

O bibliotecário e a pesquisa terminológica em prontuários na área de Saúde

The Librarian and terminology research on medical records data in the Healthcare field

Amanda Damasceno de Souza, Universidade FUMEC- amandasd81@gmail.com

Eduardo Ribeiro Felipe, Universidade Federal de Itajubá- eduardo.felipe@unifei.edu.br

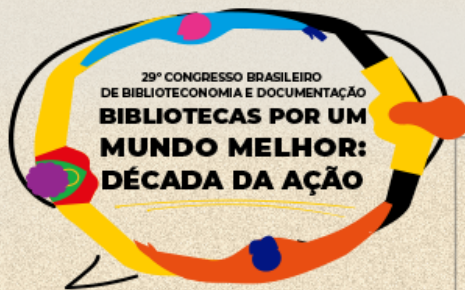
Fernanda Farinelli, Universidade de Brasília- fernanda.farinelli@unb.br

Maurício Barcellos Almeida, Universidade Federal de Minas Gerais- mba@fastmail.org

Eixo 4 - Ciência da Informação: diálogos e conexões

1 INTRODUÇÃO

No âmbito hospitalar há necessidade de padronização de terminologias clínicas das *anamneses* e *evoluções*, as quais contêm dados clínicos dos pacientes para fins de pesquisa e de comunicação efetiva entre os profissionais de saúde. Considera-se aqui “anamnese” como o registro no prontuário da história clínica completa de um paciente e “evolução” como o registro da resposta ao tratamento do paciente com informações — sinais vitais, sintomas, alteração físico-patológica, diagnóstico, medicamentos, etc. — registradas em ordem cronológica. Sem a construção de recursos que permitam a padronização para intercâmbio de terminologias, pode haver falhas na comunicação de informação sobre o paciente (SHORTLIFFE; BARNETT, 2014). Para melhor compreensão do artigo, vale diferenciar sinais e sintomas. Entende-se como sintomas, qualquer alteração que a própria pessoa percebe em seu organismo, já os sinais são as manifestações identificadas por um terceiro. Os sinais vitais incluem a frequência respiratória, frequência cardíaca, pressão arterial, e temperatura corporal. Neste contexto, destaca-se a importância a atuação do Bibliotecário em pesquisas no âmbito das terminologias clínicas devido à sua experiência com os vocabulários controlados da área médica. Estes vocabulários são essenciais para fins de recuperação de informação em saúde.



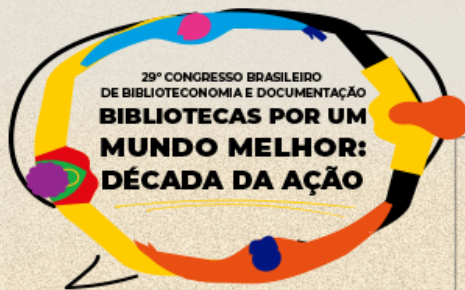
A informação em saúde é um tema fundamental para auxiliar no tratamento dos pacientes, devido à necessidade constante de pesquisa por informação satisfatória, em momento de necessidade. A Ciência da informação (CI), por sua contribuição nas áreas de recuperação, representação e organização da informação, possui um papel importante na saúde. Assim, o Bibliotecário precisa desenvolver habilidades de atuar com diferentes suportes de informação, saber manusear as Tecnologias de Comunicação e Informação (TIC), recuperar dados a partir de Processamento de Linguagem Natural (PLN) e técnicas de inteligência artificial e, com isso, contribuir para a melhoria do tratamento da informação, no cuidado em saúde. (SOUZA, 2021).

Esta pesquisa tem por objetivo descrever a atuação do Bibliotecário na pesquisa terminológica em prontuários na área de saúde. Justifica-se pela importância em buscar formas de interoperabilidade entre terminologias clínicas e, com isso, possibilitar a geração de produtos de inovação para auxiliar o médico a notificar melhor no Prontuário Eletrônico do Paciente (PEP). Com tais melhorias na notificação do PEP, contribui-se para a qualidade da assistência continuada e bem-estar do paciente. A pesquisa é oriunda de Doutorado no âmbito da Pós-graduação em Ciência da Informação (CI), aplicada em PEP no domínio da Ginecologia e Obstetrícia em Hospital Privado de Belo Horizonte, MG, Brasil. (SOUZA, 2021).

2 FUNDAMENTAÇÃO TEÓRICA

Com os avanços das TICs na área de saúde, multiplicaram-se os tipos de recursos sejam lexicais ou terminológicos, para processamentos de dados estruturados, não-estruturados, incluindo ontologias (SCHULZ *et al.*, 2017). Nas últimas décadas, o problema de padronização de linguagem e terminologias clínicas têm sido uma grande preocupação da Informática Médica. O desafio é criar "vocabulários controlados" realmente úteis para a área e os profissionais de saúde (RECTOR, 1999).

A Recuperação da Informação (RI) tem por princípio possibilitar ao usuário o acesso à informação (CAMPOS, 2001). Tem a função de recuperar todo tipo de informação em computadores, especificamente a informação textual em base de



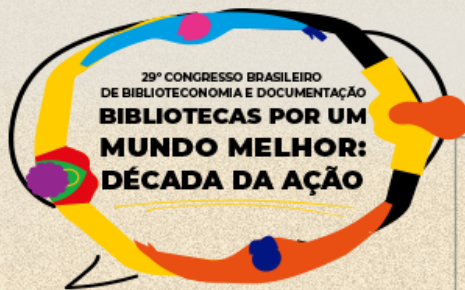
dados (HERSH, 2003), se configurando como uma interseção entre a CI e a Ciência da Computação (CC). Os sistemas de recuperação da informação consistem em mecanismos para armazenar e acessar conteúdo, mais o *software* para processar as solicitações de entrada e saída de informação com objetivo de recuperar (HERSH, 2003). O campo, de maneira geral, envolve a aquisição, a organização e a busca de informação, sendo que na biomedicina, é tradicionalmente concentrada na recuperação de textos daquela área.

Na saúde, a RI é utilizada tanto pela equipe de assistência quanto pelos próprios pacientes e usuários da Internet, em geral. Estima-se que entre os indivíduos que usaram a Internet, nos Estados Unidos no ano de 2012, que 81% são americanos adultos e destes cerca de 80% usaram-na para pesquisar informação em saúde (FOX; DUGGAN, 2003).

Dentre as possibilidades de atuação do profissional da informação no cenário de análise de dados e informação destacam-se algumas capacidades: compreender a demanda pela informação; identificar as melhores fontes de informação; descrever, organizar e representar os dados necessários para a análise (CONEGLIAN *et al.*, 2017). Nesse contexto, o Bibliotecário desempenha um papel importante na organização e recuperação de informação em saúde, ao prover suporte à decisão clínica. Em um contexto no qual a informação em medicina aumenta exponencialmente, dificultando a busca por informações de qualidade pelos médicos que não conseguem acompanhar o crescimento da literatura em saúde (HERSH, 2015).

3 TERMINOLOGIAS CLÍNICAS

Schulz *et al.* (2017) citam três tipos de terminologias em saúde e propõem uma metodologia para realizar conexão entre elas: Terminologias de Interface, que corresponde aos dados e texto clínico do prontuário, normalmente no jargão médico; Terminologias de Referência, que consiste em vocabulários controlados e ontologias; e Terminologias de Agregação que se constituem em sistemas baseados em classificação para fins diversos, por exemplo, Classificação Estatística Internacional



de Doenças e Problemas Relacionados à Saúde (CID), *Systematized Nomenclature of Medicine - Clinical Terms* (SNOMED-CT), dentre outras. Nesta pesquisa, adota-se a denominação “Terminologia de Interface” utilizada por Schulz *et al.* (2017), para representar as anamneses e evoluções clínicas do PEP.

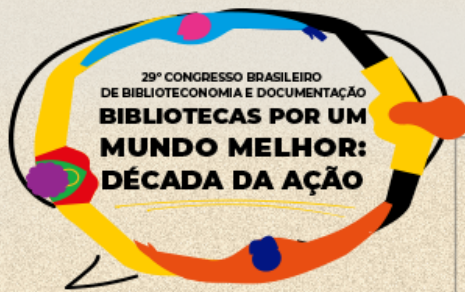
As terminologias clínicas padronizadas podem ser usadas para identificar o significado e a semântica dos conceitos lexicais no texto clínico que compõe os prontuários. Além disso, terminologias são importantes ao expandir o conceito na direção de sinônimos, na identificação de abreviações e acrônimos e também no mapeamento de conceitos para terminologias de outros sistemas (DALIANIS, 2018, p.35). Tal expansão é de grande auxílio para RI, qual, quando executada em textos em linguagem natural, revela-se uma tarefa árdua envolvendo técnicas computacionais como o Processamento de linguagem natural (PLN). O PLN diz respeito ao processamento inteligente de texto em linguagem natural, com utilização de métodos computacionais e linguísticos (MANNING; SCHÜTZE, 1999).

4 MÉTODO DA PESQUISA

A pesquisa foi classificada como sendo de natureza *aplicada* e quanto a de abordagem ao problema como: *quantitativa* e do ponto de vista dos objetivos: *descritiva* (GIL, 1994).

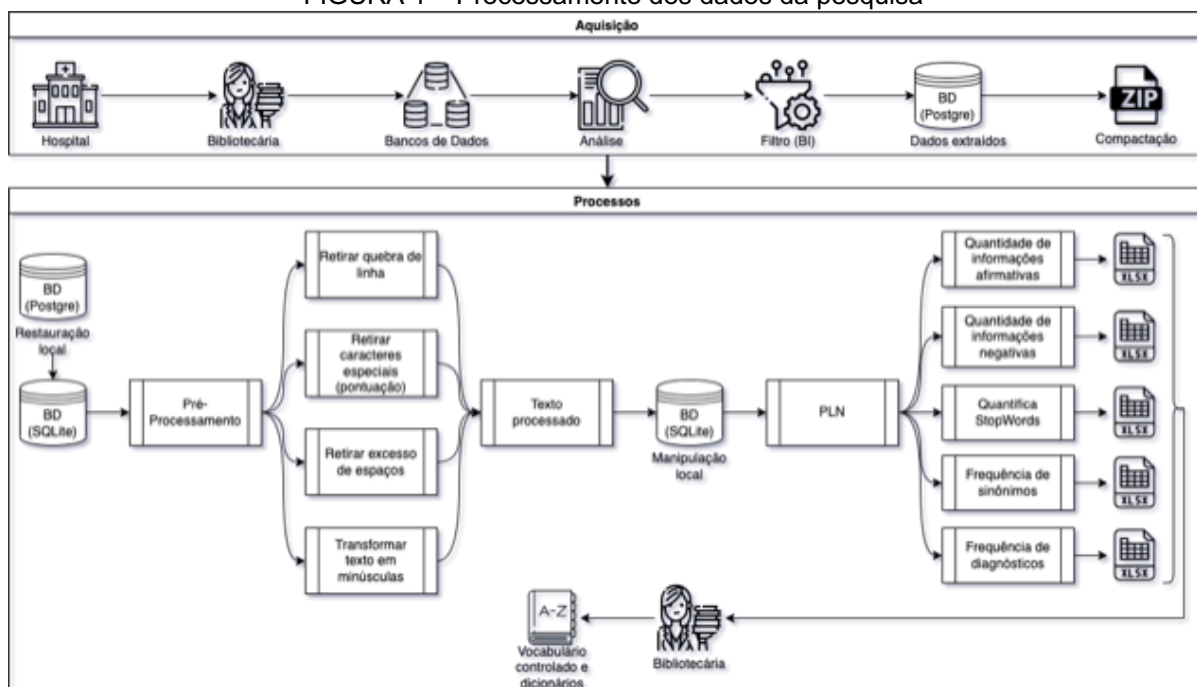
Para esta pesquisa terminológica em prontuários da área de saúde, analisou-se a Terminologia de Interface — no jargão médico ou texto e dado clínico — disponível nas anamneses e evoluções clínicas, especificamente do domínio da Ginecologia e Obstetrícia. A amostra da pesquisa foi composta de 18.256 anamneses e 14.035 evoluções de PEPs do ano de 2018. A pesquisa contou com aprovação do Comitê de Ética em Pesquisa (CEP) pelo número do CAAE:03384418.0.0000.51259.

Foram desenvolvidos algoritmos com funções independentes, escritos em linguagem de programação Python, os quais objetivavam processar os dados advindos do hospital e gravar os resultados em planilhas Excel. Durante o processamento dos dados, estes foram armazenados no sistema gerenciador de banco de dados SQL Lite. A figura 1 representa a metodologia de aquisição e



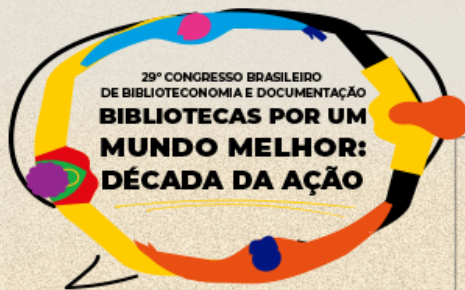
processamento dos dados clínicos pelo Bibliotecário. No processamento de dados, destacam-se de forma sintetizada: a) a extração da informação do banco de dados do Hospital, b) a restauração dos dados em ambiente local, c) a adequação do formato do banco de dados para manipulação. Por fim, a partir dos resultados, o Bibliotecário poderá auxiliar na criação de vocabulários controlados, terminologias, ontologias, dicionários para compreensão dos dados e nas linguagens documentárias em geral.

FIGURA 1 – Processamento dos dados da pesquisa



Fonte: Dados da pesquisa (2022).

A função dos algoritmos foi identificar termos relacionados que envolviam: a) presença de sinais, b) sintomas, c) expressões negativas, d) expressões afirmativas, e) termos únicos, f) termos mais frequentes, g) siglas, h) abreviaturas, i) erros ortográficos, j) diagnósticos e k) procedimentos. (Figura 1). Posteriormente ao PLN, foi realizado o tratamento estatístico dos termos extraídos pelos algoritmos, para fins de análise de frequência absoluta (f_a que reflete a quantidade de vezes que cada termo aparece) e frequência relativa (f_r que consiste no percentual de vezes que cada f_a aparece em relação ao total da amostra n). Os resultados da extração foram apresentados em tabelas e na forma de nuvem de palavras.



No processo para ajustar os algoritmos, na busca por dados e termos para a análise terminológica, foram criadas listas preliminares de termos da Ginecologia e Obstetrícia em relação a: *stop words*¹, siglas e abreviaturas, expressões negativas e afirmativas, sinais e sintomas, diagnóstico e procedimentos (Figura 1). As listas de termos preliminares foram criadas junto à equipe do grupo de pesquisa do *Núcleo Integrado de Pesquisa e Tratamento da Endometriose (NIPTE)*², médicos da Clínica de Ginecologia e Obstetrícia do Hospital do Estudo e com informações de termos de protocolos clínicos e manuais: SES/MG (2013), Peixoto (2014), Brasil (2004), Brasil (2016), CONITEC (2016), Matos *et al.* (2017). Especificamente, em relação às expressões de negação, procurou-se identificar termos negativos em anamneses e evolução médicas, o que envolve identificar uma proposição atribuindo uma condição clínica a uma pessoa, além de determinar se a proposição é de fato negada no texto (CHAPMAN *et al.*,2001). Em relação as terminologias de doenças e diagnóstico, utilizou-se a CID-10.

5 RESULTADOS

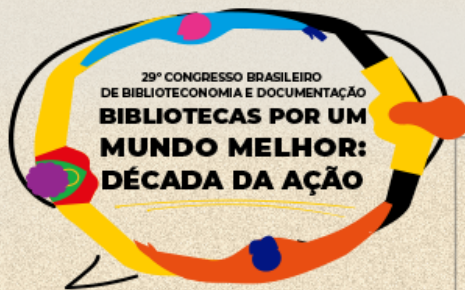
Os termos mais frequentes na Terminologia de Interface foram as siglas e as abreviaturas e os termos menos frequentes foram as expressões afirmativas (Tabela 1). A Tabela 1 apresenta os resultados quantitativos e extrações realizadas automaticamente pelos algoritmos.

TABELA 1 – Frequência absoluta (n) e frequência relativa (%) da Terminologia de Interface

Variáveis	Terminologia de Interface	
	n	%
Diagnóstico	14.652	1,07
Sinais e sintomas	15.862	1,16
Expressões afirmativas	11.163	0,82
Expressões negativas	69.315	5,08
<i>Stop words</i>	207.414	15,20

¹ Palavras comuns no texto, mas não importantes na RI pois não agregam semântica, por exemplo, artigos, preposições e advérbios. A expressão tem uso comum na área de TIC e usualmente não é traduzida, mesmo em textos acadêmicos.

² Disponível na Internet em: <<https://www.felicio-rocho.org.br/endometriose>>. Acesso em 01/09/2022.



Cabe ressaltar, que a Figura 2 é importante para demonstrar a quantidade e variedade de siglas e abreviaturas utilizadas pelos médicos na notificação em anamneses e evoluções da Ginecologia.

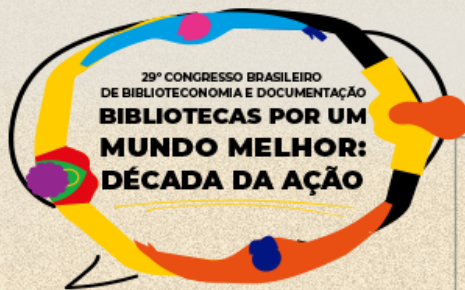
A Figura 3, também no formato de nuvem de palavras, representa quais são os termos afirmativos presentes na Terminologia de Interface. Os termos de maior destaque são aqueles de expressões que demonstram o estado do paciente, ou seja, se o paciente está bem clinicamente, sem sinais de doenças ou com presença de sintomas. Destaque também para os termos que indicam diagnósticos dos pacientes, por exemplo: “normal”, “sinais de...”, “presença”, etc. Note que na Figura 3 fica evidente um dos problemas encontrados no processamento, por exemplo, erros ortográficos. Na figura 3 encontramos com frequência a expressão “presença de”, mas também encontramos “presença de” sem a escrita do cedilha.

FIGURA 3– Nuvem de palavras dos termos afirmativos da Terminologia de Interface



Fonte: Dados da pesquisa (2021).

Os principais problemas encontrados na extração de dados e termos da Terminologia de Interface foram: ambiguidade gramatical, sinonímias, abreviaturas excessivas e incompreensíveis e erros ortográficos. Adicionalmente, existem siglas formada por letras e números, como por exemplo “Asa 1”, “G1P1A0”. A sigla



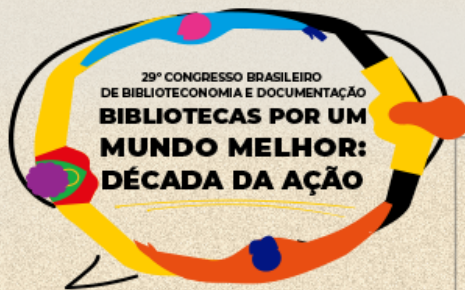
“G1P1A0”, por exemplo, ela indica que a paciente teve uma gestação (G1), um parto (P1), e nenhum aborto (A0). Estas siglas possuem seu significado relacionado diretamente aos valores numéricos. Durante o pré-processamento ou limpeza de dados os valores numéricos foram removidos o que ocasionou a não identificação de alguns termos que podem ser relevantes, sendo esta uma limitação a reportar.

6 DISCUSSÃO

A pesquisa por meio da aplicação de técnicas de PLN possibilitou delimitar e testar algoritmos de tarefas em anotações de um grande volume de anamneses e evoluções do PEP, aqui denominados “Terminologia de Interface” (Tabela 1). Com isso, foi possível apresentar a equipe de Ginecologia os problemas causados pela natureza não estruturada e “ruidosa” do texto narrativo, por exemplo: ambiguidade gramatical, sinônimos, abreviatura de termos e erros ortográficos ou expressões de negação, para citar alguns. A pesquisa tem potencial para gerar um corpus terminológico no domínio da Ginecologia e Obstetrícia além de desenvolver e aprimorar algoritmos para extração de dados nesse domínio.

Pelos resultados da análise terminológica foi possível também apresentar a equipe médica a frequência dos tipos de informações presentes no PEP. A alta frequência de siglas e abreviaturas (71,23%) de anamneses e evoluções demonstrou que a equipe médica faz uso excessivo de siglas e abreviaturas no texto clínico, o que o torna ambíguo e de difícil compreensão (Figura 2). Esta forma de escrita pode impactar na RI automática levando a uma baixa taxa de revocação — quando poucos documentos relevantes são recuperados — além de impactar diretamente na taxa de precisão.

Uma possibilidade para a atuação do Bibliotecário seria o trabalho de identificar possíveis fontes de informação em saúde que sirvam de guias ou tradutores das siglas para os termos significativos, para fins de substituição automática destas siglas durante a recuperação. As siglas podem ser usadas como termos alternativos para um determinado termo. O Bibliotecário pode ainda contribuir ao criar um dicionário ou



vocabulário controlado para a instituição para representar os termos e siglas comuns usadas nas Terminologia de Interface dos profissionais de saúde.

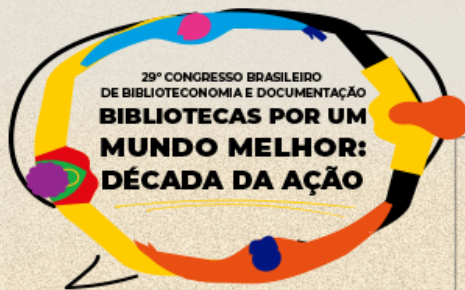
7 CONCLUSÕES

A pesquisa ressalta a importância do trabalho do Bibliotecário junto à equipe médica, pois além de buscar a melhor evidência científica para o tratamento do paciente, atua nas pesquisas sobre as terminologias clínicas.

A principal dificuldade em analisar a Terminologia de Interface, referiu-se aos seus aspectos epistemológicos que dependem fortemente do contexto médico, ou seja, muitos termos utilizados tem seus significados conhecidos somente pelos médicos contemporâneos. A intenção desta análise de termos clínicos foi entender suas características, possibilidades de extração, além de obter informação para tomada de decisão em saúde, estudos clínicos e epidemiológicos.

Pelas dificuldades de construção sintática do PEP, sugere-se que os sistemas computacionais possam auxiliar o médico, com mecanismos de sugestão de digitação inteligente (*intelligence writing, autocomplete*) que ajudam a evitar alta frequência de erros ortográficos. A escrita inteligente e o recurso de autocompletar as palavras tem potencial de proporcionar uma RI mais assertiva no futuro, e conseqüentemente, a melhoria na qualidade informação em saúde e na tomada de decisão.

Como perspectivas futuras, especificamente no domínio da Ginecologia e Obstetrícia, sugere-se a criação de terminologias e diretrizes médicas para preenchimento de anamnese e evoluções, e mesmo dicionários ou vocabulários controlados de siglas e abreviaturas. O PEP é uma fonte rica de informação no contexto hospitalar, mas a qualidade dos registros clínicos carece de aperfeiçoamento para evitar erros com os apontados nesta pesquisa e uso excessivos de termos incompreensíveis. Neste contexto, o Bibliotecário se configura como um profissional importante na organização e recuperação do conhecimento em saúde, devido a seu papel em atividades diversas, como: identificação e seleção de fontes de dados, indexação de artigos em base de dados da saúde, organização de bibliotecas hospitalares, classificação da informação em saúde para tomada de decisão em



saúde, suporte a condução de revisão sistemática, nas atividades de representação da informação como elaboração de tesouros, terminologias, vocabulários e ontologias, dentre várias outras atividades, reafirmando a importância da interlocução entre a CI e a área de Saúde.

REFERÊNCIAS

BRASIL. Ministério da Saúde. **Protocolos da Atenção Básica: Saúde das Mulheres**. Brasília: Ministério da Saúde, 2016. 230 p.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Ações Programáticas Estratégicas. **Política nacional de atenção integral à saúde da mulher: princípios e diretrizes**. Brasília: Ministério da Saúde, 2004. 82 p. (C. Projetos, Programas e Relatórios). Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/politica_nac_atencao_mulher.pdf. Acesso em 08 jan. 2020.

CAMPOS, M.L.A. **Linguagem documentárias: teorias que fundamentam sua elaboração**. Niteroi (RJ): EdUFF, 2001. 133p.

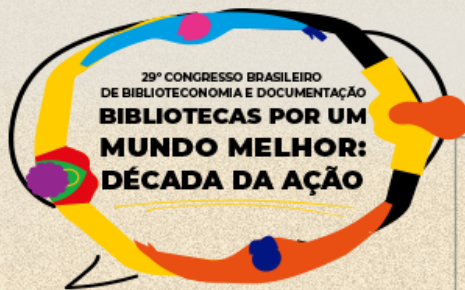
CHAPMAN, W.W.; *et al.* A simple algorithm for identifying negated findings and diseases in discharge summaries. **Journal of Biomedical Semantics**, London, v.34, n.5, p.301-10, 2001.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O Profissional da Informação na Era do Big Data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v.22, n. 50, p. 128, 2017.

COMISSÃO NACIONAL DE INCORPORAÇÃO DE TECNOLOGIAS (CONITEC) **Diretrizes Nacionais de Assistência ao Parto Normal**. Brasília: Ministério da Saúde, maio 2016. 399p. (Relatório de recomendação, nº 211).

DALIANIS, H. Medical Classifications and Terminologies. *In*: DALIANIS, H. **Clinical Text Mining: Secondary Use of Electronic Patient Records**. [s.n.], 2018a. Cap. 5 Disponível em: <<http://link.springer.com/10.1007/978-3-319-78503-5>>. Acesso em: 2 jan. 2019.

FOX, S.; DUGGAN, M. **Health Online 2013**. Pew Research Center's Internet & American Life Project. January 15, 2013. Disponível em: <https://www.pewinternet.org/2013/01/15/health-online-2013/>. Acesso em: 20 out. 2019.



GIL, A.C. **Métodos e técnicas de pesquisa social**. 4 ed. São Paulo: Atlas, 1994. 207 p.

HERSH, W.R. Information Retrieval for Healthcare. *In*: REDDY, Chandan K.; AGGARWAL, Charu C (Eds.). **Healthcare Data Analytics**. Boca Raton, Florida (USA): Chapman and Hall/CRC, 2015. 760p.cap. 14.

HERSH, W.R. **Information Retrieval: a Health and biomedical Perspective**. 2nd ed. New York: Springer, 2003. (Health Informatics, formerly computers in Health Care).

MANNING, C. D.; SCHÜTZE, H. **Foundations of statistical natural language processing**. Cambridge Massachusetts: MIT press, 1999.620p.

MATOS, M. S.; *et al.* **Manual de Ginecologia**. Salvador: EBMS, 2017.

PEIXOTO, S. **Manual de assistência pré-natal**. 2ed. São Paulo: Federação Brasileira das Associações de Ginecologia e Obstetrícia (FEBRASGO), 2014.

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

SCHULZ, S.; *et al.* Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in Health Technology and Informatics**, Amsterdam, v.245, p.940-944.2017.

SECRETARIA DE ESTADO DE SAÚDE DE MINAS GERAIS – SES/MG. ASSOCIAÇÃO DE GINECOLOGISTAS E OBSTETRIZAS DE MINAS GERAIS – SOGIMIG. ATENÇÃO À SAÚDE DA GESTANTE. **Novos Critérios para Estratificação de Risco e Acompanhamento da Gestante**: PROGRAMA VIVA VIDA Projeto Mães de Minas, maio 2013.

SHORTLIFFE, E.H.; BARNETT, G.O. Biomedical Data: Their Acquisition, Storage, and Use. *In*: SHORTLIFFE, E.H.; CIMINO, J.J. (Editors). **Biomedical Informatics: Computer Applications in Health Care and Biomedicine**. 4th Ed. London: Springer-Verlag, 2014. cap.2, p.46-79.

SOUZA, A.D. **O discurso na prática clínica e as terminologias de padronização: investigando a conexão**. 2021. Tese (Doutorado em Gestão e Organização do Conhecimento). Pós-Graduação em Gestão e Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2021. Disponível em: <http://hdl.handle.net/1843/38044>. Acesso em: 13 jul. 2022.