

## Mapeamento terminológico no domínio da radiologia obstétrica: estudo estatístico a partir de ontologias

Terminological mapping in the domain of the obstetric radiology: a statistical study based on ontologies

Mapeamiento terminológico en el dominio de la radiología obstétrica: estudio estadístico a partir de ontologías

Igor Soares Amorim<sup>1,2,a</sup>

[amorim.igors@gmail.com](mailto:amorim.igors@gmail.com) | <https://orcid.org/0000-0003-2606-6000>

Eduardo Beckhauser<sup>1,2,b</sup>

[edubeckha@gmail.com](mailto:edubeckha@gmail.com) | <https://orcid.org/0000-0002-2001-5635>

Vinícius Andreóli Petrolini<sup>1,2,b</sup>

[vinicius.petrol@gmail.com](mailto:vinicius.petrol@gmail.com) | <https://orcid.org/0000-0003-1511-3465>

Alexandre Savaris<sup>1,c</sup>

[alexandre.savaris@gmail.com](mailto:alexandre.savaris@gmail.com) | <https://orcid.org/0000-0003-0161-6408>

Aldo von Wangenheim<sup>1,2,d</sup>

[awangenh@gmail.com](mailto:awangenh@gmail.com) | <https://orcid.org/0000-0003-4532-1417>

<sup>1</sup> Instituto Nacional de Convergência Digital, Laboratório de Telemedicina. Florianópolis, SC, Brasil.

<sup>2</sup> Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Ciência da Computação. Florianópolis, SC, Brasil.

<sup>a</sup> Mestrado em Ciência da Informação pela Universidade Federal de Santa Catarina.

<sup>b</sup> Graduação em Ciência da Computação pela Universidade Federal de Santa Catarina.

<sup>c</sup> Doutorado em Ciência da Computação pela Universidade Federal do Paraná.

<sup>d</sup> Doutorado em Ciência da Computação pela Universidade de Kaiserslautern.

### Resumo

Este artigo tem como base um estudo em que foram levantados e organizados termos do domínio da radiologia obstétrica e, então, foi identificado se os mesmos estão compreendidos em quatro distintos vocabulários controlados: OntoNeo, RadLex, LOINC e SNOMED. É apresentado o Sistema Integrado Catarinense de Telemedicina e Telessaúde (STT/SC) e o projeto de estruturação de laudos de exames de radiologia obstétrica, bem como o contexto teórico da ciência da informação sobre vocabulários controlados. Foram realizadas uma pesquisa de campo para o levantamento dos termos junto a um especialista da área e uma pesquisa documental para o levantamento estatístico dos termos em vocabulários controlados. Constituiu-se uma hierarquia dos termos levantados e verificou-se a cobertura de cada um dos vocabulários controlados em relação aos termos. O SNOMED é o vocabulário controlado com maior potencial de uso para a indexação de laudos no domínio da radiologia obstétrica.

**Palavras-chave:** vocabulário controlado; telemedicina; troca de informação em saúde; armazenamento e recuperação da informação; ultrassonografia; radiologia obstétrica.

## Abstract

This article bases on a study in which terms of the obstetric radiology domain were collected and arranged, and then we identified whether they are comprised in four distinct controlled vocabularies: OntoNeo, RadLex, LOINC and SNOMED. We present the STT/SC – Sistema Integrado Catarinense de Telemedicina e Telessaúde (Santa Catarina's integrated system of telemedicine and telehealth) and the project of structuring diagnostic reports of tests in obstetric radiology, as well as the theoretical context of information science about controlled vocabulary. We carried out a survey of the terms jointly an expert and a documentary research to the statistical survey of the terms from controlled vocabularies. A hierarchy of the terms collected was established and the coverage of each of the controlled vocabularies in relation to the terms was verified. The SNOMED is the controlled vocabulary with greater potential of use for the indexation of diagnostic reports in the field of obstetric radiology.

**Keywords:** controlled vocabulary; telemedicine; health information exchange; information storage and retrieval; ultrasonography; obstetric radiology.

## Resumen

Este artículo se basa en un estudio en el cual fueron levantados y arreglados términos del dominio de la radiología obstétrica, y entonces fue identificado si los mismos están comprendidos en cuatro distintos vocabularios controlados: el OntoNeo, el RadLex, el LOINC y el SNOMED. Presentamos el STT/SC – Sistema Integrado Catarinense de Telemedicina e Telessaúde (sistema integrado catarinense de telemedicina y telesalud) y el proyecto de estructuración de resultados de exámenes de radiología obstétrica, así como el contexto teórico de la ciencia de la información sobre vocabularios controlados. Una investigación de campo fue realizada para el levantamiento de los términos junto a un especialista y una investigación documental para el levantamiento estadístico de los términos en vocabularios controlados. Se ha constituido una jerarquía de los términos levantados y se ha verificado la cobertura de cada uno de los vocabularios controlados en relación a los términos. El SNOMED es el vocabulario controlado con mayor potencial de uso para la indexación de los resultados de exámenes en el dominio de la radiología obstétrica.

**Palabras clave:** vocabulario controlado; telemedicina; intercambio de información en salud; almacenamiento y recuperación de la información; ultrasonografía; radiología obstétrica.

## INFORMAÇÕES DO ARTIGO

**Contribuição dos autores:** Concepção e desenho do estudo: Aldo von Wangenheim e Alexandre Savaris.  
Aquisição, análise ou interpretação dos dados: Igor Soares Amorim, Eduardo Beckhauser e Vinícius Andreóli Petrolini.  
Redação do manuscrito: Igor Soares Amorim, Eduardo Beckhauser e Vinícius Andreóli Petrolini.  
Revisão crítica do conteúdo intelectual: Aldo von Wangenheim, Alexandre Savaris e Igor Soares Amorim.

**Declaração de conflito de interesses:** Não há.

**Fontes de financiamento:** Secretaria de Estado da Saúde de Santa Catarina - SES/SC.

**Considerações éticas:** Não há.

**Agradecimentos/Contribuições adicionais:** Não há.

**Histórico do artigo:** Submetido: 13.jun.2017 | Aceito: 16.fev.2018 | Publicado: 30.mar.2018

**Apresentação anterior:** Não há.

**Licença CC BY-NC atribuição não comercial.** Com essa licença é permitido acessar, baixar (download), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

## Introdução

No Brasil, a área interdisciplinar de informação e saúde tem se desenvolvido principalmente a partir do final da década de 1980, influenciada por questões tecnológicas (inserção de tecnologias de informação e comunicação nas instituições de saúde), pela descentralização postulada pela Constituição de 1988, e pela criação de mecanismos de integração entre as esferas municipal, estadual e federal. Tudo isso foi acompanhado da demanda por maior informação, resultando na geração, em 1991, do Datusus (Departamento de Informática do Sistema Único de Saúde) para apoiar o Sistema Único de Saúde (SUS), o qual promove cada vez mais o acesso facilitado a um grande volume de dados de saúde. O desenvolvimento da área implica novas demandas práticas e teóricas<sup>1</sup>.

Nesse contexto, compreendemos o Sistema Integrado Catarinense de Telemedicina e Telessaúde (STT/SC), que visa aplicar tecnologias de informação e comunicação para que processos se tornem mais eficientes. Muitas de suas iniciativas focam a comunicação de serviços de imagem, como ultrassonografia, tomografia computadorizada, ressonância magnética e raio X. O STT/SC provê infraestrutura tecnológica ao núcleo de Telessaúde de Santa Catarina, integrante do programa Telessaúde Brasil Redes do Ministério da Saúde, promovendo serviços de teleconsultoria, tele-educação, telediagnóstico, regulação regional e segunda opinião formativa. Atualmente, todos os municípios de Santa Catarina têm acesso ao STT/SC, e o sistema é mantido pelo Instituto Nacional de Ciência e Tecnologia para Convergência Digital (INCoD) do Departamento de Informática e Estatística (INE) do Centro Tecnológico (CTC) da Universidade Federal de Santa Catarina (UFSC)<sup>2</sup>.

O STT/SC busca a aplicação do padrão Digital Imaging and Communications in Medicine (DICOM) para viabilizar a normalização de procedimentos, da comunicação e do armazenamento de documentos (imagens e/ou laudos estruturados). O padrão DICOM define guidelines para a aquisição, a comunicação e o armazenamento de imagens e laudos, facilitando a interoperabilidade de dados entre sistemas e equipamentos médicos. Tais dados necessitam ser acessíveis e, para tanto, faz-se necessário um sistema de indexação dos mesmos, considerando o uso combinado de diferentes vocabulários controlados.

O intuito da pesquisa aqui apresentada é viabilizar a interoperabilidade entre vocabulários controlados, de forma a contemplar a indexação e a recuperação de laudos da radiologia obstétrica armazenados pelo STT/SC. Para tanto, objetivamos levantar uma terminologia relativa ao escopo da radiologia obstétrica, sistematizar esse conjunto de termos e mapear a ocorrência destes em vocabulários controlados correntes especificamente na área de radiologia e, de modo mais geral, na área da saúde. Sua justificativa se dá pela otimização dos processos médicos relacionados à emissão de laudos, além da potencialidade de uso dos laudos indexados para pesquisa e produção de conhecimento científico. Ainda, prezamos a interoperabilidade como via mais econômica, visível quando se verifica o dispêndio de recursos necessários à criação de um vocabulário específico para cada subárea contemplada pelo STT/SC.

As duas próximas seções são relativas ao referencial teórico, explicando o que é o STT/SC e o padrão utilizado na estruturação dos laudos médicos e abordando, em seguida, noções e tipologias de vocabulários controlados. Na sequência, uma seção apresenta os procedimentos metodológicos que caracterizam a pesquisa aqui tratada e descreve como o estudo foi conduzido. Em seguida, são apresentados os resultados e discussões numa seção em que se descreve a elaboração de uma estrutura hierárquica de termos do domínio da radiologia obstétrica e se realiza uma comparação destes com os termos recuperados em ontologias mantidas no BioPortal, selecionadas em função dos objetivos da pesquisa. Por fim, são tecidas considerações sobre os resultados obtidos, as limitações deste estudo e a importância de pesquisas futuras.

## Informação e saúde: padronizações em conformidade com o Sistema Integrado Catarinense de Telemedicina e Telessaúde

O STT/SC foi criado em meados de 2010 com a intenção de unir dois serviços fornecidos no estado: Telemedicina e Telessaúde. Oferece a possibilidade de realização de exames, diagnósticos, cursos, teleconsultoria, conferências à distância via web, além de permitir o acesso dos pacientes aos seus dados.

Telemedicina define-se pela “adoção e utilização de recursos de telecomunicação e outras tecnologias para o compartilhamento de informações médicas referentes a um dado paciente”<sup>3</sup>, e caracteriza-se pela utilização da tecnologia em serviços médicos que podem superar barreiras geográficas e temporais. Dentre esses serviços, destacam-se o telediagnóstico e a realização de exames de telerradiologia. Quando os mesmos são aplicados corretamente, a telemedicina é vantajosa ao paciente por permitir que sua interação com os médicos seja à distância, sem que precise se locomover para ser atendido.

Telessaúde é o uso da tecnologia de comunicação para a realização dos mais diversos serviços à distância. Como é voltada para a área médica, a mesma possibilita a comunicação e difusão de informação entre os mais variados tipos de profissionais da saúde por meio do acesso remoto a recursos. As principais atividades desenvolvidas são: “a) teleconsultoria: tipo de consultoria realizada entre especialistas que, a partir de um sistema de cadastro de perguntas e respostas, permite a retirada de dúvidas entre dois ou mais profissionais da saúde; b) tele-educação: educação à distância a partir da mediação de recursos didáticos devidamente organizados, apresentados nos mais diversos parâmetros de informação; c) telediagnóstico: utilização de tecnologias para a realização de diagnósticos à distância”<sup>4</sup>. Este serviço se caracteriza pela emissão de laudos de exames sem que o médico especialista esteja presente no momento da execução do exame.

O STT/SC tem encontrado receptividade junto ao Sistema Único de Saúde e ao Ministério da Saúde, que procura desenvolver a telessaúde desde 2006. O sistema conecta cidades de todo o estado de Santa Catarina, permitindo a otimização de processos variados, tais como exames e conferências. O sistema é viabilizado pela parceria entre a Secretaria de Saúde do Estado de Santa Catarina (SES/SC) e o INCoD (Instituto Nacional de Ciência e Tecnologia para Convergência Digital) vinculado à Universidade Federal de Santa Catarina (UFSC), sendo desenvolvido pelo laboratório de Telemedicina do Hospital Universitário Polydoro Ernani de São Thiago (HU/UFSC). Os serviços do STT/SC concentram-se na coleta, no armazenamento e na disponibilização de dados e informações referentes a exames provenientes de instituições de saúde sediadas em diferentes municípios do estado<sup>3</sup>. Tais serviços podem ser classificados em quatro processos: a) execução de exames e armazenamento centralizado de seus dados em um repositório; b) disponibilização dos dados para avaliação por especialistas; c) armazenamento de laudos e das opiniões emitidas pelos especialistas sobre esses documentos; d) disponibilização dos resultados das avaliações aos pacientes.

Nesse cenário, a comunicação no STT/SC é fundamental e, neste sentido, a ciência da informação pode corroborar a organização e recuperação dos documentos, além de possibilitar padronizações terminológicas. Este artigo objetiva subsidiar os processos de comunicação no domínio da radiologia obstétrica e a pesquisa que o fundamenta pretende corroborar a estruturação do laudo nesse domínio.

Documentos semiestruturados ou não-estruturados são geralmente de livre preenchimento e idealização, e são de simples utilização. Contudo, há uma série de dificuldades quando se implanta o preenchimento de documentos não-estruturados em sistemas específicos, tais como: questões de padronização terminológica, a suscetibilidade a erros de grafia, a dificuldade de recuperar informações e a dificuldade de gerir os dados contidos nesses documentos e inferir sobre eles.

Documentos estruturados, por outro lado, remetem ao conceito de estruturas organizadas de informação que facilitam a busca do conteúdo, a comparação entre documentos e resultados e a implantação de serviços que ajudam a tomada de decisão quando esses documentos estão devidamente conectados a uma área

específica. Todo o seu processo de geração até a fase em que são analisados é facilitado pela forma como esse tipo de documento é gerado<sup>5</sup>.

Nesse contexto, laudo estruturado é um documento médico preenchido de acordo com um padrão, predisposto pelo profissional da saúde que o utilizará ou pela instituição na qual ele trabalha, e trará benefícios à comunicação e à pesquisa relacionada ao laudo. Além disso, esses laudos podem ser recuperados, tratados e apresentados de diferentes formas, pois toda a informação neles contida é baseada em dicionários de terminologia controlada que podem ser facilmente identificados em nível sintático e semântico<sup>6</sup>.

O laudo estruturado, no sentido geral de organização, pode ser definido como um conjunto de módulos<sup>5,6,7,8</sup>:

- Editor de dicionários: responsável pela visualização e edição de dicionários de termos e unidades de medida. Como será descrito, cada item é composto por um par nome-valor, no qual o nome é oriundo de um vocabulário de termos. No caso do foco no qual este trabalho está inserido, o editor de dicionários não será necessário, pois o usuário terá como base todo o conjunto de possíveis vocábulos à sua disposição.
- Editor de templates: permite que usuário construa a hierarquia de modelo de laudo e especifique, para cada item da hierarquia, seu valor, a origem deste valor (dicionário escolhido) e a sua relação com o nodo pai. Depois de construída a hierarquia, esse template é salvo em formato XML (eXtensible Markup Language) e será utilizado na etapa de edição de laudos.
- Editor de laudos: a partir do XML gerado no editor de templates, realizamos a criação de uma interface de forma dinâmica que permite o preenchimento do laudo por parte do médico. Esse preenchimento é salvo em formato DICOM SR (Structured Report), o qual pode futuramente ser visualizado e impresso.

Finalmente, com a estrutura que armazena esse volume de informações, podem-se criar diversos mecanismos facilitadores para o dia a dia do profissional e paciente, tais como:

- padronização de um grande sistema de saúde conforme o mesmo mecanismo de geração de laudos, fazendo com que um laudo possa ser gerado e lido em quaisquer centros de saúde que aderirem à ferramenta;
- eliminação de incertezas inerentes à língua na emissão dos laudos, facilitando futuros diagnósticos e leituras de históricos por parte dos médicos;
- precisão na leitura dos resultados do laudo por parte do paciente, desaparecendo assim com quaisquer dúvidas que possam ser geradas no diagnóstico por texto livre;
- mecanismos de apoio à tomada de decisão, como artifícios de data warehouse, que facilitam a criação dos indicadores citados, não abrangidos pelo modelo operacional de banco de dados<sup>7</sup>. Em outras palavras, a geração de laudos estruturados contribui com a construção de uma ótima base para a realização de estatísticas sobre os dados adquiridos.

Para a estruturação da quantidade potencialmente grande de dados que um laudo estruturado pode gerar, é importante considerar a utilização de algum padrão que resolva os seus impasses perante a formação, a organização e o arquivamento das informações nele adquiridas. Além disso, esse padrão deve dar ao usuário flexibilidade na hora de realizar o procedimento, mas mantendo a estrutura concisa e adequada para posterior manipulação.

O Digital Imaging and Communications in Medicine (DICOM) é um padrão utilizado para a geração, comunicação e arquivamento de imagens médicas digitais. Este padrão estabelece regras para troca de informações médicas entre diversos tipos de equipamentos de imagem e sistemas, estabelecendo uma linguagem comum baseada em protocolos.

No âmbito geral da pesquisa aqui apresentada, nos interessa mais especificamente a extensão chamada DICOM Structured Report (SR), que define especificamente como devem ser constituídos itens de informação que codificam dados a respeito de exames, diagnóstico e tratamento, além das informações inerentes ao contexto de aplicação, como procedimentos que devem ser executados para o sucesso do tratamento concedido, além de dados sobre os profissionais de saúde envolvidos nos trâmites do processo. Um dito objeto no padrão DICOM pode, então, conter referências a imagens de exames ou outros objetos representados no mesmo padrão<sup>8</sup>.

Uma das principais características que tornam o DICOM SR um padrão robusto é a utilização de terminologias controladas a fim de evitar ambiguidade inerentes à língua, facilitar o entendimento e a recuperação de informações de forma estruturada e simplificada. Um documento DICOM SR, portanto, se caracteriza pela presença de listas com relacionamentos hierárquicos, pelo uso de conteúdo codificado, pelos relacionamentos entre conceitos e pela presença de referências embutidas para imagens e objetos<sup>8</sup>.

Os itens recebem então um indexador codificado. Cada item apresenta no padrão DICOM três faces: seu nome, o qual é um termo para entendimento humano; uma apresentação esquemática; e um código responsável por sua identificação<sup>8</sup>.

Assim o padrão DICOM permite o emprego de vocabulários controlados distintos no mesmo grupo de documentos, pois a cada termo confere um identificador único. Deste modo, possibilita o compartilhamento e o reuso de vocabulários controlados. Buscando uma melhor adequação semântica, a pesquisa em pauta concentrou esforços no mapeamento terminológico de quatro vocabulários específicos: RadLex, LOINC, SNOMED e OntoNeo.

## Vocabulário controlado: indexação e classificação

Tornar pública as informações contidas em registros documentais por meio de algum símbolo intermediário é uma tarefa executada desde mais de dois mil anos antes de Cristo. Os símbolos intermediários “organizam em conjuntos finitos inúmeras informações que se relacionam de alguma forma, mas que se encontram dispersas em diferentes publicações”<sup>9</sup>. Era necessário um sistema que pudesse acomodar todos esses intermediários, que eram representações dos registros documentais, pois somente assim o acesso à informação poderia se consolidar de maneira ampla e eficaz. Dessa forma, foram desenvolvidos os sistemas classificatórios bibliográficos, sobretudo entre os anos de 1850 e 1950, quando essa arte adquiriu escopo científico próprio<sup>10</sup>. Inicialmente eram sistemas universais de classificação, que buscavam um ponto de partida neutro para o estabelecimento de taxonomias apriorísticas. Atualmente, a crença na neutralidade está em questão.

Dewey, nos Estados Unidos, e Otlet, na Europa, propuseram no final do século XIX sistemas hierárquicos que pretendiam englobar todo o saber humano, sistematizado em classes decimais e com profunda ramificação. Além de categorizar as representações intermediadoras, tais sistemas seriam utilizados também para ordenar o acervo físico, determinando lugares específicos e demarcados por esses mesmos intermediários, definidos como códigos que designam uma posição permanente dentro da hierarquia dos conhecimentos. Inicialmente, a classificação decimal universal (CDU), de Otlet, não pretendia exatamente isso, mas fundamentalmente, indexar os diferentes assuntos presentes em um único documento, tornando possível a recuperação da informação fragmentada, e não pela unidade documental. Com o cenário pós-Segunda Guerra e a disputa entre as duas potências na Guerra Fria, o desenvolvimento econômico, social, científico e militar necessitava ser acelerado. Um forte investimento estatal que integrava todas essas frentes foi ao que se assistiu. Assim, a comunicação científica tornou-se elementar na corrida entre Estados Unidos e União Soviética. Novas ferramentas eram exigidas para o aperfeiçoamento do tratamento e da recuperação

da informação científica. Essas demandas encontram apoio nos avanços das tecnologias eletrônicas de comunicação, que fora uma das responsáveis pela invenção do computador pessoal e da internet.

Com o aprofundamento das especializações científicas, as classificações que se limitavam às relações hierárquicas e de cunho universalista já não respondiam a essa nova realidade. Nos últimos 30 anos do século XX, a biblioteconomia e a ciência da informação voltaram a atenção para os serviços especializados, buscando tornar a comunicação na pesquisa básica e aplicada mais eficiente. Nesse ensejo, esse campo de atividade incorporou de vez o termo ‘informação’, sobretudo nos países anglo-saxões, na América Latina e África, o que contribuiu para a legitimação da área enquanto ciência. Estudos da época buscavam melhores formas de organizar as informações não só de livros, mas de patentes, de relatórios técnicos e principalmente de artigos de periódicos científicos. Assim, surge o tesouro, novo instrumento classificador, mais flexível e que permite combinações entre conceitos de diferentes ramos. Nesse instrumento, a relação entre os termos não é apenas hierárquica, mas também associativa; essas relações também não eram mais universais, mas específicas, o que representou um ganho na precisão nos processos de indexação e recuperação da informação. Além disso, os tesouros permitiram que os termos de busca fossem combinados no momento da execução da mesma pelo próprio usuário. Era um avanço importante, que facilitava a rotina de usuários/cientistas que poderiam ganhar tempo em suas investigações. Nessa altura dos acontecimentos, as buscas já eram feitas em sistemas automatizados, baseados em lógica booleana.

Com o avanço das últimas décadas na área da inteligência e da semântica artificial, novos instrumentos despontam como uma realidade no século XXI: as redes semânticas e as ontologias. Estas mantiveram as especificidades, contudo tornam o processo de recuperação da informação ainda mais inteligente com mecanismos de inferências lógicas que aprimoram os resultados de busca de informação. Atualmente, busca-se nas redes semânticas uma forma de recuperar informações por meio das relações entre conceitos que se estabelecem em nó, e não necessariamente em hierarquia. As ontologias seguem os mesmos direcionamentos, contudo seu horizonte é viabilizar o estabelecimento da web semântica.

A ontologia é considerada um vocabulário controlado de conceitos e suas relações, o qual representa um dado domínio; contudo, somam-se a isso restrições e caracterizações sobre os termos para que estes sejam compreensíveis não apenas a humanos, mas também às máquinas. Assim, as ontologias são compreendidas como elementos fundamentais na construção de uma web capaz de portar semântica sobre os dados e compartilhá-los<sup>11</sup>.

Há sob todo vocabulário controlado um sistema de conceito com algum grau de complexidade. Os sistemas com baixo grau de complexidade não compreendem grande variabilidade de relações entre os conceitos. Um sistema pode mesmo ser uma listagem alfabética de termos, sendo a única regra um critério de seleção que define quais conceitos compõem o sistema. Sistemas mais complexos incorporam outras regras, como as relações hierárquicas, relações não-hierárquicas ou ainda aspectos pragmáticos que delimitam o conceito em um contexto de uso específico.

É entendido que “a qualidade obtida na recuperação da informação depende substancialmente dos procedimentos e instrumentos utilizados para organização da informação”<sup>12</sup>. Portanto, a utilização de um vocabulário controlado mais adequado às necessidades de informação dos usuários de um sistema de informação se faz imprescindível. Entre os tipos possíveis, Bräscher e Carlan<sup>12</sup> reportam-se aos mais usuais: a classificação, os tesouros, as taxonomias e as ontologias.

A classificação é constituída pelo ato de classificar, ou seja, pela ação de ordenar os elementos diversos em agrupamentos estabelecidos por qualidades de semelhanças e diferenças específicas. Destacamos assim as classificações bibliográficas que se caracterizam como “sistemas predeterminados de conceitos logicamente estruturados e acompanhados de um código identificador”<sup>12</sup>. Os tesouros “são instrumentos de controle terminológico, utilizados em sistemas de informação para traduzir a linguagem dos documentos, dos indexadores e dos pesquisadores em uma linguagem controlada, usada na indexação e na recuperação

de informações”<sup>12</sup>. Taxonomias são instrumentos de organização lógica de conteúdos informacionais por meio de classes e subclasses, gerando metadados e padrões que são herdados na hierarquia, podendo ser utilizada para a organização de objetos e conhecimentos, visando dar suporte à navegação em ambiente web, ou ainda como instrumento de auxílio à gestão do conhecimento, sempre enfocando a recuperação da informação. Por fim, ontologias definem “conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual e promovem e facilitam a interoperabilidade entre sistemas de informação, em um processo ‘inteligente’ dos agentes (computadores)”<sup>12</sup>; por isso, a ontologia deve se apresentar de maneira formal e legível por computadores, viabilizando o reuso e o compartilhamento de sua estrutura.

No âmbito da saúde, há diversas iniciativas de vocabulários controlados. Muitos desses podem ser encontrados no BioPortal, uma plataforma de compartilhamento e reuso de ontologias desenvolvida pelo National Center for Biomedical Ontology (NCBO), instituição que objetiva apoiar o desenvolvimento dos conhecimentos biomédicos, oferecendo soluções web que promovam pesquisas científicas. O BioPortal é uma biblioteca de terminologias e ontologias biomédicas que visa facilitar o reuso de ontologias<sup>13</sup>. Atualmente o BioPortal é composto por 554 ontologias.

Este trabalho está compreendido no domínio da radiologia obstétrica; assim, buscamos alguns vocabulários no BioPortal que pudessem melhor abranger este domínio, conforme descrito na próxima seção. Há uma tese defendida recentemente por Juliatti de Andrade<sup>14</sup> que se assemelha à pesquisa realizada neste artigo.

## Procedimentos metodológicos

A abordagem da pesquisa aqui apresentada é qualitativa e quantitativa, na medida em que, por meio de trabalhos de estruturação de uma hierarquia junto a um especialista, recuperamos estruturas de vocabulários controlados e realizamos uma análise estatística baseada na concordância terminológica. É de natureza aplicável e de objetivos exploratórios, uma vez que propomos uma hierarquia e exploramos vocabulários selecionados no BioPortal. Ainda quanto aos procedimentos ela é considerada pesquisa de campo, pois constituiu uma seleção de termos e um arranjo segundo consultas a um especialista, e pesquisa de fontes documentais, já que exploramos vocabulários existentes na área médica<sup>15, 16</sup>.

Com o intuito de identificar aspectos semânticos e diferentes vocabulários controlados que podem sanar as necessidades semânticas de indexação de laudos na radiologia obstétrica, desenvolvemos a pesquisa em duas etapas principais: a primeira, na qual foi realizada a pesquisa de campo, e a segunda, baseada na pesquisa documental.

Em um primeiro momento, foram realizadas algumas reuniões que serviram para harmonizar nossas intenções com as expectativas do especialista consultado. No decorrer desses encontros, entramos em contato com uma base de dados, desenvolvida pelo próprio especialista, que continha em suas tabelas uma terminologia rica e focava o domínio da ginecologia e da obstetrícia. Foram mapeados os principais termos e, em conjunto com o especialista, selecionamos aqueles relacionados aos exames obstétricos e os organizamos em uma hierarquia, conforme se vê no Apêndice B.

Após essa definição (que passou pela validação da hierarquia pelo especialista), partimos para a segunda etapa, na qual buscamos os termos em vocabulários controlados recuperados no BioPortal. Definimos quatro vocabulários para consulta: RadLex (por ser o principal no âmbito da radiologia), o LOINC (por ser um vocabulário transversal e que contempla diversas áreas), o SNOMED (por sua abrangência) e o OntoNeo (pelo escopo específico e compatível com o nosso foco). O objetivo é analisar quão da nossa hierarquia é contemplado nos vocabulários eleitos no BioPortal. Para tanto, além da identificação dos termos, realizamos uma análise estatística sobre tais resultados. A seguir detalhamos os resultados.

## Resultados e discussões

Como já mencionado, para realização da primeira etapa da pesquisa de campo, foi organizada uma primeira reunião com um profissional específico, um médico especialista em ginecologia e obstetrícia atuante no Hospital Universitário da UFSC. O médico relatou a necessidade de um sistema automatizado na elaboração de laudos obstétricos, grifando que grande parte das informações contidas em tais laudos são comparações com medidas-padrão. Isso facilitaria a automatização dos laudos que, após estruturados, podem evidenciar as informações padronizadas que são referência para os médicos avaliarem os exames obstétricos.

O especialista nos disponibilizou um arquivo elaborado por ele próprio, um banco de dados em Microsoft Access, o qual contempla toda uma nomenclatura de uso corrente. A partir desse programa, foram listados os termos (Apêndice A). Em seguidas reuniões com o mesmo especialista, os termos foram filtrados. Foram eliminados os que eram relativos à ginecologia e aqueles relativos às patologias (que não estão contempladas no planejamento do atual estágio da pesquisa em pauta). Assim, organizamos a hierarquia de termos validados pelo especialista (Apêndice B).

A hierarquia formulada (Apêndice B) tem uma estrutura já direcionada à estruturação do laudo. Ela apresenta cinco categorias gerais que são subdivididas em subclasses:

- **Morfologia:** esta categoria destina-se às subclasses relativas às medidas do feto, comumente abordadas como sinais do seu desenvolvimento.
- **Doppler:** categoria que contempla as medidas de fluxos que são realizadas a partir dos equipamentos Doppler.
- **Estática e bem-estar fetal:** relativa ao bem-estar geral do feto, como condição muscular (tônus), batimentos cardíacos e posição.
- **Anexo fetal:** trata dos elementos no entorno do feto, verificando as condições essenciais à saúde do mesmo.
- **Exame:** dados sobre o exame, relativos aos modos e motivos de aplicação, à sua qualidade, e ao número de fetos.

Na formulação da hierarquia foi tomada uma decisão a fim de tornar o processo de busca dos termos nos vocabulários controlados previamente selecionados mais efetiva. Optamos pela decomposição dos termos em partes menores, porque foi constatado que a busca realizada no BioPortal é dotada de uma inteligência que, a partir da menor parte do termo, sugere outros termos alternativos que apresentam terminologia semelhante. O processo de decomposição é evidenciado por uma comparação entre o apêndice A e B. Com isso, obtivemos resultados mais efetivos na busca, possibilitando recuperar informações consideradas aproximadas.

Tal decomposição dos termos resultou em uma ramificação das subclasses findada em termos conclusivos. Assim, observamos que, em diversos ramos, os elementos chamados de folha na estrutura terminam em termos ‘alterado’ ou ‘normal’, tal como no exemplo extraído do Apêndice B:

Quadro 1. Fragmento da hierarquia 1

A2.3	biparietal
A2.3.1	diâmetro
A2.3.1.1	alterada
A2.3.1.2	normal
A2.4	cefálica
A2.4.1	circunferência
A2.4.1.1	alterado
A2.4.1.2	normal

Fonte: Os autores (2018).

No Quadro 1 notamos que a medida da circunferência cefálica pode estar alterada ou normal. Quando alterada, interessa saber qual o motivo e o nome da patologia a ser tratada.

No atual estágio do trabalho, não exploramos os termos relativos às patologias, o que será feito em uma próxima etapa. Por isso, mantemos por hora a nomenclatura ‘alterado’ e ‘normal’, o que acontece nas subclasses A1.2.1.1; A1.2.1.2; A1.3.1.1; A1.3.1.2; A1.4.1.1; A1.4.1.2; A1.5.1.1; A1.5.1.2; A2.1.1.1; A2.1.1.2; A2.2.1.1; A2.2.1.2; A2.3.1.1; A2.3.1.2; A2.4.1.1; A2.4.1.2; A2.5.1.1; A2.5.1.2; A2.5.2.1; A2.5.2.3; A2.7.1.1; A2.7.1.2; A2.8.1.1; A2.8.1.2; A2.9.1.1; A2.9.1.2; A2.10.1.1; A2.10.1.2; A2.11.1.1; A2.11.1.1; A2.12.1.1; A2.12.1.2; A2.13.1.1; A2.13.1.2; A2.14.1.1; A2.14.1.2; A2.15.1.1; A2.15.1.2; B1.1.1; B1.1.2; B1.2.1; B1.2.2; B2.1.1; B2.1.2; B2.2.1; B2.2.2; B3.1.1; B3.1.2; B3.2.1; B3.2.2; B4.1.1; B4.1.2; B4.2.1; B4.2.2; C8.1; C8.2; D1.2.1.4; D1.2.1.5.

O mesmo se aplica aos termos ‘ausente’ e ‘presente’, conforme as subclasses C2.1; C2.2; C5.1; C5.3; C7.1.1; C7.1.2; C7.2.1; C7.2.2; com os termos ‘regular’ e ‘irregular’ designados pelos códigos A2.6.1.1; A2.6.1.2; C5.2; C5.4; e, por fim, os termos ‘aumentado’ e ‘diminuído’, que se encontram em D2.1.5; D2.1.6 (ver Apêndice B).

Além disso, o processo de decomposição nos permitiu a identificação de termos que poderão ser considerados como ‘modificadores’ na estruturação do laudo, a depender da avaliação das vantagens e desvantagens de incluir tais termos como indexadores presentes em vocabulários controlados ou como elementos que condicionam os substantivos designadores dos campos. Nesse conjunto, os mais recorrentes são ‘comprimento’, que ocorre 12 vezes, encontrando-se nas subclasses A1.5.1; A2.1.1; A2.2.1; A2.7.1; A2.8.1; A2.9.1; A2.10.1; A2.11.1; A2.12.1; A2.13.1; A2.14.1; A2.15.1; ‘pulsatilidade’, que aparece em B1.1; B2.1; B3.1; B4.1, ou seja, quatro vezes; os termos ‘distância’ e ‘resistência’, que se revelam respectivamente em três subclasses, A1.2.1; A1.3.1; A1.4.1 e B1.2; B2.2; B4.2, portanto, três vezes cada um. Ainda há outros termos com apenas uma ocorrência e que se enquadram nesse conjunto, tal como ‘diâmetro’ (A2.3.1), ‘condição’ (A2.6.1) e ‘circunferência’ (A2.4.1) (ver Apêndice B).

Os conceitos principais da hierarquia proposta se concentram nas classes de segundo nível da classificação expostas no Quadro 2, pois são dessas classes que as outras informações são atributos, derivadas ou são as classes às quais elas se referem.

Quadro 2. Fragmento da hierarquia 2

A	Morfologia
A1	Medida Facial
A2	Biometria Fetal
B	Doppler
B1	umbilical
B2	cerebral
B3	Ducto Venoso
B4	Artéria uterina
C	Estática e Bem-estar Fetal
C1	Dorso
C2	tônus
C3	Apresentação
C4	Posição
C5	Batimento Cardíaco fetal
C6	Cardiotocografia
C7	Movimento
C8	Genital
D	Anexo Fetal
D1	Placenta
D2	Líquido amniótico
E	Exame
E1	via de exame
E2	qualidade de exame
E3	Gestação
E4	Indicação

Fonte: Os autores (2018).

Embora as informações principais se refiram às classes expostas no Quadro 2, destacamos ainda as subclasses derivadas de A1 – ‘Medida facial’ e A2 – ‘Biometria fetal’, pois as classes filhas de A1 e A2 representam partes do sistema ósseo e outros órgãos dos quais a medida em relação ao padrão médio é fundamental para os exames de radiologia obstétrica.

Essa hierarquia foi estabelecida mediante o contato direto com o especialista, que a validou. Após essa etapa, adentramos na pesquisa documental para o reconhecimento nos vocabulários controlados que estabelecemos como viáveis: o RadLex, o LOINC, o SNOMED e o OntoNeo.

Desenvolvido pela Radiological Society of North America (RSNA), o RadLex é um vocabulário extenso, com cerca de 68 mil termos, que visa servir de base para a comunicação online de imagens, relatórios de imagem e registros médicos no âmbito da radiologia<sup>17</sup>. O LOINC é um vocabulário voltado para as observações clínicas e laboratoriais, e é estruturado a partir de seis categorias principais: component, property, time, system, scale, and method<sup>18</sup>. O SNOMED é uma ferramenta internacional e multilíngue bastante abrangente que visa subsidiar a indexação de registros clínicos do paciente que contenham sinais, sintomas, diagnósticos, procedimentos, anatomia, patologias, entre outras manifestações<sup>19</sup>. Já o OntoNeo é um projeto recente que busca a superação de falhas relativas à interoperabilidade semântica entre sistemas de informação por meio do uso de diferentes padrões e terminologias<sup>20</sup>.

Em seguida, traduzimos os termos para o inglês e realizamos a sua busca no BioPortal. As buscas foram registradas em planilhas específicas para cada um dos vocabulários. Assim, geramos quatro planilhas relativas ao RadLex, LOINC, SNOMED e OntoNeo.

Constatamos que nenhum dos quatro vocabulários atendem completamente aos termos levantados. Abaixo segue o Quadro 3, que traz a cobertura que os vocabulários consultados oferecem em relação à hierarquia que construímos:

Quadro 3. Porcentagem geral de termos recuperados

	ONTONEO	RADLEX	LOINC	SNOMED
Total de termos	218	218	218	218
Termos recuperados	86	136	146	167
Termos não encontrados	132	82	72	51
Cobertura dos termos recuperados	39,45%	62,39%	66,96%	76,61%

Fonte: Os autores (2018).

No Quadro 3, apresentamos uma coluna para cada vocabulário controlado em que buscamos os termos. Do total de termos (218), foram recuperados 86 no OntoNeo (aproximadamente 39,45%); 136 no RadLex (aproximadamente 62,39%); 146 no LOINC (aproximadamente 66,96%); e, por fim, recuperamos 167 no SNOMED (aproximadamente 76,61%). Dos 218, 30 (cerca de 13,75%) não foram identificados em nenhum vocabulário controlado; são eles: A1; A1.4; A2.6.1; B1.2; B2.2; B4.2; C; C4.1; C7.1; D; D1.1; D1.2; D1.2.1.1; D1.2.1.2; D1.2.1.3; D1.3.1; D1.3.2; D1.3.3; D1.3.4; D2.1.1; D2.1.3; D2.1.4; D2.2.2; D2.2.3; D2.2.4; E1; E2.2; E2.3; E3.1.1; E4.5. Para fins de aplicação, este problema poderá ser solucionado via sugestões aos mantenedores das ontologias visitadas para incorporação de tais termos em seus respectivos produtos, ou ainda podem ser realizadas novas buscas em outros vocabulários controlados.

Se, por um lado, apenas 13,75% dos termos não foram identificados em nenhum vocabulário consultado, por outro, apenas 78 termos estiveram presentes nas quatro ontologias, isto é, cerca de 35,78%. Neste caso, encontramos os termos ‘alterado’ e ‘normal’ que se repetem 29 vezes cada (A1.2.1.1; A1.2.1.2; A1.3.1.1; A1.3.1.2; A1.4.1.1; A1.4.1.2; A1.5.1.1; A1.5.1.2; A2.1.1.1; A2.1.1.2; A2.2.1.1; A2.2.1.2; A2.3.1.1; A2.3.1.2; A2.4.1.1; A2.4.1.2; A2.5.1.1; A2.5.1.2; A2.5.2.1; A2.5.2.3; A2.7.1.1; A2.7.1.2; A2.8.1.1; A2.8.1.2; A2.9.1.1; A2.9.1.2; A2.10.1.1; A2.10.1.2; A2.11.1.1; A2.11.1.1; A2.12.1.1; A2.12.1.2; A2.13.1.1; A2.13.1.2; A2.14.1.1; A2.14.1.2; A2.15.1.1; A2.15.1.2; B1.1.1; B1.1.2; B1.2.1; B1.2.2; B2.1.1; B2.1.2; B2.2.1; B2.2.2; B3.1.1; B3.1.2; B3.2.1; B3.2.2; B4.1.1; B4.1.2; B4.2.1; B4.2.2; C8.1; C8.2; D1.2.1.4; D1.2.1.5), encontramos o termo ‘comprimento’ que se repete 11 vezes (A1.5.1; A2.2.1; A2.7.1; A2.8.1; A2.9.1; A2.10.1; A2.11.1; A2.12.1; A2.13.1; A2.14.1; A2.15.1), e, finalmente, os que apresentam apenas uma ocorrência: ‘abdome’ (A2.1); ‘cerebelo’ (A2.5); ‘umbilical’ (B1); ‘genital’ (C8); ‘líquido amniótico’ (D2); ‘volume’ (D2.1); ‘qualidade de exame’ (E2); ‘datar gestação’ (E4.4); ‘sangramento’ (E4.7).

Como já relatado nesta seção do artigo, a decomposição dos termos nos permitiu identificar os mais relevantes (aqueles dos quais derivam outras informações) e os menos relevantes (aqueles que só apresentam sentido junto ao termo relevante, pois se referem a ele). Também há termos que revelam condições de manifestação daqueles mais relevantes e, ainda, termos que são categorias abstratas cuja função não é representar algo empírico, mas organizar logicamente nossa hierarquia. Assim verificamos que a contagem da porcentagem de cobertura dos termos por ontologia em função do tipo de termo teria impacto na análise do potencial representativo de cada uma das ontologias.

A seguir apresentaremos quatro quadros, cada um relativo à busca realizada em uma ontologia. Nos quadros, há quatro colunas que representam o tipo de termo estabelecido da seguinte maneira: CT representa as categorias abstratas cuja principal função é organizar logicamente a hierarquia; P representa os termos centrais mais relevantes dos quais fornecemos informações nos exames de radiologia obstétrica; M representa os modos, ou como os termos relevantes podem se manifestar; e, finalmente, F diz respeito aos termos menos relevantes em nossa hierarquia.

*Quadro 4. Recuperação dos termos em tipos - OntoNeo*

<b>Tipologia dos termos em nossa classificação – OntoNeo</b>	<b>Categorias CT</b>	<b>Classes Principais P</b>	<b>Modos M</b>	<b>Elementos Finais F</b>
Total de termos na hierarquia	5	44	63	106
Identificados nos vocabulários	1	8	14	63
Termos não recuperados	4	36	49	43
Cobertura dos termos recuperados em porcentagem	20	18,1818182	22,2222222	59,4339623

Fonte: Os autores (2018).

O Quadro 4 apresenta o vocabulário controlado que menos contempla nossas necessidades. Embora tal ontologia se proponha especificamente ao domínio de obstetrícia e neonatal, são poucas as suas contribuições. Talvez por ser ainda um vocabulário recente<sup>20</sup> ou por ter um conjunto de termos reduzido (1343 segundo os dados no BioPortal). Verificamos que, em todas as tipologias de termos, a ontologia OntoNeo apresentou um baixo índice de recuperação, alcançando a quantidade de 1 de 5, 8 de 44, 14 de 63 e 63 de 106 possíveis em relação às tipologias CT, P, M e F, respectivamente. Mesmo o OntoNeo se propondo a um escopo mais aproximado da radiologia obstétrica, seus resultados são limitados na cobertura terminológica, correspondendo apenas a 39,45%, como citamos anteriormente.

O RadLex apresentou índices mais interessantes, conforme explícito no Quadro 5.

*Quadro 5. Recuperação dos termos em tipos - RadLex*

<b>Tipologia dos termos em nossa classificação – RadLex</b>	<b>Categorias CT</b>	<b>Classes Principais P</b>	<b>Modos M</b>	<b>Elementos Finais F</b>
Total de termos na hierarquia	5	44	63	106
Identificados nos vocabulários	2	25	37	72
Termos não recuperados	3	19	26	34
Cobertura dos termos recuperados em porcentagem	40	56,8181818	58,7301587	67,9245283

Fonte: Os autores (2018).

Relativo ao RadLex, o Quadro 5 revela que ele tem uma cobertura mais ampla que o OntoNeo. Essa ontologia focada na radiologia como um todo concentrou o maior número de termos recuperados na categoria A (ver Apêndice B), pois é nesta categoria que se encontra o conjunto de termos relacionados a partes do corpo, sobretudo os ossos que são fundamentais no acompanhamento do desenvolvimento do feto. A radiologia também é centrada na produção de exames relacionados a este tecido. O Quadro 5 revela um aumento nos índices de recuperação em todos os tipos de termos em relação ao OntoNeo. Os números são 2 de 5, 25 de 44, 37 de 63 e 72 de 106 em relação às tipologias CT, P, M e F, respectivamente.

Os resultados obtidos ao verificarmos a cobertura dos termos de nossa hierarquia no LOINC seguem no Quadro 6.

Quadro 6. Recuperação dos termos em tipos - LOINC

Tipologia dos termos em nossa classificação – LOINC	Categorias CT	Classes Principais P	Modos M	Elementos Finais F
Total de termos na hierarquia	5	44	63	106
Identificados nos vocabulários	2	33	40	71
Termos não recuperados	3	11	23	35
Cobertura dos termos recuperados em porcentagem	40	75	63,4920635	66,9811321

Fonte: Os autores (2018).

Apresentando 66,96% de cobertura total, o LOINC é uma grande ontologia, com 192372 classes, e desde 2009 sofre constantes atualizações, segundo o BioPortal. Os resultados relativos aos tipos dos termos CT, P, M e F são respectivamente 2 de 5, 33 de 44, 40 de 63 e 71 de 106. As porcentagens de cada um desses resultados apresentam certo equilíbrio, assim como já aconteceu no RadLex; contudo, com índices mais altos, apresentando 40% de recuperação de CT, 75% de P, 63,5% de M e praticamente 67% de F, enquanto que o RadLex apresentou 40%, 56,7%, 58,6% e 63,8% em relação aos mesmos tipos de termos.

O Quadro 7 abarca os dados relativos à cobertura dos termos de nossa hierarquia no SNOMED.

Quadro 7. Recuperação dos termos em tipos - SNOMED

Tipologia dos termos em nossa classificação – SNOMED	Categorias CT	Classes Principais P	Modos M	Elementos Finais F
Total de termos na hierarquia	5	44	63	106
Identificados nos vocabulários	2	35	47	83
Termos não recuperados	3	9	16	23
Cobertura dos termos recuperados em porcentagem	40	79,5454545	74,6031746	78,3018868

Fonte: Os autores (2018).

Por fim, o SNOMED. Com cerca de 327128 classes, entre as quais 32114 apresentam um único filho (subclasse) e 2500 apresentam mais que 25 filhos, sua amplitude interfere também em nosso resultado. O SNOMED foi o vocabulário com maior cobertura de nossas necessidades, recuperando 2 de 5, 35 de 44, 47 de 63 e 83 de 106, relativos às categorias CT, P, M e F, respectivamente. As porcentagens se apresentam bastante

equilibradas. Seguindo a mesma ordem de apresentação dos índices de recuperação, as porcentagens foram 40%, 79,4%, 74,5% e 78,2%. Embora o foco deste artigo não seja a questão semântica, destacamos que, no SNOMED, o índice de recuperação do tipo de termo P é o mais alto, o que pode ser um indicativo de convergências mais profícuas entre nossas necessidades e a ontologia em questão.

O SNOMED é destacado como um vocabulário diferenciado por sua abrangência, recomendando-o no uso conjunto com o padrão<sup>7</sup>. Nossos resultados revelam que tal vocabulário mostrou-se mais adequado para sanar nossas necessidades terminológicas; contudo, ainda necessitamos aprofundar os estudos no que concerne aos aspectos semânticos.

## Considerações finais

Em um primeiro momento, a pesquisa aqui apresentada centrou-se no levantamento de termos a partir de um banco de dados em Microsoft Access e de reuniões com um especialista. Os termos foram selecionados e arranjados hierarquicamente. Nesse arranjo, não nos preocupamos com o critério de redundância pois, além da hierarquia, a função final da pesquisa foi estruturar laudos da radiologia obstétrica no Sistema Integrado Catarinense de Telemedicina e Telessaúde (STT/SC). Essa etapa aconteceu com constantes reuniões com especialista da referida área no Hospital Universitário da Universidade Federal de Santa Catarina. Essa hierarquia apresentada no Apêndice B revela categorias abstratas cuja função é apenas permitir o arranjo lógico dos termos. As subclasses derivadas dessas hierarquias são elementos relativos aos exames de radiologia obstétrica, e representam instâncias empíricas ou manifestações possíveis dessas instâncias. Ainda há as últimas subclasses, chamadas de folhas por estarem na extremidade da árvore, em que grande parte representa um conjunto de resultados possíveis dos exames. Assim, termos ‘normal’ e ‘alterado’, ‘presente’ e ‘ausente’ e ‘regular’ e ‘irregular’ como um conjunto de termos serão especificados em pesquisas futuras.

O padrão DICOM permite o uso conjunto de vários vocabulários controlados. Em uma segunda etapa foram selecionados quatro vocabulários específicos: o OntoNeo, o RadLex, o LOINC e o SNOMED. Todos são acessíveis via o BioPortal, no qual há um sistema de busca que utilizamos com o objetivo de realizar a pesquisa documental. Nessa busca, constatamos que o SNOMED cobriu 76,61% dos termos levantados na primeira etapa da pesquisa. O LOINC foi o segundo vocabulário com maior cobertura, abrangendo cerca de 66,96%, seguido pelo RadLex, que cobriu 62,39% e, por fim, o OntoNeo com cobertura de 39,45%. Constatamos que o volume do vocabulário interferiu de forma decisiva, suplantando outros critérios como escopo ou idade do vocabulário.

Embora o objetivo deste artigo não passe pelo estudo semântico dos termos, verificamos que em nossa hierarquia de referência há tipos distintos de termos que indicam maiores possibilidades de convergência ou divergência entre eles e os vocabulários consultados via BioPortal. Assim, geramos os tipos ‘categoria’ (CT), ‘classe principal’ (P), ‘modo’ (M) e ‘elementos finais’ (F), entre os quais constatamos que P é o tipo de maior valor, seguido por M, pois são termos que representam os objetos empíricos, focos da ação médica. Buscando a cobertura por tipos de termos, identificamos que o SNOMED é o vocabulário com maior revocação dos termos P e M, seguido pelo LOINC, RadLex e, por último, o OntoNeo.

Os baixos índices de revocação no OntoNeo são um resultado pouco esperado, devido ao domínio a que a ontologia se destina, próximo semanticamente ao domínio selecionado para essa pesquisa. É provável que tal resultado se justifique pela recente formalização desse vocabulário<sup>18</sup>. Contudo, outro fator que pode interferir nos resultados é a divergência terminológica entre os termos procurados e os contidos na estrutura do vocabulário. Seria relevante aprofundar um estudo nessa direção.

Ao longo da pesquisa, as dificuldades enfrentadas foram: a questão terminológica, no que diz respeito à definição dos termos de modo que estes correspondam ao uso corrente; a questão de tradução desses

termos para o inglês para efetivação da busca; as variações morfológicas; variações de número; variações de classes gramaticais, variações sintáticas, a ordenação das palavras em termos compostos, entre outras manifestações que interferem não só na identificação e reconhecimento, mas também na validação semântica de cada termo. A questão semântica será aprofundada em trabalhos futuros.

Tais questões serão aprofundadas em futuras publicações decorrentes dessa mesma pesquisa. Esperamos que este artigo traga contribuições na interface comum entre ciência da informação, ciência da computação e a área médica, ramo interdisciplinar do conhecimento que vem sendo reconhecido pelo desígnio de informação em saúde. E ainda, que contribua também com a comunicação de conteúdo transferido cotidianamente no STT/SC.

## Referências

1. Moraes IH, Santos SR. Informação em saúde: os desafios continuam. *Ci Saúde Coletiva*. 1998 [citado em 2017 jul. 13];3(1):37-51. <http://dx.doi.org/10.1590/1413-812319983102772014>
2. Telessaúde Santa Catarina [Internet]. Histórico; 2017 [citado em 2018 fev. 27]. Disponível em: <http://telessaude.ufsc.br/historico/>
3. Savaris A, Andrade R, Macedo DD, Von Wangenheim A. O uso da telemedicina assistencial assíncrona em larga escala no setor público de saúde. In: 11º Congresso Brasileiro de Informática em Saúde; 2008 29 nov.-03 dez.; Campos do Jordão, SP. São Paulo: SBI; 2008.
4. Shimizu S, Nakashima N, Okamura K, Tanaka M. One hundred case studies of Asia-Pacific telemedicine using a digital video transport system over a research and education network. *Telemed J E-Health*. 2009 [citado em 2017 jul. 13];15(1):112-7. <https://doi.org/10.1089/tmj.2008.0067>
5. Picchi Netto O, Macedo AA, de Azevedo Marques PM, Baranauskas JA. Uma metodologia para estruturação de laudos médicos usando ontologias. In: 11º Workshop de Informática Médica; 30º Congresso da Sociedade Brasileira de Computação; 2011 19-20 jul.; Natal, RN. Natal: UFRN; 2011.
6. Morioka CA, Sinha U, Taira R, El Saden SU, Duckwiler G, Kangaroo H. Structured reporting in neuroradiology. *Ann NY Acad Sci*. 2002 Dec 1 [citado em 2017 jul. 13];980(1):259-66. <https://doi.org/10.1111/j.1749-6632.2002.tb04902.x>
7. Chennamsetty H, Chalasani S, Riley D. Predictive analytics on Electronic Health Records (EHRs) using Hadoop and Hive. In: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies; 2015 05-07 Mar; Coimbatore, India. Nova Jersey: IEEE; 2015.
8. Clunie DA. DICOM structured reporting. Bangor, Pennsylvania: PixelMed Publishing; 2000.
9. Tálamo MD, de Lara ML, Kobashi NY. Vamos perseguir a informação. *Comun Educ*. 1995 dez. [citado em 2017 jul. 13];30(4):52-7. <http://dx.doi.org/10.11606/issn.2316-9125.v0i4p52-57>
10. Dahlberg I. Teoria da classificação, ontem e hoje. In: Anais da Conferência Brasileira de Classificação Bibliográfica; 1976 set. 12-17; Rio de Janeiro: IBICT/ABDF, 1979. p.352-370.
11. Santarem Segundo JE, Coneglian CS. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. *Inf Inf*. 2016 [citado em 2017 jul. 13]; 21(2): 217-44. <http://dx.doi.org/10.5433/1981-8920.2016v21n2p217>
12. Bräscher M, Carlan E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: Robredo J, Bräscher M. Passeios no Bosque da Informação: estudos sobre representação e organização da informação e do conhecimento. Brasília, DF: IBICT; 2010. p.147-76.
13. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*. 2009; 37(suppl 2):W170-3. <https://doi.org/10.1093/nar/gkp440>
14. Andrade JD. Interoperabilidade e mapeamentos entre sistemas de organização do conhecimento na busca e recuperação de informações em saúde: estudo de caso em ortopedia e traumatologia [tese]. São Paulo: Universidade de São Paulo; 2015.
15. Creswell JW. Projeto de pesquisa métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre: Artmed; 2010.

16. Lakatos EM, Marconi MD. Fundamentos da metodologia científica. São Paulo: Atlas; 1991.
17. Radiological Society of North America [Internet]. RadLex: about. 2017 [citado em 2017 maio 13]. Disponível em: <https://www.rsna.org/RadLex.aspx>
18. Chen ES, Melton GB, Engelstad ME, Sarkar IN. Standardizing clinical document names using the HL7/ LOINC document ontology and LOINC codes. AMIA Annu Symp Proc; 2010 Nov [citado em 2017 jul. 13];2010(2010):101-5. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041326/>
19. Girão IPT; Rabelo CRO; Pinto VB. SNOMED-CT as Standard Language for Organization and Representation of the Information in Patient Records. Knowl Organ [Internet]. 2014 [citado em 2017 maio 04];41(4):311-8. Disponível em: <http://www.repositorio.ufc.br/ri/handle/riufc/19136>
20. Farinellia F, Almeida MB, Elkin PL, Smith B. Medical encounters in the obstetric and neonatal domain: an approach based on ontological realism. In: Joint International Conference on Biological Ontology and BioCreative; 2016 Aug 1-6; Corvallis, Oregon. Oregon State University: ICBO; 2016.

## Apêndice A

Lista dos termos
Apresentação fetal cefálica
Apresentação fetal córmica
Apresentação fetal oblíqua
Apresentação fetal pélvica
Batimentos fetais ausentes
Batimentos fetais irregulares
Batimentos fetais presentes
Batimentos fetais regulares
Bexiga cheia
Bexiga parcialmente cheia
Bexiga vazia
Binocular
Binocular alterado
Binocular normal
Cérebro
Cérebro - alterado
Cérebro - normal
Circunferência abdominal
Circunferência abdominal - alterada
Circunferência abdominal - normal
Circunferência cefálica
Circunferência cefálica - alterada
Circunferência cefálica - normal
Circunferência cerebral/fêmur
Circunferência cerebral/fêmur alterado
Circunferência cerebral/fêmur normal
Circunferência cerebral/circunferências abdominal
Circunferência cerebral/circunferências abdominal alterado
Circunferência cerebral/circunferências abdominal normal
Clavícula
Clavícula - alterada
Clavícula - normal
Colo uterino - outros
Colo uterino alterado
Colo uterino com presença de cisto de Naboth
Colo uterino normal
Comprimento cabeça-nádega
Comprimento cabeça-nádega - alterada
Comprimento cabeça-nádega - normal
Comprimento do fêmur
Comprimento do fêmur alterado

Comprimento do fêmur normal
Comprimento do fêmur/diâmetro biparietal
Comprimento do fêmur/diâmetro biparietal alterado
Comprimento do fêmur/diâmetro biparietal normal
Conclusão - biometria pela USG concordante com DUM
Conclusão - biometria pela USG discordante com DUM
Conclusão - cardiocografia - CTG anormal - conduta urgente
Conclusão - cardiocografia - CTG atípica - nova avaliação necessária
Conclusão - cardiocografia - CTG normal - nova avaliação opcional
Conclusão - cardiocografia - outros
Conclusão - controle da ovulação - ocorrência da ovulação evidenciada
Conclusão - controle da ovulação - ocorrência da ovulação evidenciada - repetir o exame
Conclusão - controle da ovulação - ocorrência da ovulação evidenciada - outros
Conclusão - controle da ovulação - ocorrência da ovulação não evidenciada
Conclusão - exame alterado
Conclusão - exame dentro da normalidade
Conclusão - gestação com crescimento fetal abaixo da idade gestacional de referência
Conclusão - gestação com crescimento fetal acima da idade gestacional de referência
Conclusão - gestação com crescimento fetal adequado para a idade gestacional de referência
Conclusão - morfologia - alterada
Conclusão - morfologia - biometria fetal pela USG compatível com IDADE GESTACIONAL DE REFERÊNCIA
Conclusão - morfologia - normal
Conclusão 1 trimestre - biometria pela USG concordante com DUM
Conclusão 1 trimestre - biometria pela USG discordante com DUM
Conclusão 1 trimestre - exame alterado
Conclusão 1 trimestre - exame dentro da normalidade
Conclusão doppler 2 e 3 trimestre - fluxo materno alterado, rastreio para DHEG positivo, fluxo fetal normal
Conclusão doppler 2 e 3 trimestre - fluxo materno e fetal normais
Conclusão doppler 2 e 3 trimestre - Fluxo materno normal e fetal alterado
Conclusão doppler 2 e 3 trimestre - outros
Contorno uterino - outros
Contorno uterino irregular
Contorno uterino regular
Contração uterina - ausente
Contração uterina - irregular
Contração uterina - outros
Contração uterina - regular
Contração uterina - tipo Braxton-Hicks
Corionicidade - outros
Corionicidade dicoriônica
Corionicidade incerta
Corionicidade monocoriônica, diamniótica
Corionicidade monocoriônica, monoamniótica
Dedo alterado
Dedo normal

Desaceleração-cardiotocografia nenhuma ou ocasional/variável <30seg
Desaceleração-cardiotocografia variável >60seg
Desaceleração-cardiotocografia variavel, 30 a 60seg
Diâmetro biparetal
Diâmetro biparetal alterado
Diâmetro biparetal normal
Dorso fetal anterior
Dorso fetal direita
Dorso fetal esquerdo
Dorso fetal posterior
Eco endometrial - outros
Eco endometrial centrado
Eco endometrial desviado
Eco endometrial hormonal - outros
Eco endometrial hormonal espessado
Eco endometrial hormonal fino
Eco endometrial hormonal linear
Eco endometrial hormonal proliferativo
Eco endometrial hormonal secretor
Eco endometrial presença de DIU deslocado
Eco endometrial presença de DIU normoposicionado
Eixo uterino - outros
Eixo uterino centrado
Eixo uterino desviado para direita
Eixo uterino desviado para esquerda
Etnia branca
Etnia cabocla
Etnia cafusa
Etnia indígena
Etnia mulata
Etnia negra
Etnia parda
Fêmur, comprimento feto
Fêmur, comprimento feto - alterado
Fêmur, comprimento feto - normal
Feto/embrião - número - gemelar
Feto/embrião - número - outros
Feto/embrião - número - ovo anembrionado
Feto/embrião - número - quádruplos
Feto/embrião - número - quádruplos
Feto/embrião - número - quádruplos
Feto/embrião - número - trigemelar
Feto/embrião - número - único
Fíbula
Fíbula alterado
Fíbula normal

Fundo de saco - outros
Fundo de saco livre
Fundo de saco preenchido
Indicação de exame 1 trimestre - avaliação do crescimento fetal
Indicação de exame 1 trimestre - datar gestação
Indicação de exame 1 trimestre - outros
Indicação de exame 1 trimestre - rastreio de aneuploidias
Indicação de exame 1 trimestre - sangramento
Indicação de exame 2 e 3 trimestre - avaliação fetal
Indicação de exame 2 e 3 trimestre - datar gestação
Indicação de exame 2 e 3 trimestre - sangramento
Indicação exame doppler gestacional - avaliação da vitalidade fetal
Indicação exame doppler gestacional - diagnosticar gestação
Indicação exame doppler gestacional - outras
Indicação exame doppler gestacional - outros
Indicação exame doppler gestacional - rastreio de DHEG
Índice cefálico
Índice cefálico alterado
Índice cefálico normal
Índice de líquido amniótico
Índice de líquido amniótico alterado
Índice de líquido amniótico não mesurado
Índice de líquido amniótico normal
Índice de pulsatilidade da artéria cerebral média
Índice de pulsatilidade da artéria cerebral média alterado
Índice de pulsatilidade da artéria cerebral média normal
Índice de pulsatilidade umbilical
Índice de pulsatilidade umbilical alterado
Índice de pulsatilidade umbilical normal
Índice de resistência umbilical
Índice de resistência umbilical alterado
Índice de resistência umbilical normal
Interocular
Interocular alterado
Interocular normal
Líquido amniótico - normodramnia
Líquido amniótico - oligodramnia grave
Líquido amniótico - oligodramnia leve
Líquido amniótico - oligodramnia moderada
Líquido na cavidade - outros
Líquido na cavidade ausente
Líquido na cavidade presente muita quantidade
Líquido na cavidade presente pouca quantidade
Maturidade placentária - bicorno
Maturidade placentária - Di Delphos

Maturidade placentária - sub-septado
Maturidade placentária Grau 0
Maturidade placentária Grau I
Maturidade placentária Grau II
Maturidade placentária Grau III
Morfologia uterina - arqueado
Morfologia uterina - normal
Morfologia uterina - outros
Morfologia uterina - septado
Movimento fetal ausente
Movimento fetal outros
Movimento fetal ovo anembrionado
Movimento fetal presente
Movimento respiratório outros
Movimento respiratório ausente
Movimento respiratório presente
Ocular
Ocular alterado
Ocular normal
Osso nasal
Osso nasal alterado
Osso nasal normal
Ovário direito apresentando corpo lúteo
Ovário direito compatível com ooforectomia
Ovário direito não visível
Ovário direito outros
Ovário direito visível
Ovário esquerdo apresentando corpo lúteo
Ovário esquerdo compatível com ooforectomia
Ovário esquerdo não visível
Ovário esquerdo outros
Ovário esquerdo visível
Pé
Pé alterado
Pé normal
Peso
Peso alterado
Peso normal
Placenta - posição - anterior
Placenta - posição - fúndica
Placenta - posição - inserção baixa
Placenta - posição - lateral direita
Placenta - posição - lateral esquerda
Placenta - posição - outros
Placenta - posição - posterior



Útero - posição - outros
Útero - posição - retroversão
Útero - posição - retroversoflexão
Variabilidade - cardiocografia 5 bpm ou 25bpm
Variabilidade - cardiocografia 5 bpm por 40 a 80min
Variabilidade - cardiocografia 6 a 25 bpm
Via de exame abdominal
Via de exame transvaginal
Volume uterino - outros
Volume uterino alterado
Volume uterino aumentado
Volume uterino diminuído
Volume uterino normal

## Apêndice B

<b>A</b>	<b>Morfologia</b>
A1	Medida facial
A12	Binocular
A121	distância
A1211	alterado
A1212	normal
A13	Interocular
A131	distância
A1311	alterada
A1312	normal
A14	Ocular
A141	diâmetro
A1411	alterada
A1412	normal
A15	Osso nasal
A151	comprimento
A1511	alterada
A1512	normal
A2	Biometria fetal
A21	Abdome
A211	circunferência
A2111	alterado
A2112	normal
A22	Cabeça-nádega
A221	comprimento
A2211	alterada
A2212	normal
A23	Biparietal
A231	diâmetro
A2311	alterada
A2312	normal
A24	Cefálica
A241	circunferência
A2411	alterado
A2412	normal
A25	Cerebelo
A251	diâmetro transveso
A2511	alterada
A2512	normal
A252	cisterna magna
A2521	alterada
A2522	diminuído
A2523	normal

A26	Saco gestacional
A261	condição
A2611	irregular
A2612	regular
A27	Fêmur
A271	comprimento
A2711	alterada
A2712	normal
A28	Clavícula
A281	comprimento
A2811	alterada
A2812	normal
A29	Fíbula
A291	comprimento
A2911	alterada
A2912	normal
A210	Dedo
A2101	comprimento
A21011	alterada
A21012	normal
A211	Pé
A2111	comprimento
A21111	alterada
A21111	normal
A212	Rádio
A2121	comprimento
A21211	alterado
A21212	normal
A213	Tíbia
A2131	comprimento
A21311	alterada
A21312	normal
A214	Ulna
A2141	comprimento
A21411	alterada
A21412	normal
A215	Umero
A2151	comprimento
A21511	alterada
A21512	Normal
<b>B</b>	<b>Doppler</b>
B1	Umbilical
B11	pulsatilidade
B111	alterada
B112	normal

B12	resistência
B121	alterada
B122	normal
B2	Cerebral
B21	pulsatilidade
B211	alterada
B212	normal
B22	resistência
B221	alterada
B222	normal
B3	Ducto venoso
B31	pulsatilidade
B311	alterada
B312	normal
B32	relação atrioventricular
B321	alterada
B322	normal
B4	Artéria uterina
B41	pulsatilidade
B411	alterada
B412	normal
B42	resistência
B421	alterada
B422	normal
<b>C</b>	<b>Estática e bem estar fetal</b>
C1	Dorso
C11	anterior
C12	direito
C13	esquerdo
C14	posterior
C2	Tônus
C21	ausente
C22	presente
C3	Apresentação
C31	cefálica
C32	córmica
C33	oblíqua
C34	pélvica
C4	Posição
C41	córmica
C42	transversal
C43	oblíqua
C44	longitudinal
C5	Batimento cardíaco fetal
C51	ausentes

C52	irregulares
C53	presentes
C54	regulares
C6	Cardiotocografia
C61	desaceleração
C611	nenhuma ou ocasional/variável <30seg
C612	variável >60seg
C613	variável 30 a 60seg
C62	variabilidade
C621	5 bpm ou 25bpm
C622	5 bpm por 40 a 80min
C623	6 a 25 bpm
C63	conclusão
C631	CTG anormal - conduta urgente
C632	CTG atípica - nova avaliação necessária
C633	CTG normal - nova avaliação opcional
C7	Movimento
C71	corporal
C711	ausente
C712	presente
C713	ovo anembrionado
C72	respiratório
C721	ausente
C722	presente
C8	Genital
C81	alterado
C82	normal
<b>D</b>	<b>Anexo fetal</b>
D1	Placenta
D11	corionicidade
D111	dicoriônica
D112	incerta
D113	monocoriônica
D12	Amniocidade
D121	quantidade
D1211	monoamniótico
D1212	diamniótico
D1213	triamniótico
D1214	alterada
D1215	normal
D13	Maturidade
D131	granum 0
D132	granum I
D133	granum II
D134	granum III

D2	Líquido amniótico
D21	volume
D211	normodramnia
D212	oligodramnia grave
D213	oligodramnia leve
D214	oligodramnia moderada
D215	aumentado
D216	diminuído
D22	Índice de Líquido Amniótico (ILA)
D221	oligodramnia grave
D222	oligodramnia leve
D223	oligodramnia moderada
D224	polidraminia
<b>E</b>	<b>Exame</b>
E1	via de exame
E11	abdominal
E12	transvaginal
E2	qualidade de exame
E21	prejudicado por adiposidade
E22	prejudicado por estática fetal
E23	prejudicado por interposição de alças
E24	prejudicado por movimentação fetal
E25	satisfatória
E3	gestação
E31	quantidade
E311	único
E312	gemelar
E313	trigemelar
E314	quádruplos
E315	quíntuplos
E316	sêxtuplos
E317	óctuplos
E4	indicação
E41	avaliação da vitalidade fetal
E42	avaliação do crescimento fetal
E43	diagnosticar gestação
E44	datar gestação
E45	rastreio de DHEG
E46	rastreio de aneuploidias
E47	sangramento