela literatura, a *Ontology Web Language* (OWL) (atualmente na segunda versão). A OWL tem o propósito de, sendo uma linguagem de representação baseada em Lógica de Descrições (*Description Logics* – DL), evidenciar o sentido dos termos em vocabulários e suas relações na forma de ontologias.

Nesse âmbito, com o grande aumento do interesse da comunidade científica (mais especificamente das comunidades biomédica e computacional) pelo estudo das ontologias, um tópico vem ganhando destaque pelo impacto causado na população e pela complexidade de interações apresentada pelas entidades relacionadas: o estudo de doenças transmissíveis. Estas, em grande maioria ocorrem em regiões tropicais e são negligenciadas pela pouca contribuição das indústrias farmacêuticas.

As Doenças Tropicais Negligenciáveis (DTN) são constituídas por um grupo de doenças as quais representam as infecções crônicas mais relevantes para a população do planeta. Alguns exemplos dessas podem ser citados, como ancilostomíase, ascaridíase, tricuríase, doença de Chagas, febre do dengue, leishmaniose, esquistossomose, filariose linfática, tracoma, lepra e febre amarela selvagem. Dentre as DTN, um grupo em especial merece destaque: as doenças transmitidas por vetores artrópodes. Estas apresentam considerável relevância, pois no século XX, essas causaram mais morte em humanos do que todas as outras causas de doenças juntas.

Evidenciar os pontos das DTN (geografia, transmissão, epidemiologia)

Consequentemente, o estudo de ontologias sobre as DTN´s poderá promover um melhor conhecimento sobre os eventos de doença as quais subsistem na população. Esse conhecimento engloba o *status* epidemiológico, medidas de prevenção e controle e informações sobre as áreas de ocorrência, incluindo as condições naturais necessárias para a manutenção de tais processos patológicos. Assim, esses componentes são de grande importância para o estabelecimento de metas e prioridades de saúde, proporcionando o direcionamento e indicação de ações de saúde definidas por órgãos superiores de suporte nacional (e internacionais) à saúde. Esse conjunto poderá trazer mais eficiência, eficácia e efetividade para o sistema nacional de saúde

## REVISÃO DA LITERATURA

## WEB SEMÂNTICA E ONTOLOGIAS

A web semântica apresenta-se como uma extensão da web atual (Berners-Lee, Hendler, & Lassila, 2001), na qual a informação é produzida com um sentido bem definido, possibilitando pessoas e computadores trabalharem cooperativamente. Para isso, as informações na web precisam ser anotadas como meta-dados para a descrição de conteúdo. Adicionalmente, algumas linguagens foram criadas para suportar tais requisitos e raciocinadores para a realização de inferências, sobre um dado conteúdo representado semanticamente (Mishra & Kumar, 2010).

Uma das formas de representar a informação para a web semântica é na forma de ontologia. A palavra ontologia, segundo sua utilização, resguarda duas denominações. Uma denominação dada pela filosofia e escrita com o “O” em caixa alta (*Ontologia*), descreve as ontologias como o estudo dos atributos os quais pertencem a coisas pela sua própria natureza. A segunda denominação trata as ontologias, do ponto de vista computacional escrita com “o” em caixa baixa (*ontologia*), como um objeto de informação ou um artefato computacional (Guarino, Oberle, & Staab, 2009), constituído de um vocabulário específico para a descrição de certa realidade, adicionados de um conjunto de pressupostos explícitos relacionados ao sentido pretendido para as palavras de tal vocabulário. Esses conjuntos de pressupostos são comumente formalizados como uma teoria em lógica de primeira ordem (*First Order Logics* – FOL), ou algum outro sistema lógico, sendo conhecido como ontologia formal (Guarino, 1998) (Hofweber, 2004).

Adicionalmente, ontologias computacionais são o meio utilizado para modelar formalmente a estrutura de um sistema, relacionando entidades relevantes e as relações as quais emergem da observação, sendo representados como predicados unários e binários, respectivamente (Guarino, Oberle, & Staab, 2009).

* 1. **ONTOLOGY WEB LANGUAGE – *OWL***

Com a visão da web semântica, a OWL é construída segundo a abilidade do *eXtensible Markup Language* (XML) de definir marcações personalizadas e da abordagem de representação de dados flexível do *Resource Description Framework* (RDF). A utilização do RDF reside na criação de uma linguagem de representação para ontologias a qual pode descrever formalmente o sentido das terminologias utilizadas em documentos, e havendo a necessidade de realizar raciocínio, a linguagem deve ir além do *Resource Description Framework Schema* (RDFS). Nesse âmbito, a *Ontology Web Language* (OWL) foi concebida para tomar parte de todos esses requisitos, tornando-se uma recomendação do *World Wide Web Consortium* (W3C) relacionada a web semântica. Em suma, a OWL é um vocabulário que extende o RDF (McGuinness & F. van Harmelen, 2004).

Nesse sentido, a OWL foi concebida com 3 sublinguagens, em ordem crescente de expressividade, relacionadas ao uso final: OWL-Lite, OWL-DL e OWL-Full. A OWL-Lite oferece suporte a construção de hierarquias e restrições simples. Já a OWL-DL apresenta expressividade considerável sem perder a completude e decidibilidade, e sua denominação (DL) reside nesta ser baseada em Lógica de Descrições. Por fim, a OWL-Full é destinada aos usuários os quais desejam o máximo de expressividade sem preocupação com completude e decidibilidade. Nesse sentido, cada fragmento da OWL é mais simples que seu sucessor, resultando em: toda conclusão retirada de uma OWL-Lite é válida em OWL-DL, e está terá sua conclusão válida em OWL-Full (McGuinness & F. van Harmelen, 2004).

Comparando as três sublinguagens da OWL (*Full*, DL e *Lite*), tem-se que a OWL-DL é uma linguagem decidível, sendo relacionada como um meio termo entre a expressividade e a inferência eficiente. Além disso, a OWL é suportada pela grande maioria dos editores de ontologias, se apresentado sintaticamente mais compacta (ao contrário da *Full*)(McGuinness & F. van Harmelen, 2004). Alguns benefícios relacionados à modelagem de ontologias com a OWL são: compartilhamento de ontologias, evolução, interoperabilidade, detecção de inconsistências, balanço entre expressividade e escalabilidade, facilidade de utilização, compatibilidade com outros padrões (utilizados pela indústria ou na web) e internacionalização (Zhihong & Mingtian, 2003).

Em 2009, a W3C criou a OWL2, como uma extensão e uma revisão da OWL. Assim, foram criadas algumas novas funcionalidades, do ponto de vista sintático e de expressividade, como chaves, cadeias de propriedades, *datatypes* mais ricos, imagem de dados, restrições qualificadas de cardinalidade, assimetria, reflexividade e disjunção de propriedades e uma melhor capacidade de anotação. Além disso, foram definidos novos perfis e uma nova sintaxe. Os novos perfis são o OWL2 EL (algoritmos em tempo polinomial), OWL2 QL (consultas com conjunções respondidas em *LogSpace,* utilizando tecnologias de bancos de dados relacionais), e OWL2 RL (implementação de algoritmos de raciocínio em tempo polinomial). Com relação a sintaxe, para a OWL foi desenvolvida a *Manchester Syntax*, com o objetivo de introduzir não-logicistas a uma sintaxe para facilitar a construção de ontologias (Horridge, Drummond, Goodwin, Rector, & Wang, 2006), e para a OWL2, foi a *Manchester OWL2 Syntax*, não havendo mudanças significativas (Horridge & Patel-Schneider, 2009).

Alguns trabalhos utilizaram-se da OWL para a representação de suas respectivas ontologias. (Yan & Zha, 2010) apresentaram, em seu trabalho, uma ontologia para gerenciar conhecimento sobre consumidores, demonstrando a capacidade expressiva da OWL em representar tal conhecimento. (Thomas, Pappu, & Baker, 2009) desenvolveram uma ontologia em OWL, segundo os preceitos da *Basic Formal Ontology (*BFO), para relacionar entidades físicas, químicas e descrições funcionais de nanoparticulas as quais foram formuladas e testadas para aplicações no diagnóstico do câncer e para fins terapêuticos. Outro trabalho, desenvolvido por (Bouamrane, Rector, & Hurrell, 2010), construiu e programou uma ontologia em OWL para auxiliar na avaliação, como suporte a decisão, em eventos cirúrgicos.

# LÓGICA DE DESCRIÇÕES – DL

A Lógica de Descrições (DL) é uma família de linguagens para representação do conhecimento a qual pode ser utilizada para descrever um domínio de aplicação. O nome lógica de descrições advém das expressões criadas com conceitos atômicos (predicados unários) e papéis atômicos (predicados binários ou relações). Um ponto crucial da DL é que esta é construída com semântica baseada em lógica. A DL possui uma sintaxe e uma semântica básica e é construída sob uma série de construtores booleanos - a conjunção (), a disjunção (e a negação (¬) - restrições existenciais (Ǝ*r.*C) e restrições de valor (*r*.C), sendo esse conjunto de caracteres denominados como constituintes da ALC (*Attribute concept Language with Complements*) (AL adicionado de negação “¬” resulta em ALC) (Baader, Horrocks, & Sattler, 2008).

Descrições sobre conceitos podem ser utilizadas para criar declarações em uma base de conhecimento construída em DL. Comumente, esta base de conhecimento pode ser dividida em dois grandes blocos: a *Tbox* e a *Abox*. A *Tbox* é responsável por englobar os pontos relevantes de determinado domínio, definindo propriedades de conceitos e papeis, além dos respectivos relacionamentos entre eles. Já a *Abox*, *assertional*, é utilizada para descrição de situações concretas, por definir propriedades para indivíduos (Baader, Horrocks, & Sattler, 2008).

Os sistemas modernos baseados em DL provêem os seus usuários, por meio de raciocínio automatizado, deduzir conhecimento implícito a partir da descrição explicita, produzindo respostas concretas em tempo finito. Nesse âmbito, o raciocínio é construído com base em subsunção, algoritmos os quais determinam relações entre subconceitos-superconceitos (ou subclasse-superclasse) para computar uma hierarquia (Baader, Horrocks, & Sattler, 2008), ou seja, se os relacionamentos entre as classes são coerentes (satisfazíveis) do ponto de vista de como é especificado, não haverá inconsistência na construção da base de conhecimento.

Nos últimos anos, a modelagem ontológica com base na DL, e na ALC, fez com que a DL tenha sido estendida para incluir restrições numéricas (qualificadas), relações inversas, transitivas, subrelações, domínios concretos, e nominais. Sendo cada uma dessas adições a AL representando um acréscimo em expressividade, uma denominação com uma letra: a letra *S* para *basic* (AL propriamente dita), *H* para subrelações, *O* para nominais*, I* para relações inversas, *N* para restrições numéricas, e *Q* para restrições numéricas qualificadas, foi adicionada. A lógica de descrições a qual corresponde a OWL-DL que apresenta todos esses acréscimos de expressividade é conhecida como SHOIQ (Baader, Horrocks, & Sattler, 2008) (Quadro 1)

|  |  |  |
| --- | --- | --- |
| Classe | Descrição | Sintaxe |
| *S* | Top (conceito mais geral) | ┬ |
| Bottom (conceito mais específico) | ┴ |
| Combinação de dois conceitos | *C*1 ∩ *C*2 |
| Disjunção de dois conceitos | *C*1 U *C*2 |
| Negação de um conceito | ⌐ *C*2 |
| Qualificador de quantificação universal | *r.C* |
| Qualificador de quantificação existencial | *r.C* |
| *O* | Pelo menos um dos objetos | {o1...on} |
| *Q* | Cardinalidade mínima | > *n r.C* |
| Cardinalidade máxima | < *n r.C* |
| Cardinalidade exata | = *n r.C* |

Quadro 1: Conjunto de construtores da lógica de descrição SHOIQ e sua sintaxe. Adaptado de (Allard & Ferré, 2008).

## DOENÇAS TROPICAIS NEGLIGENCIÁVEIS

As Doenças Tropicais Negligenciáveis (DTN) são um conjunto de doenças que persistem exatamente em condições físicas, psicossociais e econômicas específicas dos mais pobres, em sua maioria populações marginalizadas de regiões mundiais em desenvolvimento (Molyneux, Hotez, & Fenwick, 2005). A maioria dessas doenças são afecções crônicas, debilitantes, relacionadas à pobreza como promotora de infecções parasitárias (helmínticas e protozoárias), bacterianas, e, também, virais ou fungicas (Hotez et al., 2007). Para a maioria das DTN´s, existem soluções preventivas para o controle e/ou erradicação, tais como o tratamento em massa da população e atividades de iniciativa global em torno dessas doenças (Molyneux, Hotez, & Fenwick, 2005). Contrastando com essa situação, em países desenvolvidos, as DTN´s são vistas como um evento raro (King & Bertino, 2008).

Entre as décadas de 1980 e 1990, foi estabelecida uma colaboração entre a Organização Mundial de Saúde e o Banco Mundial (Mathers, Ezzati, & Lopez, 2007), sendo criada uma escala para mensurar o impacto criado pelas doenças no estado de saúde dos indivíduos. A essa mensuração foi dado o nome de *disability-adjusted life year* (DALY) – anos de vida com deficiência-ajustados – sendo possível uma aproximação e o nivelamento das doenças. O principal objetivo da DALY é comparar e quantificar agregados regionais e distribuições mundiais da situação de saúde criada por diferentes estados de doença. Este dado pode propiciar uma noção sobre a distribuição dessas de forma não-subjetiva, refletindo consensos sociais e possibilitando o conhecimento de vieses presentes em avaliações de doenças individuais (Murray, 1994).

De acordo com a Organização Mundial de Saúde, os casos de malária identificados no mundo (casos agudos) são responsáveis por, aproximadamente, 46,5 milhões de anos de vida com deficiência-ajustados. A Filariose Linfática e a Leishmaniose (visceral e tegumentar) são responsáveis por 5,777 milhões e 2,09 milhões de anos (DALY), respectivamente. Outras doenças como a Febre Hemorrágica do Dengue e a doença de Chagas também são relacionadas como doenças de alto impacto na situação de saúde da população mundial (World Health Organization, 2004).

Em muitos países da América Latina, Ásia e África Sub-Sahariana há uma larga distribuição das DTN pelas condições climáticas (entre 15°C e 40°C) e pela altitude (abaixo de 2.200 metros). Nessas regiões, grade parte da população (por volta de 2,7 milhões de pessoas) recebe apenas US$ 2,00 por dia, podendo ser considerados os mais pobres do mundo sendo também relacionados como os mais vulneráveis as DTN´s (Hotez et al., 2007).

Assim como em muitas áreas do mundo, no Brasil, as DTN´s são mais elevadas em áreas de pobreza sendo relacionada ao índice de desenvolvimento humano (IDH). As regiões do país mais afetadas são o Norte e o Nordeste, apresentando altos níveis de prevalência da malária, doença de Chagas, leishmaniose, esquistossomose, dengue, oncocercose e a filariose linfática. No Brasil, existe uma estimativa que 40 milhões de pessoas estão infectadas com uma ou mais DTN. O controle da DTN´s pode representar o mecanismo mais eficiente e efetivo para promover a saída desses 40 milhões de brasileiros das margens da linha da pobreza (Hotez, 2008). Um aspecto importante dessas doenças é que elas não só ocorrem junto à pobreza, mas a promovem (Hotez et al., 2007).

Na região da América Latina e Caribe pode ser encontrada a maior disparidade de renda entre classes sociais, com 48% de toda a renda alocada no 1/3 mais rico da população e o terço menos favorecido apenas 1,6%(The World Bank, 2003). Dentre a população com menor potencial financeiro, 1/3 desses vivem em zona rural (praticando agricultura de subsistência, pesca, entre outros) (Ault, 2007) sendo peculiar a presença indígena e de afro-descendentes, com um alto nível de exclusão e iniqüidade social, privados de acesso a água para consumo de qualidade e serviços de saúde. Os outros dois terços da população pobre podem ser distribuídos em favelas, assentamentos, bairros pobres ou em áreas periféricas das cidades. Nessas áreas há uma combinação de água de qualidade duvidosa para o consumo humano, baixos níveis sanitários e proliferação de roedores, reservatórios e vetores de doenças (Ault, 2007) (Riley, Ko, Unger, & Reis, 2007).

Dentre as DTN´s, as doenças transmitidas por vetores (dengue, leishmaniose, doença de Chagas (tripanossomíase americana), tripanossomíase africana, filariose linfática, febre amarela, entre outras) resultam em um alto nível de mortalidade em regiões pouco favorecidas (HOTEZ *et al.,* 2009). Essas infecções perduram por um longo tempo e podem causar deficiências graves e desfiguramento, resultando em profundas consequencias econômicas, sociais e políticas(Hotez et al., 2007)(Beyrer et al., 2007). Dentre todos os casos de doenças transmitidas por vetores que ocorrem na América latina, 67% da leishmaniose visceral, 46% dos casos de leishmaniose cutânea (Organização Pan-Americana da Saúde, 2007), 80,8% dos casos de dengue (Pan American Health Organization, 2008) e 8% dos casos de filariose linfática ocorrem no Brasil (World Health Organization, 2006)

# JUSTIFICATIVA

A demanda de utilização por uma forma diferenciada de modelagem de informações e conhecimento foi visualizada a partir da estabilização dos sistemas de gerenciamento de bancos de dados (SGBD) e a continuação da problemática relacionada à modelagem conceitual. Já a comunidade de engenharia de software, motivados pelas linguagens orientadas a objetos, começaram a reconhecer a importância do que é conhecido com *modelagem de domínio,* além do crescimento e da complexidade dos programas criando dificuldades para a manutenção e a disponibilização de tais para novos usuários (Smith & Welty, 2001); entre vários outros problemas relacionados na literatura.

Voltados para a solução de tais impasses, em 2001 foi introduzido o conceito de web semântica por Tim Berners-Lee e colaboradores (Berners-Lee, Hendler, & Lassila, 2001), impulsionados pela possibilidade de transformar o conteúdo da internet, do que é construído apenas para humanos lerem, em algo que seja manipulado por máquinas e possa ser entendido por estas. Sendo assim, como modo de representar conhecimento e um dos componentes da web semântica, as ontologias foram introduzidas com o objetivo de integrar informações provenientes de diversas fontes, sem controle sobre as formas de aquisição e de como serão guardadas.

No campo da ciência da computação, as ontologias foram adotadas pela inteligência artificial com o objetivo de facilitar o compartilhamento e o reuso. Adicionalmente, elas são chave para possibilitar o desenvolvimento de tecnologias para a web semântica, sendo estudadas por várias comunidades dentro da Inteligência Artificial, como engenharia de conhecimento, processamento de linguagem natural e representação do conhecimento(Davies, Fense, & F. V. Harmelen, 2003).

Como um *framework* para representação formal, o estudo das ontologias foi relacionado, como um dos artefatos utilizados para representar as entidades presentes na biomedicina, além de seus termos e suas relações(Bodenreider, 2008). Nesse sentido, o estudo das ontologias na área da saúde é motivado pela quantidade imensa de informações coletadas, demandando esforços para analisar e estruturar novos dados, além de extrair o conhecimento inerente a estes (Stenzhorn, Beisswanger, & Schulz, 2007).

Referenciando o estudo da saúde como um todo, as ontologias apresentam algumas vantagens quando se trata do estudo de doenças. Por exemplo, ontologias auxiliam a aquisição de informação sobre tópicos específicos relacionados às doenças e provê acesso inteligente a informações científicas; compartilhamento de conhecimento aumenta a eficiência e a eficácia para a pesquisa, pois evita as redundâncias e pode auxiliar a direcionar caminhos para a pesquisa científica; como base da interoperabilidade, ontologias proporcionam fontes heterogêneas de informação tornarem-se cooperativas; e por fim, ontologias podem auxiliar no estudo de doenças e desordens confluindo diversos fatores em busca de um denominador (Hadzic & Chang, 2005).

Uma vez que o estudo de doenças e dos fatores que corroboram para o seu estabelecimento nas populações baseado em ontologias se revela importante, o estudo desse artefato conjuntamente com as doenças tropicais negligenciáveis podem trazer reflexos e resultados inimagináveis. Uma razão importante para tal iniciativa reside nesse conjunto de doenças subsistir concomitante a condição de pobreza, apresentando maior desenvolvimento em ambientes socioeconomicamente empobrecidos de regiões tropicais(World Health Organization, 2010).

As Doenças Tropicais Negligenciáveis (DTN) são tradicionalmente deixadas de lado nas agendas de saúde internacionais, mesmo essas afetando a saúde de mais de 1 bilhão de pessoas em todo o mundo, fato esse resultante do baixo potencial de voz política dessa população e da baixa visibilidade de tais doenças pelas condições de ocorrência(World Health Organization, 2010).

Portanto, é nesse cenário que o presente estudo se resguarda, levando o estudo das DTN a uma representação ontológica. Pelos benefícios listados para o estudo das ontologias relacionando doenças e pelo impacto social causado pelas doenças tropicais negligenciáveis, é visível a importância de tal artefato no para o estudo desta. Iniciativa semelhante está sendo desenvolvida por (Topalis et al., 2010), a qual visa à construção de uma ontologia segunda junção de várias outras, criando um conjunto variado de informações.

Entretanto, a descrição completa do mecanismo de transmissão, de maneira que possibilite o estudo de qualquer outra DTN, aliado a uma descrição completa de representação geográfica, não apenas baseada na distribuição espacial de países ou cidades, mas também do ambiente o qual exerce função primordial no mecanismo de transmissão, foco do presente trabalho.

Sendo assim, o estudo das DTN por meio de ontologias poderá propiciar o estudo das condições que a subsistência de tais eventos depreciadores de saúde resulta na população, de maneira inteligente e baseada no conhecimento prévio sobre a problemática, auxiliando a indicação de medidas de intervenção, como um sistema de suporte a decisão, as quais possibilitem, pelo menos, diminuir o impacto das DTN às populações expostas.

# OBJETIVOS

## OBJETIVO GERAL

O objetivo do presente estudo é o desenvolvimento de uma ontologia genérica sobre as doenças tropicais negligenciáveis, a qual oferecerá suporte para o estudo de tais eventos de doença, auxiliando no processo de planejamento e metas de saúde, segundo o padrão de morbidade da população brasileira.

## OBJETIVOS ESPECÍFICOS

* Definição e representação dos conceitos-chave com base em publicações científicas (livros, periódicos, entre outros);
* Definição e representação de padrões de transmissão e desenvolvimento das DTN nas populações afetadas e representação de tais padrões segundo um modelo de lógica de descrições;
* Representação de entidades geográficas e do ambiente, de maneira genérica, os quais exercem efeito no desenvolvimento das DTN;
* Representação de conhecimento epidemiológico para auxiliar na identificação de de dados provenientes de fontes heterogêneas relacionadas a estatísticas de morbidade e mortalidade, como o Sistema Nacional de Agravos (SINAM) e o Sistema Nacional de Mortalidade (SIM), respectivamente, e;
* Construção de um modelo ontológico, utilizando uma ou mais DTN, para aquisição, exemplificação e testes sobre o modelo genérico apresentado.

# MÉTODOS

A metodologia de trabalho empregada reside no estabelecimento de padrões genéricos para construção de ontologias relacionadas a doenças transmissíveis, focando as DTN como modelo genérico. Para verificação da aplicabilidade dos modelos, corretude e validação dos padrões de representação, algumas das DTN, focando as transmitidas por vetores, servirão como modelo para testes, pois essas apresentam uma maior complexidade na interação entre os organismos e com as necessidades ambientais.

Assim, está em desenvolvimento um conjunto de ontologias, genéricas o suficiente para representar DTN´s transmitidas por vetores, conhecimentos epidemiológicos e um modelo de representação geográfica/ambiental, tendo como base materiais de publicações provenientes de periódicos, publicações do Ministério da Saúde, livros, entre outros. Pela inexistência de uma metodologia formal para construção de ontologias, essas serão construídas segundo instruções de (Fernandez, Gomez-Perez, & Juristo, 1997) e (Noy & Mcguinness, 2000) e utilizando o editor OWL Protégé v.4.1 (Stanford University School of Medicine, Stanford, Califórnia, EUA) como ferramenta de modelagem.

Como base do modelo ontológico central, o modelo de construção para ontologias biomédicas provenientes da BioTop (Stenzhorn, Beisswanger, & Schulz, 2007) será seguido, pois reúne definições padronizadas para disposição de classes e utilização de relação, evitando intercorrências relacionadas a erros de representação.

Posteriormente, serão traçados pontos de integração na ontologia para busca em bases de dados com informações relacionadas aos casos referentes a doenças e agravos, provenientes do Ministério da Saúde (MS), relacionadas aos locais de ocorrência de cada evento de doença associado.

Finalmente, com a associação das informações sobre eventos de doença, presente nos bancos de dados do MS, será possível traçar e indicar indicar o local medidas de saúde necessárias para a condição encontrada, dependentes da região de ocorrência.

Assim, todas as informações necessárias à modelagem ontológica serão representadas por meio de Lógica de Descrições (DL), utilizando como linguagem de representação a *Ontology Web Language* (OWL). Testes serão realizados sobre a ontologia seguindo um conjunto de questões de competência, devidamente representadas em DL, e construídas segundo a necessidade dos módulos pertencentes à ontologia (transmissão, doenças, epidemiologia e geografia/ambiente).

Alguns aspectos de desenvolvimento das ontologias englobam a reunião de conhecimento a partir de diversas fontes diferentes (artigos científicos, livros, entre outros). A ligação dos diferentes módulos das futuras ontologias será realizada tal qual as rotinas de importação de repositórios externos. Isso também incluirá a criação de um instrumento para visualização baseado em web/ontologias e recuperação com interface, tendo como ponto de vista a re-utilização dos aplicativos existentes.

# GRUPO DE ESTUDO

Este trabalho se desenvolve por parceria entre o Centro de Informática da Universidade Federal de Pernambuco, o Centro de Pesquisas Aggeu Magalhães, o Departamento de Informática Médica da Albert-Ludwigs-Universität Freiburg (Freiburg im Breisgau, Alemanha) e o Instituto de Informática Médica, Estatística e Documentação da Graz General Hospital and University Clinics (Graz, Austria).

|  |  |  |
| --- | --- | --- |
| Filipe Santana da Silva | Mestrando | Centro de Informática UFPE |
| Robera Fernandes | Mestranda | Centro de Informática UFPE |
| Frederico L. G. Freitas | Orientador | Centro de Informática UFPE |
| Zulma M. Medeiros | Co-Orientadora | Aggeu Magalhães - FIOCRUZ |
| Stefan Schulz | Co-Orientador | Inst. of Med. Inf., Stat. and Doc. Graz General Hospital and University Clinics |
| Daniel Schober | Colaborador | Albert-Ludwigs-Universität Freiburg |

# ATIVIDADES E CRONOGRAMA

No quadro 2 estão descritas as atividades propostas a serem executadas até o final da dissertação:

|  |  |
| --- | --- |
| **Atividade** | **Descrição** |
| A1 | Listagem de termos candidatos ao domínio |
| A2 | Avaliação de ontologias reutilizáveis, vocabulário e dicionário geográfico |
| A3 | Representação das entidades geográficas e ambientais |
| A4 | Modelagem de Padrões de Transmissão |
| A5 | Primeira Versão - NTDO |
| A6 | Refinamento – Testes das questões de competência com *DL Query* |
| A7 | Versão Final - NTDO |
| A8 | Resultados publicados |
| A9 | Dissertação |

O cronograma proposto para esta dissertação de mestrado está descrito no quadro 3. Utilizando como referências os semestres letivos.

Quadro 3 **–** Cronograma

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2011 | | | | | | | | | | 2012 |
| MAR | ABR | MAI | JUN | JUL | AGO | SET | OUT | NOV | DEZ | JAN |
| A1 | X |  |  |  |  |  |  |  |  |  |  |
| A2 | X |  |  |  |  |  |  |  |  |  |  |
| A3 | X | X |  |  |  |  |  |  |  |  |  |
| A4 | X | X |  |  |  |  |  |  |  |  |  |
| A5 |  | X | X | X |  |  |  |  |  |  |  |
| A6 |  |  | X | X |  |  |  |  |  |  |  |
| A7 |  |  |  |  | X | X |  |  |  |  |  |
| A8 | X |  |  | X |  |  | X | X |  |  |  |
| A9 |  |  |  |  |  |  |  | X | X | X | X |

# RESULTADOS PARCIAIS

# REFERÊNCIAS BIBLIOGRÁFICAS

Allard, P., & Ferré, S. (2008). Dynamic Taxonomies for the Semantic Web. *2008 19th International Conference on Database and Expert Systems Applications*, 382-386. Ieee. doi: 10.1109/DEXA.2008.71.

Ault, S. K. (2007). Pan American Health Organizationʼs Regional Strategic Framework for addressing neglected diseases in neglected populations in Latin America and the Caribbean. *Memórias do Instituto Oswaldo Cruz*, *102*, 99-107. doi: 10.1590/S0074-02762007005000094.

Baader, F., Horrocks, I., & Sattler, U. (2008). Description Logics. In F. van Harmelen, V. Lifschitz, & B. Porter (Eds.), *Knowledge Creation Diffusion Utilization* (1st ed., Vol. 6526, pp. 135-179). Elsevier Inc. doi: 10.1016/S1574-6526(07)03003-9.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34-43. doi: 10.1038/scientificamerican0501-34.

Beyrer, C., Villar, J. C., Suwanvanichkij, V., Singh, S., Baral, S. D., & Mills, E. J. (2007). Neglected diseases, civil conflicts, and the right to health. *The Lancet*, *370*(9587), 619-627. doi: 10.1016/S0140-6736(07)61301-4.

Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. *IMIA: Yearbook of medical informatics*, 67-79. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2592252&tool=pmcentrez&rendertype=abstract.

Bouamrane, M.-M., Rector, A., & Hurrell, M. (2010). Using OWL ontologies for adaptive patient information modelling and preoperative clinical decision support. *Knowledge and Information Systems*. doi: 10.1007/s10115-010-0351-7.

Davies, J., Fense, D., & Harmelen, F. V. (Eds.). (2003). *TOWARDS THE SEMANTIC WEB: Ontology-driven Knowledge Management*. West Sussex, England: John Wiley & Sons Ltd.

Fernandez, M., Gomez-Perez, A., & Juristo, N. (1997). METHONTOLOGY : From Ontological Art Towards Ontological Engineering. *AAAI Technical Report*, *SS-97-06*, 33-40.

Guarino, N. (1998). Formal Ontology and Information Systems. *Formal Ontology in Information Systems. Proceedings of FOIS’98*, (June), 3-15.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an Ontology? In S. Staab & R. Studer (Eds.), (Second Edi., pp. 1-17). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-92673-3.

Hadzic, M., & Chang, E. (2005). Ontology-Based Support for Human Disease Study. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, *00*(C), 143a-143a. Ieee. doi: 10.1109/HICSS.2005.472.

Hofweber, T. (2004). Logic and Ontology. *Stanford Encyclopedia of Philosophy*. Retrieved February 1, 2011, from http://plato.stanford.edu/entries/logic-ontology/#3.

Horridge, M., Drummond, N., Goodwin, J., Rector, A., & Wang, H. H. (2006). The Manchester OWL Syntax. *Proc. of the 2006 OWL Experiences and Directions Workshop (OWL-ED2006)*. Athens, Georgia.

Horridge, M., & Patel-Schneider, P. (2009). OWL 2 Web Ontology Language: Manchester Syntax. Retrieved from http://www.w3.org/TR/owl2-manchester-syntax/.

Hotez, P. J. (2008). The giant anteater in the room: Brazilʼs neglected tropical diseases problem. *PLoS neglected tropical diseases*, *2*(1), e177. doi: 10.1371/journal.pntd.0000177.

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., et al. (2007). Control of neglected tropical diseases. *The New England journal of medicine*, *357*(10), 1018-27. doi: 10.1056/NEJMra064142.

King, C. H., & Bertino, A.-M. (2008). Asymmetries of poverty: why global burden of disease valuations underestimate the burden of neglected tropical diseases. *PLoS neglected tropical diseases*, *2*(3), e209. doi: 10.1371/journal.pntd.0000209.

Mathers, C. D., Ezzati, M., & Lopez, A. D. (2007). Measuring the burden of neglected tropical diseases: the global burden of disease framework. *PLoS neglected tropical diseases*, *1*(2), e114. doi: 10.1371/journal.pntd.0000114.

McGuinness, D. L., & Harmelen, F. van. (2004). OWL Web Ontology Language: Overview. *W3C*. Retrieved from http://www.w3.org/TR/owl-features/.

Mishra, R. B., & Kumar, S. (2010). Semantic web reasoners and languages. *Artificial Intelligence Review*. doi: 10.1007/s10462-010-9197-3.

Molyneux, D. H., Hotez, P. J., & Fenwick, A. (2005). "Rapid-impact interventions": how a policy of integrated control for Africaʼs neglected tropical diseases could benefit the poor. *PLoS medicine*, *2*(11), e336. doi: 10.1371/journal.pmed.0020336.

Murray, C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization*, *72*(3), 429-45.

Noy, N. F., & Mcguinness, D. L. (2000). Ontology Development 101 : A Guide to Creating Your First Ontology. *Development*. Stanford University. Retrieved from http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.

Organização Pan-Americana da Saúde. (2007). *Saúde nas Américas* (Publicação.). Washignton: Organizaçao Pan-Americana de Saúde.

Pan American Health Organization. (2008). 2008: Number of Reported Cases of Dengue & Dengue Hemorrhagic Fever (DHF), Region of the Americas (by country and subregion). Retrieved March 21, 2011, from http://www.paho.org/english/ad/dpc/cd/dengue-cases-2008.htm.

Riley, L. W., Ko, A. I., Unger, A., & Reis, M. G. (2007). Slum health: diseases of neglected populations. *BMC international health and human rights*, *7*, 2. doi: 10.1186/1472-698X-7-2.

Smith, B., & Welty, C. (2001). Ontology: Towards a New Synthesis. *Proceedings of the international conference on Formal Ontology in Information Systems - FOIS ’01*. New York, New York, USA: ACM Press. doi: 10.1145/505168.505201.

Stenzhorn, H., Beisswanger, E., & Schulz, S. (2007). Towards a top-domain ontology for linking biomedical ontologies. *Studies in health technology and informatics*, *129*(Pt 2), 1225-9.

The World Bank. (2003). Inequality in Latin America & the Caribbean: Breaking with History? Retrieved March 17, 2011, from http://go.worldbank.org/TFJHCL2B30.

Thomas, D. G., Pappu, R. V., & Baker, N. a. (2009). Ontologies for cancer nanotechnology research. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2009*, 4158-61. doi: 10.1109/IEMBS.2009.5333941.

Topalis, P., Dialynas, E., Mitraka, E., Deligianni, E., Siden-Kiamos, I., & Louis, C. (2010). A set of ontologies to drive tools for the control of vector-borne diseases. *Journal of biomedical informatics*. doi: 10.1016/j.jbi.2010.03.012.

World Health Organization. (2004). *The World Health Report: 2004 : Changing History*. *World Health* (p. 96p). Geneva: World Health Organization.

World Health Organization. (2006). Global Programme to Eliminate Lymphatic Filariasis. *Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*, *81*(22), 221-32.

World Health Organization. (2010). *First WHO report on neglected tropical diseases 2010: Working to overcome the global impact of neglected tropical diseases*. *World Health* (p. 186). Geneva.

Yan, Y., & Zha, X. (2010). Applying OWL to Build Ontology for Customer Knowledge Management. *Journal of Computers*, *5*(11), 1693-1699. doi: 10.4304/jcp.5.11.1693-1699.

Zhihong, Z., & Mingtian, Z. (2003). Web Ontology Language OWL and its description logic foundation. *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies, 2003. PDCATʼ2003.*, 157-160. Ieee. doi: 10.1109/PDCAT.2003.1236278.