An Application of Natural Language Processing and Ontologies to Electronic Healthcare Records in the Field of Gynecology

***Amanda Damasceno de Souza*[1] -**
amanda.dsouza@fumec.br
Fernanda Farinelli
Eduardo Ribeiro Felipe
Armando Sérgio de Aguiar Filho[1]
and Mauricio Barcellos Almeida

[1]*FUMEC University, Graduate Program in Information and Communication Technology and Knowledge Management (PPGTICGC), Belo Horizonte, MG, Brazil.*

Research Background

- Electronic Health Records (EHR) are an essential source of real-world health information for several purposes.
- Information in EHRs is often recorded in an **unstructured format**, which poses challenges to using it for computational purposes.
- Therefore, an effective means of **connecting the ordinary terms found in EHRs** with standard medical terminologies could improve IR processes.
- One option is to map the EHR's terms to **standardized terminologies.**

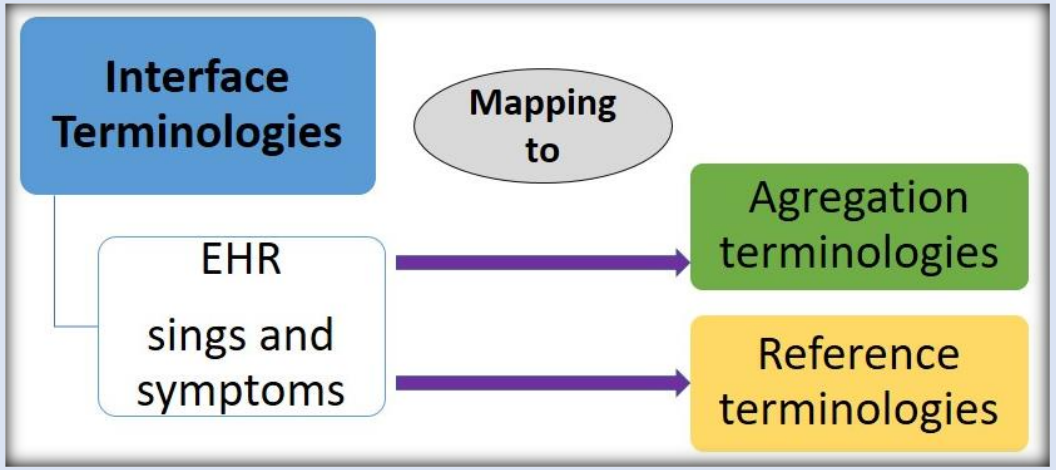
Research Background

Our investigation draws on a study by Schulz et al. (2017), who analyze terminology standardization and propose a methodology to connect three types of clinical terminologies:

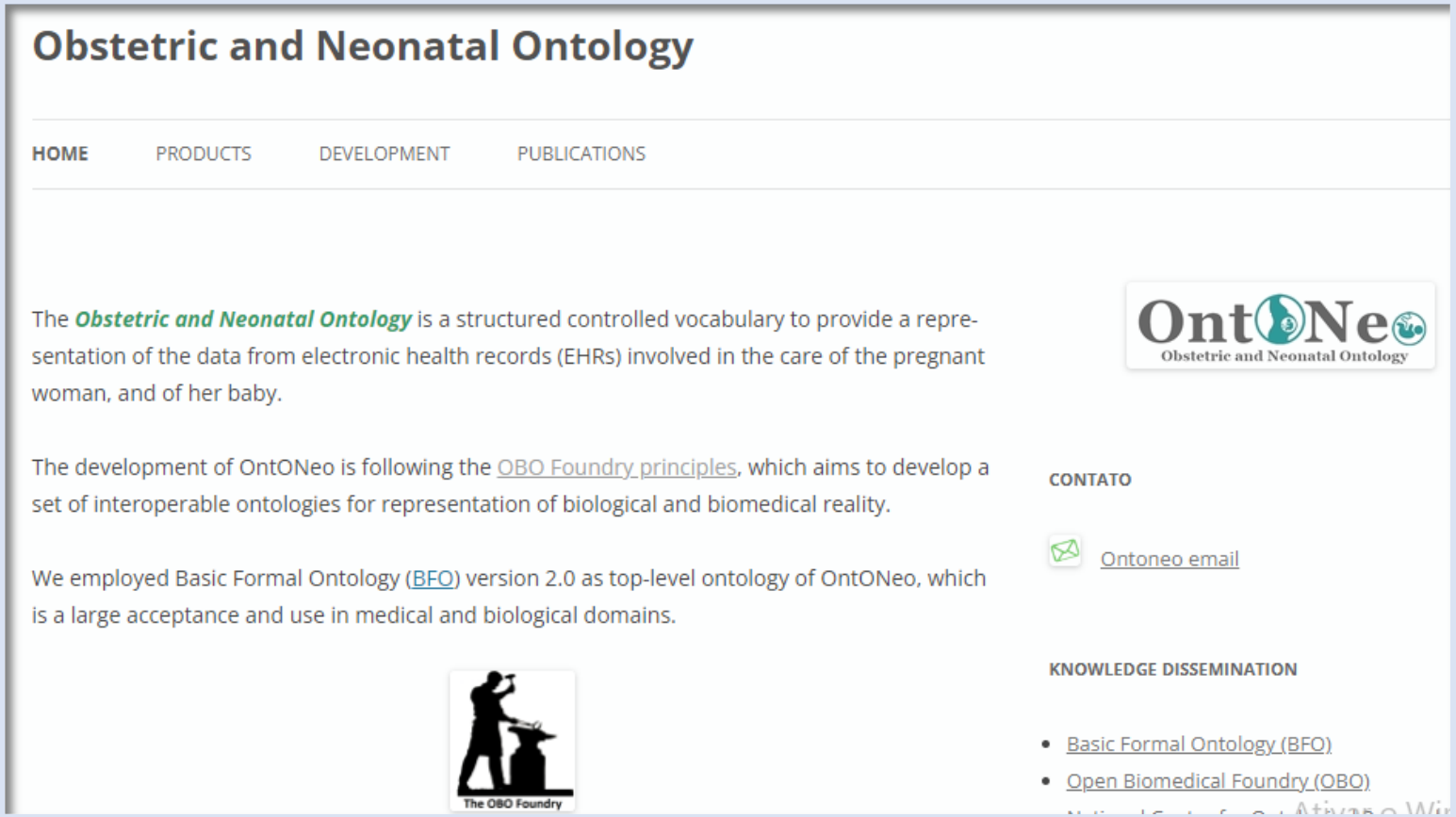
1. ***Interface terminologies*****:** namely, medical chart text or medical jargon (EHR);
2. ***Reference terminologies:*** which are controlled vocabularies and ontologies;
3. ***Agregation terminologies:*** which include the International Classification of Diseases (ICD), Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and others.

Proposal

- ✓The gap posed by Schulz et al.(2017) requires finding a way to connect the clinical data in an EHR's clinical texts to **standardized clinical terminologies**.
- ✓As its principal contribution, our research verified medical jargon terms that do not correspond to OntONeo Ontology (***reference terminology***) and verified medical jargon terms that do not correspond to ICD-10 (***agregation terminology***).

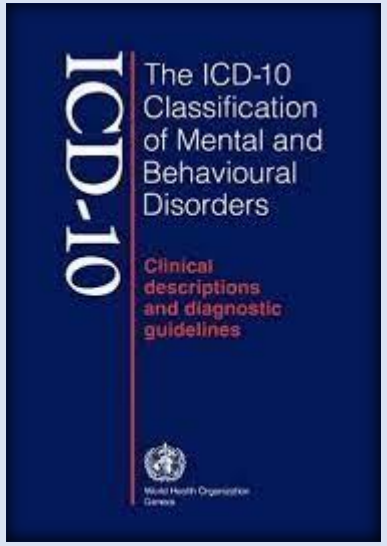


Clincal terminologies



https://ontoneo.com/

DeCS/MeSH
Descritores em Ciências da Saúde



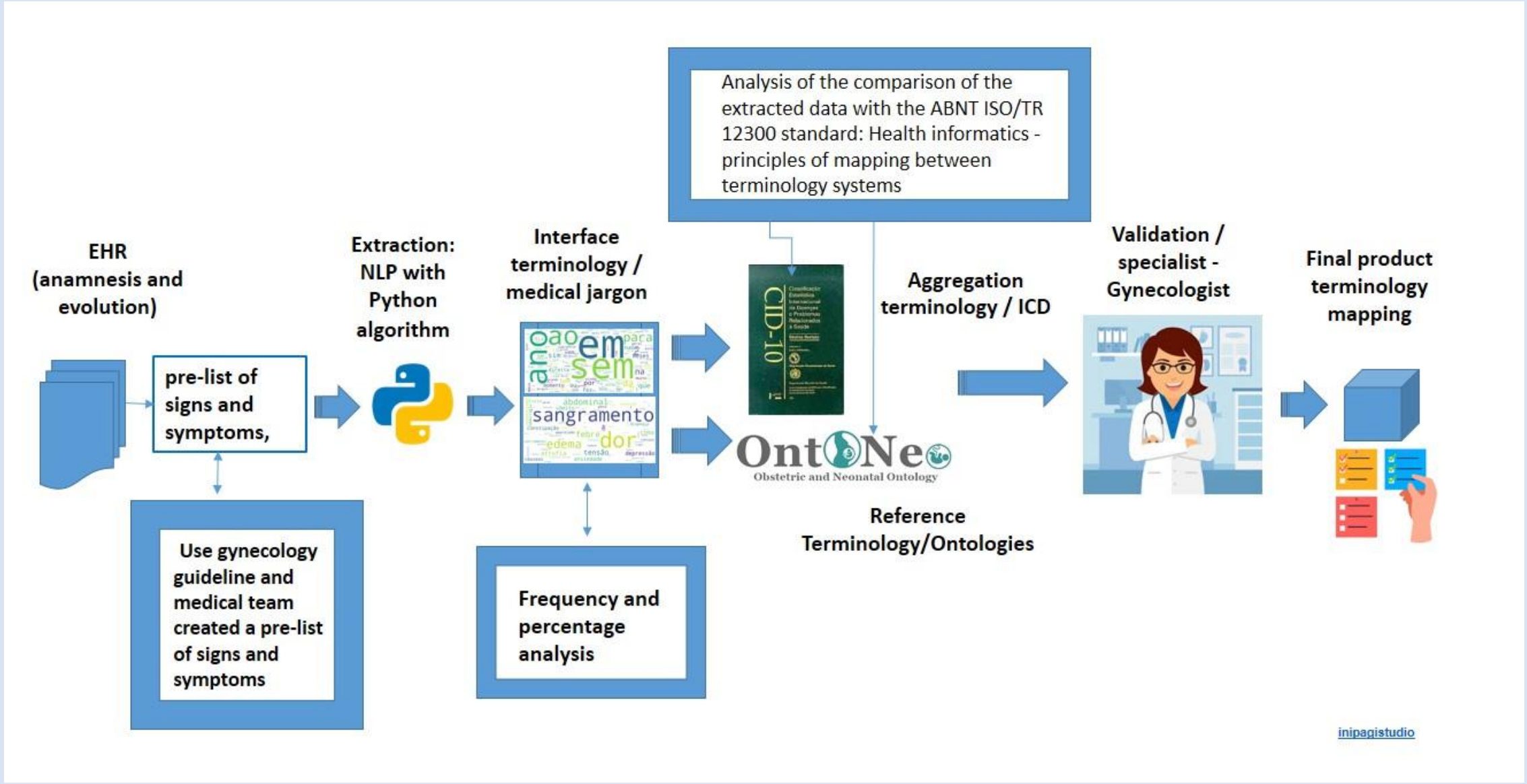
https://www.medicinanet.com.br/cid10.htm

https://decs.bvsalud.org/

Methodology

- We applied natural language processing (NLP) techniques and ontologies specifically OntONeo.
- **NLP** to extract and analyze **signs and symptoms** from clinical texts.
- The information extraction was performed in a large private hospital, which provided a sample of **32,291** real EHRs containing medical notes in free text.
- Notes cover the **evolution and medical history of patients (anonymized)** from the gynecology department during the year **2018.**
- Ethical approval was by the local committee!

Methodology



The steps for mapping

1) Document the mapping process between clinical terminologies.

2) Verify the semantic equivalence between terms.

3) Utilize a source mapping for terms with **multiple synonyms**.

4) Analyze risk factors and document ways to ensure consistency in mapping.

5) Clarify the meaning and fully use the form for abbreviations in the interface terminology.

6) Map the target terms of the reference terminology selected from (**DeCS**) and (**MeSH**), and **OntONeo**.

7) Create a mapping table to demonstrate the types of interoperability verification:

- **one term for one**
- **one term for many terms**
- **many terms for one term**
- **many terms for many terms**
- **do not interoperate**.

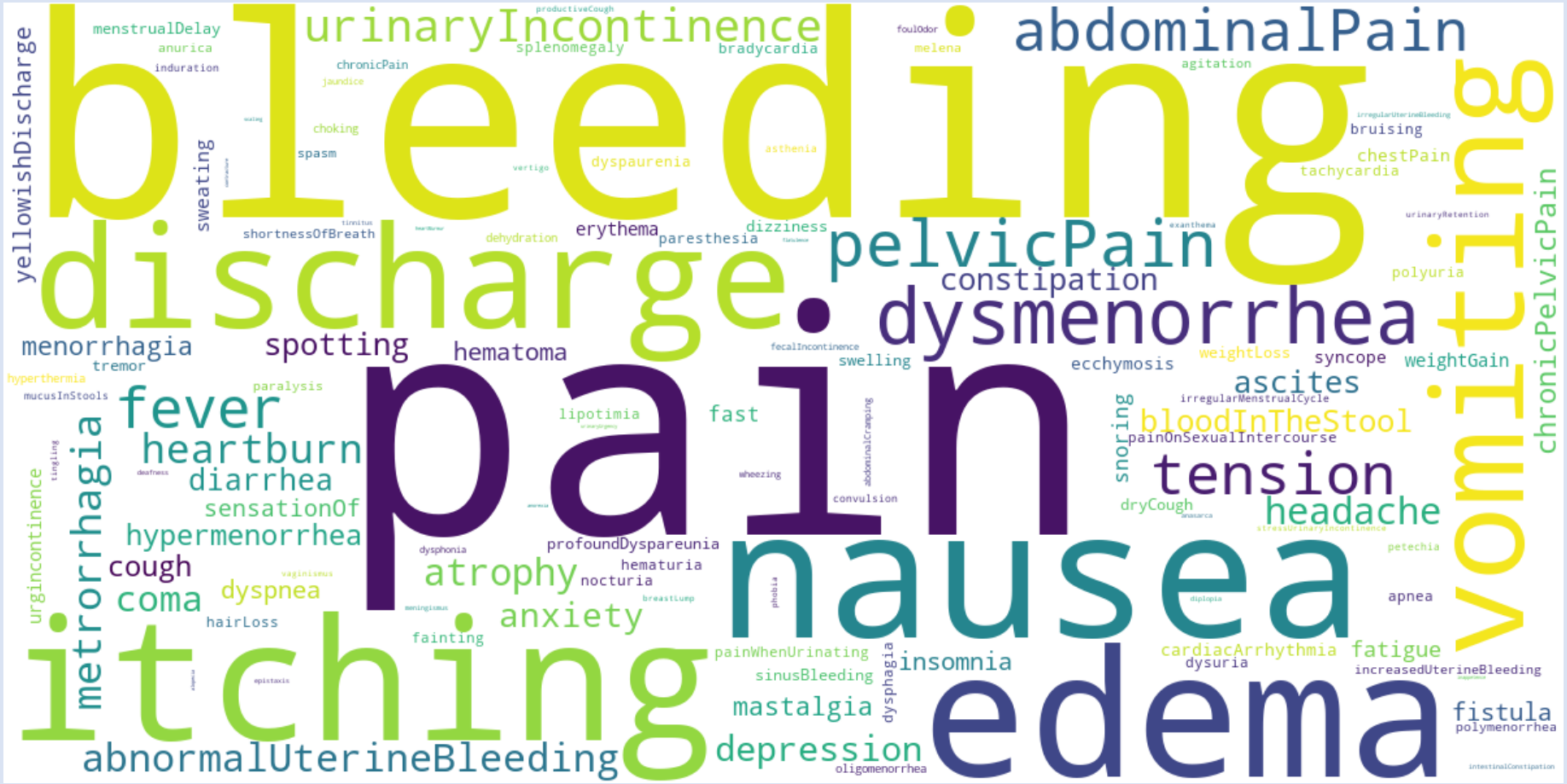
Table of Mapping of Terms

Mapping	Relation	Final decision
Interoperate one term for one	A single source class is linked to a single target class or term	Retain
Interoperate one term for many terms	A single source class is linked to multiple target classes or terms	Define a class according to basic formal ontology (BFO) and choose term that poses **no** clinical risk
Interoperate many terms for one term	Multiple source classes are linked to a single target class or term	Define a class according to BFO and choose term that poses **no** clinical risk
Interoperate many terms for many terms	Multiple source classes are linked to multiple target classes or terms	Define a class according to BFO and choose a term that poses **no** clinical risk

ABNT ISO/TR 12300

Results

Figure 1: Word Cloud of Most Frequent Signs and Symptoms.



Examples of correlated terms found compared with signs and symptoms of OntoNeo, DeCS/MeSH, and ICD-10

EHRs	OntONeo	DeCS/MeSH	ICD-10
Irregular menstrual cycle	-Process - biological_process - reproductive process - single organism reproductive process - ovulation cycle - menstrual cycle - Quality - Phenotypic abnormality - Abnormal genital system morphology -Abnormality of the menstrual cycle	Menstrual cycle	–
Itching	–	Pruritus	L29.0 Pruritus ani L29.2 Pruritus vulvae L29.3 Anogenital pruritus, unspecified L29.8 Other pruritus L29.9 Pruritus, unspecified Itch NOS
Dysmenorrhea	- Quality - information carrier- sintoma - nervous system symptom - sensation perception - pain	Dysmenorrhea	R10 Abdominal and pelvic pain R10.1 Pain localized to upper abdomen
Painful urination	- Quality - information carrier- sintoma - nervous system symptom - sensation perception - pain - renal colic	–	R30 Pain associated with micturition

- The term "**irregular menstrual cycle**" is correlated to the OntoNeo Ontology and DeCS/MeSH terms but did not show a corresponding term in the ICD-10.
- The term "**itching**" is absent in the ontology.
- "Dysmenorrhea" is already included in the three terminologies.

Mapping Interface Terminology Terms to the Reference Terminology (OntONeo)

Interoperability	Signs and Symptoms	
	n	%
Interoperate one term for one	**27**	**20,30**
Interoperate one term for many terms	5	3,76
Interoperate many terms for one term	18	13,53
Interoperate many terms for many terms	3	2,26
Non-interoperable	**80**	**60,15**
Total	**133**	**100**

Mapping Interface Terminology Terms to Aggregation Terminology (ICD)

Interoperability	Signs and Symptoms	
	n	%
Interoperate one term for one	**43**	**30,07**
Interoperate one term for many terms	13	9,09
Interoperate many terms for one term	6	4,20
Interoperate many terms for many terms	5	3,50
Non-interoperable	**76**	**53,15**
Total	143	100

Limitations

- ✓ This research described some differences in syntax and semantics that posed obstacles to achieving interoperability between clinical terminologies.

- ✓ To reduce these differences, we propose using existing **knowledge representation resources** in the Information Science field and the assistance of Clinical Medical Librarians.

- ✓ We identified several issues with ***spelling, punctuation, and typographical errors*** in the analyzed text from EHR.

Final Considerations

- ✓ We modified the second step of the proposal by Schulz *et al.* (2017), instead of the reconciliation step between **reference and aggregation terminologies**, we mapped **interface terminologies to aggregation terminologies**.
- ✓ This modification was necessary because we focused on analyzing the mappings between **interface terminology and clinical terminologies**.
- ✓ The medical jargon (***interface terminology***) used in clinical practice proved to be different and distant from standardized terminologies such as ontologies (**reference terminologies**) and even from ICD-10 (**aggregation terminology**).
- ✓ A primary difficulty in analyzing the medical jargon used in interface terminology, namely, its **epistemological aspects**, which depend heavily on the medical context.
- ✓ Thus, **ontology is an artifact** that should be used in seeking a solution to this difficulty.

References

S. Schulz, J. M. Rodrigues, A. Rector, C. G. Chute. Interface Terminologies, Reference Terminologies, and Aggregation Terminologies: A Strategy for Better Integration. Stud Health Technol Inform, 245(2017):940-944.

A. L. Rector. Clinical Terminology: Why is it so Hard? Methods of Information in Medicine, Stuttgart, 38:147-157, 1999.

F. Farinelli. et al. OntONeo: The Obstetric and Neonatal Ontology. In: Conference: International Conference on Biomedical Ontology 2016, ICBO, Corvallis, Oregon, USA, August 2016. Available at: https://www.researchgate.net/publication/304254064_OntONeo_The_Obstetric_and_Neonatal_Ontology.

A. D. Souza. Clinical Practice Discourse and Standardization Terminologies: Investigating the Connection. Pós-Graduação em Gestão e Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte. 2021. [Portuguese]. Available at: http://hdl.handle.net/1843/38044. Accessed on: 03 Jul. 2023.

Associação Brasileira de Normas Técnicas. Relatório técnico ISO/TR 12300: Informática em saúde – princípios de mapeamento entre sistemas terminológicos. 28.11.2016. Rio de Janeiro: ABNT, 2016.[Portuguese]

Health Sciences Descriptors: DeCS [Internet]. 2017 ed. São Paulo (SP): BIREME / PAHO / WHO. 2017 [updated 2017 May; cited 2017 Jun 13]. Available at: http://decs.bvsalud.org/I/homepagei.htm

National Center for Biotechnology Information, U.S. National Library of Medicine. (2022). MeSH (Medical Subject Headings). Bethesda: NLM. Available at: https://www.ncbi.nlm.nih.gov/mesh/.

R. Arp, B. Smith, and A. D. Spear. Building Ontologies with Basic Formal Ontology. Cambridge, MA: The MIT Press, 2015.

ICD-10. (2022). Symptoms and signs involving the genitourinary system R30-R39. Available at: https://www.icd10data.com/ICD10CM/Codes/R00-R99/R30-R39.

Amanda Damasceno de Souza

PhD in Knowledge Management and Organization - ECI/UFMG;
Master in Information Science - ECI/UFMG; Specialist in Strategic Information Management - ECI/UFMG; Graduated in Library Science - ECI/UFMG
Vice-coordinator of the Research Ethics Committee at Felício Rocho Hospital - BH/MG
Coordinator of the Research Ethics Committee at FUMEC University - BH/MG;
Member of the research groups: Knowledge Representation, Ontologies and Language (ReCOL) and Center for Studies and Research on Information Resources, Services and Praxis (NERSI)
Lecturer in the Bachelor's Degree in Aesthetics and the Postgraduate Program in Information and Communication Technology and Knowledge Management (PPGTICGC), Universidade FUMEC

UNIVERSIDADE FUMEC

NERSI

ReCOL CIÊNCIA DA INFORMAÇÃO - UFMG

ResearchGate: https://www.researchgate.net/profile/Amanda_De_Souza2
LinkedIn: https://www.linkedin.com/in/amanda-damasceno-de-souza-aa509452/
Instagram: @amandasouza81

Thank you! Obrigada! Danke