

Relato e Considerações sobre o Desenvolvimento de uma Ontologia para Avaliação de Sites da Área de Saúde

Daniel Lichtnow (Organizador)¹

José Palazzo Moreira de Oliveira (Orientador)¹

Adriana Jouris

Alexandro Bordignon

Ana Marilza Pernas Fleischmann

Felipe Hoppe Levin

Gleison Samuel do Nascimento

Isabel Cristina Siqueira da Silva

Isabela Gasparini

Juliano Machado Teixeira

Luiz Henrique Longhi Rossi

Olinto Eliziano David de Oliveira

Paulo Schreiner

Sílvia Regina Vargas Gomes

Resumo: Este artigo descreve desenvolvimento de uma ontologia que visa avaliar a qualidade de *sites*/páginas com assuntos relacionados à área de saúde. O processo de desenvolvimento da ontologia foi realizado ao longo do segundo semestre de 2008 pelos alunos da disciplina CMP234 - Modelagem Conceitual e

¹ Instituto de Informática, UFRGS, Caixa Postal 15064 {dlichtnow, palazzo @inf.ufrgs.br}

Ontologia do Programa de Pós-Graduação do Instituto de Informática da Universidade Federal do Rio Grande do Sul, ministrada pelo Professor Dr. José Palazzo Moreira de Oliveira. O desenvolvimento da ontologia foi feito com base na Metodologia 101 [1]. Assim, estrutura do documento observa, em grande parte, as etapas da Metodologia 101 e procura manter grande parte das informações e observações feitas ao longo do desenvolvimento visando o uso destas em futuras versões da ontologia. Para o desenvolvimento da ontologia foi utilizado o editor *Protégé*. As instâncias da ontologia foram criadas mediante extração dos dados dos *sites*/páginas, extração esta que foi feita manualmente pelos alunos da disciplina.

1 Introdução

Este artigo descreve o processo de construção de uma ontologia voltada para avaliação da qualidade de *sites*/páginas da área de saúde. O trabalho está relacionado aos Projetos *SALUS – CYTED – Qualidade em Sites na Área de Saúde* e *PROSUL – AVAL-SAÚDE*.

Os projetos têm por objetivo discutir e criar instrumentos para avaliação de *sites* da área de saúde visando fornecer a usuários com perfis distintos indicativos de qualidade dos *sites*.

O processo de desenvolvimento da ontologia foi realizado ao longo do segundo semestre de 2008 pelos alunos da disciplina CMP234 - Modelagem Conceitual e Ontologia do Programa de Pós-Graduação do Instituto de Informática da Universidade Federal do Rio Grande do Sul, ministrada pelo Professor Dr. José Palazzo Moreira de Oliveira.

O desenvolvimento da ontologia foi feito com base na Metodologia 101 [1]. Assim, a estrutura do documento observa, em grande parte, as etapas da Metodologia 101:

- **Passo 1.** Determinar o domínio e escopo da ontologia;
- **Passo 2.** Considerar o reuso de ontologias existentes;
- **Passo 3.** Enumerar termos importantes na ontologia;
- **Passo 4.** Definir as classes e hierarquia de classes;
- **Passo 5.** Definir as propriedades das classes (*slots*);
- **Passo 6.** Definir as restrições das propriedades das classes;

- **Passo 7.** Criar instâncias.

Seguindo as etapas da Metodologia 101, a seção 2 irá descrever o escopo de domínio da ontologia fazendo uso de questões de competência – *competency questions*.

Já a seção 3 apresenta algumas ontologias que são incorporadas a versão atual da ontologia ou que poderão ser utilizadas nas próximas versões.

A seção 4 apresenta os termos relevantes para a ontologia juntamente com comentários que procuram esclarecer o significado de cada termo. Estes termos foram definidos pelos alunos a partir especialmente da observação de critérios de qualidade para *sites*/páginas da área de saúde como HON - *Health On the Net Foundation* [2] e *Web Médica Acreditada* [3].

A seção 5 apresenta a definição da hierarquia de classes da ontologia desenvolvida. Estas classes foram definidas a partir dos termos identificados na seção 4.

A seção 6 apresenta as principais restrições das propriedades das classes. Os critérios de qualidade estabelecidos a partir da observação das propriedades identificadas são descritos. A seção descreve ainda aspectos relacionados à implementação da ontologia no *Protégé* [4].

A seção 7 descreve a criação de instâncias. Estas instâncias foram criadas a partir da extração manual que foi feita a partir da observação de alguns *sites*/páginas recuperados usando o *Google*.

A seção 8 apresenta as considerações finais e aponta possíveis passos a serem observados na construção de futuras versões da ontologia.

Cabe ressaltar que este documento foi elaborado pensando em manter informações levantadas ao longo do processo de desenvolvimento da versão da ontologia aqui descrita. Neste sentido, eventualmente, alguns dos termos relacionados na seção 4 e das ontologias mostradas na seção 3 não foram efetivamente utilizadas na versão da ontologia descrita aqui. No entanto, a opção por manter estas informações reside no fato de que estas poderão ser úteis em futuras discussões nos projetos.

Para implementação da ontologia foi utilizada *OWL - Web Ontology Language*, o editor *Protégé* [4] e o *Pellet* [5] como motor de inferência.

2 Domínio e Escopo da Ontologia

O domínio da ontologia apresentada neste documento está relacionado a *sites*/páginas que abordem temas da área de saúde. De forma mais específica a ontologia irá procurar representar aspectos relacionados à qualidade do *site* sendo a ênfase dada para avaliação da qualidade do conteúdo. No entanto, alguns aspectos relacionados à acessibilidade são considerados.

Uma questão importante, surgida durante o desenvolvimento do trabalho, refere-se a se a ontologia iria avaliar um *site* ou páginas específicas de um *site*, como as retornadas por um motor de busca.

Cabe considerar que, eventualmente, informações relevantes para avaliação da qualidade de uma página podem estar em outras páginas do *site* ao qual esta página pertence. Um exemplo está em um *site* com diversas páginas que contém informações sobre os autores do conteúdo de todas as páginas do *site* em uma página específica. Nesta situação, avaliar a página isoladamente pode conduzir a uma avaliação errônea. Ainda páginas dentro de um mesmo *site* podem apresentar, eventualmente, qualidade distinta.

Em função disto, foi criada na ontologia uma classe *Site* (ver seção 5) de forma que toda página retornada por um motor de busca é caracterizada como um *site*, embora isto possa não ser desejável em alguns casos, como no caso do retorno de um motor de busca ser um documento *.pdf*.

Assim, a ontologia desenvolvida será utilizada para:

- Verificar se o *site*/página apresenta qualidade;
- Representar as propriedades que um *site*/página apresenta, isto é, características que evidenciem a qualidade do *site*/página;
- Permitir representar não apenas a qualidade inerente ao *site*/página quando este é analisado isoladamente, mas também a sua adequação a um determinado público e/ou contexto.

Considerando o uso da ontologia, serão realizados três processos básicos de inferência:

1. Um para determinar a qualidade do *site*/página sem levar em conta o contexto de uso e o público alvo;
2. Outro para indicar o quanto é aconselhável o uso do *site*/página para determinada classe de usuários;
3. Outro para definir o grau de acessibilidade de um *site*/página.

2.1 Considerações sobre qualidade de *sites*

Qualidade é freqüentemente associada à adequação para uso - "*fitness for use*". Assim, o que determina qualidade, é quem faz uso do produto. Esta definição vem sendo utilizada em diversos contextos, estando aí incluídos trabalhos que lidam com qualidade da informação [6].

Tendo em vista o fato de que a ênfase está em avaliar a qualidade do conteúdo de um *site*/página assume especial importância aspectos relacionados à avaliação da Qualidade da Informação e Qualidade dos Dados.

Em [7] a Qualidade dos Dados - *Data Quality* é definida enfatizando que dados de qualidade são dados que estão aptos a serem usados por consumidores de dados. Ainda [7] define Dimensão de Qualidade dos Dados - *Data Quality Dimension* como um conjunto de atributos de qualidade de dados que representam um único aspecto ou conceito relacionado à Qualidade de Dados.

Seguindo [7] as dimensões de qualidade são classificadas em quatro categorias:

- **Dimensões Intrínsecas - *Intrinsic Dimensions*.** São independentes do contexto do usuário. Estão relacionadas ao fato da informação corresponder a realidade e ser consistente;
- **Dimensões Contextuais - *Contextual Dimensions*.** Refletem como a qualidade da informação deve ser medida em um determinado contexto, levando em conta, por exemplo, o perfil do consumidor das informações;
- **Dimensões Representacionais - *Representational Dimensions*.** Estas dimensões abordam questões relacionadas ao formato e a facilidade de interpretação dos dados;
- **Dimensões de Acessibilidade - *Accessibility Dimensions*.** Estas dimensões fazem referência a aspectos relacionados ao acesso a informação. Esta dimensão é tratada normalmente como algo distinto em relação a qualidade, mas por ser uma dimensão percebida pelos consumidores de informação deve ser levada em conta.

Em [8] são utilizadas estas características, com algumas poucas adaptações, para avaliar a qualidade dos dados no contexto da *Web*.

Na definição da ontologia são consideradas especialmente três destas dimensões: Intrínsecas, Contextuais e de Acessibilidade.

No que se refere às propriedades intrínsecas, são levados em conta especialmente alguns dos critérios estabelecidos por algumas organizações certificadoras como autoria e existência de propaganda dentre outras [2] [3].

No que se refere às propriedades contextuais, na avaliação dos *sites/páginas* é levando em conta de um *site/página* é voltado para médicos ou para leigos.

No que se refere à acessibilidade é considerado o documento *Web Content Accessibility 1.0* [9] que oferece uma série de recomendações ou princípios gerais para o design de páginas *web-acessíveis*.

Assim, seguindo [9] são considerados os pontos de verificação distribuídos entre três níveis de prioridade. Para verificação é fornecido um *checklist*.

Os *sites* avaliados por este *checklist* podem ter três classificações de acessibilidade: *A*, *AA* e *AAA*. A classificação *A* significa que o *site* atende as recomendações de prioridade 1, que são as mais básicas. A classificação *AA*, indica que ele atende as recomendações de prioridade 1 e 2, e finalmente, a classificação *AAA* indica o cumprimento de todas as prioridades, sendo a classificação de acessibilidade mais completa.

A partir do *checklist* da *W3C* [10] foram definidas algumas propriedades que foram representadas na ontologia de forma a ilustrar a avaliação de um *site*/página quanto à acessibilidade. Algumas medidas a serem adotadas pelos desenvolvedores do *site*, visando garantir acessibilidade, são apresentadas a seguir.

Classificação A

- Prover texto alternativo para todos os elementos não-textuais;
- Assegurar que todas as informações conduzidas por cores sejam também disponíveis sem cores;
- Organizar os documentos de forma que eles sejam lidos sem folhas de estilo;
- Utilizar linguagem apropriada e mais clara e simples possível para o conteúdo do *site*;
- Se, apesar de todos os esforços, o *site* não puder ser criado com acessibilidade, então, neste caso, disponibilizar um *site* alternativo, totalmente acessível.

Classificação AA

- Assegurar que as cores de texto e do fundo possuem contraste suficiente;
- Utilizar folhas de estilo para controlar o layout e a apresentação;
- Identificar claramente para onde vai cada link;
- Utilizar os mecanismos de navegação de maneira consistente.

Classificação AAA

- Explicar o que cada sigla e abreviatura significam, na sua primeira ocorrência;
- Identificar a linguagem do documento;
- Fornecer informações para que os usuários possam receber documentos de acordo com suas preferências;
- Fazer distinção entre informações de começo de parágrafo, listas, etc.

2.2 Questões de Competência

Seguindo a Metodologia 101, foram elaboradas uma série de questões de competência. Questões que devem ser respondidas mediante o uso da ontologia. Três questões principais devem ser respondidas por meio do uso da ontologia:

- 1) Este *site*/página é confiável?
- 2) Para que perfil de usuário o *site*/página é recomendado?
- 3) Qual o nível de acessibilidade de um *site*/página?

Além destas três questões principais, a seguir são apresentadas uma série de questões fornecem respostas parciais às questões principais. Na seção 6 são elencadas as propriedades da ontologia que fornecem suporte para responder estas questões.

- 1) De que assunto específico trata o *site*?
- 2) Para qual público o *site* está voltado?
- 3) Qual o propósito do *site*?
- 4) O *site* possui selos de certificação?
- 5) Quais são selos de certificação na área da saúde?
- 6) A que instituição o *site* está vinculado?
- 7) Quais instituições são conceituadas atualmente?
- 8) Existem propagandas no *site*?
- 9) O *site* possui apoio/patrocínio de algum fabricante ou organização?
- 10) Qual o nível de compreensão que o usuário deve ter para entender o conteúdo do *site*?
- 11) Usuários podem interagir com o *site*? Alterar/Acrescentar conteúdos?
- 12) Quem são os autores do conteúdo do *site*? Quais são as qualificações destes autores?
- 13) O *site* é assinado (ou revisado) por algum médico credenciado?
- 14) O *site* apresenta forma de contato com os responsáveis?
- 15) O *site* possui *links* externos para outros *sites*/páginas?
- 16) O *site* é referenciado? Por quem? Quantas referências?
- 17) O conteúdo do *site* está em que idioma?

18) O *site* possui conteúdo atualizado?

19) Considerando os critérios de acessibilidade, onde o *site* se enquadra *A*, *AA*, *AAA* ou nenhuma?

3 Reutilização de Ontologias

Esta seção descreve sucintamente algumas ontologias (a maioria relacionada à área de saúde) que podem vir a ser utilizadas na definição de futuras versões da ontologia.

- **FOAF – Friend of a Friend.** É uma ontologia que descreve agentes (pessoas, grupos e organizações), suas atividades e o relacionamento entre agentes e outros objetos (como documentos, por exemplo) [11]. A ontologia *FOAF* permite descrever uma rede social sem que seja necessária a existência de uma base centralizada. Definições sobre o vocabulário da ontologia podem ser encontradas em [12]. É possível importar a ontologia para o *Protégé*. Para a versão 3.3.1 do *Protégé* a ontologia foi importada do endereço http://www.mindswap.org/2005/foaf_cleaner/ (ocorreram problemas na importação a partir do *site* indicado no *Protégé*).
- **MeSH – Medical Subject Headings.** É um vocabulário controlado usado para indexar artigos e livros que também pode ser utilizado para facilitar buscas (neste último caso *MeSH* é utilizado como um thesaurus). O *MeSH* é mantido pela *United States National Library of Medicine (NLM)*, e usada na *MEDLINE - Medical Literature Analysis and Retrieval System Online* que é uma banco de dados de literatura relacionada a ciências da vida e informação biomédica. Uma ontologia baseada em *MeSH* pode ser encontrada em <http://onto.eva.mpg.de/obo/mesh.owl>. Ainda cabe ressaltar a existência dos Descritores de em Ciências da Saúde (disponível em: <http://decs.bvs.br/homepage.htm>) que baseados no *MeSH* apresentam também um vocabulários que inclui inglês, português e espanhol. No *MeSH* existem descritores para doenças, pessoas e organizações, o que pode eventualmente ser útil. As Figuras 1 e 2 apresentam parte da ontologia.
- **GALEN.** É uma ontologia que representa um modelo para terminologia na área médica. Ela pode ser acessada em <http://www.opengalen.org/>. Detalhes sobre o desenvolvimento da ontologia podem ser encontrados em [13]. Na Figura 3 é apresentado o resultado de uma consulta feita a ontologia.
- **OpenCyc.** Esta ontologia é uma versão open source da ontologia *Cyc*, que é uma ontologia de topo <http://sw.opencyc.org/>. A Figura 4 apresenta parte da ontologia que foi obtida a partir de uma consulta.

Persons [M01]
Occupational Groups [M01.526]
Administrative Personnel [M01.526.070] +
Clergy [M01.526.213]
Ethicists [M01.526.276]
Faculty [M01.526.339] +
Foreign Professional Personnel [M01.526.407] +
▶ Health Personnel [M01.526.485]
Allied Health Personnel [M01.526.485.067] +
Anatomists [M01.526.485.133]
Caregivers [M01.526.485.200]
Coroners and Medical Examiners [M01.526.485.230]
Dental Staff [M01.526.485.290] +
Dentists [M01.526.485.330] +
Faculty, Dental [M01.526.485.360]
Faculty, Medical [M01.526.485.375]

Figura 1: Parte Ontologia *MeSH* – *Persons*

Health Care Economics and Organizations [N03]
Economics [N03.219] +
Health Planning [N03.349] +
▶ Organizations [N03.540]
Academies and Institutes [N03.540.052]
Congresses as Topic [N03.540.199] +
Consumer Organizations [N03.540.245] +
Government [N03.540.348] +
Government Agencies [N03.540.400]
Health Planning Organizations [N03.540.452] +
Home Care Agencies [N03.540.483]
International Agencies [N03.540.514] +
Labor Unions [N03.540.571] +
National Academy of Sciences (U.S.) [N03.540.608] +
Organizations, Nonprofit [N03.540.630] +
Public-Private Sector Partnerships [N03.540.706]
Self-Help Groups [N03.540.782] +
Societies [N03.540.828] +
Social Control, Formal [N03.706] +

Figura 2: Parte Ontologia *MeSH* - *Organizations*

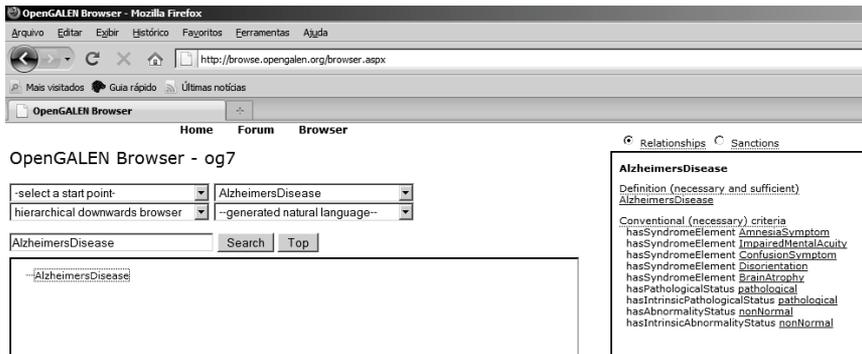


Figura 3: Parte Ontologia *Galen*

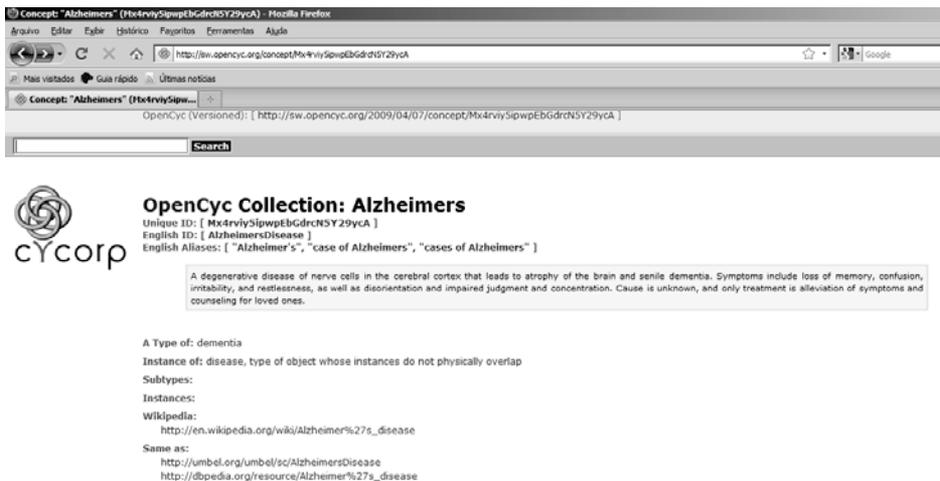


Figura 4: Parte Ontologia *OpenCyc*

Sobre o uso das ontologias cabe destacar que é possível aproveitar apenas alguns segmentos destas ontologias. Para tanto pode ser utilizado algum *software* como *Segmenter* [14].

4 Termos Relevantes para Ontologia

Esta seção enumera e descreve uma série de termos relevantes para o desenvolvimento da ontologia. Seguindo a Metodologia 101, não há, nesta seção, a preocupação de definir exatamente quais e de que forma estes termos irão para a ontologia.

Assim, alguns dos termos presentes não foram incorporados a ontologia descrita ao final deste documento, mas foram mantidos aqui pelo fato de que no futuro possam vir a fornecer informações relevantes para novas versões da ontologia.

Assim esta seção corresponde ao registro de um processo de *brainstorm* feito pelos alunos da disciplina e visa manter documentadas algumas das considerações feitas pelos alunos.

Para cada termo foi feita uma pequena descrição que tem por objetivo esclarecer o seu significado e eventualmente salientar alguns aspectos. **Os termos estão agrupados de acordo com sua afinidade.** Para alguns termos são fornecidos sinônimos e/ou a tradução para inglês.

Termo: *Site*.

Sinônimos/Tradução: *Web Site*.

Descrição/Observações.

Seguindo as definições dadas em [15]:

“A collection of interlinked Web pages, including a host page, residing at the same network location. “Interlinked” is understood to mean that any of a Web site’s constituent Web pages can be accessed by following a sequence of references beginning at the site’s host page; spanning zero, one or more Web pages located at the same site; and ending at the Web page in question.”

Um *site* é formado por várias páginas. A questão, já ressaltada na seção 2, é se deve ser medida a qualidade de uma página ou de um *site*. Neste sentido, ao medir de uma página, devem ser consideradas informações disponíveis em outros locais (política editorial, dono e/ou autores podem ser mencionados apenas na página principal). Outro aspecto é relacionado ao fato de que páginas podem variar em qualidade dentro de um mesmo *site*. Neste sentido cabe salientar que os selos dados aos *sites* são selos dados a todo *site*, e não a páginas específicas.

Termo: *Web Page*.

Sinônimos/Tradução: *Webpage*, *Página Web*.

Descrição/Observações.

Seguindo as definições dadas em [15]:

“A collection of information, consisting of one or more Web resources, intended to be rendered simultaneously, and identified by a single URI. More specifically, a Web page consists of a Web resource with zero, one, or more embedded Web resources intended to be rendered as a single unit, and referred to by the URI of the one Web resource which is not embedded.

Examples: *An image file, an applet, and an HTML file identified and accessed through a single URI, and rendered simultaneously by a Web client.*

Note: *The components of a Web page can reside at different network locations. The location of the Web page, however, is determined by the URI identifying the page.*

Note: *The scope of a Web page is limited to the collection of Web resources which are displayed simultaneously by requesting the Web page's URI. The components of a Web page actually rendered in a page view is client-dependent.”*

A partir destas definições convém destacar também outras que as complementam [15]:

“Web Resource. A resource, identified by a URI, that is a member of the Web Core. Note: The URI identifying the Web Resource does not itself have to be found within the Web Core. That is, a URI written on a bus identifying a resource that is a member of the Web Core identifies a Web Resource.

Web Core. The collection of resources residing on the Internet that can be accessed using any implemented version of HTTP as part of the protocol stack (or its equivalent), either directly or via an intermediary. Notes: By the term "or its equivalent" we consider any version of HTTP that is currently implemented as well as any new standards which may replace HTTP (HTTP-NG, for example). Also, we include any protocol stack including HTTP at any level, for example HTTP running over SSL.”

Termo: Finalidade do Site.

Sinônimos/Tradução: *Purpose of Site*.

Descrição/Observações.

Um *site* pode ser voltado para uma finalidade específica ou para várias. São exemplos de finalidades: diagnóstico, prevenção, tratamento, divulgação de artigos científicos, dentre outras.

Termo: Assunto do Site.

Sinônimos/Tradução: Tema do *Site*.

Descrição/Observações.

Pode ser representado por um texto (como um resumo), *keywords*, metadados como *Dublin Core*, descritores da área médica como *MESH* ou até mesmo *CID 10*.

Termo: Usabilidade do Site.

Sinônimos/Tradução: *Usability*.

Descrição/Observações.

Segundo [16], usabilidade refere-se ao esforço necessário para usar um produto de software, bem como o julgamento individual de tal uso para um conjunto explícito ou implícito de usuários.

Termo: Público Alvo.

Sinônimos/Tradução: *Target Public*, Audiência.

Descrição/Observações.

Define a quem se destina o *site*.

Termo: Agente.

Sinônimos/Tradução: *Agent*.

Descrição/Observações.

Segue a definição existente em *foaf:Agent* [12].

Termo: Pessoa.

Sinônimos/Tradução: *Person*.

Descrição/Observações.

Segue a definição existente em *foaf:Person* [12].

Termo: Grupo.

Sinônimos/Tradução: *Group*.

Descrição/Observações.

Segue a definição existente em *foaf:Group* [12].

Termo: Organização.

Sinônimos/Tradução: *Organization*.

Descrição/Observações.

Segue a definição existente em *foaf:Organization* [12].

Termo: Médico.

Descrição/Observações.

Pessoa habilitada a praticar medicina. Um médico pode assumir o papel de usuário (alguém que acessa o *site*), dono, editor, revisor e/ou autor.

Termo: Leigo.

Descrição/Observações.

Pessoa sem formação na área de medicina. Um leigo pode assumir o papel de usuário (alguém que acessa o *site*), dono, editor, revisor e/ou autor.

Termo: Dono.

Sinônimos/Tradução: *Owner*.

Descrição/Observações.

Dono do *site*. É em última análise, o responsável pelo conteúdo do *site*. Pode ser uma Organização ou uma Pessoa ou um Grupo. Não necessariamente tem formação na área de medicina.

Termo: e-mail Dono.

Descrição/Observações.

Cabe a observação de que o *e-mail* do dono está ligado ao *site*, mas ele é do dono. Ainda, um agente pode estar vinculado a um *site*, possuir *e-mail*, mas não ter seu *e-mail* publicado. É preciso ainda diferenciar o e-mail do dono de um *e-mail* que pode ser o *e-mail* publicado no *site* que visa o registro de reclamações ou o esclarecimento de dúvidas. Cabe salientar que qualquer agente pode possuir um *e-mail*.

Termo: Endereço Dono.

Descrição/Observações.

Propriedade do *site* e do Dono - vide obs. *e-mail* Dono.

Termo: Telefone Dono.

Descrição/Observações.

Propriedade do *site* e do Dono - vide obs. *e-mail* Dono.

Termo: Webmaster.

Sinônimos/Tradução: *Web Architect*, *Web Developer*, *Site Author*, *Website Administrator*

Descrição/Observações.

É um Agente - Pode ser uma pessoa, organização ou grupo de pessoas responsável pelo projeto e manutenção de um *site*. Não é necessariamente o responsável pelo conteúdo ou dono, embora possa ser.

Termo: Editor.

Descrição/Observações.

É um Agente - Pode ser uma pessoa, organização ou grupo de pessoas responsável pelos aspectos editoriais de uma publicação. Determina o conteúdo final de um texto.

Termo: Autor.

Descrição/Observações.

É um Agente - Pode ser uma pessoa, organização ou grupo de pessoas responsável pelo conteúdo ou parte do conteúdo de um *site*.

Termo: Revisor.

Descrição/Observações.

É um Agente - Pode ser uma pessoa, organização ou grupo de pessoas responsável por avaliar (revisar) o conteúdo do *site*

Termo: Número de Acessos.

Sinônimos/Tradução: *Page View*.

Descrição/Observações. É uma requisição de uma página. Em princípio apenas um *webmaster* ou dono pode ter acesso a esta informação.

Termo: Número de Downloads.

Descrição/Observações. É uma requisição feita sobre um recurso presente em uma página de um *site*. Em princípio apenas um *webmaster* ou dono pode ter acesso a esta informação.

Termo: Inlinks.

Sinônimos/Tradução: *Incoming Links, Inbound Links e Inward Links*.

Descrição/Observações. São *hyperlinks* de outros *sites*/páginas que fazem referência ao *site*/página.

Termo: *Outlinks*.

Sinônimos/Tradução: *External Links*, *Links* Externos.

São *hyperlinks* para outros *sites*/páginas que o site referencia (for do domínio onde está uma página). Cabe ressaltar que algoritmos como *HITS* [17] fazem análise dos *outlinks* para determinar a qualidade de uma página. O problema é que *sites*/páginas ruins podem referenciar bons *sites*/páginas propositadamente (*spam*).

Termo: *Internal Link*.

Sinônimos/Tradução: *Link* Interno.

Descrição/Observações. É um *hyperlink* que é uma referência para outra parte de um documento, página ou *site*.

Termo: *Broken Link*.

Sinônimos/Tradução: *Link* quebrado.

Descrição/Observações. É um *hyperlink* que aponta para algo que não existe. Algumas ferramentas permitem fazer esta verificação <http://validator.w3.org/checklink>.

Termo: Referências na Imprensa.

Descrição/Observações. Eventualmente um *site*/página pode ser referenciado na imprensa. Difícil de detectar se a referência não é feita mediante *hyperlinks*.

Termo: Referências em Publicações Científicas.

Descrição/Observações. Eventualmente um *site*/página pode ser referenciado em algumas publicações científicas. Difícil de detectar se a referência não é feita mediante *hyperlinks*.

Termo: *PageRank*.

Descrição/Observações. É um valor numérico calculado a partir da análise dos *InLinks* [18].

Termo: Data da Criação.

Descrição/Observações. Data em que um *site*/página foi criado.

Termo: Data da Última Atualização/Revisão.

Descrição/Observações. Data em que um *site*/página foi atualizado/revisado.

Termo: Tipo de Usuário.

Descrição/Observações. Em [19] são identificados três tipos básicos de usuários ou visitantes de *sites*:

- **Casual - Casual.** É alguém que acessa o *site* quase que por uma eventualidade e durante um curto espaço de tempo.
- **Intentional - Intencional.** É alguém que possui ao menos algum conhecimento relacionado ao tema do *site*.
- **Specialist – Especialista.** É alguém que conhece bem os temas abordados no *site* e terá interesse em acessar um vasto conteúdo relacionado.

Termo: Especialidade.

Descrição/Observações. Referente à especialidade de um Médico ou a Especialidade a qual está relacionada o *site*. Considerar *MeSH* e [20].

Termo: Acessibilidade.

Descrição/Observações. Na *web*, acessibilidade significa que qualquer pessoa (usando um tipo específico de tecnologia), deve ser capaz de visitar e interagir com diversos *sites*, compreendendo as informações nele apresentadas. Os principais instrumentos para acessibilidade na *web* são os guias de recomendação, as listas de verificação (*checklists*) e as ferramentas automáticas.

Termo: Patrocinador.

Sinônimos/Tradução: *Sponsor*.

Descrição/Observações. É um Agente. Pode Patrocinar um *site*. Pode ser o dono do *site* ou alguém (Agente) que o financie. Pode ou não colocar propaganda de seus produtos e/ou serviços no *site*.

Termo: Enfermeiro.

Descrição/Observações. É uma pessoa com conhecimentos na área de saúde.

Termo: Acompanhante.

Sinônimos/Tradução: Cuidador, *Carer*.

Descrição/Observações. É uma Pessoa, que não tem formação na área de saúde, mas cuida de um paciente.

Termo: Paciente.

Descrição/Observações. É uma pessoa com problemas de saúde.

Termo: Estudante da Área de Saúde.

Descrição/Observações. É uma pessoa que cursa algum curso da área de saúde.

Termo: Agente de Saúde.

Sinônimos/Tradução: Agente Comunitário de Saúde.

Descrição/Observações. É uma pessoa. Tem por objetivo auxiliar e controlar os indicadores das manifestações epidemiológicas e outras funções dentro da comunidade.

Termo: Hospital.

Descrição/Observações. É uma organização da área de saúde.

Termo: Pronto-Socorro.

Descrição/Observações. É uma organização da área de saúde.

Termo: Faculdade de Medicina.

Descrição/Observações. É uma organização de ensino relacionada à área de saúde.

Termo: Universidade.

Descrição/Observações. É uma organização. Pode estar, ou não, relacionada a área de saúde.

Termo: Curso.

Descrição/Observações. Curso vinculado a alguma instituição de ensino.

Termo: Curso de Medicina.

Descrição/Observações. Curso vinculado a alguma instituição de ensino que forma médicos.

Termo: Centro de Pesquisa.

Descrição/Observações. Centro de pesquisa que pode estar ou não a área de saúde.

Termo: Instituição Certificadora.

Descrição/Observações. Organização que fornece um selo de qualidade a um *site*.

Termo: ONG – Organização Não Governamental.

Sinônimos/Tradução: *NGO - Non-Governmental Organization*.

Descrição/Observações. É uma organização pode estar vinculada ou não a área de saúde.

Termo: Política de Privacidade.

Sinônimos/Tradução: *Privacy Politics*.

Descrição/Observações. Descreve o que é feito com os dados obtidos junto aos usuários do *site*.

Termo: Política Editorial.

Descrição/Observações. Critérios para publicação de conteúdo.

Termo: Resumo.

Sinônimos/Tradução: *Abstract*.

Descrição/Observações. Descreve de forma sucinta o conteúdo de um *site*.

Termo: Glossário.

Sinônimos/Tradução: *Glossary*.

Descrição/Observações. Uma lista de termos com significado. Em páginas *web* pode-se pensar em *hyperlinks* dos termos para páginas (externas ou não) que contenham a definições. Relevante especialmente para *sites* destinados ao público em Geral.

Termo: Medicamento.

Sinônimos/Tradução: *Medicine, Drug*.

Descrição/Observações. Algo que trata previne ou alivia os sintomas de uma doença.

Termo: Equipamento Médico.

Descrição/Observações. Refere-se a quaisquer equipamentos que possam vir a ser utilizados na realização de diagnósticos ou tratamentos.

Termo: Fabricante.

Descrição/Observações. É uma organização que produz algo. No contexto do trabalho importam os fabricantes de medicamentos e equipamentos médicos.

Termo: Tratamento.

Sinônimos/Tradução: *Treatments*

Descrição/Observações. Procedimentos médicos ou aplicações que se destinam a aliviar doenças ou lesões.

Termo: Tratamentos Alternativos.

Sinônimos/Tradução: *Alternative Treatments*.

Descrição/Observações. Por tratamentos alternativos entende-se alternativas a um tratamento. Isto é para um tratamento existem outros que podem ser utilizados. Não se refere à medicina alternativa.

Termo: Efeitos Colaterais.

Sinônimos/Tradução: *Colateral Effects*.

Descrição/Observações. Resultado não desejado do tratamento e/ou medicamentos. Na descrição de um tratamento/diagnóstico devem estar presentes.

Termo: Diagnóstico.

Sinônimos/Tradução: *Diagnosis*.

Descrição/Observações. Identificação da razão/causa de algum problema.

Termo: Doença.

Sinônimos/Tradução: *Disease*.

Descrição/Observações. Comprometimento ou problema relacionado a saúde. Considerar CID-10.

Termo: Exame Médicos.

Sinônimos/Tradução: *Medical Exam, Checkup, Medical Checkup, Medical Examination, Health Check*.

Descrição/Observações. Procedimento de avaliação de saúde.

Termo: Código de Ética.

Descrição/Observações. O *site* observa (ou afirma observar) algum código de ética.

Termo: Simulação de Procedimentos.

Descrição/Observações. No Brasil é recomendado que um *site* não tenha simulações de procedimentos [21].

Termo: *Blog*.

Descrição/Observações. Sobre a presença de *blogs* é necessário verificar quem escreve, os comentários que são feitos e a moderação.

Termo: CRM – Conselho Regional de Medicina.

Descrição/Observações. É uma organização formada e mantida por médicos no Brasil. No Brasil, cada médico é identificado por um número que é referenciado como CRM.

Termo: Idioma.

Descrição/Observações. Pode ser propriedade do *site* e de pessoa. No caso do *site* indica em que idioma está o conteúdo. No caso de pessoas indica que idiomas a pessoa domina.

Termo: Qualidade Geral do *Site*.

Descrição/Observações.

Este termo indica a qualidade geral do *site*. A qualidade geral de um *site*, não necessariamente precisa ser representada como uma propriedade do *site*. Para simplificar, uma possibilidade seria se trabalhar com três níveis de qualidade, como em [22]:

- *Unsatisfactory* – Insatisfatório;
- *Marginal* - Marginal – Mediano;
- *Satisfactory* – Satisfatório.

Termo: *Código Internacional de Doenças - CID.*

Sinônimos/Tradução: *ICD- International Classification of Diseases*

Descrição/Observações. Na prática, no Brasil, é conhecida como *CID-10*. Cabe verificar qual a relação entre *MeSH* e *CID*. Descritores *MeSH do tipo C* referem-se a doenças.

Termo: *MESH - Medical Subject Headings*

Descrição/Observações. É um vocabulário controlado usado para indexar artigos na área de medicina. Foi criado pela *United States National Library of Medicine (NLM)*.

Termo: Termos da Área Médica.

Descrição/Observações. São termos da área médica que são pouco compreensíveis por leigos.

Termo: Organização Reconhecida na Área de Saúde.

Descrição/Observações. É Organização da Área de Saúde reconhecida como de excelência.

Termo: Organização da Área de Saúde.

Descrição/Observações. É uma Organização da Área de Saúde (Hospital, Clínica, etc).

Termo: Currículo.

Sinônimos/Tradução: *Curriculum Vitae*.

Descrição/Observações. Documento que contém informações acadêmicas e/ou as experiências profissionais de uma pessoa. Pode estar presente de forma sucinta em um documento de autoria de uma pessoa.

Termo: Publicações.

Descrição/Observações. Publicações da pessoa (especialmente médico ou pessoas que estejam envolvidas com elaboração ou revisão do conteúdo de *sites*). Pode ser considerado propriedade do *site* (o *site* pode referenciar publicações dos autores do conteúdo).

Termo: Propaganda.

Sinônimos/Tradução: *Advertising*.

Descrição/Observações. A presença de propaganda em um *site* pode determinar a sua qualidade especialmente em termos de isenção.

Termo: Propaganda em Excesso.

Descrição/Observações. Propaganda pode ser permitida, dentro de certos limites.

Termo: Propaganda Inserida no Conteúdo.

Descrição/Observações. Não se consegue distinguir o que é propaganda do que não é ou existem referências a medicamentos de determinado fabricante.

Termo: Referência as Fontes de Dados.

Descrição/Observações. Faz referência a outras fontes de informação que não são *sites*. Não necessariamente estão na *Web*, mas podem ser *outLinks*.

Termo: Metadados.

Descrição/Observações. Faz referência especialmente ao uso de descritores (*MESH*, *CID-10*) ou *keywords* para referenciar conteúdo.

Termo: Tipo de Acesso.

Descrição/Observações. Propriedade do *site*. Especifica qual tipo de acesso é dado a que usuários de um *site*. Relacionado a questões de segurança.

5 Classes e Hierarquia de Classes

A partir dos termos identificados na seção 4, foram definidas as classes que compõe a ontologia. Na hierarquia foram consideradas classes presentes na *FOAF*. Desta forma a classe *Site* está definida como subclasse da classe *foaf:Document*.

A hierarquia das classes é mostrada nas Figuras 5, 6, 7, 8 e 9. Pelo fato da ontologia ser construída no *Protégé*, as classes superiores das hierarquias de classes mostradas nas Figuras 5, 6, 7, 8 e 9 são todas sub-classes da classe *owl:Thing*.

Na Figura 5, convém ressaltar que um *site* pode estar classificado ou como *Satisfatório* ou como *Mediano* ou como *Insatisfatório*. Estas classes são disjuntas entre si.

Considerando esta mesma hierarquia de classes um *site* pode ser considerado voltado para público Leigo ou para Médicos.

Já quanto aos critérios de acessibilidade, um *site* pode estar classificado como um *site* da classe *ClasseA*, *ClasseAA* ou *ClasseAAA*. Estas três classes estão organizadas de maneira hierárquica pelo fato de que um *site* que atende os critérios relacionados a *ClasseAAA* deve necessariamente atender todos os critérios relacionados as classes *ClasseA* e *ClasseAA*, conforme pode ser constatado pela avaliação dos critérios de acessibilidade disponíveis em [9]. De forma geral a cada classe são associados os critérios do *triplo A*.

Isto não acontece com as classes *Satisfatório*, *Mediano* e *Insatisfatório*, pelo fato de que existe uma série de combinações que permitem associar um *site* a cada uma destas classes.

Já na hierarquia de classes apresentada na Figura 6 um Agente pode ser usuário e ao mesmo tempo, mas não obrigatoriamente, ser Leigo ou Médico. No que diz respeito a classificação dos usuários, estes podem estar classificados como Casual, Intencional e Especialista, seguindo [19].

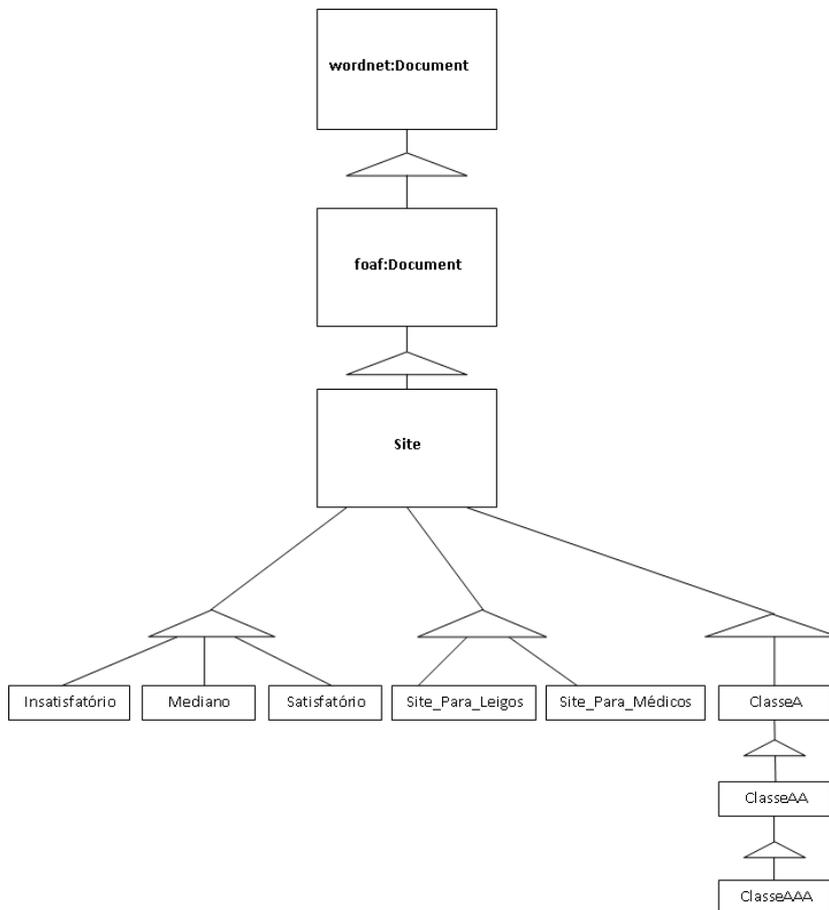


Figura 5: Classes de avaliação dos *sites*/páginas.

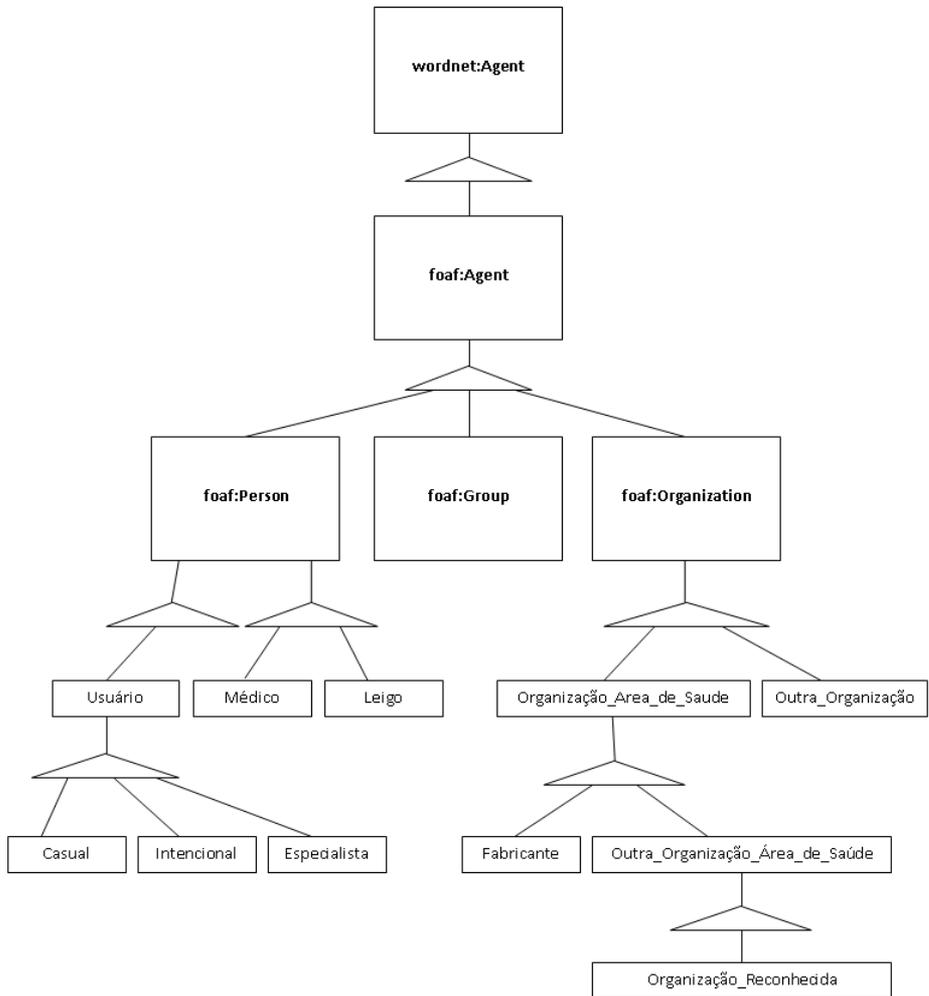


Figura 6: Classes relacionadas aos Agentes (*foaf:Agent*).

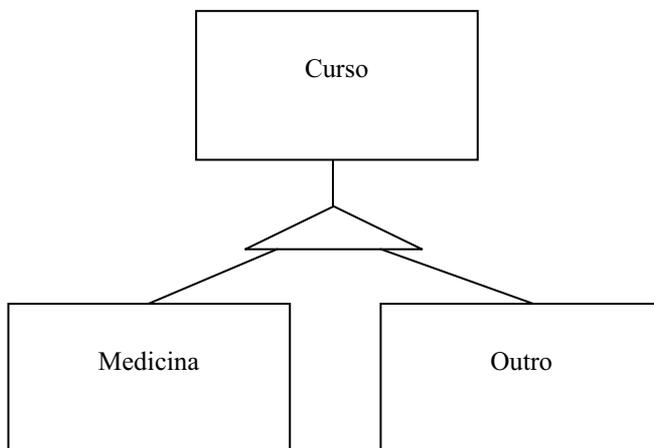


Figura 7: Classes relacionadas a cursos.

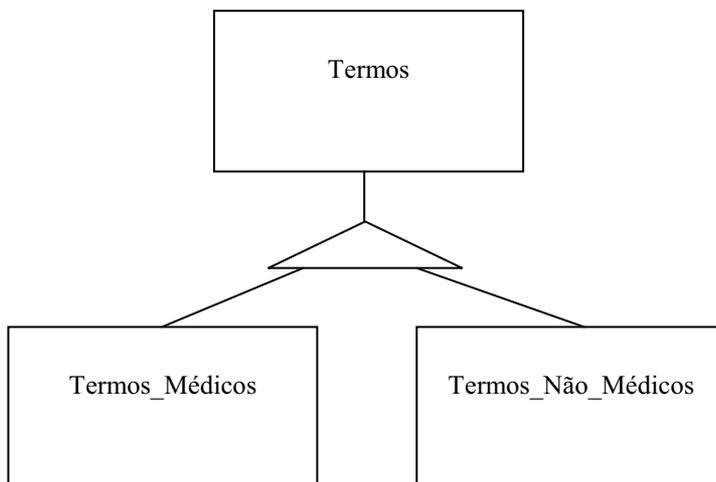


Figura 8: Classes relacionadas aos termos presentes em um *site*/página.

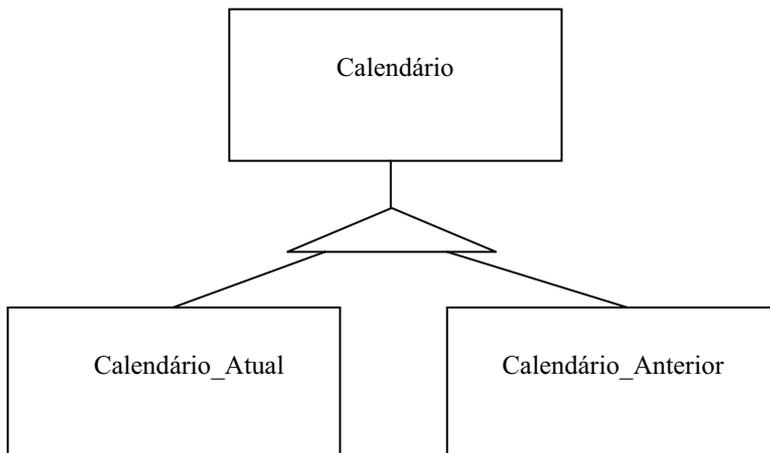


Figura 9: Classes relacionadas ao calendário.

Na Figura 7 são mostrados os cursos. As classes desta hierarquia são utilizadas para classificar uma pessoa como sendo Médico (ver propriedade *fezfazCurso* na seção 6).

A Figura 8 apresenta outra hierarquia de classes que classifica os termos presentes na ontologia entre termos médicos e termos não-médicos. Esta informação permite identificar se o *site/página* é voltado ou não para o público Leigo.

Por fim, a Figura 9 apresenta a hierarquia de classes que classifica datas como do calendário atual ou do calendário anterior. As datas são previamente armazenadas nestas classes com o objetivo de determinar se um *site* é atualizado.

6 Propriedades das Classes

A seguir são definidas as propriedades e *facets* relacionados a cada uma das propriedades das classes. Cabe ressaltar que em função de algumas limitações existentes no mecanismo de inferência utilizado *Protégé* e de forma a favorecer a inferência, algumas das definições apresentadas nesta seção foram modificadas.

Em especial cabe considerar aquelas propriedades que usam *DataType* e são usadas para inferência. No caso daquelas que possuem o *DataType Boolean*, estas propriedades foram transformadas para uma *Object property* na implementação da ontologia de forma a permitir a inferência.

Algo semelhante acontece com o atributo *Site:temDataAtualizacao* que está associado a classe *Calendário*. Isto foi feito objetivando facilitar o processo de inferência que não seria possível definindo a propriedade como *DataType*.

Outra questão é relacionada à cardinalidade. Propriedades com cardinalidade 0,1 ou 1 foram definidas como funcionais - *functional*.

Como as classes, as propriedades estão relacionadas aos termos presentes na seção 4.

Tabela 1. Propriedades da classe *Site*

Propriedades	Tipo (Range)	Cardinalidade	Obs.
<i>Site:URL</i>	<i>String</i>	1	
<i>Site:nomeSite</i>	<i>String</i>	1	
<i>Site:finalidade</i>	<i>Boolean</i>	0,1	
<i>Site:assunto</i>	<i>MeSH</i>	0,N	A ser incorporado. Não foi implementado na versão atual.
<i>Document:foaf:primaryTopic</i>	<i>MeSH</i>	0,1	A ser incorporado. Não foi implementado na versão atual.
<i>Site:responsavelDono</i>	<i>foaf:Agent</i>	0,1	
<i>Site:autor</i>	<i>foaf:Agent</i>	0,N	
<i>Site:temConselhoEditorial</i>	<i>foaf:Group</i>	0,1	
<i>Site:temPoliticaEditorial</i>	<i>Boolean</i>	0,1	
<i>Site:dadosContatoAutor</i>	<i>Boolean</i>	0,1	
<i>Site:dadosContatoResponsavel</i>	<i>Boolean</i>	0,1	

<i>Site:referenciaFontes</i>	<i>Boolean</i>	0,1	
<i>Site:publicoAlvo</i>	TipoPublico	1	Pode ser <i>PMedico</i> ou <i>PLeigo</i> . Esta classe foi criada para representar a indicação explícita, que existe eventualmente em um <i>site</i> , relacionada ao público ao qual é direcionado o <i>site</i> .
<i>Site:seloDeQualidade</i>	<i>Boolean</i>	0,N	
<i>Site:apoio</i>	<i>Foaf:Agent</i>	0,N	Patrocinadores
<i>Site:temPropaganda</i>	<i>Boolean</i>	0,1	
<i>Site:temPropagandaInseridaConteudo</i>	<i>Boolean</i>	0,1	
<i>Site:idiomaDocumento</i>	<i>Idioma</i>	1,N	
<i>Site: possuiTermos</i>	<i>Termos</i>	1,N	Os termos podem ser termos médicos ou não.
<i>Site:glossario</i>	<i>Boolean</i>	0,1	
<i>Site:artigosCientificos</i>	<i>Boolean</i>	0,1	
<i>Site:procedimentoCirurgico</i>	<i>Boolean</i>	0,1	
<i>Site:formulasDeMedicamentos</i>	<i>Boolean</i>	0,1	
<i>Site:descricaoDeEquipamentos</i>	<i>Boolean</i>	0,1	
<i>Site:controleDeAcesso</i>	<i>Boolean</i>	0,1	
<i>Site:comentariosUsuarios</i>	<i>Boolean</i>	0,1	

<i>Site:contribuicoesUsuarios</i>	<i>Boolean</i>	0,1	
<i>Site:descreveTratamento</i>	<i>Boolean</i>	0,1	
<i>Site:descreveDiagnostico</i>	<i>Boolean</i>	0,1	
<i>Site:descreveTratamentoAlternativo</i>	<i>Boolean</i>	0,1	
<i>Site:descreveFormasPrevencao</i>	<i>Boolean</i>	0,1	
<i>Site:descreveSintomas</i>	<i>Boolean</i>	0,1	
<i>Site:referenciaMedicamentos</i>	<i>Boolean</i>	0,1	
<i>Site:outLinks</i>	<i>Site</i>	0,N	
<i>Site:inLinks</i>	<i>Site</i>	0,N	
<i>Site:distincaoLinksInternosExternos</i>	<i>Boolean</i>		
<i>Site:pageRank</i>	<i>Integer</i>	0,1	
<i>Site:temDataAtualizacao</i>	<i>Calendario</i>	1	Esta propriedade aponta para datas classificadas como sendo do calendário atual ou do calendário anterior, o que permite determinar se o conteúdo foi ou não atualizado

Tabela 2. Propriedades da classe *Site* relacionadas à acessibilidade

Propriedades	Tipo (Range)	Cardinalidade	Obs.
<i>Site:textoAlternativoDadosNaoTextuais</i>	<i>Boolean</i>	1,1	
<i>Site:coresTextoFundoPossuemContraste</i>	<i>Boolean</i>	1,1	
<i>Site:siglasTemExplicacao</i>	<i>Boolean</i>	1,1	

Tabela 3. Propriedades da classe *foaf:Agent* e subclasses

Propriedades	Tipo (Range)	Cardinalidade	Obs.
<i>foaf:Person:nivelInstrucao</i>	<i>String</i>	1	
<i>foaf:Person:vinculo</i>	<i>foaf:Organization</i>	0,N	
<i>foaf:Person:idiomas</i>	<i>Idioma</i>	1,N	
<i>foaf:Agent:endereço</i>	<i>String</i>	0,N	
<i>foaf:Agent:telefone</i>	<i>String</i>	0,N	
<i>foaf:Person:nacionalidade</i>	<i>String</i>	0,1	
<i>foaf:Person:knows</i>	<i>Foaf:Person</i>	0,N	
<i>foaf:Group:member</i>	<i>Foaf:Group</i>	0,N	
<i>Foaf:Person:fezFazCurso</i>	<i>Curso</i>	0,N	

Os *sites/páginas* são classificados mediante processos de inferência nas classes mostradas na Figura 5. Para cada classe foram definidas as condições que devem ser satisfeitas para que um *site* seja classificado em cada uma das classes. Estas condições são apresentadas a seguir. Cabe salientar que as condições expressas nas classes não são exaustivas, servindo apenas para exemplificar como por meio das propriedades um *site/página* poderia ser classificado.

Inicialmente são consideradas as classes *Insatisfatório*, *Mediano*, *Satisfatório* que avaliam a qualidade intrínseca do *site/página*. As condições para um *site/página* ser classificado dentro de uma destas classes são apresentadas a seguir na Figura 10.

Sobre estas classes cabem algumas considerações:

- Foi assumido que a presença de um selo de qualidade (*HON Code*, por exemplo) é suficiente para determinar que a qualidade de um *site* ou página é satisfatória;
- Para permitir o processo de inferência foram criadas algumas classes. Por exemplo, foi criada uma classe *Sim* e outra *Não* para implementar propriedades que seriam *DataType* com *range=booleano*. Algo similar foi feito em relação a data de atualização. Existe uma classe *Calendário* que possui as subclasses *Calendario_Atual* e *Calendario_Anterior* onde são classificadas as datas.
- Outras propriedades poderiam ser consideradas para classificar um *site/página* dentro das classes. Dentre estas está o fato de que um *site* pode não ter a autoria do seu conteúdo declarada, mas pertencer a uma organização reconhecida o que poderia lhe aferir igual qualidade.

Ainda outras propriedades podem ser utilizadas para classificarem *sites/páginas* dentro destas classes relacionadas à qualidade. No caso da classe *Satisfatório*, isto é mostrado na Figura 10 que considera o fato de que o vínculo a uma organização reconhecida pode ser tão relevante quando a identificação da autoria.

Já as classes *Site_Para_Médicos* e *Site_Para_Leigos* objetivam classificar um *site/página* como sendo adequado a um tipo de público específico (ver Figura 11).

Insatisfatório

*(seloDeQualidade only Nao) and
(temDataAtualizacao only Calendario_Anterior) and
(autor only Leigo)*

Mediano

*(seloDeQualidade only Nao) and
(((autor only Leigo) and (temDataAtualizacao only Calendario_Atual)) or
((autor only Medico) and (temDataAtualizacao only Calendario_Anterior)))*

Satisfatório

*(seloDeQualidade only Sim) or
((autor only Medico) and (temDataAtualizacao only Calendario_Atual))*

Figura 10: Regras para inferência.

Site_Para_Leigos

*(possuiTermos only Termos_Nao_Medicos) or
(possuiTermos some Termos_Medicos) and (glossario some Sim)) or
(publicoAlvo only PLeigo)*

Site_Para_Medicos

*((possuiTermos some Termos_Medicos) and (glossario some Nao)) or
(publicoAlvo only PMedico)*

Figura 11: Regras para inferência - Perfil.

Cabem algumas observações relacionadas a classificação de um *site* como sendo voltado para profissionais da área ou para médicos:

- Eventualmente o tipo de público pode ser mais detalhado. Existem vários profissionais da área de saúde não somente médicos para os quais um *site* pode estar voltado. Além disto, mesmo o público leigo pode ser classificado segundo um diferente grau de conhecimento;
- Outro ponto é que embora possa existir a indicação no *site*/página acerca do público para o qual é voltado um *site*, isto nem sempre corresponde a realidade. Assim, um *site* pode estar com a indicação de ser voltado para o público leigo, mas isto não corresponder à realidade;

- Outras características podem ser consideradas. Um *site*/página para leigos não deveria conter, em princípio, artigos científicos, fórmulas de medicamentos, descrição de equipamentos e descrição de procedimentos cirúrgicos (no caso dos procedimentos cirúrgicos existem recomendações de que estes não sejam descritos em *sites* [21]).

7 Criação de Instâncias

A criação das instâncias é o último passo na definição de uma ontologia pelo Método 101. No caso da ontologia desenvolvida neste documento, as instâncias correspondem a *sites* da área de saúde. Para obtenção das instâncias foi usado um motor de busca (*Google*), e a palavra-chave "*alzheimer*", para busca dos *sites*/página.

Pela própria funcionalidade do motor de busca *Google*, são retornados os *sites*/páginas na ordem em que este reconhece como relevantes para pesquisa, assim, foram considerados os primeiros três *sites* apresentados como resultado.

A criação das instâncias foi feita de forma manual, uma vez que não foram desenvolvidas ferramentas para extração. Uma vez criadas as instâncias, usando os recursos do *Protégé*, foi utilizado o *Pellet* para realizar as inferências e classificar o *site*/página dentro das classes mostradas nas Figuras 5 e 6 . A seguir é apresentada a avaliação de dois *sites*/páginas que demonstram o uso da ontologia e dos critérios de qualidade nela definidos.

- <http://www.alzheimermed.com.br/>. Feita a extração dos dados, este *site*/página foi classificado na classe *Insatisfatório*. Apesar de referenciar um autor o *site* não é atualizado (a data de atualização é de 2005). O *site* foi classificado, corretamente, como *Site Para Leigos*, isto está explicitamente declarado. O *site*/página não atende critérios básicos de acessibilidade (não possui texto alternativo para imagens). Para fins de teste, foi alterada a data de atualização/revisão para 2008. Com isso, a ontologia inferiu o *site*/página como *Satisfatório*, pois a data de atualização é o único fator negativo, considerando os critérios da ontologia, do *site*/página.
- http://pt.wikipedia.org/wiki/Mal_de_Alzheimer. O *site*/página foi classificado como *Mediano*. Apesar de ser atualizado o autor não é declarado e o *site*/página não possui nenhum selo de qualidade associado. O *site*/página foi classificado como *Site Para Leigos*, uma vez que embora existam termos médicos existem, para a maioria destes, *links* que permitem seu entendimento (o que é esperado dentro da *Wikipedia*). Além disto, é preciso considerar que a *Wikipedia* dificilmente será utilizada por especialistas. O *site*/página não atende critérios básicos de acessibilidade (não possui texto alternativo para imagens, por exemplo).

8 Considerações Finais

Este artigo descreve o processo de criação de uma ontologia seguindo a Metodologia 101. A ontologia foi implementada utilizando o *Protégé* e testada mediante a extração manual de informações de *sites*/páginas retornadas pelo *Google*.

A partir do trabalho realizado, podem ser feitas algumas considerações relacionadas a sua continuidade:

- É preciso definir o que deve ser avaliado: uma página ou um *site*. Pelas avaliações parece claro que a qualidade global de um *site* deve ser considerada na avaliação de uma página pertencente a este *site*. É preciso considerar que os selos de qualidade são dados para *sites* não para páginas. Por outro lado, um usuário acessa normalmente uma página ou algumas páginas de um *site*;
- Alguns dos critérios de qualidade utilizados na versão da ontologia aqui descrita são dependentes do contexto no qual será feito uso das informações, e isto deve ser melhor avaliado em futuras versões da ontologia. Atualização, por exemplo, pode não ser um critério tão relevante se o que se está buscando são informações gerais sobre uma doença, por exemplo. Já para notícias sobre epidemias em curso, a atualização torna-se mais crítica;
- Ainda sobre critérios, avaliar *sites*/páginas sem autoria declarada como aquelas presentes na *Wikipedia* pode requerer o uso de outros critérios. Na *Wikipedia* o fornecimento de referências que dêem suporte ao conteúdo inserido é considerado indicador de qualidade. Assim, futuras versões da ontologia talvez possam incorporar este critério, que é denominado Verificabilidade – *Verifiability*, e que mede a proporção em que a informação é verificável e facilmente atribuída a uma fonte [23];
- No que se refere à implementação da ontologia com o uso de *Protégé* e *OWL* algumas características do denominado “mundo aberto” *Open World Assumption* – *OWA* [24] dificultaram a representação de alguns critérios e o processo de inferência;
- A criação de instâncias na ontologia mediante processo automático de extração de informação representa também um desafio a ser enfrentado na continuidade do trabalho. O problema é que, em princípio, o objetivo é extrair informações de quaisquer páginas, o que não é um processo fácil dado o grau de heterogeneidade existente na *Web*;
- Outro ponto se refere à incorporação de um mecanismo que permita mostrar a um usuário de um *browser* a avaliação feita. Neste sentido cabe considerar especialmente [8].

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq, SALUS Cyted e Pro-Sul AVAL-SAÚDE. À colega Mariusa Warpechowski pelo apoio dado no uso da ferramenta *Protégé* para construção da ontologia.

Referências

1. NOY, N. F. & MCGUINNESS, D. **Ontology Development 101: A Guide to creating your first Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, Stanford, 2001.
2. HONCODE. **Health on The Net Foundation**, 2008. Disponível em: <<http://www.hon.ch/>>. Acesso em: ago 2008.
3. WMA. **Web Médica Acreditada**, 2009. Disponível em: <<http://wma.comb.es/>>. Acesso em: ago 2008.
4. PROTÉGÉ. **Protégé Ontology Editor**, 2008. Disponível em: <<http://protege.stanford.edu/>>. Acesso em: ago 2008.
5. PELLET. **The Open Source OWL Reasoner**, 2008. Disponível em: <<http://clarkparsia.com/pellet>>. Acesso em: ago 2008.
6. BIZER, C.; CYGANIAK, R. Quality-driven information filtering using the WIQA policy framework. **Journal of Web Semantics** v. 7 n.1 p. 1–10, 2009.
7. WANG, R. Y.; STRONG, D. M. Beyond accuracy: what data quality means to data consumers. **J. Manage. Inf. Syst.**, Armonk, NY, USA, v.12, n.4, p.5–33, 1996.
8. BIZER, C. **Quality-Driven Information Filtering in the Context of Web-Based Information Systems**. 2007. Tese (Doutorado em Ciência da Computação)— Freie Universität FB Wirtschaftswissenschaft, Berlin, Germany.
9. W3C. WCAG - **Web Content Accessibility Guidelines 1.0**, 1999. Disponível em: <<http://www.w3.org/TR/WAI-WEBCONTENT/>>. Acesso em: nov. 2008.
10. W3C. **List of Checkpoints for Web Content Accessibility Guidelines 1.0**, 1999. Disponível em: <<http://www.w3.org/TR/WCAG10/checkpoint-list.html>>. Acesso em nov. de 2008.
11. FOAF. **Friend of a Friend Project**, 2008. Disponível em: <<http://www.foaf-project.org/>>. Acesso em: set 2008.
12. BRICKLEY, D.; MILLER, L. **FOAF Vocabulary Specification 0.91**. Disponível em: <<http://xmlns.com/foaf/spec/>>. Acesso em: set 2008.

13. RECTOR, A., L.; ROGERS, J., E.; POLE, P. The Galen High Level Ontology **Studies Health Technol. Informatics**, v. 34, p. 174–178, 1996.
14. SEIDENBERG, J.; RECTOR, A. Web ontology segmentation: analysis, classification and use. In: 15TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB. EDINBURGH, Scotland. **Proceedings..** ACM, 2006. p.13-22.
15. W3Ca. **Web Characterization Terminology & Definitions Sheet**, 1999. Disponível em: <<http://www.w3.org/1999/05/WCA-terms/>>. Acesso em: nov. 2008.
16. ROCHA, H. V. and BARANAUSKAS, M. C. C. Design e avaliação de interfaces humano-computador. NIED/UNICAMP, Campinas, SP, 2003.
17. KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. In: SODA '98: NINTH ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, 1998, Philadelphia, PA, USA. **Proceedings. . .** Society for Industria and Applied Mathematics, 1998. p.668–677.
18. BRIN, S.; PAGE, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. **Computer Networks**, [S.l.], v.30, n.1-7, p.107–117, 1998.
19. GARZOTTO, F., MAINETTI, L., and PAOLINI, P. Designing model hypermedia applications. In: EIGHTH ACM CONFERENCE ON HYPERTEXT. Southampton, United Kingdom. **Proceedings. . .** ACM, New York, NY, p. 38-47.
20. CEF – CONSELHO FEDERAL DE MEDICINA. **Resolução 1845 - convênio de reconhecimento de especialidades médicas**. Diário Oficial da União; Poder Executivo, Brasília, DF, 16 jul. 2008. Seção I, p. 164-8. Disponível em: <www.cremesp.org.br/library/modulos/legislacao/versao_impressao.php?id=7856>. Acesso em: nov. 2008.
21. CREMESP – CONSELHO REGIONAL DE MEDICINA DE SÃO PAULO. **Manual de Ética Médica na Internet – Anexo e Parecer**. 2001. Disponível em: <<http://www.cremec.com.br/medicinainternet.htm>>. Acesso em: dez. 2008
22. COVELLA, G. J.; OLSINA, L. A. Assessing quality in use in a consistent way. In: ICWE '06: 6TH INTERNATIONAL CONFERENCE ON WEB ENGINEERING, 2006, New York, NY, USA. **Proceedings. . .** ACM, 2006. p.1–8.
23. PIPINO, L.; LEE, Y.W.; WANG, R.Y. Data quality assessment. **Commun. of ACM**. NY, USA v. 45 n.4 p. 211–218, 2002.
24. GRIMM, S. and MOTIK. B. Closed-World Reasoning in the Semantic Web through Epistemic Operators. In: CEUR OWL EXPERIENCES AND DIRECTIONS WORKSHOP, 2005. Galway, Ireland, 2005 **Proceedings....**[s.l.], 2005.