



Mapeamento e desambiguação de reduções léxicas no contexto da saúde: contribuições da Informática para estudos e pesquisa

MAPPING AND DISAMBIGUATION OF LEXICAL REDUCTION IN THE HEALTH'S CONTEXT: CONTRIBUTIONS OF INFORMATICS FOR RESEARCH AND STUDIES

Virgínia Bentes Pinto¹, José Álvaro Lima Silva de Arruda², Anderson Melo Duarte³, Andréa Soares Rocha da Silva⁴

¹ Doutora em Sciences de l'information et de la communication. Université Stendhal, Grenoble3-França. Universidade Federal do Ceará. ORCID: <https://orcid.org/0000-0003-1283-8292>
Email: vbentes@ufc.br

² Graduando em Biblioteconomia pela UFC/ Bolsista de Iniciação Científica (CNPq). Universidade Federal do Ceará. ORCID: <https://orcid.org/0000-0001-8310-3731>
Email: alvarolima093@gmail.com

³ Graduando em Biblioteconomia pela UFC/ Bolsista de Iniciação Científica (CNPq). Universidade Federal do Ceará. ORCID: <https://orcid.org/0000-0002-1586-2717>
Email: andersomello@alu.ufc.br

⁴ Doutora em Educação pela FACED/UFC. Universidade Federal do Ceará. ORCID: <https://orcid.org/0000-0001-5584-7071>
Email: andreasrs07@gmail.com

Correspondência: Av. da Universidade, 2762- 2º Andar - Fortaleza-CE, Brasil - CEP: 60020-180

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Conflito de interesses: os autores declaram que não há conflito de interesses.

Como citar este artigo

Bentes Pinto V; Arruda, JASL; Duarte AD; Silva ASR. Mapeamento e desambiguação de reduções léxicas no contexto da saúde:

contribuições da Informática para estudos e pesquisa. [online], volume 7, número especial III. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, fevereiro de 2022, p.143-160. Disponível em: <http://periodicos.ufc.br/resdite/index>. Acesso em "dia/mês/ano".

Data de recebimento do artigo: 12/08/2021

Data de aprovação do artigo: 09/11/2021

Data de publicação: 14/02/2022

Resumo

Introdução: O uso excessivo de reduções léxicas atinge todas as áreas do conhecimento e pode trazer interferências na comunicação entre os indivíduos. **Objetivo:** verificar as contribuições da Informática para minimizar as interferências oriundas do emprego de reduções léxicas no contexto da saúde. **Metodologia:** pesquisa exploratória com abordagem descritiva e pautada na Revisão Bibliográfica Integrativa. Empíria nas bases de dados *MEDLINE*, *Brapi*, *PubMed*, e *Scopus*. A pesquisa analisou títulos, resumos, palavras-chave ou textos completos, cobrindo o período de janeiro de 1989 a julho de 2019. **Resultados:** 570 artigos na MEDLINE, 572 na PubMed, 217 na Scopus e 2 na Brapi, dos

quais 51 foram concernentes ao objeto da pesquisa. Os textos são de pesquisadores da área da Ciência da Informação, Bioinformática, Medicina, entre outras, separados nas categorias de mapeamento e desambiguação definidas pelos pesquisadores. Os softwares Alice e Allie, destacam-se como boas ferramentas visando as reduções léxicas na literatura científica e para a criação de outros bancos de dados de processamento em linguagem natural. **Conclusão:** a Revisão Bibliográfica Integrativa traz grandes contribuições para o avanço do conhecimento relativo às reduções léxicas no contexto da saúde, evidenciando a necessidade de padronização e desambiguação visando reduzir as interferências no processo de comunicação.

Palavras-chave: Reduções Léxicas. Processamento de Linguagem Natural. Desambiguação de reduções léxicas. Revisão Integrativa.

Abstract

Introduction: *The excessive use of lexical reductions affects all areas of knowledge and can bring interference to those involved in the communication process.* **Goals:** *to verify the contributions of Informatics to minimize the*

interference arising from the use of lexical reductions in the health context. **Methodology:** *Exploratory research with a descriptive approach and guided by the Integrative Bibliographic Review. Based on MEDLINE, Brapci, PubMed, and Scopus databases. The research analyzed titles, abstracts, keywords or full texts, in the period from January 1989 to July 2019.* **Results:** *570 articles in MEDLINE, 572 in PubMed, 217 in Scopus and 2 in Brapci, which 51 were related to the object of the research. The texts were made by researchers in the areas of Information Science, Bioinformatics, Medicine, among others, separated in the categories mapping and disambiguation, defined by the researchers. Alice and Allie softwares stands out as good tools for lexical reductions in scientific literature and for the creation of other natural language processing databases.* **Conclusion:** *the Integrative Bibliographic Review brings significant contributions to the advancement of knowledge related to lexical reductions in the context of health, highlighting the need for standardization and disambiguation in order to reduce interferences in the communication process.*

Keywords: *Lexical Reductions. Natural Language Processing. Disambiguation of Lexical Reductions. Integrative Review.*

1. Introdução

Em um contexto de explosão informacional de proporções inimagináveis relacionadas ao aumento constante da produção documental, saltam aos olhos as iniciativas desenvolvidas e implementadas pelas áreas de Ciência da Informação (CI), Biblioteconomia, Informática, Ciência da Computação, entre outras, propondo alternativas para a produção, representação, organização, acesso e recuperação da informação.

No meio dessa exorbitância documental, percebe-se igualmente certa necessidade da adoção de reduções léxicas, na perspectiva de encurtar a escrita ou redação de textos, bem como a comunicação, adotando para tal, siglas, acrônimos, truncamentos ou símbolos. Citamos como exemplo as Ciências da Saúde (CS), Ciência da Informação (CI), Química, Física, Matemática, Informática e no meio ambiente do *ciberespaço*. Tais estratégias podem contribuir do ponto de vista positivo ou negativo. No primeiro aspecto, o emprego dessas reduções viabiliza os processos comunicativos ou a recuperação da informação (RI) em campos de conhecimentos, atuações profissionais e no cotidiano dos sujeitos em suas

buscas de informação. Contrariamente, no aspecto negativo, essas reduções, a exemplo das palavras do léxico assim como das linguagens de especialidades, são polissêmicas e, portanto, trazem interferências no processo de comunicação e prejuízos inimagináveis para os envolvidos nesses processos.

Estudos vêm sendo empreendidos a respeito dessas práxis, particularmente nos campos da CI, Terminologia, Linguística, Bioinformática, Ciências da Saúde, Informática e Direito. Na perspectiva de construção de vocabulários controlados que visam à padronização, bem como analisando os inconvenientes ou prejuízos oriundos dessa conduta. Diante de tal fato, questionamos: quais iniciativas oriundas da Informática estão sendo realçadas na literatura científica, que visem contribuir para reduzir interferências resultantes do emprego de reduções léxicas no contexto da saúde?

A partir desse problema, definimos como **objetivo geral**: verificar as contribuições da Informática para minimizar as interferências provenientes do emprego de reduções léxicas no contexto da saúde. Os **objetivos específicos** são: i) mapear na literatura científica estudos e pesquisas que abordem a temática das reduções léxicas no contexto da saúde; ii) identificar as ferramentas de mapeamento automático e de desambiguação das reduções léxicas relacionadas à saúde.

Desde a institucionalização da Informática, suas contribuições para que as pesquisas avancem, particularmente, nos campos da Ciência da Informação e suas subáreas (Biblioteconomia, Museologia, Arquivologia), da Terminologia, da Linguística e na área da Saúde, são perceptíveis. Da mesma forma não se pode negar seu papel no cotidiano laboral dos profissionais dessas áreas. Destacam-se neste íterim as ferramentas voltadas para a representação, organização, acesso e recuperação da informação, bem como para a construção de vocabulários, tesouros, ontologias e no contexto do tratamento automático da língua.

Nesse cenário, vale ressaltar os trabalhos apresentados nas *Text Retrieval Conference* (TREC), evento que congrega uma vasta literatura referente a aplicabilidade da Informática e da Computação para a representação e organização da informação. Na mesma linha, Chang et al¹ criaram um “método robusto para identificar abreviações usando aprendizado de máquina supervisionado. O método adota um conjunto de recursos que descrevem diferentes padrões vistos comumente em abreviações”.

Destarte, neste artigo, por meio de uma Revisão Bibliográfica Integrativa, discutimos e apresentamos algumas ferramentas disponíveis que auxiliam no processamento automático para o mapeamento e desambiguação de reduções léxicas no contexto da

saúde. Esperamos que os achados despertem para novas pesquisas neste cenário, trazendo à baila as contribuições da Informática para o desenvolvimento de ferramentas que, dentre outras coisas, venham a auxiliar a equipe multiprofissional da área da Saúde a elucidar e manusear as reduções léxicas, resolvendo os problemas provenientes dessa prática.

2. Fundamentação teórica

2.1 Algumas palavras sobre redução léxica

As reduções léxicas relacionadas às abreviações (abreviaturas, siglas, truncamentos e acrônimos) e símbolos são empregadas na redação de textos verbais e também na comunicação oral em todas as áreas de saberes. Tal fato não é recente, ele se constata desde os tempos mais remotos da História da Humanidade e conforme Bloom² afirma, os “[...] romanos, logo descobriram que quando o espaço e o tempo eram limitados, longas inscrições poderiam ser mais facilmente acomodadas em pergaminho (ou laboriosamente cinzeladas na pedra) se as palavras comuns fossem abreviadas”. Tal prática não foi perdida, muito pelo contrário, “[...] Como a língua latina e a cultura formaram as ferramentas básicas para a ciência e a tecnologia europeia (incluindo a maioria da terminologia médica) o hábito não foi perdido”. Nessa esteira, o autor argumenta que “A redação médica incorporou muitas mudanças visando compactar cada vez mais as palavras no meio analógico e em páginas eletrônicas”. Para esse autor, o fenômeno “Inicialismos e suas variantes pronunciáveis, o acrônimo, são cada vez mais favorecidos nesses meios de comunicação, muitas vezes ao preço da clareza”². Também, no Cristianismo são encontradas reduções nos manuscritos religiosos, NSJC (*Notre Seigneur Jésus-Christ*), INRI (*Jesus Nazarenus Rex Iudaeorum*) ou do poder SAR (*Son Altesse Royale*) etc.

Conforme defende Garner³, “Um dos tipos mais irritantes de pedantismo na escrita moderna é o uso excessivo de abreviações, especialmente nomes abreviados [...]”. Em suas origens, “com certeza, as abreviações visavam servir à conveniência do leitor, encurtando os nomes para que frases complicadas não precisassem ser repetidas na íntegra. As supostas simplificações realmente simplificaram”. Contudo, muitos escritores, “particularmente técnicos parecem ter perdido de vista esse objetivo: eles permitem que termos abreviados proliferem, e sua prosa rapidamente se torna um sistema híbrido-ínglês de hieróglifos exigindo que o leitor se refira constantemente aos usos originais de termos para compreender o significado”.

As pesquisas em que os *corpora* se constituem de prontuários do paciente autenticam tais reflexões, fato que dificulta a compreensão por parte dos profissionais da saúde, particularmente, em administração de medicamentos ou na realização de exames, e também pelos pacientes ou seus responsáveis que precisam compreender os protocolos a serem seguidos em tratamentos de rotina ou especializados.

No entendimento de Calvet⁴, “A siglação às vezes carrega, conotado, ironia ou zombaria, como no (RPR, a Religion Prétendue Réformée), ou o NC, os Novos Conversos, Acrônimo do século 17”. No entanto, afora esse fato, “a sigla é rara até o século 19 e, sobretudo, não aparece nos campos onde o encontramos principalmente hoje”. Em compensação, no âmbito político e no sindicalismo, o autor afirma que elas começaram a aparecer no início do século XX, sendo as primeiras: Confederação Geral dos Trabalhadores (CGT), criada em 23 de setembro de 1895, na cidade de Limoges-França, e SFIO (Secção Francesa da Internacional Operária).

Em nossas observações empíricas, na área da Saúde, consideramos que, de certa forma, há exagero na adoção de reduções léxicas no momento das anotações nos registros dos prontuários do paciente e, inclusive, nas orientações médicas ao paciente e receitas que são entregues a eles e/ou familiares. A literatura enuncia que esse hábito pode trazer consequências danosas tanto do ponto de vista das ações de cuidados da pessoa enferma, bem como aos profissionais que poderão, inclusive, responder processos pelas interferências causadas pela tradição dessas reduções léxicas na redação de terapêuticas, diagnósticos definitivos ou nas orientações dadas ao doente ou de alta hospitalar.

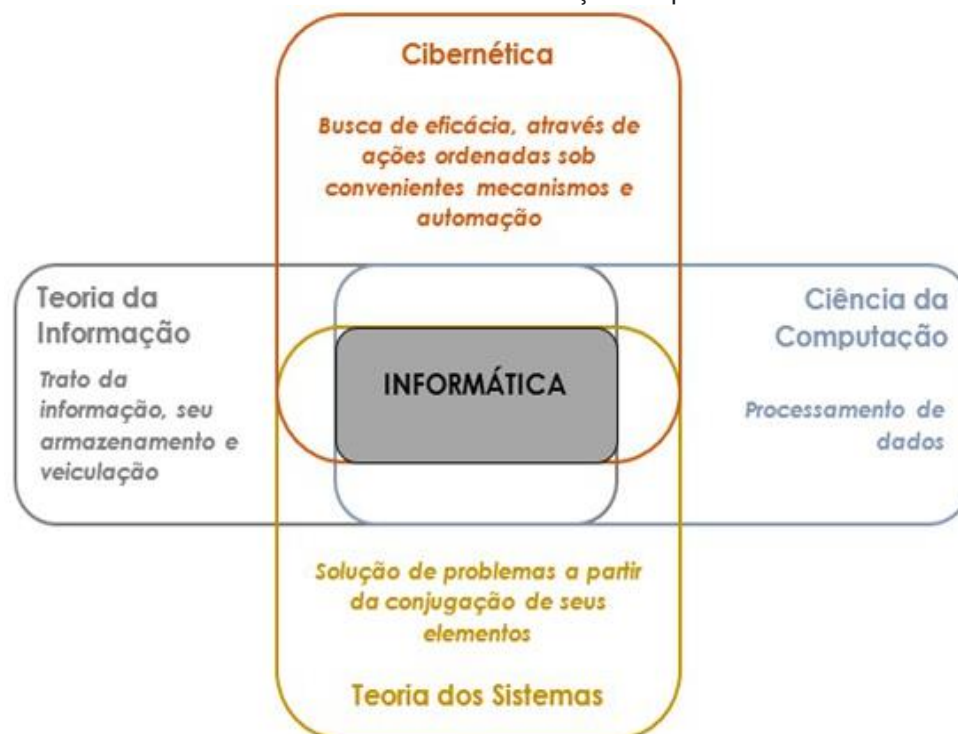
2.2 Informática e processamento automático de informação

A informática revolucionou o desenvolvimento e o funcionamento da sociedade de forma inquestionável. Todavia, embora estejamos todos de alguma forma já acostumados com seu uso nas nossas atividades cotidianas, a ponto de até abstrairmos sua presença, seu conceito, bem como a identificação e categorização de seus componentes e como estes funcionam, permanece um “mistério” para muitas pessoas.

Velloso⁵ inicia sua obra definindo a informática, simplesmente, como “informação automática”. Todavia, dedica às demais 390 páginas de sua obra a descrever, minimamente, conceitos, componentes e processos envolvidos nessa automação informacional. O autor afirma que o neologismo “informática” (*informatique*) originou-se na França em abril de 1966, como expressão alternativa a *information science*, que era então reconhecida pela

comunidade internacional, e que o termo inglês *informatic* surgiu como consequência. Analisando pelo terreno da semiótica, a informática é considerada comprometida com as ciências exatas e com as ciências sociais, situando-a na interseção de quatro áreas do conhecimento, conforme suas atividades características, como ilustrado na figura 1.

Figura 1: A Informática e suas atividades na intersecção de quatro áreas do conhecimento.



Fonte: Adaptada de Velloso (2011).

O tratamento automático da informação pressupõe o manuseio de computadores, que evoluem de forma contínua e acelerada com a redução progressiva do tamanho de seus componentes, assim como de preço, desde o período da Segunda Guerra Mundial nos anos de 1940 até o presente tempo. As atividades relacionadas à Informática sofreram diferentes designações ao longo do tempo, acompanhando a evolução do *hardware* (componente físico, incluindo dispositivos e acessórios) e dos processos lógicos que lhes dão vida e conferem efetividade ao seu propósito, e os *softwares* (algoritmos implementados em linguagens de programação específicas). Assim, derivações tais como “Processamento Automático de Dados” e “Processamento Eletrônico de Dados” foram simplificadas para “Processamento de Dados”, ou se tornaram mais específicas e especializadas, a depender do tipo ou formato midiático do dado manipulado; dos quais posteriormente surgiram as áreas de “Processamento de textos”, “Processamento de imagens”, dentre outras.

As especialidades da Informática foram naturalmente sendo percebidas e aplicadas às demais áreas do conhecimento, algumas por aproximações óbvias, como no caso da Ciência da Informação; outras, aparentemente mais distantes, têm progressivamente reconhecido sua utilidade e se apropriado de suas ferramentas e recursos, encontrando uma diversidade de aplicações para estas, como atualmente observamos na área da Saúde.

A Informática possui inúmeras aplicações na área da Saúde, nos dias atuais. Sistemas de registro eletrônicos em saúde, sistemas de monitoramento remoto de pacientes, sistemas de apoio à decisão clínica, telemedicina, sistemas de informação em saúde, sistemas de processamento de imagens radiológicas, tecnologias aplicadas à educação em saúde e processamento de textos e linguagem natural são apenas algumas das muitas aplicações da Informática nesse contexto⁶. Ademais, não podemos deixar de mencionar a grande contribuição da Informática para o Tratamento Automático da Língua (TAL).

3. Procedimentos Metodológicos

Pesquisa exploratória com abordagem descritiva e pautada na Revisão Bibliográfica Integrativa (RBI). Conforme as orientações de Cooper⁷, a RBI é constituída pelas seguintes fases: i) formulação do problema; ii) coleta de dados ou pesquisa na literatura; iii) avaliação dos dados, a fim de definir quais documentos serão selecionados; iv) análise e interpretação; e v) apresentação dos resultados.

A etapa de formulação do problema nos possibilitou definir a seguinte questão-problema: quais iniciativas da Informática estão sendo evidenciadas na literatura científica visando contribuir para reduzir interferências oriundas do emprego de reduções léxicas no contexto da saúde?

Para a segunda fase, que corresponde à coleta ou busca na literatura, foi executada pesquisa bibliográfica nas bases de dados *MEDLINE*, *Brapci*, *PubMed* e *Scopus*. Consultamos o tesouro *Medical Subjects Headings* (MeSH), que possui uma subdivisão específica para abreviações ("*Abbreviations as Topic*"). No caso da *Brapci*, que não utiliza esse tesouro, foi definida a seguinte palavra-chave: *acr?n?m**. Nas demais bases de dados, foi adotado o termo "*Abbreviations as Topic*", com a utilização de filtros para publicações feitas no período de janeiro de 1989 a julho de 2019, além de limitar a busca para *journal articles*.

Na terceira fase, avaliação dos dados, selecionamos os artigos científicos recuperados na fase anterior cuja leitura dos títulos, resumo ou texto completo, ratificaram

a contribuição da Informática no tratamento automático da língua, contemplando as reduções léxicas.

A quarta fase, análise e interpretação, possibilitou a definição de duas temáticas principais: mapeamento e desambiguação de reduções léxicas por meio do processamento automático da linguagem natural. E, na quinta fase, apresentação dos resultados, que estão espelhados nos quadros 1 e 2 referentes aos artigos que contemplam as temáticas de mapeamento e desambiguação, respectivamente.

4. Resultados e Discussões

Os resultados da pesquisa trouxeram 570 artigos recuperados na *MEDLINE*, 572 na *PubMed*, 217 na *Scopus* e 2 na *Brapci*. Ressaltamos que, no caso da *Brapci*, nenhum dos documentos recuperados se enquadrava no escopo do nosso trabalho.

Após a análise, foram selecionados 51 documentos concernentes ao objeto da pesquisa e que possibilitaram que os textos selecionados na *RBI* fossem estruturados pelos pesquisadores em duas categorias, quais sejam: mapeamento e desambiguação. Desse modo, ficou evidente que 28 artigos são referentes a *softwares* de mapeamento e 23 à desambiguação. Ressaltamos que dos 23 documentos referentes à desambiguação, 9 não deixaram claro nos resumos e nem nos textos completos quais recursos/*software* foram utilizados. Dos 28 alusivos ao mapeamento, 3 deles também não apresentaram os recursos/*softwares* empregados.

a) Categoria Mapeamento

Foram classificados nesta categoria os trabalhos que demonstraram esforços para desenvolver ou analisar ferramentas capazes de, por meio de recursos tecnológicos, localizar na literatura pesquisada as reduções léxicas, e identificar suas formas expandidas. Os achados estão enunciados no quadro 1 (nos anexos).

Os documentos categorizados aqui enunciam a necessidade de contornar as interferências oriundas do das reduções léxicas no processo de comunicação que, à primeira vista, parecem simples, contudo, podem trazer sérios problemas para todo o entorno relacionado à atenção de cuidados com o paciente. Uma das alternativas é a criação de algoritmos para identificar se determinado ponto final (sinal gráfico) sinaliza o fim de uma frase ou faz parte de uma abreviação. No entendimento de Kreuzthaler, Schulz⁸, “se as funções dos caracteres de ponto não forem devidamente esclarecidas, as frases serão divididas de maneira inadequada, o que tem um impacto severo na interpretação do texto”.

Os autores afirmam outrossim que “[...] se um sistema não consegue identificar abreviaturas, sua interpretação por mapeamento para formulários completos é prejudicada”.

Por sua vez, do ponto de vista da terminologia e da linguística, quando da criação de uma sigla, necessariamente o protocolo normatiza que seja adotado um ponto após cada letra maiúscula. Entretanto, conforme argumenta Kreuzthaler, Schulz⁸, no contexto da língua germânica, tal fato se concretiza de forma diferente.

Nossos textos são em língua alemã, os quais as abreviaturas são quase que obrigatoriamente seguidas de um ponto - ao contrário das siglas, que chamam a atenção pela colocação de várias letras maiúsculas e, ocasionalmente, dígitos. Abreviações não acronômicas sem ponto (como “Prof”) geralmente não são permitidas.

Outro aspecto observado na pragmática da pesquisa, em relação aos mapeamentos, é que boa parte dos documentos utiliza o *machine learning* como ferramenta para tal ação, haja vista que permite trabalhar com grande quantidade textual. Ademais, Wu *et al.*⁹ chama atenção para os desafios de se utilizar essa ferramenta, pois muitas vezes as abreviações e seus sentidos necessitam ser anotados manualmente (processo chamado de *machine learning* supervisionado). Como alternativa, pode-se utilizar bases de dados de abreviações com seus significados já existentes, mas nem sempre elas conseguem abranger a variedade dessas reduções léxicas. Contudo, o autor continua suas reflexões afirmando que bases de dados, como por exemplo, “*Unified Medical Language System (UMLS)*, *Metathesaurus* e *SPECIALIST Lexicon*, têm baixas taxas de cobertura para abreviações usadas em textos clínicos”⁹, pois a criação de novas unidades contraídas sem nenhum critério é uma atividade comum em alguns profissionais da área da Saúde. Seguindo a linha de pensamento de Wu *et al.*⁹, a cobertura mínima dessas unidades linguísticas nas grandes bases de dados é um desafio quando se busca lidar com abreviações clínicas, além da ambiguidade presente nelas:

O segundo desafio para lidar com abreviações clínicas é que elas podem ser ambíguas (por exemplo, “AA” pode significar afro-americano, aorta abdominal, Alcoólicos Anônimos etc.). Resolver a ambiguidade das palavras é uma tarefa fundamental da pesquisa da PLN, chamada de desambiguação do sentido das palavras (*Word Sense Disambiguation - WSD*).⁹

No cenário do mapeamento, a literatura apresenta o desenvolvimento do *Abbreviation Lifter using Corpus-based Extraction (ALICE)*. Esse *software* foi desenvolvido pelos pesquisadores Hiroko Ao e Toshihisa Takagi do *Department of Computational Biology*,

University of Tokyo. Eles argumentam que a motivação para o desenvolvimento dessa ferramenta de mapeamento é advinda da dificuldade encontrada para identificar e localizar abreviações da literatura biomédica, particularmente na base de dados *MEDLINE*. No entendimento desses pesquisadores, o crescente abuso de abreviações nas áreas da medicina dificulta a compreensão em alguns casos por serem ambíguas, principalmente aquelas que refletem genes e proteínas.¹⁰

O reconhecimento de reduções léxicas utilizando o *ALICE*, conforme Ao Takagi¹⁰, permite a identificação dos pares (forma reduzida e expandida). “Ele procura parênteses, identifica e extrai pares de abreviações e suas expansões usando regras heurísticas de correspondência de padrões.” Além do mais, “[...] o algoritmo usa padrões, regras e listas de palavras vazias (*stop words lists*) adicionais expandidas manualmente, que são baseadas em investigação e heurísticas completas”.

b) Categoria Desambiguação

Foram reunidos nesta categoria a literatura que envida esforços para solucionar os problemas de ambiguidade nas reduções léxicas, particularmente, no contexto da área da Saúde. No quadro 2 (nos anexos) são expostos os achados.

Ficou constatado nas fontes estudadas que existem vários problemas relacionados à adoção de reduções léxicas na área da Saúde. Um deles é que, segundo Wren, Garner¹¹, “[...] quando siglas são usadas para identificar termos relevantes dentro do texto, a definição usada no corpo do texto pode ser diferente daquela correspondente à entrada do banco de dados [...]”

No contexto da RBI, foram evidenciadas diversas alternativas de processamento automático da informação para contornar os problemas relativos à polissemia concernente à utilização das reduções léxicas, nas anotações de prontuários e nos textos referentes às pesquisas clínicas.

Destaca-se no âmbito da desambiguação o *SPECIALIST Lexicon and Lexical Tools*, desenvolvido pela Biblioteca Nacional de Medicina dos Estados Unidos (NLM), em 1994, pautado na UMLS. O *SPECIALIST Lexicon* fornece informações léxicas para sistemas de Processamento da Linguagem Natural (PLN), no vocabulário biomédico em inglês.¹²

Nessa continuidade da desambiguação, outro realce é para o *Naïve Bayes Classifier*, um algoritmo baseado no teorema de *Bayes*, e que “tem uma longa história de sucesso na desambiguação de sentido da palavra, pois assume a forma paramétrica de um modelo de

independência condicional (que descreve as relações entre os recursos) e por isso só precisa aprender os parâmetros desse modelo”.¹³

Citamos nesse íterim da desambiguação a base de dados *Allie*, criada por Yamamoto *et al.*¹⁴, que utiliza o dicionário de nomes de genes, denominado GENA, desenvolvido pelos próprios autores para normalizar termos cujos conceitos são idênticos, porém, com significados diferentes. Além do que, utiliza o *Metathesaurus* da *UMLS*.

Excetuando-se os resultados apresentados nos quadros 1 e 2 (anexos), inclusive merece destaque o *software Clinical Abbreviation Recognition and Disambiguation* (CARD), desenvolvido por Wu *et al.*⁹, pesquisadores da área de Informática Biomédica e de Medicina que atuam em instituições americanas. Utilizando-se o *CARD* é possível fazer o mapeamento e a desambiguação.

5. Conclusões

Encontramos muitos avanços na literatura, no sentido de solucionar problemas de identificação de reduções léxicas no processamento automático de texto e na desambiguação destes termos, particularmente analisando informações disponíveis na *MEDLINE* e *PubMed*.

O banco de dados *Allie* é notado como grande recurso para elucidar ambiguidades de reduções léxicas na literatura científica e deve servir de referência para a criação de outros bancos de dados voltados à mesma problemática de tratamento automático da linguagem natural, particularmente, de prontuário do paciente, e em outros idiomas como o português, que não foi muito destacado em nossa pesquisa.

Portanto, a RBI é uma metodologia que traz grandes contribuições para o conhecimento sobre o avanço de pesquisas relativas, entre outras coisas, às reduções léxicas no contexto da saúde salientando a necessidade de padronização e desambiguação de modo a reduzir as interferências no processo de comunicação nessa área. Nesse sentido, a literatura investigada evidencia a preocupação com o crescente uso de reduções léxicas nas Ciências da Saúde e incentiva o desenvolvimento de pesquisas com essa temática de caráter interdisciplinar com a Ciência da Informação e da Comunicação. Da mesma forma, destaca-se a grande contribuição da Informática, tanto para o desenvolvimento de ferramentas de mapeamento, quanto nas estratégias e ferramentas que realizam a desambiguação das reduções léxicas. Ressaltamos, nesse caráter da interdisciplinaridade, o papel da Biblioteca Nacional de Medicina dos Estados Unidos (NLM) conjuntamente com

profissionais da saúde e da Informática. Salientamos, mais uma vez, a necessidade de estudos da área da Ciência da Informação nessa temática, posto que é de interesse desta área pesquisas envolvendo linguagens documentárias e terminologias de especialidade, particularmente contemplando a área da Saúde.

Agradecimentos

Agradecemos ao CNPq o apoio e financiamento desta pesquisa.

Referências

1. Chang JT, Chutze H, Altman RB. Creating an Online Dictionary of Abbreviations from MEDLINE. JAMIA [Internet]. 2002 [acesso em 2020 ago 17]; 9 (6): 612-620. Disponível em <http://europepmc.org/article/PMC/2576267>.
2. Bloom DA. Acronyms, abbreviations and initialisms. BJU International [Internet]. 2000 [acesso em 2020 dez 22]; 86 (1): 1-6. Disponível em: <https://bjui-journals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1464-410x.2000.00717.x>.
3. Garner BA. The Oxford Dictionary of American Usage and Style. 1. Oxford: Oxford University Press; 2000.
4. Calvet LJ. Les Sigles. Paris: Presses Universitaires de France; 1980.
5. Velloso FC. Informática: conceitos básicos. 8. ed. Rio de Janeiro: Elsevier; 2011.
6. Shortliffe EH; Cimino JJ. Biomedical Informatics: computer Applications in Health Care and Biomedicine. 3. ed. [local desconhecido]: Springer; 2006.
7. Cooper HM. Synthesizing Research: a guide for literature. 3. ed. Thousand Oaks: Sage Publications; 1998.
8. Kreuzthaler M, Schulz S. Detection of sentence boundaries and abbreviations in clinical narratives. BMC Medical Informatics and Decision Making [Internet]. 2015 [acesso em 2020 nov 23]; 15 (S2). Disponível em: <https://doi.org/10.1186/1472-6947-15-s2-s4>
9. Wu Y, Denny JC, Trent Rosenbloom S, Miller RA, Giuse DA, Wang L, Blanquicett C, et al. A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). JAMIA [Internet]. 2016 [acesso em 2021 jan. 22]; 24(e1): e79-e86. Disponível em: <https://doi.org/10.1093/jamia/ocw109>
10. Ao H, Takagi T. ALICE: An Algorithm to Extract Abbreviations from MEDLINE. Journal of the American Medical Informatics Association [Internet]. 2005 [acesso em 2020 out. 12]; 12 (5): 576-86. Disponível em: <https://doi.org/10.1197/jamia.m1757>
11. Wren JD, Garner HR. Heuristics for identification of acronym-definition patterns within text: towards an automated construction of comprehensive acronym-definition dictionaries. Methods Inf Med [Internet]. 2002 [acesso em 2020 out. 23]; 41(05): 426-34. Disponível em: <https://doi.org/10.1055/s-0038-1634373>
12. NATIONAL LIBRARY OF MEDICINE. UMLS® Reference Manual [Internet]. Bethesda: National Library of Medicine (US); 2009. [acesso em 2021 mar 10] Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK9676/>. Acesso em: 10 mar. 2021.
13. Joshi M, Pakhomov S, Pedersen T, Chute CG. A comparative study of supervised learning as applied to acronym expansion in clinical reports. In: AMIA Annual Symposium, 2006, Washington. AMIA Annual Symposium Proceedings. 2006 [acesso em 2020 out 20]; 399-403. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839635/>.

14. Yamamoto Y, Yamaguchi A, Bono H, Takagi T. Allie: a database and a search service of abbreviations and long forms. Database [Internet]. 2011 [acesso em 2021 jan 19]; 2011; 2011 (0): bar013. Disponível em: <https://doi.org/10.1093/database/bar013>

Anexos

Quadro 1- Literatura que aborda o mapeamento automático de reduções lexicais.

Título da publicação	Autores	Área de atuação dos autores	Base de dados	Recursos/software
A comparative study of current Clinical Natural Language Processing systems on handling abbreviations in discharge summaries	Wu Y; Denny JC; Rosenbloom ST; Miller RA; Giuse DA; Xu H	Informática Biomédica, Medicina	Scopus e PubMed	MetaMap, MedLEE, cTAKES
A language independent acronym extraction from biomedical texts with hidden Markov models	Osiek BA; Xexeo G; Vidal de Carvalho LA	Engenharia da computação, Biomedicina	Scopus	Modelo oculto de Markov (Hidden Markov Model - HMM)
A new clustering method for detecting rare senses of abbreviations in clinical notes	Xu H; Wu Y; Elhadad N; Stetson PD; Friedman C	Informática biomédica	Scopus, PubMed e MEDLINE	Tight Clustering for Rare Senses (TCRS), Expectation Maximization (EM)
A sense inventory for clinical abbreviations and acronyms created using clinical notes and medical dictionary resources	Moon S; Pakhomov S; Liu N; Ryan JO; Melton GB	Informática médica	Scopus, PubMed e MEDLINE	Unified Medical Language System (UMLS), Another Database of Abbreviations in MEDLINE (ADAM), Stedman's (Dictionary, Medical Abbreviations, Acronyms & Symbols, 4th edition), Lexical Variant Generation (LVG).
A simple algorithm for identifying abbreviation definitions in biomedical text	Schwartz AS; Hearst MA	Ciência da computação	PubMed e MEDLINE	Não identificado
A study of abbreviations in clinical notes	Xu H; Stetson PD; Friedman C	Informática biomédica	Scopus, PubMed e MEDLINE	Unified Medical Language System (UMLS), Another Database of Abbreviations in MEDLINE (ADAM)
A study of abbreviations in the UMLS	Liu H; Lussier YA; Friedman C	Ciência da computação	PubMed e MEDLINE	Unified Medical Language System (UMLS)
ALICE: an algorithm to extract abbreviations from MEDLINE	Ao H; Takagi T	Biologia computacional	PubMed e MEDLINE	Abbreviation Lifter using Corpus-based Extraction (ALICE)
Automatic construction of gene relation networks using text mining and gene expression data	Karopka T; Scheel T; Bansemmer S; Glass A	Informática médica e biometria	MEDLINE	Gene Relation Finder (GRF), GATE environment (General Architecture for Text Engineering)
Automatic extraction of acronym-meaning pairs from MEDLINE databases	Pustejovsky J; Castañó J; Cochran B; Kotecki M; Morrell M	Informática médica e Linguística	PubMed e MEDLINE	AcroMed

Automatic resolution of ambiguous terms based on machine learning and conceptual relations in the UMLS	Liu H; Johnson SB; Friedman C	Sistemas de informação	PubMed e MEDLINE	MetaMap, UMLS abbreviation extraction program
BioABACUS: a database of abbreviations and acronyms in biotechnology and computer science	Rimer M; O'connell M	Biologia molecular	PubMed e MEDLINE	BioABACUS (Biotechnology Abbreviation and ACronym Uncovering Service). http://www.nmsu.edu/~molbio/bioABACUShome.htm
Biomedical term mapping databases	Wren JD; Chang JT; Pustejovsky J; Adar E; Garner HR; Altman RB	Biologia	PubMed e MEDLINE	Acronym Resolving General Heuristic (ARGH), Stanford Biomedical Abbreviation Database, AcroMed, Simple and Robust Abbreviation Dictionary (SaRAD)
Building an abbreviation dictionary using a term recognition approach	Okazaki N; Ananiadou S	Informática médica	PubMed e MEDLINE	Acromine
Creating an online dictionary of abbreviations from MEDLINE	Chang JT; Schütze H; Altman RB	Genética	PubMed e MEDLINE	Não identificado
Detecting abbreviations in discharge summaries using machine learning methods	Wu Y; Rosenbloom ST; Denny JC; Miller RA; Mani S; Giuse DA; Xu H	Informática biomédica	Scopus, PubMed e MEDLINE	Decision Trees (DT), Support Vector Machines (SVMs) Random Forests (RF).
Detection of sentence boundaries and abbreviations in clinical narratives	Kreuzthaler M; Schulz S			ETL (Extract Transform Load) workflow with Talend Open Studio
Developing and evaluating an auto-retrieval algorithm for abbreviations in academic articles	Wakoh M; Nishimoto N; Uesugi M; Terashita T; Ogasawara K	Ciências da Saúde	Scopus, PubMed e MEDLINE	Java software
Enhancing acronym/abbreviation knowledge bases with semantic information	Torii M; Liu H	Bioestatística, Bioinformática e Biomatemática	Scopus, PubMed e MEDLINE	ADAM (an existing collection of SF (acronyms/abbreviations), LF (definitions) pairs extracted from MEDLINE
Finding abbreviations in biomedical literature: three BioC-compatible modules and four BioC-formatted corpora	Islamaj Doğan R; Comeau DC; Yeganova L; Wilbur WJ	Informação Biotecnológica	PubMed	BioC (http://www.biocreative.org/)
Heuristics for identification of acronym-definition patterns within text: towards an automated construction of comprehensive acronym-definition dictionaries	Wren JD; Garner HR	Genética e Desenvolvimento, Ciência Biomédica	PubMed e MEDLINE	Acronym Resolving General Heuristics (ARGH)
Machine learning with naturally labeled data for identifying abbreviation definitions	Yeganova L; Comeau DC; Wilbur W J	Informação Biotecnológica	Scopus, PubMed e MEDLINE	Ab3P, BIOADI and Medstract

Mapping abbreviations to full forms in biomedical articles	Yu H; Hripcsak G; Friedman C	Informática Médica	PubMed e MEDLINE	Abbreviation recognition and extraction (Abbre)
Mapping of medical acronyms and initialisms to Medical Subject Headings (MeSH) across selected systems	Shultz M	Ciências da Saúde	PubMed e MEDLINE	PubMed MeSH database, PubMed Automatic Term Mapping, NLM Gateway Term Finder, Ovid MEDLINE
MBA: A literature mining system for extracting biomedical abbreviations	Xu Y; Wang Z; Lei Y; Zhao Y; Xue Y	Ciências da Computação e Tecnologia	Scopus, PubMed e MEDLINE	MBA
Pathology Abbreviated A Long Review of Short Terms	Berman JJ	Informática Patológica	MEDLINE	Perl, UMLS
Proposed three-letter abbreviations for homologues of the coded amino acids	Burton JA	Farmacêutica Psiquiátrica	PubMed e MEDLINE	Não identificado
Term identification in the biomedical literature	Krauthammer M; Nenadic G	Informática Biomédica	MEDLINE	O texto trata da questão de identificação de acrônimos e abreviações de maneira teórica, trazendo algumas ferramentas somente para discussão.

Fonte: Elaborado pelos autores

Quadro 2 - Literatura que aborda a desambiguação das reduções lexicais

Título da publicação	Autores	Área de atuação dos autores	Base de dados	Recursos/software
[A system for generating standardized abbreviations in ophthalmologic information systems]	Hanne W		PubMed e MEDLINE	Não identificado
A comparative study of supervised learning as applied to acronym expansion in clinical reports	Joshi M; Pakhomov S; Pedersen T; Chute CG	Ciência da computação, Informática biomédica	PubMed e MEDLINE	Weka Data Mining suite, the naïve Bayes classifier, the C4.5 decision tree learner, Support Vector Machine
A long journey to short abbreviations: Developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD)	Wu Y; Denny JC; Trent Rosenbloom S; Miller RA; Giuse DA; Wang L; Blanquicett C; Soysal E; Xu J; Xu H	Informática biomédica	Scopus, PubMed, MEDLINE	clinical abbreviation recognition and disambiguation (CARD)
A multi-aspect comparison study of supervised word sense disambiguation	Liu H; Teller V; Friedman C	Sistemas de informação	PubMed e MEDLINE	Naïve Bayes learning (NBL), Decision List Method (DLL), Instance-based Learning (IBL)

A preliminary study of clinical abbreviation disambiguation in real time	Wu Y; Denny JC; Rosenbloom ST; Miller RA; Giuse DA; Song M; Xu H	Informática biomédica	Scopus, PubMed, MEDLINE	real-time Clinical Abbreviation Recognition and Disambiguation (rCARD)
A randomized-controlled trial of computerized alerts to reduce unapproved medication abbreviation use	Myers JS; Gojraty S; Yang W; Linsky A; Airan-Javia S; Polomano RC	Medicina	Scopus, PubMed, MEDLINE	Alteração no software de notas de progresso
A study of abbreviations in MEDLINE abstracts	Liu H; Aronson AR; Friedman C	Ciência da Computação	PubMed e MEDLINE	UMLS
AAC menu interface: effectiveness of active versus passive learning to master abbreviation-expansion codes	Gregory E; Soderman M; Ward C; Beukelman DR; Hux K	Comunicação	PubMed e MEDLINE	Augmentative and alternative communication (AAC)
Abbreviation and acronym disambiguation in clinical discourse	Pakhomov S; Pedersen T; Chute CG	Informática biomédica	PubMed e MEDLINE	Não identificado
Abbreviations for device names: A proposed methodology with specific examples	Alam M; Dover JS; Alam M; Goldman MP; Kaminer MS; Orringer J; Waldorf H; Alam, M; Avram M; Cohen JL; Draelos ZD; Dover JS; Hruza G; Kilmer S; Lawrence N; Lupo M; Metelitsa A; Nestor M; Ross EV	Cirurgia dermatológica	Scopus, PubMed e MEDLINE	Não identificado
Allie: A database and a search service of abbreviations and long forms	Yamamoto Y; Yamaguchi A; Bono H; Takagi T	Ciências da vida	Scopus, PubMed, MEDLINE	Allie, ALICE, SPECIALIST Lexicon (UMLS)
An easily implemented method for abbreviation expansion for the medical domain in Japanese text: A preliminary study	Shinohara EY; Aramaki E; Imai T; Miura Y; Tonoike M; Ohkuma T; Masuichi H; Ohe K	Planejamento, informação e gestão	Scopus e MEDLINE	Não identificado
Automated disambiguation of acronyms and abbreviations in clinical texts: window and training size considerations	Moon S; Pakhomov S; Melton GB	Informática médica	Scopus, PubMed, MEDLINE	Não identificado
Challenges and practical approaches with word sense disambiguation of acronyms and abbreviations in the clinical domain	Moon S	Informática biomédica e Ciências da computação	Scopus, MEDLINE	Não identificado
Combining corpus-derived sense profiles with estimated frequency information to	Xu H; Stetson PD; Friedman C	Informática biomédica	Scopus, PubMed	The profile-based disambiguation method vector space model

disambiguate clinical abbreviations				
Distinction between medical and non-medical usages of short forms in clinical narratives	Moon S; Ihrke D; Zeng Y; Liu H	Ciências da Saúde	PubMED e MEDLINE	Unified Medical Language System (UMLS)
Methods for Building Sense Inventories of Abbreviations in Clinical Notes	Xu H; Stetson PD; Friedman C	Informática Biomédica	Scopus e PubMed	Expectation Maximization(EM), Farthest First(FF)
Resolving abbreviations to their senses in MEDLINE	Gaudan S; Kirsch H; Rebholz-schuhmann D	Bioinformática	PubMed e MEDLINE	Não identificado
SaRAD: a Simple and Robust Abbreviation Dictionary ?	Adar E		PubMed e MEDLINE	Simple and Robust Abbreviation Dictionary (SaRAD)
Statistical semantic and clinician confidence analysis for correcting abbreviations and spelling errors in clinical progress notes	Wong W; Glance D	Prática de Software	Scopus e PubMed	Não identificado
Using UMLS lexical resources to disambiguate abbreviations in clinical text	Kim Y; Hurdle J; Meystre SM	Computação	Scopus, PubMed e MEDLINE	SPECIALIST Lexicon LRABR (UMLS)
Word Sense Disambiguation of clinical abbreviations with hyperdimensional computing	Moon S; Berster BT; Xu H; Cohen T	Informática Biomédica	Scopus e PubMed	Binary Spatter Code Word Sense Disambiguation(BSC-WSD)
Towards Comprehensive Clinical Abbreviation Disambiguation Using Machine-Labeled Training Data	Finley GP; Pakhomov SV; McEwan R; Melton GB	Informática para Saúde	Scopus, PubMed e MEDLINE	Não identificado

Fonte: Elaborado pelos autores