

## Representações formais do conhecimento e métodos linguísticos aplicados à integração de terminologias clínicas.

Linha de Pesquisa : Gestão e Tecnologia da Informação e Comunicação  
Grupo de Pesquisa: Representação do Conhecimento, Ontologias e Linguagem

Doutoranda : **Ma. Jeanne Louize Emygdio**  
Orientador : Prof. Dr. Maurício Barcellos Almeida

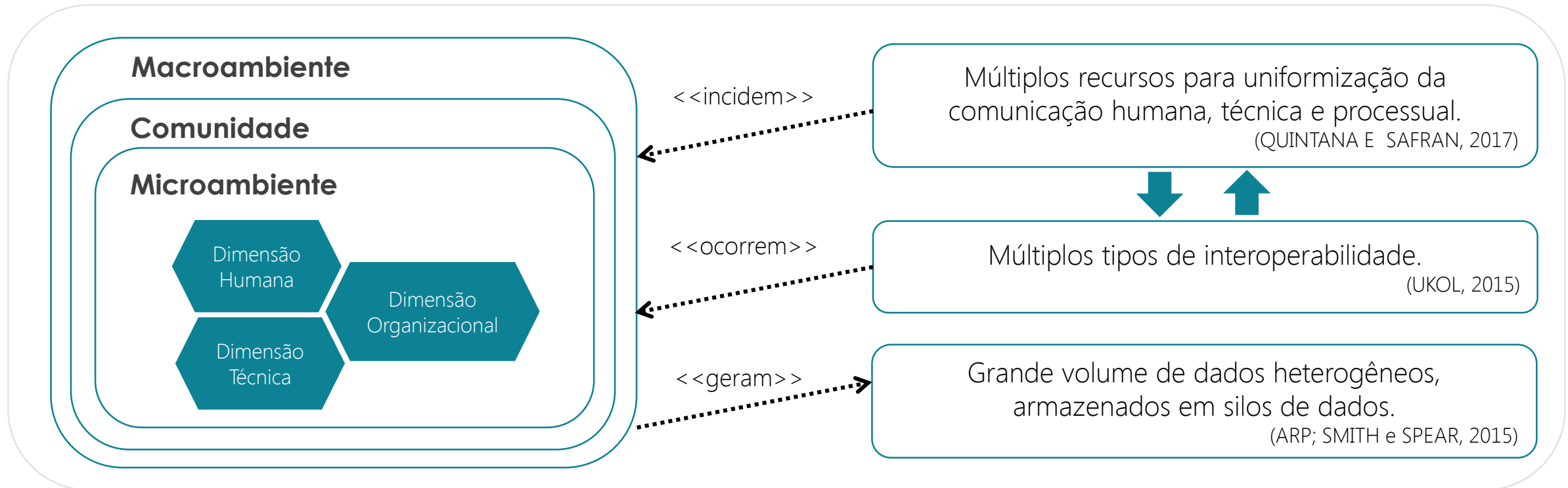
Apresenta-se uma pesquisa em andamento cujo propósito é o de **demonstrar** por meio de um **experimento**, que abordagens ontológicas podem ser melhores do que as puramente linguísticas para fins de **interoperabilidade semântica**. A metodologia envolve uso de métodos léxicos e estruturais para combinar, mapear, alinhar e integrar terminologias clínicas, sob o enfoque linguístico e sob o enfoque ontológico. Prevê-se uma **análise qualitativa dos dados**, através da mensuração do grau de eficiência de cada abordagem para prover integração entre as terminologias clínicas. Espera-se obter **contribuições teóricas** que permitam a descoberta e distinção dos aspectos ontológicos e epistêmicos inerentes às terminologias clínicas, os desafios epistêmicos à interoperabilidade semântica e a relevância da abordagem ontológica para mitigar tais problemas. Espera-se obter uma ontologia computacional de integração entre o SNOMED CT e a CID no âmbito da obstetrícia, a ser distribuída publicamente, em formato aberto e documentada.

**Esta pesquisa possui término previsto para Fev./2021.**

# Justificativas

- i. **Elementos distintos** compõem os ecossistemas de saúde e participam das **práticas de interoperabilidade sob diversas formas**, nem sempre conhecidas ou bem compreendidas pelos atores do domínio (ABNT, 2016; SCHULZ et al, 2017);

## Ecossistemas de saúde



Fonte: Elaborado pelos autores, 2019.

**Esta pesquisa aborda questões que inviabilizam a interoperabilidade semântica no contexto.**

## ii. Padrões terminológicos (Schulz et al, 2017; Access CT, 2016):

- **Se proliferam e se sobrepõem;**
- **Natureza semântica diversificada;**
- **Devem ser usados de forma integrada.**

Terminologias clínicas	Nº de TCs* hospedadas	Tipos de estruturas	Tipo de semântica
Repositório BioPortal <sup>1</sup>	970	Taxonomias, Tesouros, Meta-tesouros, Frameworks, Padrões Técnicos e Ontologias.	Linguagem humana, Formal, Mista.

<sup>1</sup> National Center for Biomedical Ontology (NCBO), Leland Stanford Junior University, California, USA.

\* Adota-se na pesquisa o termo genérico Terminologias Clínicas (Schulz et al, 2017).

# Justificativas

A **Ciência da Computação** produz importante e expressiva quantidade de pesquisa sobre alinhamento semântico de terminologias clínicas.

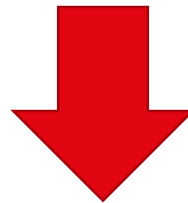
(Ochieng e Kyanda, 2018;

Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez, 2015).

- Técnicas de combinação de ontologias (*ontology matching*), mapeamento de ontologias (*ontology mapping*), alinhamento de ontologias (*ontology alignment*), integração de ontologias (*ontology integration*), etc. (Euzenat e Shvaiko, 2007; Noy e Musen, 1999;2000).
- OAEI – Iniciativa de Avaliação de Alinhamento de Ontologias - evolução longitudinal das técnicas.

Para a **Ciência da Informação** a ênfase, nesta questão recai sobre:

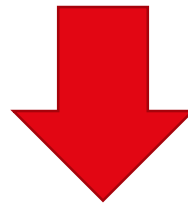
- a distinção entre os aspectos ontológicos e epistêmicos inerentes às classificações (Bodenreider, Smith e Burgun 2004);
- os impactos à interoperabilidade semântica;
- a forma como os fundamentos ontológicos podem mitigar tais problemas.



**Pesquisas interdisciplinares no âmbito das ontologias aplicadas** (Jansen, 2008; Smith, 2003).  
**Contribuição para organização da informação** (Saracevic, 1996) **antes da aplicação das técnicas de *matching* e *posteriores*.**

## Exemplo de presença epistêmica no SNOMED CT:

- i) "**Morte natural com (suspeita) de causa provável**, possibilita a comunicação de uma imprecisão clínica sobre um diagnóstico de morte;
- ii) "**Doença do coração (excluída)**", reflete uma convicção momentânea do médico e não a natureza ou a gravidade de um diagnóstico;
- iii) "**Operação no coração, (re-agendada)**", comunica a intenção de mudança na situação de um processo que ainda não ocorreu.



**Consequências:** definição de classes que não atendem aos princípios de classificação, impactando negativamente na comparação dos termos para fins de combinação, alinhamento, mapeamento, integração e evolução (Rector 1999; Bodenreider, Smith e Burgun 2004).

# Trabalhos correlatos

Na literatura científica encontram-se estudos relacionados à adoção do escrutínio ontológico em terminologias clínicas para fins de:

- i) melhorias na representação do conhecimento a partir da correção de anomalias estruturais, definição de conceitos, relações e meta-classes (Schulz; Suntisrivaraporn e Baader, 2007; Bodenreider; Smith e Burgun, 2004);
- ii) práticas de alinhamento indireto que demonstram eficiência quanto ao número de combinações terminológicas encontradas, bem como economia de recursos pela redução do número de mapeamentos diretos necessários para cobrir um conjunto mínimo de termos a serem mapeados (Zhang e Bodenreider, 2005);
- iii) abordagens integradas para identificação de fronteiras de conhecimento compartilhado entre TCs visando sua harmonização para fins de mapeamento e evolução, tendo por base a definição de axiomas em linguagem OWL (Rodrigues et al, 2015).

Aspectos relacionados à avaliação de qualidade e interoperabilidade de TCs indicam que o uso de técnicas híbridas de métodos léxicos e estruturais são efetivos para a detecção de erros na SNOMED CT, além de sugestivos para correção (Bodenreider, 2018).



# Questão de pesquisa

De que forma ontologias podem contribuir para resolver questões que impedem a interoperabilidade semântica?

Testar a integração de terminologias clínicas a partir de uma abordagem ontológica e a partir de uma abordagem linguística, verificando qual é a mais eficiente à interoperabilidade semântica.

# Objetivos Específicos

- i. **identificar e distinguir os aspectos ontológicos e epistêmicos** presentes nas terminologias clínicas, a partir da aplicação dos princípios de classificação ontológicos, eliminando informações não-essenciais à execução dos processos subsequentes de interoperabilidade;
- ii. **identificar âncoras léxicas** que estabeleçam fronteiras compartilhadas de conhecimento entre as terminologias clínicas, a partir da aplicação da técnica de criação de índices léxicos e da geração de mapeamentos entre as âncoras identificadas e as relações que estabelecem entre si;
- iii. **identificar relações semânticas** que possibilitem uma representação explícita do conhecimento que permeia as relações entre as âncoras léxicas através da extração de relações hierárquicas e técnicas complementares de descoberta de conhecimento implícito;
- iv. **identificar âncoras estruturais**, através da comparação das relações semânticas previamente identificadas, que constituam evidências positivas para a construção de alinhamentos entre as ontologias bem como os conflitos que possam prevenir alinhamentos indevidos;
- v. **demonstrar o grau de eficiência** alcançado por cada tipo de abordagem, tendo por base o número de combinações terminológicas encontradas e o percentual de evidências positivas para interoperabilidade que estes resultados representam.

# Metodologia

**Universo da pesquisa:** Terminologias clínicas.

**Unidade de análise....:** Conjuntos de termos relacionados à área obstétrica e neonatal

## Terminologias clínicas e repositórios de ontologias pesquisados:

Qtd.	Tipo	Quais	Semântica
5	<b>Taxonomias</b>	ATC/DDD, CPT-4, DSM, HCPCS, <b>ICD</b> , NANDA, ISO9999	<b>Linguagem natural</b>
6	Tesauros	ICECI, ICF, ICHI, MedDRA, MeSH, NCIt	Linguagem natural
1	<b>Meta-tesauro</b>	<b>SNOMED CT</b>	<b>Mista</b>
3	Frameworks	openEHR, HL7, IHE-PIX	Formal
4	Padrões Técnicos e terminológicos	TISS, DICOM, LOINC, <i>ISBT 128</i>	Mista
13	<b>Ontologias</b>	BWW, DOLCE, UFO, GFO, SOWA, <b>BFO</b> , <b>OntONeo</b> , GO, CL, <b>FMA</b> , <b>IAO</b> , OBI, <b>OGMS</b> .	<b>Formal</b>
3	Repositórios de ontologias	<b>OBO Foundry</b> , UMLS, Bioportal	Formal
35	Total		

## Justificativas:

- Contribuição social;
- Especialista para validação dos resultados;
- Ontologista brasileira responsável pela construção da OntONeo/RECOL;
- OntONeo possui labels em inglês e português.
- Replicação do experimento base (Zhang e Bodenreider, 2005);
- Maturidade do repositório OBO Foundry em construção de ontologias biomédicas;
- Ortogonalidade ao repositório.

Fonte: Elaborado pelos autores, 2019.

# Metodologia

**Definição das amostras estratificadas por classes e propriedades relacionadas ao domínio obstétrico e neonatal nas terminologias clínicas selecionadas:**

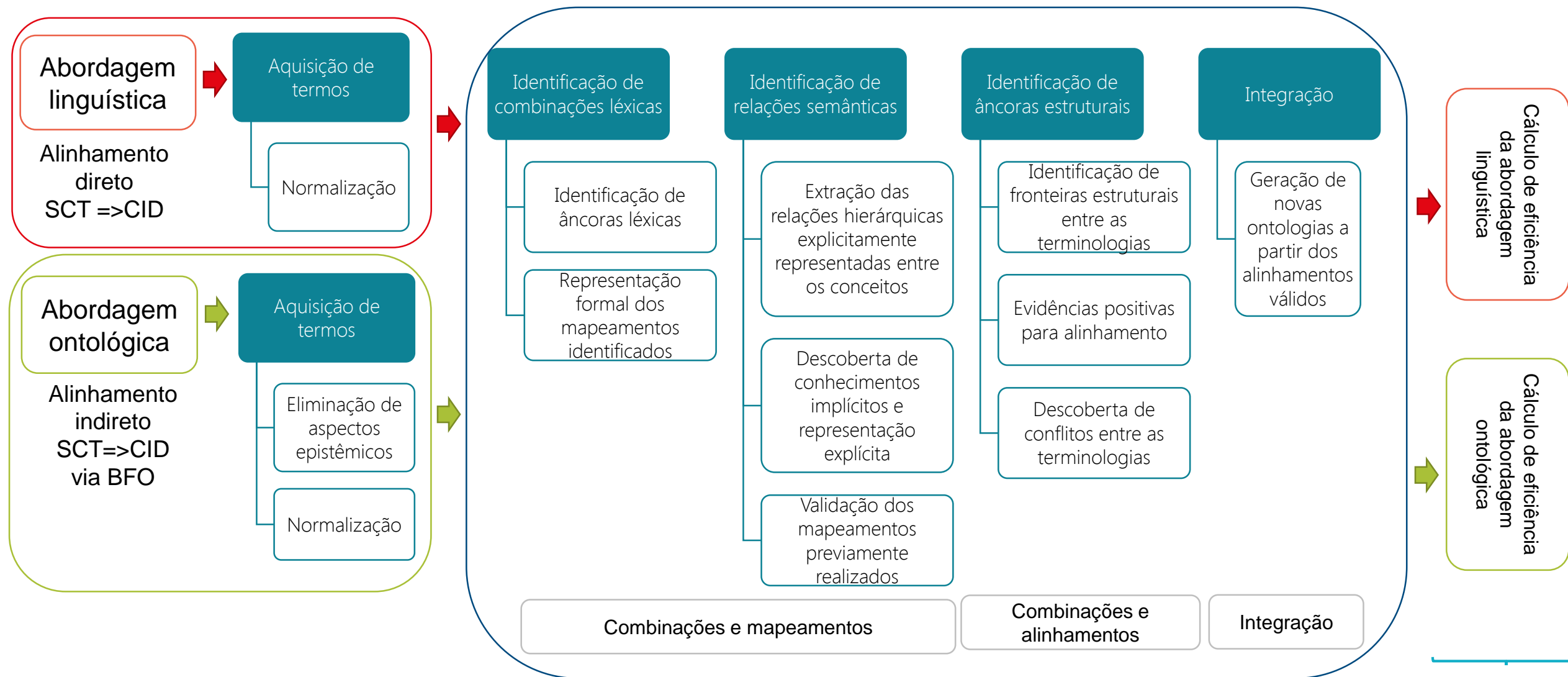
Terminologias clínicas	Gerais		Selecionadas	
	Classes	Propriedades	Classes	Propriedades
<b>SNOMED CT</b> <i>Systematized Nomenclature of Medicine–Clinical Terms</i>	347,077	216	<b>Iniciando estudos</b>	
<b>ICD-10</b> <i>International Classification of Diseases</i>	12.445 (Bioportal) 14.502 (Protegé)	1	3 Capítulos (XV, XVI, XVII) 222 Categorias (O00-O99;P00-P96;Q00-Q99) 1.381 Sub-categorias	
<b>BFO</b> <i>Basic Formal Ontology</i>	35	0	As que forem reutilizadas pela OntONeo	
<b>OntONeo</b> <i>Obstetric and Neonatal Ontology</i>	<b>1.767*</b>	<b>451</b>	<b>1.767</b>	<b>451</b>
<b>FMA</b> <i>Foundational Model of Anatomy</i>	107.721	168	Em caso de necessidade serão elencadas futuramente	
<b>IAO</b> <i>Information Artifact Ontology</i>	220	54		
<b>OGMS</b>	143	0		

\* Classes próprias e reutilizadas de outras ontologias OBO Foundry

Amostra inicial: 2.218 termos

# Metodologia

Processo de execução da pesquisa:



IDENTIFICAÇÃO DA MELHOR TÉCNICA POR COMPARAÇÃO DOS RESULTADOS

## Cálculo de eficiência de cada método de alinhamento

A eficiência de cada abordagem para prover interoperabilidade semântica entre as terminologias clínicas será calculada como uma medida da proporção de evidências positivas para alinhamentos sobre o número total de combinações encontradas, conforme mostrado na equação abaixo:

$$\text{Eficiência} = \left( \frac{\text{número de evidências positivas para alinhamentos}}{\text{total de combinações encontradas}} \right) \times 100$$

Simulação:

	SCT-CID 715 combinações	BFO-CID 1.353 combinações	BFO-SCT 2.173 combinações
Sem evidência	62 (8,7%)	66 (4,9%)	205 (9,4%)
Evidência positiva	653 (91,3%)	1.283 (94,8%)	1.958 (90,1%)
Evidência negativa	0	4 (0,3%)	10 (0,5%)

# Metodologia

## Ferramentas identificadas:

- **Logmap, Yam++, Yam-Bio e Reference:** ferramentas para suporte às práticas de *matching* e *mapping*.
- **Prompt:** suporte às práticas de interoperabilidade entre ontologias
  
- **Protegé:** editor de ontologias
- **Plugins para Protegé:**
  - Bioportal Import Plugin (importação de ontologias do BioPortal)
  - HermiT (verificação de inconsistências por reasoning)
  - OWL2 Query (consultas em SPARQL)
  - Snow OWL (browser do Snomed para Protegé)
  - OWL Diff (comparação de ontologias)
  
- **Ontofox:** importação de termos e propriedades das terminologias clínicas
  
- **Processamento de Linguagem Natural**
  - Linguagem Python
  - Bibliotecas NLTK, OWLReady2



# Referências

- ABNT. (2016) **Informática em saúde: princípios para mapeamento entre sistemas terminológicos**. Norma ABNT ISO/TR 12300:2016: Norma ABNT ISO/TR. [s.l.] Associação Brasileira de Normas Técnicas - ABNT.
- Almeida, M. B.; Farinelli, F. (2017) Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. **Journal of the Association for Information Science and Technology**.
- Arp, R.; Smith, B.; Spear, A. D. (2015) **Building Ontologies with Basic Formal Ontology**. Cambridge, Massachusetts: The MIT Press, 2015.
- Bodenreider, O. (2018) **Evaluating the Quality and Interoperability of Biomedical Terminologies**. [s.l.] LHNCBC Board of Scientific Counselors.
- Bodenreider, O.; Smith, B.; Burgun, A. (2004) The Ontology-Epistemology Divide: A Case Study in Medical Terminology. **Formal ontology in information systems : proceedings of the ... International Conference. FOIS (Conference)**,v.2004,p.185.
- Euzenat, J.; Shvaiko, P. (2007) **Ontology Matching**. Berlin, Heidelberg: Springer-Verlag, 2007.
- Jansen, L. (2008) Chapter 7: Classifications. In: **Applied Ontology: An Introduction [Hardcover]**. 8. ed. German: Katherine Munn e Barry Smith, 2008. v. 9p. 342.
- Noy, N. F.; Musen, M. A. (2000) **PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment**. Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence. **Anais...AAAI Press**, 2000. Disponível em: <<http://dl.acm.org/citation.cfm?id=647288.721118>>. Acesso em: 24 maio. 2018.
- Ochieng, P.; Kyanda, S. (2018) Large-Scale Ontology Matching. **ACM Computing Surveys**, v. 51, n. 4, p. 1–35.
- Otero-Cerdeira, L.; Rodríguez-Martínez, F. J.; Gómez-Rodríguez, A. (2015) Ontology matching: A literature review. **Expert Systems with Applications**.
- Quintana, Y.; Safran, C. (2017) Global Health Informatics—An Overview. In: **Global Health Informatics**. [s.l.] Elsevier, 2017. p. 1–13.

# Referências

- Rector, A. L. (1999) Clinical terminology: why is it so hard? **Methods of Information in Medicine**, v. 38, n. 4–5, p. 239–252.
- Rodrigues, J-M. et al.(2015) Semantic Alignment between ICD-11 and SNOMED CT. *Studies in Health Technology and Informatics*, v. 216, p. 790–794.
- Saracevic, T. (1996) *Ciência da informação: origem, evolução e relações*. v. 1, n. 1, p. 22, 1996.
- Schulz, S.; Baader, F.; Suntisrivaraporn, B. (2007) SNOMED CT's problem list: Ontologists' and logicians' therapy suggestions. *Studies in health technology and informatics*, v. 129, p. 802–806.
- Schulz, S. et al. (2017) Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in health technology and informatics**, v. 245, p. 940–944.
- Schulz, S.; Stegwee, R.; Chronaki, C. (2019) Standards in Healthcare Data. In: Kubben, P; Dumontier, M.; Dekker, A. (Eds.). . **Fundamentals of Clinical Data Science**. Cham: Springer International Publishing, 2019. p. 19–36.
- Smith, B. (2003) Ontology. In: **The Blackwell Guide to the Philosophy of Computing and Information**. Oxford: Luciano Floridi, 2003.
- Smith, B. et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature biotechnology**, v. 25, n. 11, p. 1251.
- UKOLN. (2015) **UKOLN - Interoperability Focus - About**. UKWA -UK Web Archive. Disponível em: <<https://www.webarchive.org.uk/wayback/en/archive/20150806095947/http://www.ukoln.ac.uk/interop-focus/about/leaflet.html>>. Acesso em: 25 fev. 2019.
- Zhang, S.; Bodenreider, O. (2005) Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference. **AMIA ... Annual Symposium proceedings. AMIA Symposium**, p. 864–868.