





## **USO DE FERRAMENTAS COMPUTACIONAIS COMO AUXÍLIO AO MÉTODO DE MAPEAMENTO CRUZADO ENTRE TERMINOLOGIAS CLÍNICAS**

Denilsen Carvalho Gomes<sup>1</sup>   
Lucas Emanuel Silva e Oliveira<sup>1</sup>   
Marcia Regina Cubas<sup>1</sup>   
Claudia Maria Cabral Moro Barra<sup>1</sup> 

<sup>1</sup>Pontifícia Universidade Católica do Paraná, Programa de Pós-Graduação em Tecnologia em Saúde, Curitiba, PR, Brasil

### **RESUMO**

**Objetivo:** refletir sobre o uso de ferramentas computacionais no método de mapeamento cruzado entre terminologias clínicas.

**Método:** estudo de reflexão.

**Resultados:** o método de mapeamento cruzado consiste na obtenção de listagem de termos, por meio de extração e normalização; ligação entre os termos da listagem e os da base de referência, mediante regras previamente definidas; e agrupamento dos termos em categorias: combinação exata ou parcial ou, de maneira mais detalhada, termo similar, termo mais abrangente, termo mais restrito e termo não concordante. Realizado manualmente em muitos estudos, pode ser automatizado com a utilização do *Unified Medical Language System* (UMLS). A obtenção da listagem de termos pode ocorrer de forma automática por algoritmos de processamento de linguagem natural, sendo que a utilização de regras para identificação de informação em textos permite que o conhecimento do especialista seja acoplado ao algoritmo, podendo ser realizada por técnicas baseadas em *Machine Learning*. Quando se trata de mapeamento de termos utilizando o modelo de sete Eixos da Classificação Internacional para a Prática de Enfermagem (CIPE<sup>®</sup>), o processo também pode ser automatizado por meio de algoritmos de processamento de linguagem natural, como o *POS-tagger* e o *parser sintático*.

**COMO CITAR:** Gomes DC, Oliveira LES, Cubas MR, Barra CMCM. Uso de ferramentas computacionais como auxílio ao método de mapeamento cruzado entre terminologias clínicas. *Texto Contexto Enferm* [Internet]. 2019 [acesso ANO MÊS DIA];28:e20170187. Disponível em: <https://dx.doi.org/10.1590/1980-265X-TCE-2017-0187>

**Conclusão:** o método de mapeamento cruzado pode ser intensificado pelo uso de algoritmos de processamento de linguagem natural. No entanto, mesmo em casos de mapeamento automático, a validação dos resultados por especialistas não deve ser descartada.

**DESCRITORES:** Terminologia. Enfermagem. Informática. Vocabulário controlado. Métodos.

## USE OF COMPUTATIONAL TOOLS AS SUPPORT TO THE CROSS-MAPPING METHOD BETWEEN CLINICAL TERMINOLOGIES

### ABSTRACT

**Objective:** to reflect on the use of computational tools in the cross-mapping method between clinical terminologies.

**Method:** reflection study.

**Results:** the cross-mapping method consists of obtaining a list of terms through extraction and normalization; the connection between the terms of the list and those of the reference base, by means of predefined rules; and grouping of the terms into categories: exact or partial combination or, in more detail, similar term, more comprehensive term, more restricted term and non-agreeing term. Performed manually in many studies, it can be automated with the use of the Unified Medical Language System (UMLS). Obtaining the terms list can occur automatically by natural language processing algorithms, being that the use of rules to identify information in texts allows the expert's knowledge to be coupled to the algorithm, and it can be performed by techniques based on Machine Learning. When it comes to mapping terms using the 7-Axis model of the International Classification for Nursing Practice (ICNP®), the process can also be automated through natural language processing algorithms such as POS-tagger and the syntactic parser.

**Conclusion:** the cross-mapping method can be intensified by the use of natural language processing algorithms. However, even in cases of automatic mapping, the validation of the results by specialists should not be discarded.

**DESCRITORS:** Terminology. Nursing. Informatics. Controlled vocabulary. Methods.

## USO DE HERRAMIENTAS COMPUTACIONALES COMO AYUDA AL MÉTODO DE MAPEO CRUZADO ENTRE TERMINOLOGÍAS CLÍNICAS

### RESUMEN

**Objetivo:** reflexionar sobre el uso de herramientas computacionales en el método de mapeo cruzado entre terminologías clínicas.

**Método:** estudio de reflexión.

**Resultados:** el método de mapeo cruzado consiste en la obtención de listado de términos, por medio de extracción y normalización; conexión entre los términos del listado y los de la base de referencia, mediante reglas previamente definidas; y agrupación de los términos en categorías: combinación exacta o parcial o, de manera más detallada, término similar, término más amplio, término más restringido y término no concordante. Realizado manualmente en muchos estudios, puede ser automatizado con el uso del *Unified Medical Language System* (UMLS). La obtención del listado de términos puede ocurrir de forma automática por algoritmos de procesamiento de lenguaje natural, siendo que la utilización de reglas para identificación de información en textos permite que el conocimiento del especialista sea acoplado al algoritmo, pudiendo ser realizado por técnicas

basadas en *Machine Learning*. Cuando se trata de mapeo de términos utilizando el modelo de siete Ejes de la Clasificación Internacional para la Práctica de Enfermería (CIPE®), el proceso también puede ser automatizado a través de algoritmos de procesamiento de lenguaje natural, como el *POS-tagger* y el *parser sintático*.

**Conclusión:** el método de mapeo cruzado puede ser intensificado por el uso de algoritmos de procesamiento de lenguaje natural. Sin embargo, incluso en casos de asignación automática, la validación de los resultados por expertos no debe descartarse.

**DESCRIPTORES:** Terminología. Enfermería. Informática. Vocabulario controlado. Métodos.

## INTRODUÇÃO

No desenvolvimento de terminologias em saúde, é necessário harmonizar conceitos para assegurar a interoperabilidade dos dados e informar os pesquisadores da área sobre possíveis atualizações a serem realizadas.<sup>1</sup> O processo de elaboração, desenvolvimento e harmonização de terminologias em saúde compreende um grande esforço por parte de seus desenvolvedores, que possuem limites para efetivá-lo de forma individual. A *International Standards Organization* (ISO) e a *International Health Terminology Standards Organization* (IHTSDO) são organizações mundiais que se dedicam a esse processo,<sup>2-3</sup> enquanto a primeira investe esforços no sentido de normalização para o desenvolvimento de terminologias no domínio da saúde,<sup>2</sup> a segunda é responsável pelo desenvolvimento da *Systematized Nomenclature of Medicine - Clinical Terms* (SNOMED CT), coordenando projetos de harmonização dessa nomenclatura com outras terminologias, vocabulários controlados e classificações.<sup>3-4</sup>

A SNOMED CT compreende uma terminologia clínica de âmbito global, abrangendo inúmeras especialidades, disciplinas e requisitos. Por essa razão, minimiza o uso de diferentes terminologias ou sistemas clínicos, o que permite um maior compartilhamento e reutilização de informações clínicas estruturadas.<sup>3</sup> No Brasil, o Ministério da Saúde, por meio da Portaria nº 2073, de 31 de agosto de 2011, definiu sua utilização para a codificação de termos clínicos e mapeamento das terminologias nacionais e internacionais em uso no país, visando a suportar a interoperabilidade semântica entre os sistemas.<sup>5</sup>

Especificamente, na área de enfermagem, a utilização de uma terminologia padronizada proporciona um método claro para a documentação de suas práticas, fornece orientação e apoio para o enfermeiro em seu raciocínio clínico e nomeia os fenômenos de interesse da profissão, contribuindo para a construção do conhecimento específico.<sup>6</sup> Assim, a implementação de uma terminologia em ambientes de cuidado pressupõe que, antecipadamente, seja realizada uma comparação entre os registros do prontuário do paciente e a linguagem padronizada, o que pode ser feito por meio da metodologia de mapeamento cruzado.<sup>7</sup>

Considerando a recomendação do Ministério da Saúde quanto ao emprego da SNOMED CT, o mapeamento cruzado entre essa terminologia e as de enfermagem pode ampliar a representatividade dos fenômenos de enfermagem nas bases de dados nacionais, com a possibilidade de comparação com bases internacionais. Ainda, a utilização desse método contribui para a evolução e disseminação de terminologias pelos diversos países e especialidades de atuação dos enfermeiros<sup>8</sup> e seus resultados colaboram para que os profissionais reflitam sobre os termos que utilizam cotidianamente e não são registrados de maneira uniformizada.

O mapeamento cruzado é um procedimento metodológico referenciado pela área de enfermagem desde a década de 1990, tendo como objetivo principal a determinação de semelhanças e diferenças entre termos<sup>9</sup> e sendo uma das etapas para a construção de subconjuntos de diagnósticos, resultados e intervenções de enfermagem da Classificação Internacional para a Prática de Enfermagem (CIPE®).<sup>10</sup>

Diversas pesquisas nacionais e internacionais que têm por finalidade contribuir para a implementação de uma linguagem padronizada por meio de terminologias vêm utilizando essa metodologia.<sup>7,11-15</sup> Pesquisa que realizou o mapeamento cruzado dos termos da CIPE® 1.0 e da SNOMED CT identificou que 80% dos termos daquela estão presentes nesta.<sup>11</sup> Nesse sentido, atualmente, o *Internacional Council of Nurses (ICN)* disponibiliza tabelas de equivalência entre os enunciados de diagnósticos, resultados e intervenções de enfermagem da CIPE® e da SNOMED CT.<sup>16-17</sup>

Nos estudos citados, o mapeamento cruzado foi realizado de forma manual e, paulatinamente, foram incorporadas ferramentas computacionais para auxiliar sua operacionalização, de modo a reduzir tempo e diminuir inconsistências humanas. Outra justificativa para inclusão de ferramentas computacionais para ajudar no mapeamento cruzado é o uso de terminologias em múltiplos idiomas, que resulta na demanda de estudos visando à tradução automática ou semiautomática de um conjunto de dados.<sup>18-19</sup> Por sua vez, ferramentas computacionais são importantes recursos para elaboração e aprimoramento de subconjuntos terminológicos e criação de glossários e ontologias terminológicas completas.<sup>20-21</sup>

Apesar da incorporação de recursos computacionais para o mapeamento cruzado entre terminologias,<sup>22</sup> entre textos clínicos e terminologias<sup>23</sup> e entre elementos presentes em arquétipos para terminologias,<sup>24</sup> o potencial das ferramentas ainda não é completamente aproveitado e não se apresenta consenso sobre a automatização do método, tampouco sobre a sua efetividade, o que justifica a reflexão proposta neste artigo, que tem como objetivo refletir sobre o uso de ferramentas computacionais no método de mapeamento cruzado entre terminologias clínicas.

## REFLEXÃO

No contexto das linguagens padronizadas, o *cross-mapping*, traduzido como mapeamento cruzado, consiste em um método que permite a comparação de uma linguagem padronizada com a linguagem utilizada no cotidiano dos serviços de saúde ou, ainda, entre diferentes sistemas de classificação existentes.<sup>25</sup> O método constitui-se pela obtenção da listagem de termos, por meio de extração e normalização; ligação entre os termos da listagem e os da base de referência (terminologias estruturadas), mediante regras previamente definidas; e agrupamento dos termos em categorias. Os termos extraídos devem representar a amplitude das práticas de enfermagem num determinado espaço de cuidado; portanto, as pesquisas nesse domínio usam bases e temporalidades distintas. Sendo um processo humano, é passível de falhas pela quantidade de dados a ser processada.

Como exemplo das distintas bases e temporalidades, para o mapeamento entre a CIPE® 1.0, os diagnósticos de enfermagem contidos em prontuários de crianças e a nomenclatura de diagnósticos e intervenções de enfermagem do município de Curitiba PR, foi necessário recuperar, manualmente, 20% dos prontuários dos pacientes atendidos em seis meses - considerou-se uma consulta de cada prontuário selecionado, num total de 80.<sup>7</sup>

A transcrição integral das informações contidas em prontuários, mediante leitura profunda e exaustiva, foi relatada em estudo que mapeou diagnósticos de enfermagem de pacientes de uma Unidade de Terapia Intensiva (UTI) com a *North American Nursing Diagnosis Association International (NANDA I)*. A base foi constituída por 256 prontuários de pacientes que estiveram hospitalizados na UTI, num período de seis meses.<sup>12</sup> Por sua vez, estudos similares utilizaram uma ferramenta computacional, denominada Poronto,<sup>21</sup> para extração de termos de evoluções de enfermagem contidas em prontuário eletrônico de um hospital universitário<sup>26</sup> e para identificação de termos, em artigos científicos, relacionados à prática de enfermagem direcionada à criança e ao adolescente em situação de violência doméstica.<sup>27</sup> No primeiro estudo, foi utilizada uma base de 115.760 evoluções do paciente e uma temporalidade de dois anos, tendo sido extraídos 257.893 termos dos registros.<sup>26</sup> No segundo, a base foi composta por 40 artigos na íntegra, dos quais foram extraídos 17.365 termos.<sup>27</sup>

A extração automática de termos em meio aos textos é uma tarefa de algoritmos de Processamento de Linguagem Natural (PLN), a qual envolve a resolução de termos simples e compostos e pode ser baseada em estatística, linguística e/ou conhecimento.<sup>28</sup> Entre as ferramentas\* disponíveis para utilização, podem-se citar: CoGrOO, *Natural Language Toolkit* (NLTK), OpenNLP, *Stanford Core NLP* e GATE. Portanto, a inclusão de ferramenta computacional possibilita o processamento de um grande quantitativo de textos em busca de informação, em um tempo menor se comparado à atividade manual. No processo automático, utilizando como base empírica prontuários de pacientes, percebe-se a utilização de bases de dados e temporalidades maiores, quando comparado à extração de termos pelo processo manual. Além disso, a quantificação de termos, ou seja, a frequência com que um termo aparece no *corpus* de análise, é realizada automaticamente pela ferramenta, demonstrando a relevância de um termo ou conceito de enfermagem, em um determinado espaço de cuidado.

A quantificação dos termos de grandes bases de dados é uma atividade complexa se realizada de forma manual, devido ao número de ocorrência dos termos; por exemplo, evoluções de enfermagem de um hospital universitário apresentaram mais de 50.000 ocorrências dos termos “horário” e “abdome”, em um total de 115.760 registros.<sup>26</sup> Isso, em partes, explica a utilização de bases de dados menores por estudos que extraem termos pelo processo manual. No entanto, ressalta-se que, quanto maiores forem a base de dados e a temporalidade, maior será a possibilidade de representação dos fenômenos da prática de enfermagem.

O processo de normalização do conteúdo consiste na retirada de termos duplicados e na sua adequação quanto à grafia, gênero, número e tempo verbal.<sup>12,26–27</sup> Na forma automática, os métodos de normalização e adequação de textos originam-se das etapas de pré-processamento utilizadas em algoritmos de PLN. O pré-processamento e a tokenização, por sua vez, dizem respeito à transformação do texto de entrada em algo que o algoritmo computacional possa entender e manipular, podendo incluir a remoção de *stop words* (palavras sem função relevante em determinado contexto, não sendo necessárias para o processamento dos textos) e capitalizações do texto (letras maiúsculas para minúsculas), a normalização de palavras por meio de redutores linguísticos ou outras formas de padronização e, por fim, a separação das sentenças e palavras do texto em unidades individuais, denominadas *tokens*.<sup>29</sup>

Os algoritmos de PLN podem utilizar análise morfológica, sintática, semântica e pragmática. No que se refere à análise morfológica dos textos, pode-se citar o *POS-Tagger*,<sup>29</sup> o qual define a morfologia das palavras e suas classes gramaticais; por exemplo, na frase “paciente relata dor reduzida”, o algoritmo definiria que “paciente” é um substantivo masculino singular, “relata” é um verbo no presente do indicativo na terceira pessoa do singular, assim por diante. Ainda, casos comuns de normalização morfológica abrangem alterações na classe gramatical das palavras (substantivo, adjetivo e verbo), alterações de suas variações (tempo verbal, gênero, número e grau) ou até mesmo redução léxica de um conjunto de palavras que representam significado similar em um único termo, utilizando *stemming* ou *lemmatization* (exemplo: as palavras “organizar”, “organizado” e “organizando” são transformadas em “organ”).<sup>29–34</sup>

Quando realizada de forma automática, a normalização morfológica pode reduzir o tempo empregado pelo pesquisador durante esta etapa. No entanto, o conhecimento do especialista é de extrema importância, a exemplo da normalização dos termos “direita” e “direitos do paciente” – o primeiro refere-se a uma localização e o segundo, a um foco de atenção da enfermagem –; em caso de normalização automática, os dois termos seriam normalizados para “direito”.<sup>26</sup> Isso remete

---

\* CoGrOO NLP tools - <http://cogroo.sourceforge.net/download/current.html>; Natural Language Toolkit (NLTK) - <http://nltk.org>; OpenNLP - <https://opennlp.apache.org/>; Stanford Core NLP - <https://stanfordnlp.github.io/CoreNLP/>; GATE - <https://gate.ac.uk/>

à reflexão de que o processo de normalização necessita ser realizado de forma semiautomática, ou seja, é necessário o conhecimento do especialista para que a semântica dos termos seja preservada.

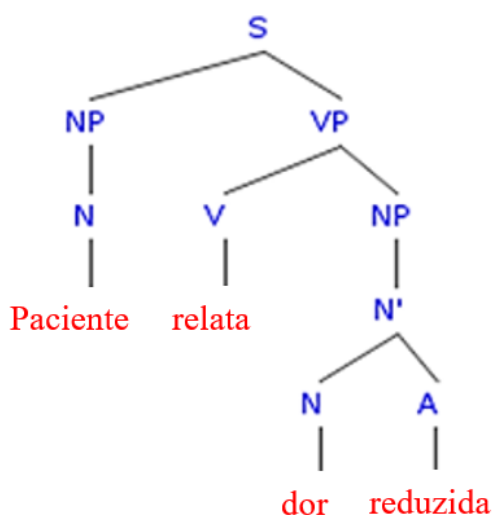
Com relação à análise sintática dos textos, pode-se citar o *parser sintático*,<sup>29</sup> o qual é separado em duas categorias: o *constituency parser* (Figura 1), que demarca a estrutura das sentenças de um texto, e o *dependency parser* (Figura 2), que estabelece as relações de dependência entre as palavras de um texto.

Outro aspecto de normalização do conteúdo extraído que pode ser realizado de forma automática, além da remoção de *stop words* e capitalização, é a expansão de abreviaturas utilizadas pelos profissionais de enfermagem no registro de suas atividades.<sup>29,31</sup>

A análise semântica consiste na descoberta do significado de palavras ou conceitos em meio ao texto. Entre os problemas que a análise semântica objetiva solucionar, destacam-se a resolução de ambiguidade<sup>35</sup> e o reconhecimento de entidades nomeadas,<sup>36</sup> o qual compreende a identificação e classificação de entidades, como nome de pessoas, organizações e localizações, em um texto.

As regras para o mapeamento cruzado podem ser determinadas de acordo com o delineamento do estudo, com base nas características da estrutura dos dados do sistema de informação e na terminologia a ser utilizada.<sup>37</sup> A quantidade de regras, se, por um lado, garante a exatidão do

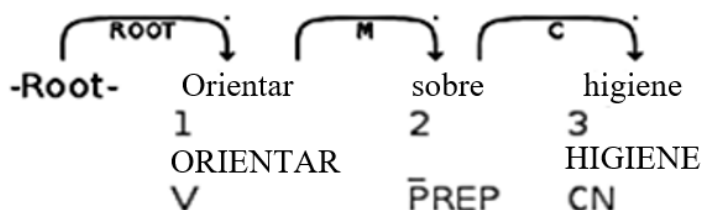
**Figura 1** - Demarcação da estrutura de uma sentença pelo constituency parser



S: sentença; VP: frases verbais; NP: frases nominais; N: nomes – substantivos; V: verbo; A: adjetivo; S: sentença; N': subfrase nominal constituinte.

Fonte: Adaptado de constituency parser, 2017.

**Figura 2** - Relação de dependência entre as palavras pelo dependency parser



Root: raiz; M: modificador; C: complemento; V: verbo; PREP: preposição; CN: nome comum.

Fonte: Adaptado de dependency parser, 2017.



mapeamento, por outro, exige do pesquisador um esforço e conhecimento que vai além da sua especialidade: necessita de base teórica e prática do sistema classificatório e de equivalências semânticas e transculturais.

Estudos que realizaram o mapeamento cruzado de forma manual entre termos e diagnósticos de enfermagem contidos em prontuários de pacientes e a NANDA I<sup>12,38</sup> incluíram regras como: garantir o sentido dos termos, verificando o contexto e o significado e não somente as palavras; comparar os termos com os enunciados de diagnósticos e os focos de atenção; comparar os termos com as características definidoras e os fatores relacionados e de risco; identificar e descrever os possíveis conceitos diagnósticos de enfermagem; e mapear os diagnósticos de enfermagem nos domínios e classes da NANDA I.

Do mesmo modo, estudos que fizeram o mapeamento cruzado de forma manual entre intervenções de enfermagem e a *Nursing Interventions Classification* (NIC)<sup>13</sup> incluíram regras como: utilizar os verbos das intervenções para realizar o mapeamento para a NIC; mapear a intervenção partindo do título da intervenção NIC para a atividade; manter a consistência entre a intervenção mapeada e a definição da intervenção na classificação; usar o título da intervenção NIC mais específico; e mapear as intervenções que tinham dois ou mais verbos para duas ou mais intervenções NIC correspondentes.

A utilização de regras para identificação automática de informação em meio ao texto (*rule-based information extraction*) é uma metodologia muito utilizada em ferramentas computacionais, pois permite que o conhecimento do especialista seja incorporado ao algoritmo. Essa abordagem tem algumas limitações conhecidas<sup>30</sup> e pode ser aprimorada se usada em conjunto com técnicas baseadas em estatística, como o *Machine Learning* (ML).

No caso de algoritmos de ML com aprendizado supervisionado, o conhecimento do especialista é passado ao algoritmo por meio da anotação de dados nos textos, podendo ser morfológica, sintática ou semântica. Esse processo é muito lento e custoso e demanda especialistas engajados na execução da tarefa, *guidelines* de anotação bem definidos e ferramentas computacionais que acelerem e auxiliem no processo.<sup>39</sup>

Quando se trata de algoritmos de aprendizado não supervisionado, o conhecimento do especialista não é necessário, pois o próprio algoritmo pode agrupar dados por similaridade e extrair as informações necessárias. Além disso, o uso de métodos baseados em estatística pode ampliar o alcance do algoritmo, não ficando limitado ao conhecimento do especialista e à geração de regras.

O estabelecimento de categorias para a disposição dos termos, última fase do mapeamento, deve seguir critérios capazes de possibilitar comparações posteriores ou reutilização dos resultados. De modo geral, quando o termo encontrado corresponde exatamente ao termo do sistema de classificação, é categorizado como combinação exata e, quando apresenta conceitos similares, sinônimos ou relacionados, como combinação parcial.<sup>13</sup>

No domínio da enfermagem, é comum o uso de critérios estabelecidos por Leal,<sup>40</sup> os quais indicam categorias mais detalhadas para o mapeamento, entre elas: termo similar, quando não existe concordância da grafia do termo, mas o significado é idêntico; termo mais abrangente, quando o termo identificado tem um significado maior do que o da terminologia; termo mais restrito, quando o termo identificado tem um significado mais limitado do que o da terminologia; e termo não concordante, quando não há concordância entre o termo identificado e o da terminologia.

No tocante ao mapeamento cruzado entre terminologias de enfermagem, a utilização do *Unified Medical Language System* (UMLS) pode ancorar a realização do mapeamento automático, uma vez que compreende uma fonte de conhecimento que integra centenas de terminologias ou classificações relacionadas à saúde mediante uma plataforma unificada.<sup>1</sup> Além disso, já existe uma iniciativa para a tradução do UMLS para o português brasileiro.<sup>41</sup>

O UMLS utiliza vários processos para integrar as terminologias, como o uso de ferramentas lexicais, para normalização de conceitos e preservação de significados e relações nos vocabulários de origem.<sup>42</sup> Entretanto, o processo automatizado pode apresentar limitações. O mapeamento automático, pelo UMLS, entre a *Logical Observation Identifiers Names and Codes* (LOINC), terminologia para testes laboratoriais e observações clínicas, e a SNOMED CT demonstrou-se insatisfatório, apesar de as duas terminologias abrangerem simultaneamente o domínio de procedimentos laboratoriais e utilizarem formalismos de representação de conhecimento semelhantes. O estudo considerou que, para melhorar o desempenho do processo de mapeamento automático, técnicas adicionais são necessárias.<sup>43</sup>

Correspondências imprecisas também foram observadas no mapeamento entre terminologias de enfermagem, indicando uma série de complexidades a ser abordadas no UMLS, sendo necessária a colaboração entre especialistas para resolver problemas nos mapeamentos semânticos.<sup>1</sup> Esse fato foi abordado em um mapeamento cruzado entre a CIPE<sup>®</sup> e a Classificação de Cuidados Clínicos (CCC) e entre a CIPE<sup>®</sup> e a NANDA I, em que houve 97% de correspondências exatas quando o mapeamento foi realizado por especialistas; quando processado pelo UMLS, a análise de comparação apresentou uma precisão global de 33,6% no mapeamento semântico.<sup>1</sup>

Por outro lado, quando se trata de mapeamento de termos utilizando o modelo de 7 Eixos da CIPE<sup>®</sup>, o processo pode ser automatizado por meio de algoritmos de PLN, como o *POS-tagger* e o *parser sintático*.<sup>29</sup> Isso se justifica pelo fato de, no modelo de 7 Eixos, os termos pertencentes ao eixo foco consistirem, em sua maioria, em substantivos; os termos do eixo julgamento corresponderem a adjetivos; e os termos do eixo ação referirem-se a verbos no infinitivo, o que possibilita uma disciplina para a semântica dos termos.

Considerando que a análise do contexto dos termos extraídos de bases empíricas é de extrema importância em trabalhos terminológicos,<sup>44</sup> sendo muitas vezes necessário considerar trechos de registros de enfermagem para identificar o contexto dos termos de enfermagem,<sup>38</sup> o *dependency parser* constitui ferramenta que pode apoiar a metodologia de mapeamento cruzado, uma vez que as dependências entre as palavras auxiliam a compreender o contexto em que os termos estão inseridos.

Embora o *POS-tagger*, o *parser sintático*<sup>29</sup> e outras técnicas de NLP sejam identificados como facilitadores para o método de mapeamento cruzado, com ferramentas amplamente disponíveis para utilização, ainda não se identificam estudos que os utilizaram, sendo esse um limite para a reflexão proposta neste artigo.

## CONCLUSÃO

A operacionalização da metodologia de mapeamento cruzado pode ser prejudicada pela quantidade de dados oriundos das bases empíricas e pela limitação humana no processo de comparação. Nesse sentido, as ferramentas computacionais constituem recursos para maximizar o tempo e minimizar erros da inspeção manual, mas são apoiadoras do especialista.

É necessário que os pesquisadores que se dedicam ao desenvolvimento de terminologias da área de enfermagem conheçam ferramentas computacionais capazes de auxiliar o processo de mapeamento cruzado, para que possam avaliá-las e utilizá-las potencialmente.

Ademais, as etapas de obtenção e normalização de termos são as que mais exploram o potencial de recursos computacionais, podendo o método de mapeamento cruzado ser intensificado pelo uso de algoritmos de PLN. Entretanto, mesmo em casos de mapeamento automático, a validação dos resultados por especialistas não deve ser descartada, em especial no que se refere à equivalência transcultural.



## REFERÊNCIAS

1. Kim TY, Coenen A, Hardiker N, Bartz CC. Representation of nursing terminologies in UMLS. *Amia Annu Symp Proc* [Internet]. 2011 [acesso 2016 Dez 20];2011:709-14. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243214/>
2. International Standards Organization (ISO). ISO/TS 17117 de 2002: controlled health terminology: structure and high-level indicators. 2002 [acesso 2016 Dez 06]. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:ts:17117:ed-1:v1:en>
3. International Health Terminology Standards Development Organisation (IHTSDO). Systematized Nomenclature of Medicine Clinical Terms - SNOMED CT [Internet]. Starter Guide. 2014 [acesso 2017 Jan 02]. Disponível em: <http://www.ihtsdo.org/snomed-ct/what-is-snomed-ct>
4. Campbell JR, Brear H, Scichilone R, White S, Giannangelo K, Carlsen B, *et al.* Semantic interoperability and electronic health records: context sensitive mapping from SNOMED CT to ICD-10. *Stud Health Technol Inform* [Internet]. 2013 [acesso 2016 Dez 21];192:603-7. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/23920627>
5. Ministério da Saúde (BR). Portaria 2073 de 31 de agosto de 2011: regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar. Brasília (DF): MS; 2011 [acesso 2016 Dez 21]. Disponível em: [http://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073\\_31\\_08\\_2011.html](http://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073_31_08_2011.html)
6. Carvalho EC, Cruz DALM, Herdman TH. Contribuição das linguagens padronizadas para a produção do conhecimento, raciocínio clínico e prática clínica da enfermagem. *Rev Bras Enferm* [Internet]. 2013 [acesso 2016 Dez 22];66(esp):134-41. Disponível em: <https://doi.org/10.1590/S0034-71672013000700017>
7. Luciano TS, Nóbrega MML, Saporoli ECL, Barros ALBL. Cross mapping of nursing diagnoses in infant health using the International Classification of Nursing Practice. *Rev Esc Enferm USP* [Internet]. 2014 [acesso 2016 Dez 23];48(2):250-6. Disponível em: <https://doi.org/10.1590/S0080-623420140000200008>
8. Barra DCC, Dal Sasso GTM. The nursing process according to the international classification for nursing practice: an integrative review. *Texto Contexto Enferm* [Internet]. 2012 [acesso 2017 Mar 13];21(2):440-7. Disponível em: <http://dx.doi.org/10.1590/S0104-07072012000200024>
9. Delaney C, Moorhead S. Synthesis of methods, rules and issues of standardizing nursing intervention language mapping. *Nurs Diagn*. 1997;8(4):152-156. <https://doi.org/10.1111/j.1744-618X.1997.tb00471.x>
10. Carvalho CMG, Cubas MR, Nóbrega MML. Brazilian method for the development terminological subsets of ICNP®: limits and potentialities. *Rev Bras Enferm* [Internet]. 2017 [acesso 2017 Mar 17]; 70(2):430-35. Disponível em: <http://dx.doi.org/10.1590/0034-7167-2016-0308>
11. Park HA, Lundberg C, Coenen A, Konicek D. Evaluation of the Content Coverage of SNOMED CT Representing ICNP Seven-axis Version 1 Concepts. *Methods Inf Med* [Internet]. 2011 [acesso 2016 Dez 20];50(5):472-8. Disponível em: <http://dx.doi.org/10.3414/ME11-01-0004>
12. Ferreira AM, Rocha EN, Lopes CT, Bachion MM, Lopes JL, Barros ALBL. Nursing diagnoses in intensive care: cross-mapping and NANDA-I taxonomy. *Rev Bras Enferm*. [Internet]. 2016 [acesso 2016 Dez 18]; 69(2):307-15. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-71672016000200307&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672016000200307&lng=en&nrm=iso&tlng=en)
13. Silva TG, Santana RF, Souza PA. Nursing interventions for elderly who aged in psychiatric institutions: crossed mapping. *Rev Eletr Enf* [Internet]. 2016 [acesso 2016 Dez 19];18:e1185. Disponível em: <https://www.revistas.ufg.br/fen/article/view/39049/21988>

14. Kim TY, Hardiker N, Coenen A. Inter-terminology mapping of nursing problems. *J Biomed Inform* [Internet]. 2014 [acesso 2016 Dez 19];49:213-20. Disponível em: <http://dx.doi.org/10.1016/j.jbi.2014.03.001>
15. Park HA, Lundberg C, Coenen A, Konicek D. Mapping ICNP version 1 concepts to SNOMED CT. *Stud Health Technol Inform* [Internet]. 2010 [acesso 2017 Feb 19];160(Pt 2):1109-13. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/20841856>
16. International Council of Nurses (ICN). ICNP para SNOMED CT: equivalency table for diagnosis & outcome statements. [Internet] Geneva: ICN; 2018. [cited 2019 Jan 18]. Available from: <https://www.icn.ch/sites/default/files/inline-files/ICNP-SCT%20Cross-Mapping%202018-DC.pdf>
17. International Council of Nurses (ICN). ICNP para SNOMED CT: equivalency table for intervention statements [Internet]. Geneva: ICN; 2018 [cited 2019 Jan 18]. Available from: <https://www.icn.ch/sites/default/files/inline-files/ICNP-SCT%20Cross-Mapping%202018-IC.pdf>
18. Cimino JJ, Barnett GO. Automated Translation between Medical Terminologies using Semantic Definitions. *MD Comput* [Internet]. 1990 [acesso 2017 Mar 10];7(2):104-9. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/2186251>
19. Silva J, Chaves T. An ontology-based approach for SNOMED CT translation [Internet]. 2015 [acesso 2017 Mar 05];1-5. Disponível em: <http://ceur-ws.org/Vol-1515/regular4.pdf>
20. Matney SA, Warren JJ, Evans JL, Kim TY, Coenen A, Auld VA. Development of the nursing problem list subset of SNOMED CT. *J Biomed Inform* [Internet]. 2012 [acesso 2017 Mar 05];45(4):683-8. Disponível em: <http://dx.doi.org/10.1016/j.jbi.2011.12.003>
21. Zahra FM, Carvalho DR, Malucelli A. Poronto: ferramenta para construção semiautomática de ontologias em português. *J Heal Informatics* [Internet]. 2013 [acesso 2017 Mar 05];5(2):52-9. Disponível em: <http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/232/167>
22. Lamy JB, Tsopra R, Venot A, Duclos C. A semi-automatic semantic method for mapping SNOMED CT concepts to VCM icons. *Stud Health Technol Inform* [Internet]. 2013 [acesso 2017 Mar 12];92:42-6. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861883/>
23. Stenzhorn H, Pacheco EJ, Nohama P, Schulz S. Automatic mapping of clinical documentation to SNOMED CT. *Stud Health Technol Inform* [Internet]. 2009 [acesso 2017 Mar 12];150:228-32. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19745302>
24. Yu S, Berry D, Bisbal J. Clinical coverage of an archetype repository over SNOMED-CT. *J Biomed Inf* [Internet]. 2012 [acesso 2017 Mar 24];45(3):408-18. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1532046411002152>
25. Nonino FOL, Napoleão AA, Carvalho EC, Petrilli Filho JF. A utilização do mapeamento cruzado na pesquisa de enfermagem: uma revisão da literatura. *Rev Bras Enferm* [Internet]. 2008 [acesso 2018 Jan 25];61(6):872-7. Disponível em: <https://doi.org/10.1590/S0034-71672008000600013>
26. Gomes DC, Cubas MR, Pleis LE, Shmeil MAH, Peluci APVD. Terms used by nurses in the documentation of patient progress. *Rev Gaúcha Enferm* [Internet]. 2016 [acesso 2017 Jan 08];37(1):e53927. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1983-14472016000100412&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1983-14472016000100412&lng=en&nrm=iso&tlng=en)
27. Albuquerque LM, Carvalho CMG, Apostólico MR, Sakata KN, Cubas MR, Egry EY. Nursing Terminology defines domestic violence against children and adolescents. *Rev Bras Enferm* [Internet]. 2015 [acesso 2017 Jan 08];68(3):393-400. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-71672015000300452&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672015000300452&lng=en&nrm=iso&tlng=en)
28. Conrado MS, Felippo A, Pardo TAS, Rezende SO. A survey of automatic term extraction for Brazilian Portuguese. *J Brazilian Comput Soc* [Internet]. 2014 [acesso 2017 Jan 08];20(12):1-28. Disponível em: <http://conteudo.icmc.usp.br/pessoas/taspardo/JBCS2014-ConradoEtAl.pdf>

29. Jurafsky D, Martin JH. *Speech and language processing*. 2nd ed. New Jersey, US: Pearson Prentice Hall; 2009.
30. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc* [Internet]. 2011 [acesso 2017 Feb 05];18(5):544-51. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3168328&tool=pmcentrez&rendertype=abstract>
31. Indurkha N, Damerau FJ. *Handbook of Natural Language Processing*. [Internet] USA: CRC Press Book; 2010, Disponível em: <https://karczmarczuk.users.greyc.fr/TEACH/TAL/Doc/Handbook%20Of%20Natural%20Language%20Processing,%20Second%20Edition%20Chapman%20&%20Hall%20Crc%20Machine%20Learning%20&%20Pattern%20Recognition%202010.pdf>
32. Singh J, Gupta V. A systematic review of text stemming techniques. *Artificial Intelligence Review*. Springer Netherlands [Internet]. 2017 [acesso 2018 Jan 25];48(2):157-217. Disponível em: <https://doi.org/10.1007/s10462-016-9498-2>
33. Liu H, Christiansen T, Baumgartner WA, Verspoor K. BioLemmatizer: A lemmatization tool for morphological processing of biomedical text. *J Biomed Semantics* [Internet]. 2012 [acesso 2018 Jan 25];3(1):1-29. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3359276/>
34. Soares MVB, Prati RC, Monard MC. Improvements on the Porter's Stemming Algorithm for Portuguese. *IEEE Lat Am Trans* [Internet]. 2009 [acesso 2018 Jan 25];7(4):472-7. Disponível em: <https://doi.org/10.1109/TLA.2009.5349047>
35. Ranjan Pal A, Saha D. Word Sense Disambiguation: a Survey. *Int J Control Theory Comput Model* [Internet]. 2015 [acesso 2018 Jan 25];5(3):1-16. Disponível em: <https://doi.org/10.5121/ijctcm.2015.5301>
36. Goulart RRV, Strube de Lima VL, Xavier CC. A systematic review of named entity recognition in biomedical texts. *J Brazilian Comput Soc* [Internet]. 2011 [acesso 2018 Jan 25];17(2):103-16. Disponível em: <https://link.springer.com/article/10.1007/s13173-011-0031-9>
37. Moorhead S, Delaney C. Mapping nursing intervention data into the nursing interventions classification (NIC): process and rules. *Nurs Diagn* [Internet]. 1997 [acesso 2016 Dez 19];8(4):137-44. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/9624991>
38. Tosin MHS, Campos DM, Blanco L, Santana RF, Oliveira BGRB. Mapping Nursing language terms of Parkinson's disease. *Rev Esc Enferm USP* [Internet]. 2015 [acesso 2017 Mar 03];49(3):409-16. Disponível em: <https://doi.org/10.1590/S0080-623420150000300008>
39. Oliveira LES, Gebelucá CP, Silva AMP, Moro CMC, Hasan SA, Farri O. A statistics and UMLS-based tool for assisted semantic annotation of Brazilian clinical documents. In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017 [acesso 2018 Jan 25]; Kansas City, USA, p. 1072-8. Disponível em: <https://doi.org/10.1109/BIBM.2017.8217805>
40. Leal MT. *A CIPE® e a visibilidade da enfermagem: mitos e realidade*. Lisboa (PT): Lusociência; 2006.
41. Oliveira LES, Hasan SA, Farri O, Barra CMCM. Translation of UMLS ontologies from European Portuguese to Brazilian Portuguese. In: *Anais do XV Congresso Brasileiro de Informática em Saúde*, 2016 Nov 27-30, Goiânia, Brazil. p. 373-9.
42. National Library of Medicine (NLM). *UMLS® Reference Manual*. Bethesda, MD: NLM; 2009.
43. Bodenreider O. Issues in mapping LOINC laboratory tests to SNOMED CT. *AMIA Annu Symp Proc*. [Internet] 2008 [acesso 2016 Dez 19];2008:51-5. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655945/>
44. Pavel S, Nolet D. *Manual de terminologia* [Internet] 2002 [acesso 2017 Mar 02]. Disponível em: [http://publications.gc.ca/collections/collection\\_2011/tpsgc-pwgsc/S53-28-2001-spa.pdf](http://publications.gc.ca/collections/collection_2011/tpsgc-pwgsc/S53-28-2001-spa.pdf)

## NOTAS

### CONTRIBUIÇÃO DE AUTORIA

Concepção do estudo: **Gomes DC, Oliveira LES, Cubas MR.**

Redação e/ou revisão crítica do conteúdo: **Gomes DC, Oliveira LES, Cubas MR.**

Revisão e aprovação final da versão final: **Cubas MR, Barra CMCM.**

### INFORMAÇÃO DE FINANCIAMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

### APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

### CONFLITO DE INTERESSES

Não há conflito de interesses.

### CONSENTIMENTO DE USO DE IMAGEM

Não se aplica.

### HISTÓRICO

Recebido: 19 de abril de 2017

Aprovado: 08 de fevereiro de 2018

### AUTOR CORRESPONDENTE

Denilsen Carvalho Gomes

deni.gomesc@gmail.com

