

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE**

HUGO BULEGON

**IDENTIFICAÇÃO DE DIAGNÓSTICOS CONTIDOS EM NARRATIVAS CLÍNICAS E
MAPEAMENTO PARA A CLASSIFICAÇÃO INTERNACIONAL DE DOENÇAS**

**CURITIBA
2011**

HUGO BULEGON

**IDENTIFICAÇÃO DE DIAGNÓSTICOS CONTIDOS EM NARRATIVAS CLÍNICAS E
MAPEAMENTO PARA A CLASSIFICAÇÃO INTERNACIONAL DE DOENÇAS**

Dissertação de mestrado do Programa de Pós-Graduação em Tecnologia em Saúde, Centro de Ciências Biológicas e da Saúde da Pontifícia Universidade Católica do Paraná, apresentada como requisito parcial à obtenção do título de Mestre em Tecnologia em Saúde.

Orientador: Prof^a. Dr^a. Claudia Maria Cabral
Moro Barra

**CURITIBA
2011**



Pontifícia Universidade Católica do Paraná
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Tecnologia em Saúde

ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
DO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE
DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

DEFESA DE DISSERTAÇÃO Nº-139

Aos 28 dias de fevereiro de 2011 realizou-se a sessão pública de defesa da dissertação: **“Identificação de Diagnósticos Contidos em Narrativas Clínicas e Mapeamento para a Classificação Internacional de Doenças”**, apresentada por **Hugo Bulegon**, como requisito parcial para a obtenção do título de Mestre em Tecnologia em Saúde – Área de Concentração – **Informática em Saúde** perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Drª Claudia Maria Cabral Moro,
PUCPR (Orientadora e presidente)



APROVADO
parecer (aprov/ reprov.)

Prof. Dr. Edson Emilio Scalabrin,
PUCPR (Examinador)



APROVADO

Prof. Dr. Heitor Silvério Lopes,
UTFPR (Examinador)



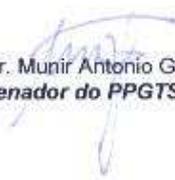
APROVADO

Prof. Dr. Angelmar Constantino Roman,
UP (Examinador)



APROVADO

Conforme as normas regimentais do PPGTS e da PUCPR, o trabalho apresentado foi considerado APROVADO (aprovada/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do Programa.


Prof. Dr. Munir Antonio Gariba,
Coordenador do PPGTS PUCPR



AGRADECIMENTOS

Agradeço primeiramente a Deus e a Santo Expedito, que me concederam entre outras graças, principalmente a saúde necessária para a realização de todas as atividades ligadas ao PPGTS.

Aos meus pais Aracelis e Derli Pedro Bulegon, que sempre incentivaram a continuidade do estudo de seus filhos, das atividades extras, da busca pelo conhecimento. Também agradeço a minha irmã – Joana Bulegon e minha tia – Vera Lúcia Alves dos Santos, família que sempre esteve me apoiando nas diversas situações e etapas do mestrado.

Agradeço à professora orientadora - Claudia Maria Cabral Moro Barra, que me recebeu muito bem desde a entrevista de seleção, atendeu as minhas expectativas sobre o mestrado e me apoiou diversas vezes. Obrigado por todo suporte dentro e fora do mestrado, inclusive pelas contribuições para a vida.

Ao PPGTS e a CAPES, por confiarem em meu trabalho, pelo suporte financeiro, pelo modelo de programa de pós-graduação multidisciplinar que me proporcionou atividades com profissionais das mais diversas áreas.

Agradeço à Prof^ª. Dra. Mariza Machado Klück e ao Hospital de Clínicas de Porto Alegre (HCPA), que forneceram parte do material de trabalho para o desenvolvimento desta pesquisa.

À Izelde Angheben do Nascimento, que sempre atendeu os mestrandos com excelência, qualidade e paciência.

Ao professor Stefan Schulz que discutiu a metodologia, entre outras contribuições importantes para o desenvolvimento deste trabalho. Aos amigos que fiz durante o mestrado e contribuíram com a dissertação, Michel Oleynik, Dennys Robson Girardi e Dandara Novakowski Spigolon.

Ao CBIS – Congresso Brasileiro de Informática na Saúde e à Sociedade Brasileira em Informática e Saúde, que despertaram a vontade de continuar os estudos relacionados em tecnologia em saúde.

Obrigado a todos!

RESUMO

Durante o atendimento de pacientes são registradas algumas informações no prontuário na forma de textos, que são denominadas narrativas clínicas. A utilização das informações contidas nestas narrativas é restrita, uma vez que é muito difícil identificá-las no texto de forma automática. Diagnósticos são algumas das informações contidas nestas narrativas e geralmente não se relacionam somente ao atendimento, que também são armazenados em outros campos não texto, estruturados da base de dados do sistema de prontuário eletrônico do paciente (PEP). São diagnósticos secundários, anteriores ao atendimento, em investigação ou até mesmo os excluídos. Além de identificar os diagnósticos, é importante facilitar a sua codificação na Classificação Internacional de Diagnósticos (CID-10), que é um dos principais padrões utilizados nos Sistemas de Informação em Saúde. Um método para extração de diagnósticos contidos nas narrativas clínicas pode auxiliar na identificação destes, aprimorando a utilização dos dados do PEP, bem como facilitando o preenchimento das informações. Desta forma, o objetivo desta pesquisa foi elaborar um método para identificar diagnósticos em narrativas clínicas, sugerindo um mapeamento para grupos e categorias da CID-10. Foi utilizada uma base de dados contendo sumários de alta referentes a atendimentos realizados no serviço de cardiologia do Hospital de Clínicas de Porto Alegre (HCPA), contendo aproximadamente 6.000 registros. O desenvolvimento da pesquisa foi realizado em seis fases, iniciando-se pela avaliação da forma da elaboração dos textos dos sumários de alta. A seguir foi realizado o processamento das narrativas aplicando-se conceitos de Processamento de Linguagem Natural (PLN). Para isto foram utilizadas as técnicas de PLN embutidas no CoGrOO, um corretor gramatical que é preparado para trabalhar com textos em português. Foi necessário incluir ao CoGrOO um *corpus* da área médica para obter melhores resultados na identificação dos diagnósticos, ao invés de se basear no *corpus* original que é relacionado a textos jornalísticos. O processamento dos sumários foi realizado a partir da elaboração de regras baseadas na classificação dos termos resultante da etapa de análise sintática do PLN. Os diagnósticos foram identificados aplicando-se as regras propostas. A partir da identificação dos diagnósticos, foi realizada uma correlação de grupo e categoria da CID-10. Médicos analisaram estas extrações e correlações, confirmando-as ou sugerindo alterações. Uma ferramenta semiautomática, NOTTUS, foi desenvolvida no intuito de permitir a integração ao PEP. Foram criadas 32 regras para extração dos diagnósticos. Com a aplicação do método proposto nesta pesquisa foram extraídos 52.379 termos (repetidos ou não) das 6.000 narrativas clínicas, onde apenas 2,5% dos termos representaram falsos diagnósticos. Num estudo específico avaliando casos de hipertensão, foram encontrados apenas 113 registros com este diagnóstico definido num campo estruturado da base de dados. Aplicando-se o método proposto nesta pesquisa foram identificados mais 70% de pacientes com o diagnóstico de hipertensão relatado no sumário de alta. O método proposto permite facilmente adicionar novas regras para identificação de outros diagnósticos, podendo se adaptar melhor a cada contexto de aplicação, tais como: hospitalar, atenção primária, epidemiológico, pesquisa, entre outros.

Palavras-chave: Narrativas Clínicas. Processamento em Linguagem Natural. Diagnósticos.

ABSTRACT

Diagnoses identification in clinical reports and correlation to the International Classification of Diseases

During patient treatment some information are registered in text format into medical records, named clinical narratives. The use of information from these narratives is restrained once it's very difficult to identify them automatically. Diagnoses are one of the information in the clinical narratives and normally there are not related only to the treatment, that are registered in structured fields, that are not free text, of the Electronic Health System (EHR). There are secondary diagnoses: before of the treatment, used to investigate diseases and even the excluded ones. Besides identifying diagnoses, it is important facilitate codification to the International Classification of Diagnoses (ICD-10), that is one of the main standards applied to Health Information Systems. A method to extract diagnoses included in clinical reports may help the identification of them, improving the use of data from EHR, as well as facilitating the filling of information. So, the purpose of this research was to develop a method to identify clinical diagnoses in narratives, suggesting a correlation to groups and categories of ICD-10. A discharge summaries database from the Cardiology Department of Clinical Hospital of Porto Alegre (HCPA) was used composed of approximately 6,000 records. The development of this research was realized in six phases, starting with the assessment to the drafting texts of the discharge summaries. Then narratives were processed applying concepts of Natural Language Processing (NLP). For this purpose NLP techniques embedded in CoGrOO were applied, a grammar that is prepared to work with texts wrote in Portuguese. To obtain better results during diagnosis identification, it was necessary to include a medical corpus in CoGrOO, instead of basing on the original corpus that is related to journalistic texts. The summaries processing was carried out from the elaboration of rules based on classification of the terms that resulted from the parsing stage of NLP. The diagnoses were identified applying the proposed rules. After the identification of diagnoses, a correlation to the groups and categories of the ICD-10 was performed. Physicians analyzed these diagnoses extraction and correlation, confirming them or suggesting changes. A semiautomatic tool, NOTTUS, was developed in order to allow integrating with EHR. 32 rules were created to extract diagnoses. Applying the method proposed in this research were extracted 52,379 words (repeated or not) were extracted from 6,000 clinical narratives, where only 2.5% of the terms represented false diagnoses. In a specific study evaluating cases of hypertension, only 113 records were found with this diagnosis defined in a structured field of the EHR database. Applying the method proposed in this study over 70% of patients with diagnosis of hypertension reported in the discharge summary were identified. The proposed method allows easy addition of new rules to identify other diagnoses, better this adapting easy to each context of application, such as: hospitals, primary care, epidemiology and research, among others.

Key-words: Clinical Narratives. Natural Language Processing. Diagnostics.

LISTA DE FIGURAS

Figura 1 – Narrativa clínica.....	14
Figura 2 – O processamento em linguagem natural.....	21
Figura 3 – Extração de informações sobre HIV	25
Figura 4 – Resumo do ciclo.....	31
Figura 5 - Expansão de acrônimos.....	34
Figura 6 – Funcionamento interno CoGrOO.....	35
Figura 7 – Preparação da narrativa.....	37
Figura 8 – Identificação de regras	39
Figura 9 – Ciclo para definição de regras para identificar diagnósticos.....	41
Figura 10 – Correlação para hipertensão.....	42
Figura 11 – NOTTUS, funcionamento.....	53
Figura 12 – Esquema da ferramenta para a extração de diagnósticos clínicos, extrair diagnósticos	54
Figura 13 – Arquivo com regras para extração de diagnóstico.....	55

LISTA DE QUADROS

Quadro 1 – Exemplo de informações contidas em sumários de alta.....	15
Quadro 2 – Extração de informação (ferramentas).....	23
Quadro 3 – Texto em maiúscula e com acrônimos.....	32
Quadro 4 – Exemplo da escrita ideal.....	33
Quadro 5 – Sumários de Alta x CoGrOO	37
Quadro 6 – Significado das etiquetas utilizadas.....	38
Quadro 7 – Regras iniciais para extração de diagnósticos	40
Quadro 8 – Classificação diagnóstico cardiopata isquêmico	45
Quadro 9 – Classificação diagnóstico cardiopatia isquêmica.....	45
Quadro 10 – Classificação diagnóstico cardiopata isquêmica.....	46
Quadro 11 – Falsos diagnósticos	46
Quadro 12 – Regras para extração de diagnósticos.....	49
Quadro 13 – Importância da ordem das regras para extrair diagnóstico	50
Quadro 14 – Sugestões e correções da validação com médicos.....	52
Quadro 15 – Tratamento das regras em sumários iniciando com diagnósticos	57
Quadro 16 – Negação de diagnósticos	60
Quadro 17 – Problemas revelados por <i>Stop Words</i>	61

LISTA DE ABREVIATURAS E SIGLAS

CID	- Classificação Internacional de Doenças
CoGrOO	- Corretor Gramatical acoplável ao OpenOffice.org
DATASUS	- Banco de dados do Sistema Único de Saúde
GATE	- <i>General Architecture for Text Engineering</i>
HAS	- Hipertensão Arterial Sistêmica
HCPA	- Hospital de Clínicas de Porto Alegre
ICP2	- Classificação Internacional do Cuidado Preliminar Versão 2
IDE	- <i>Integrated Development Environment</i>
MedLEE	- <i>A Medical Language Extraction and Encoding System</i>
MeSH	- <i>Medical Subject Headings</i>
NOTTUS	- <i>Natural and Ontology Terminology User System</i>
PLN	- Processamento de Linguagem Natural
PPGTS	- Programa de Pós-Graduação em Tecnologia em Saúde
PUCPR	- Pontifícia Universidade Católica do Paraná
SNOMED	- <i>Systematized Nomenclature of Medicine</i>
SNOMED-CT	- <i>Systematized Nomenclature of Medicine-Clinical Terms</i>
SOM	- <i>Self-Organizing Maps</i>
UMLS	- <i>Unified Medical Language System</i>
VISL	- <i>Visual Interactive Syntax Learning</i>
WHO	- <i>World Health Organization</i>
XML	- <i>eXtensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO	10
1.1 OBJETIVOS	11
1.1.1 Objetivo geral	11
1.1.2 Objetivos específicos	12
2 REVISÃO BIBLIOGRÁFICA.....	13
2.1 NARRATIVAS CLÍNICAS.....	13
2.1.1 Sumários ou resumos de alta.....	14
2.1.2 Caracterizando o diagnóstico clínico.....	16
2.1.3 Acrônimos	17
2.2 RECUPERAÇÃO DE INFORMAÇÃO EM SAÚDE.....	18
2.2.1 Processamento de linguagem natural (PLN)	19
2.2.2 Ferramentas.....	22
2.3 PROCESSAMENTO EM LINGUAGEM NATURAL APLICADA A NARRATIVAS CLÍNICAS.....	24
2.4 CORRELAÇÃO E PADRONIZAÇÃO DE DIAGNÓSTICOS.....	27
3 METODOLOGIA.....	30
3.1 ANÁLISE DE NARRATIVAS CLÍNICAS.....	31
3.1.1 Formato de preenchimento	32
3.1.2 Análise dos acrônimos	33
3.2 PROCESSAMENTO DAS NARRATIVAS CLÍNICAS.....	34
3.3 DEFINIÇÃO DAS REGRAS PARA EXTRAIR DIAGNÓSTICOS.....	38
3.4 CORRELAÇÃO DE DIAGNÓSTICOS DA CARDIOLOGIA PARA O CID-10	42
3.5 ANÁLISE DOS DIAGNÓSTICOS ENCONTRADOS	43
3.6 DESENVOLVIMENTO DE UMA FERRAMENTA PARA INTEGRAÇÃO AO PEP	44
4 RESULTADOS.....	47
4.1 MÉTODO PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS.....	47
4.2 REGRAS PARA EXTRAIR DIAGNÓSTICOS	47
4.2.1 Aplicação das regras.....	49
4.3 CORRELAÇÃO DE DIAGNÓSTICOS.....	50
4.4 ANÁLISE PELOS ESPECIALISTAS DOS DIAGNÓSTICOS IDENTIFICADOS	51

4.5 NOTTUS – NOVO SISTEMA PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS	53
5 DISCUSSÃO	56
5.1 MÉTODO PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS EM NARRATIVAS CLÍNICAS	56
5.2 MAPEAMENTO DOS DIAGNÓSTICOS PARA A CID-10	58
5.3 ANÁLISE PELOS ESPECIALISTAS DOS DIAGNÓSTICOS IDENTIFICADOS	59
5.4 NOTTUS	59
5.5 TRABALHOS FUTUROS	61
6 CONCLUSÃO	63
REFERÊNCIAS	64
APÊNDICES	68
ANEXOS	70

1 INTRODUÇÃO

Atualmente, no atendimento ao paciente, são gerados e armazenados dados em prontuários clínicos ou hospitalares. Alguns dados como nome, data de nascimento, sexo e codificação do diagnóstico são registrados em campos estruturados nas bases de dados do sistema de Prontuário Eletrônico do Paciente (PEP). Também existem dados que são preenchidos em campos de texto livre - são as narrativas clínicas, tais como descrições de cirurgias, avaliações, exames e sumários de alta. A análise dos dados contidos em narrativas clínicas não é uma tarefa trivial, sendo necessário aplicar um processo especial que envolve métodos de recuperação de informação (CHU, 2002; SHORTLIFFE; BARNETT, 2001).

Dentro do contexto do PEP, o sumário de alta é um dos documentos mais completo em comparação com os demais tipos de narrativas clínicas. Isto ocorre devido ao fato que, quando os médicos elaboram um sumário de alta, sabem que este deverá ser utilizado por diferentes profissionais e instituições de saúde. É um documento que resume o atendimento que o paciente recebeu, sendo fundamental para a continuidade do atendimento e tratamento. Os sumários contêm vários dados que geralmente não estão em nenhum outro campo da base de dados do sistema do PEP. Porém, muitas vezes os sumários também contêm dados misturados nos textos, inclusive os pessoais, como, por exemplo, nomes e sobrenomes de médicos, pacientes, números de identificação, como telefone, datas e localização geográfica. Estes são obstáculos que dificultam a extração de informações na narrativa. Este tipo de texto é caracterizado como sem estrutura (UZUNER et al., 2008). Os laudos de exames de imagens geralmente possuem uma estrutura pré-definida, o que pode facilitar a identificação dos dados.

Sabendo-se que sumários de alta geralmente contêm dados relacionados aos diagnósticos, pois os profissionais da área de saúde têm dificuldade em definir diagnósticos de forma estruturada, e que o Processamento de Linguagem Natural (PLN) pode auxiliar na extração de dados de narrativas clínicas, é importante o desenvolvimento de métodos que auxiliem na identificação de diagnósticos em narrativas clínicas, em especial nos sumários de alta (LONG, 2007; WANG et al., 2008). Porém, além dos trabalhos encontrados na literatura descreverem superficialmente o método para identificação das informações, geralmente utilizam

contextos restritos e termos bem específicos (FARKAS; SZARVAS, 2008; FRIEDLIN; MCDONALD, 2006; ZENG et al., 2006). Além disso, em nenhum deles é relatada a extração de diagnósticos em narrativas clínicas escritas em português.

O PLN é um método que permite extrair informações em narrativas clínicas, que pode ser dividido em quatro fases: análise morfológica, sintática, semântica e pragmática. A análise sintática foi a principal fase utilizada nesta pesquisa. Ela faz relações entre as palavras e termos de uma frase, definindo sujeito, predicado, substantivo, verbo, complementos, adjuntos e outros (MELTON; HRIPCSAK, 2005; RUSSEL; NORVIG, 2009; RICH; KNIGHT, 1994; SUN; SUN, 2006).

Porém, além de auxiliar na extração de narrativas dos sumários de alta, é importante incluir no método de identificação de diagnósticos em narrativas clínicas a correlação dos mesmos para a Classificação Internacional de Doenças (CID-10), uma vez que geralmente os médicos devem elaborar e codificar os diagnósticos nesta classificação. Futuramente, esta correlação, assim como o método de extração, poderá facilitar a realização de diagnósticos em CID-10 durante a elaboração das narrativas.

Desta forma, o objetivo deste trabalho foi identificar os diagnósticos contidos nos sumários de alta, sugerindo uma correlação dos mesmos com a CID-10. Utilizando a análise sintática e o modelo de identificação de informações em documentos da área de saúde propostos por Bui et al. (2010), foram construídas regras para a extração dos diagnósticos nas narrativas clínicas.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Identificar diagnósticos contidos em sumários de alta hospitalar elaborados em formato de texto livre, propondo correlações com a CID-10.

1.1.2 Objetivos específicos

- a) Elaborar um método para a identificação de diagnósticos relacionados à cardiologia presentes nos sumários de alta.
- b) Correlacionar os diagnósticos encontrados para a CID-10.
- c) Analisar os diagnósticos encontrados pelo método e a sua codificação com a CID-10.
- d) Desenvolver uma ferramenta para identificação de diagnósticos.

2 REVISÃO BIBLIOGRÁFICA

A revisão bibliográfica foi realizada com intuito de apresentar quais são os dados em saúde presentes no prontuário, como eles são inseridos e as opções para a sua extração e utilização.

Os artigos foram selecionados baseando-se no *abstract*. Embora durante a pesquisa, tenha sido dada preferência para artigos publicados nos últimos dez anos (1999–2009), aqueles que não estão nesse intervalo de tempo são justificados pela utilização de artigos que são clássicos na literatura e que auxiliam na fundamentação dos temas.

2.1 NARRATIVAS CLÍNICAS

Os dados dos prontuários são essenciais para dar continuidade ao tratamento do paciente, pois, além do cadastro, estes dados referem-se ao seu acompanhamento, podendo existir relatórios de comportamento preventivo que são utilizados em identificação de riscos, avaliação de novas intervenções, diagnósticos, entre outros até que se receba a alta (LAPELLE et al., 2006).

Estes dados integram o registro de evolução ou história clínica do paciente, que, quando preenchido de forma correta, alerta sobre variações e resultados das consultas, diagnósticos, medicamentos e comportamento do paciente (NARYSHKIN; SCHULTZ, 2009). Os dados em formatos de texto constituem as narrativas clínicas (CHU, 2002).

No prontuário, as narrativas podem ser preenchidas com textos sem estrutura (Figura 1), que permitem mais flexibilidade ao especialista e não possuem qualquer formato ou regra. As narrativas são elaboradas por um ou mais profissionais da área de saúde e podem auxiliar na identificação de um grande volume de informação.

Novo Cadastro	Monitoramento	Ajuda	Sobre			
Agenda	Adicionar Consulta	Lista de Problemas	Subjetivo	Objetivo	Avaliação	Plano
Nome do Paciente: <input type="text" value="PacienteX"/>						
Farmâcos: <input type="text" value="Hidroclortiazida 25mg
Captopril 25mg"/>						
Observação: <input type="text" value="Paciente hipertenso, internou com quadro de dor anginosa em repouso.
ECG com alterações da RV em parede anterior.
Marcadores de necrose normais.
Realizou ergometria interrompida precocemente por dor típica, sem alterações do ecg.
Realizou cat que mostrou coronárias normais. Diagnóstico final: cardiopatia hipertensiva."/>						
<input type="button" value="Limpar Farmâcos"/> <input type="button" value="Limpar Comentários"/> <input type="button" value="Salvar"/>						

Figura 1 – Narrativa clínica

As narrativas de pacientes sobre as suas doenças abordam aspectos da vida, discussões sobre a doença, depoimentos que são tomados e fazem parte de anamneses. As anamneses são sistematizadas em prontuários onde constam informações como, por exemplo: identificação, diagnósticos, queixa principal, história da doença atual, história patológica pregressa, revisão dos sistemas, história familiar e social. As narrativas clínicas também podem ser usadas como propostas educativas e materiais informativos (GROSSMAN; CARDOSO, 2006).

Os exemplos mais comuns de narrativas clínicas são: história médica, exames físicos, notas de evolução e laudos de procedimentos especiais, como os da radiologia, patologia, endoscopia, e os sumários ou resumos de alta. Alguns dos tipos de narrativas, como os laudos, possuem uma estrutura determinada e conhecida, porém os sumários de alta são frequentemente escritos sem qualquer tipo de padrão. Neles, encontram-se muitos dados e informações que podem auxiliar a equipe de saúde no acompanhamento ao paciente (TANGE et al., 1997).

2.1.1 Sumários ou resumos de alta

O sumário de alta é um dos documentos mais completos, entretanto, no seu conteúdo, podem ser encontradas informações pessoais e privadas, como por

exemplo, nomes e sobrenomes de médicos, pacientes, familiares, números de identificação como telefone, celular, fax, datas e localização geográfica, entre outros obstáculos que dificultam a extração de diagnósticos (UZUNER et al., 2008).

Existem várias informações dentro do sumário de alta, que vão desde o atendimento até os procedimentos realizados que podem beneficiar os pacientes, pois os profissionais utilizam essas informações para tomada de decisão dentro de um sistema de saúde, evitando duplicidade de exames ou procedimentos e podendo diminuir o custo da assistência (KLUCK; GUIMARÃES, 1999).

Fator	Texto retirado do Sumário de Alta
Dados pessoais, acrônimos, diagnósticos, entre outros.	"Maria, 57 anos # HAS # Dislipidemia # Cardiopatia Isquêmica - Stent DAE 1996 - Stent CD março/02 - ACTP intrastent CD setembro/02 - Cateterismo com necessidade de dissecação pós-punção de artéria braquial direita. Submetida à CRM por angina refratária em 06/05/03 com implante de ponte de safena para artéria descendente posterior. Medicamentos em uso na alta: - AAS 100 mg 1x/dia - Enalapril 5 mg 12/12 horas"
Texto totalmente escrito em maiúsculo.	"HAS, ACFA, CRM PRÉVIA, INTERNA PARA STENT EM CD (2), PROCEDIMENTO SEM INTERCORRÊNCIAS, COM SUCESSO ANGIOGRÁFICO. RECEBE ALTA ASSINTOMÁTICO E EM USO DAS MEDICAÇÕES E RETORNO PARA SEGUIMENTO."

Quadro 1 – Exemplo de informações contidas em sumários de alta

Fonte: Sumário de Alta HCPA

No Quadro 1 é possível notar diversas informações que são importantes para o tratamento do paciente e que, muitas vezes, não estão em nenhum outro lugar estruturado da base. Destaca-se em cinza, destaca-se no texto apresentado no Quadro 1, a presença de diagnósticos e acrônimos que são frequentes nas narrativas clínicas.

Os diagnósticos presentes em narrativas clínicas podem ser confundidos com fatores de risco. Por isso, existe a necessidade de conhecer mais sobre o processo de diagnosticar um paciente.

2.1.2 Caracterizando o diagnóstico clínico

Na literatura, alguns artigos descrevem o processo de como diagnosticar um paciente; porém, não existe uma definição específica do que é um diagnóstico clínico (LACQUANITI et al., 2010; SELLEVOL; STENSETH, 2010).

Quando um especialista chega ao fim do processo de diagnosticar um paciente, em alguns sistemas em saúde e prontuários eletrônicos, adotam-se classificações que, atualmente, são recomendações de órgãos importantes, como o *World Health Organization* (WHO, 2009).

A definição do diagnóstico utilizando classificadores padrões pode ser gerada através do *Systematized Nomenclature of Medicine-Clinical Terms* (SNOMED), embora seu uso seja relativamente bem menor se comparado à CID-10 - Classificação Internacional de Doença, que é o mais recomendado e utilizado (IHTSDO, 2009).

A CID-10 é um padrão de referência estabelecido pela Organização Mundial da Saúde, que cataloga doenças e problemas relacionados à saúde, permitindo que programas possam referenciar de modo padronizado as classificações e auxiliar na busca de informação diagnóstica para finalidades gerais. Classifica morfologicamente neoplasias, exibe listas especiais de tabulação para mortalidade e morbidade, além de fornecer as definições e os regulamentos da nomenclatura, através da Lista Tabular (DATASUS, 2010).

Em narrativas clínicas geralmente são utilizados acrônimos, inclusive para os diagnósticos, como é o caso de HAS (Hipertensão Arterial Sistêmica). Apesar de alguns acrônimos não serem oficiais, são frequentemente utilizados. Existe a necessidade de expandir os acrônimos para melhorar a qualidade da narrativa escrita e facilitar a identificação das informações, principalmente quando significam um diagnóstico.

2.1.3 Acrônimos

Acrônimos são abreviaturas. São muito utilizados pelos médicos durante a escrita das narrativas clínicas. É importante identificar e tratar os acrônimos, pois podem resolver problemas de ambiguidade, que são termos com mais de um significado e dependem do contexto, e também podem melhorar a qualidade do texto. Por exemplo: DAE, que pode significar “dimensão do átrio esquerdo”, “descendente anterior esquerdo” ou “drogas antiepilépticas”.

Os acrônimos são utilizados para a representação de vários conceitos, como, por exemplo: AA – após almoço, CP – comprimido, ESQ – esquerdo, LUTS - sintomas do trato urinário inferior, HAS – hipertensão arterial sistêmica, VEF1 - volume expiratório forçado 1, entre outros.

Os diagnósticos podem estar em forma de acrônimos, utilizados com frequência em narrativas clínicas, mesmo que seja uma prática não recomendável. Essa recomendação é feita porque o uso de acrônimos não oficiais, ou seja, siglas criadas por critérios pessoais podem causar ambiguidades (exemplo: “na” é sódio ou apenas “na”) e são barreiras para trabalhos como, por exemplo, o processamento de texto automático (PAKHOMOV; PEDERSEN; CHUTE, 2005).

Sohn et al. (2008) demonstram a criação de um algoritmo para identificação de abreviações na base do MEDLINE. A estratégia é identificar a definição mais provável para uma sigla, apresentando a precisão do resultado de forma automática

Xu, Stetson e Friedman (2007) relatam a extração de acrônimos, além de outras informações clínicas. Utilizaram a ferramenta Weka com um *corpus* preparado para a detecção de acrônimos. Uma lista com palavras em inglês de duas ou três letras foram inseridas em uma lista de *stop words* para evitar falsos resultados.

O artigo *A Study of Abbreviations in Clinical Notes* descreve um modelo para a construção de anotações clínicas, eliminando os chamados falsos-acrônimos. A detecção destes acrônimos ocorreu em textos livres e foram colocados em um *Thesaurus* dicionário.

Foram usados quatro métodos para a detecção: o primeiro faz uma comparação no inventário e utiliza *Tokens* para buscar as abreviaturas no texto, em que o desconhecido é descartado. O segundo utiliza um método heurístico baseado

em regras que usa informações relativas da palavra de formação, tais como letras maiúsculas, caracteres alfabéticos e numéricos, funcionando em conjunto com o inventário. No terceiro e quarto métodos, é formada uma árvore de decisão com o algoritmo J48, quando se realiza a análise pragmática do conteúdo.

O método obteve uma precisão de 91,4%. A construção do inventário usou duas fontes de conhecimento: o *Unified Medical Language System (UMLS)* e a base de dados do MEDLINE (XU; STETSON; FRIEDMAN, 2007).

Uma alternativa para evitar parte dessas abreviações contidas na narrativa clínica pode ser a codificação de elementos encontrados no texto para CID-10 e SNOMED. Casos já existentes de mapeamento estão relatados em trabalhos na área de radiologia, em terminologias biomédicas e em reações de medicamentos, conseguindo bons resultados (FARKAS; SZARVAS, 2008; ALECU; BOUSQUET; JAULENT, 2008).

Os acrônimos precisam ser tratados, pois podem representar informações importantes, como, neste caso, os diagnósticos. Por sua vez, para identificar os diagnósticos presentes em narrativas clínicas, é necessário utilizar técnicas para recuperação de informação.

2.2 RECUPERAÇÃO DE INFORMAÇÃO EM SAÚDE

Rosier, Burgun e Mabo (2008) descrevem que o PLN e a Mineração de Textos são técnicas que podem auxiliar na identificação de dados clínicos contidos em relatórios cirúrgicos. No seu trabalho, eles identificaram informações referentes à estimulação cardíaca e à desfibrilação. Isto foi realizado baseando-se numa série de regras para extrair dados dos textos dos relatórios relacionados aos dispositivos e aos procedimentos médicos. Para validar o método e criar um padrão ouro, especialistas realizaram uma extração manual da base e compararam os resultados encontrados com a utilização do método por eles especificado.

Um cardiologista especialista em marcapassos definiu uma lista de 5 categorias de dados relevantes: nome do paciente, nome do cardiologista, modelo do dispositivo, fabricante e número de série. O sistema extraiu informações, incluindo dados numéricos, texto e de combinação de números e caracteres, com

alta sensibilidade (maior que 90%) e um valor preditivo ainda maior (superior a 95%). Apresentou uma precisão superior à do registro original, povoado por inserção manual (ROSIER; BURGUN; MABO, 2008).

Outras formas de extração podem acontecer através de fontes que não estão prontas em forma de narrativa. São os casos de vídeos, em que o áudio é capturado e transformado em texto e o conteúdo auxilia na construção do sumário (GROEN et al., 2006). A interpretação da interface de forma automática, através de monitores, também é uma forma diferente de extração do conteúdo, podendo criar sumários automaticamente, mostrando a informação contida em sinais (CHARNOVSCK et al., 2004).

2.2.1 Processamento de linguagem natural (PLN)

O PLN é uma técnica para extrair informações de textos. Está dividida em quatro etapas: análise morfológica, análise sintática, análise semântica e análise pragmática.

A **análise morfológica** é responsável por definir artigos, substantivos, verbos e adjetivos, armazenando-os em um dicionário. Depois de construído o dicionário, a **análise sintática** faz uso dele, procurando mostrar relacionamento entre as palavras e em um segundo momento, verifica sujeito, predicado, complementos nominais e verbais, adjuntos e aposto (RUSSEL; NORVIG, 2009).

Na **análise semântica**, ocorre o encontro de termos ambíguos, de sufixos, afixos, ou seja, questões dos significados associados aos morfemas componentes de uma palavra e do sentido real da frase ou palavra. Para a visualização de todo este contexto, a **análise pragmática** faz a junção de todo o mecanismo e mostra visualmente o resultado. Neste caso, é possível utilizar algoritmos que mostram em forma de árvore todos os passos até se chegar à conclusão, que é a identificação dos termos de uma sentença (RICH; KNIGHT, 1994).

Quando são utilizados textos registrados em áudio, a **análise fonética** deve ser aplicada ao PLN. Segundo Ribeiro (2001), o objetivo desta análise é descobrir a sequência de sons das palavras a partir de suas letras, a maneira como estes sons

são agrupados em sílabas e qual sílaba de cada palavra recebem ênfase na pronúncia (a sílaba tônica), transformando os sons em palavras.

As análises realizadas durante o PLN são baseadas fundamentalmente em comparações estatísticas com documentos que contêm estrutura sintática e semântica. Isto é, as palavras que compõem os textos estão classificadas como artigos, verbos, substantivos, nomes próprios, adjetivos, entre outros. Os relacionamentos entre as palavras e termos. Também foram estabelecidos. O conjunto de documentos com estas definições é denominado *corpus*. A criação de um *corpus* é realizada através da aplicação de PLN em um texto que, a seguir, depois de ter sido classificado, é analisado por especialistas para confirmar ou corrigir a classificação proposta. *Corpus* que foram corrigidos por especialistas são denominados de *corpus* anotados. A maioria das ferramentas de PLN disponíveis utiliza *corpus*, baseados em textos jornalísticos (CHANLEKHA; KAWAZOE; COLLIER, 2010; SOUTH et al., 2009).

O esquema apresentado na Figura 2 representa a relação das principais fases do PLN. O termo “sintagmas” referente à semântica, expressa uma relação de dependência em que existem um elemento determinado e outro subordinado. Serve para se referir a alguma parte da sentença, seja palavra, sentença ou período.

Para corrigir erros do processamento, é possível incluir uma lista de **steeming** e de **stop words**. *Steeming* é uma ação que reduz uma palavra ao seu radical, ou seja, associa várias palavras com um mesmo significado em uma palavra radical, auxiliando na construção de regras mais exatas para a extração. *Stop words* é um conjunto de palavras que representa as palavras indesejáveis, extraídas através de regras de identificação de informação. Quando uma determinada palavra indesejada aparece nos resultados esperados e não é possível mudar a regra de extração, é adicionada em uma lista, para que seja removida dos resultados.

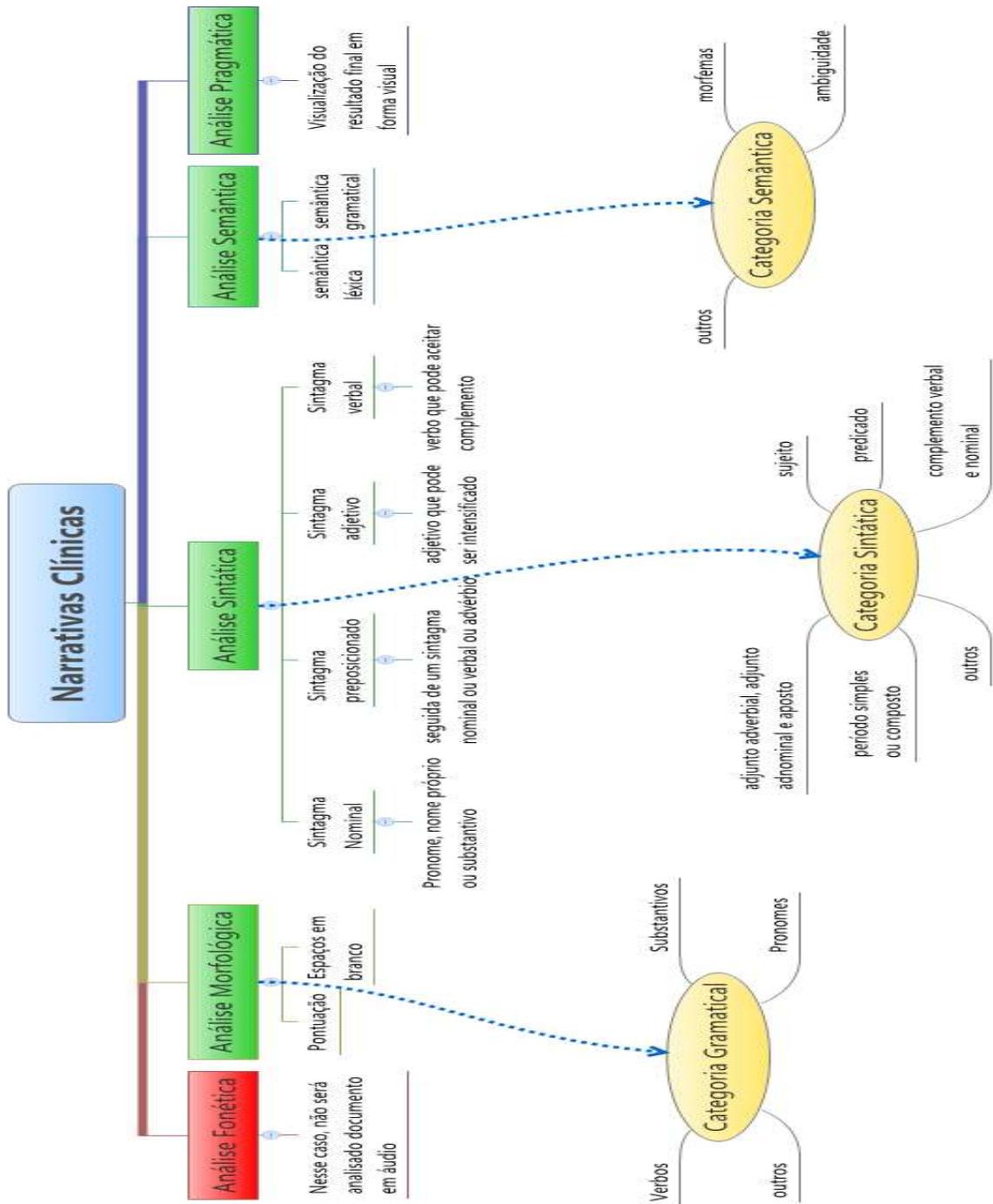


Figura 2 – O processamento em linguagem natural

Fonte: Adaptado de RICH; KNIGHT,1994 e de RUSSEL; NORVIG, 2009

Alguns trabalhos relataram o PLN como uma forma alternativa para extrair informações de narrativas clínicas, tais como medicamentos, recuperação de conteúdo de resumos de artigos e muitos outros referentes à biologia (BARNICKE et al., 2009; XU; STETSON; FRIEDMAN, 2007; ZHOUG et al., 2005).

Na saúde, as ferramentas que auxiliaram o processamento em linguagem natural foram utilizadas, na maioria das vezes, com a validação da classificação dos textos, ou seja, com a criação de um *corpus*. A necessidade da criação de um *corpus* é justificada por vários motivos, como a linguagem, os regionalismos e outros tipos de ferramentas que são preparadas para ser utilizada em outro contexto (VISL, 2009).

2.2.2 Ferramentas

Durante a busca de artigos e livros que envolvem o PLN, algumas alternativas de ferramentas para a Mineração de Textos e para o PLN foram encontradas e podem ser adaptadas ou ser a solução para a identificação de informação em textos. As ferramentas encontradas estão relacionadas no Quadro 2.

Ferramenta/ Link	<i>Steeming</i>	<i>Stop Words</i>	PLN	Mineração de Textos	Linguagem	Análise
CoGrOO/ http://cogroo.sourceforge.net	Não	Não	Sim	Não	Português	Sintática, Semântica e Pragmática
Dt-Search/ http://www.dtsearch.com	Sim	Não	Não	Sim	Inglês	Não descreve.
GATE/ http://www.gate.ac.uk	Sim	Sim	Sim	Sim	Inglês	Semântica.
MedLEE/ http://www.lucid.cpmc.columbia.edu/medlee	Não	Não	Sim	Não	Inglês	Sintática
OpenNLP/ http://opennlp.sourceforge.net/projects.html	Não	Não	Sim	Não	Inglês	Sintática, Semântica e Pragmática
Weka/ http://www.cs.waikato.ac.nz/ml/weka	Não	Não	Sim	Sim – prioridade para a Mineração de Dados	Inglês	Sintática, Semântica e Pragmática
WizDoc/	Sim	Sim	Não	Sim	Inglês	Morfológica

http://www.searchools.com/tools/wizdoc.html						
---	--	--	--	--	--	--

Quadro 2 – Extração de informação (ferramentas)

Estas ferramentas podem trazer alto poder de precisão na extração de informações, como, por exemplo, o MedLEE, que possui trabalhos publicados mostrando 92% de acerto na busca; porém, seu método é preparado para a linguagem escrita em inglês (WANG et al., 2008).

Uma ferramenta criada através de trabalhos realizados por Roger Schank e Christopher Riesbeck (1981), envolvendo teorias de Memória Dinâmica e Raciocínio Baseado em Casos aplicadas à compreensão da linguagem natural, é o DMAP.

O DMAP - *Direct Memory Access Parser* é uma ferramenta que funciona de uma forma diferente em relação às outras ferramentas citadas. Trata-se de um *parser* (programa que executa a análise da entrada de dados e que determina uma estrutura gramatical seguindo uma gramática formal). Trabalha também com lista de palavras e tem como finalidade principal a manipulação dos conceitos armazenados na memória, a fim de compreender sentenças em linguagem natural, reproduzindo as funções cognitivas do ser humano. Sua característica essencial é buscar os conceitos na memória e relacioná-los a uma frase de entrada (WOJCIECHOWSKI, 2003).

Neste trabalho, será utilizada uma ferramenta preparada para o português, conhecida como CoGrOO. A opção foi feita por que é um corretor gramatical desenvolvido em código aberto na linguagem Java (que é independente de plataforma) e tem um acoplamento ao *OpenOffice*. O CoGrOO verifica a colocação pronominal, concordância nominal, concordância entre sujeito e verbo, concordância verbal, uso de crase, regência nominal, regência verbal e erros comuns da língua portuguesa falada. Também possui como vantagem o *OpenNLP* embutido em seu mecanismo (CoGrOO, 2009).

2.3 PROCESSAMENTO EM LINGUAGEM NATURAL APLICADA A NARRATIVAS CLÍNICAS

O PLN aplicado em saúde está focado mais na recuperação de informações em narrativas clínicas, dos quais são extraídos acrônimos, medicamentos e doenças. A literatura ainda apresenta artigos que se apoiam em classificações internacionais, como o SNOMED e a CID-10 (KLING, 2007).

Zeng et al. (2006) utilizaram uma amostra revisada por médicos contendo 150 registros coletados de um sumário de alta com 97.000 pacientes com asma e doença pulmonar crônica obstrutiva para criar uma ferramenta de extração conhecida como HiTE_x, que incorpora o GATE. Foram extraídos diagnóstico principal, comorbidades e tabagismo com o HiTE_x e médicos responderam, após análise manual dos textos, se confirmavam o diagnóstico ou não, bem como se havia informação suficiente para esta confirmação. Este artigo não descreve quais são as regras utilizadas para a extração dos diagnósticos. Porém, afirma que combinando o HiTE_x com o CID-9, é possível extrair os diagnósticos com boa margem de acerto.

Bui et al. (2010) definiram regras sintáticas para extrair e combinar relações entre o HIV e os medicamentos da doença, através de resumos de artigos retirados do PubMed. A regra foi criada a partir de um estudo do texto e de como ele era classificado. A ferramenta utilizada não foi relatada, apenas foi citado que foi utilizada uma ferramenta PLN.

Os pesquisadores adicionaram nas regras sintáticas palavras-chave para adicionar regras específicas que a análise sintática não encontrou. Os principais termos são *“resistance”* e *“responsive”*.

As etapas para a elaboração foram divididas em três fases: recuperação de resumos de artigos do PubMed, processamento do texto e a criação de relações entre os termos em torno dos medicamentos que causam resistência ao HIV. Os textos coletados no PubMed foram comparados a uma lista pré-definida em XML, com o nome de medicamentos que causam resistência ao HIV, para selecionar aqueles que mais tinham a possibilidade de conter informações sobre medicamentos e reações relacionadas ao vírus.

O processamento do texto engloba 3 etapas: simplificar sentenças, analisar sentenças e gerar relações gramaticais.

Para simplificar as sentenças, alguns termos são excluídos, por tratarem de contextos biomédicos que não interessam na pesquisa, removendo caracteres especiais, como, por exemplo: +&#\$%/, e definindo palavras chaves que auxiliam a identificação da interferência do medicamento em relação ao HIV, por exemplo: mutação e resistência.

Na análise de sentenças, foi verificado como os termos definidos na simplificação da sentença estão localizados na frase e da distância para o medicamento que causou resistência, comparando com uma nova tabela, que diz o possível grau de interferência.

As relações gramaticais são definidas através da classificação do medicamento e da classificação dos seus termos antecessores e sucessores. A Figura 3 exemplifica como é realizada a identificação sintática para encontrar um medicamento que causou determinada resistência.

As classificações representadas nas cores azul, verde e rosa foram adaptações na Figura para identificar a classificação utilizada nas palavras-chave para a criação da regra. Na cor rosa é representado a classificação da palavra que deseja ser extraída (*DRUG0*), que é comparada com a classificação das palavras em verde (*TO*, *BUT*). As classificações que estão na cor azul são classificações em comum para o termo que deseja ser extraído (classificação representada na cor rosa) e para as palavras que são comparadas para a construção da regra (classificação representada na cor verde).

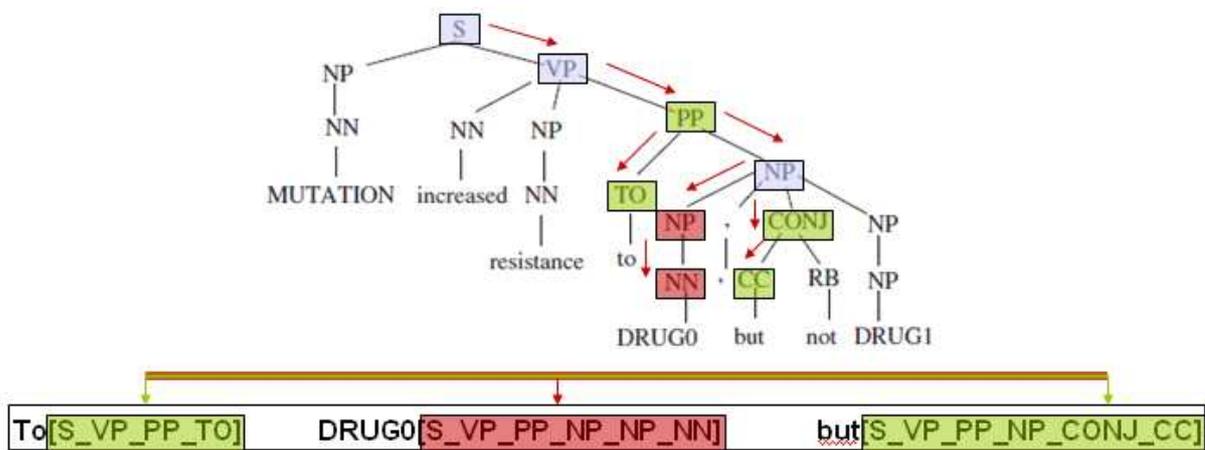


Figura 3 – Extração de informações sobre HIV

Fonte: Adaptado de BUI et al., 2010

Melton e Hripcsak (2005) extraíram informações do texto utilizando a ferramenta MedLEE, com um *corpus* anotado manualmente para procurar telefonemas, relatórios de incidentes profissionais e formulários da gestão de processos.

Groen et al. (2006) desenvolveram uma nova ferramenta para extrair informações contidas em áudio para construção de um sumário de alta. Utilizaram terminologias da endoscopia como um padrão para a extração.

Rosier, Burgun e Mabo (2008) utilizaram o GATE com um *corpus* preparado e associaram expressões regulares para extração de informações de marcapasso em relatórios de procedimento clínico.

Automatic construction of rule-based ICD-9-CM coding systems de Farkas e Szarvas (2008) relata uma extração de diagnósticos contidos em narrativas clínicas da radiologia, utilizando o padrão da CID-9. A primeira etapa foi converter o texto em minúsculo. As regras de extração foram criadas a partir de relacionamentos de sintomas que revelam possíveis diagnósticos. Os resultados que foram obtidos através de frases negativas foram automaticamente desprezados, assim como frases em que palavras antecedentes ao diagnóstico apresentaram palavras como “*can*”, “*consistent*”, “*could*”.

Friedlin e McDonald (2006) utilizaram sumários de internação para procurar doze tipos de doenças que aparecem por histórico familiar, como, por exemplo, doença arterial coronariana, diabetes e hipertensão. O artigo relata o uso de um *corpus* padrão *gold* e uma ferramenta denominada REX; porém, não mostra qual a regra utilizada para a extração dos diagnósticos.

Lessons Extracting Diseases from Discharge Summaries de Long (2007) relata o desenvolvimento de uma nova ferramenta para a extração de doenças contidas em narrativas clínicas, utilizando um *corpus* anotado e com o padrão UMLS. Neste trabalho são feitas a eliminação de preposições e a análise de verbos antes do diagnóstico. O programa divide o texto em frases usando pontuações e conjunções (“e” “ou”...) e alguns verbos (“é”, “sente”, ...) e, em seguida, as relaciona com o padrão UMLS.

2.4 CORRELAÇÃO E PADRONIZAÇÃO DE DIAGNÓSTICOS

Atualmente, o compartilhamento de dados através de sistemas de informação é um sinônimo da falta de interoperabilidade entre diferentes terminologias ou classificações. Mapear diagnósticos para alguma classificação pode viabilizar a troca de conteúdo, além de automatizar processos (WANG et al., 2008).

Wang et al. (2006) desenvolveram um sistema para mapear automaticamente uma terminologia ICPC-2 PLUS para SNOMED CT. O mapeamento utiliza um dicionário baseado na *UMLS* e cria relações entre ICPC-2 PLUS e os termos da SNOMED CT, criando uma biblioteca. Em seguida, o PLN busca semelhanças lexicais para a segunda fase do mapeamento entre as terminologias. Essa relação define um termo da ICPC-2 mapeado em uma agregação de dois ou mais termos para o SNOMED CT.

Wang et al. (2008) também relatam o mapeamento entre duas terminologias, do ICPC-2 PLUS para o SNOMED CT. Para fazer o mapeamento, em um primeiro momento são comparados termos do ICPC-2 PLUS e SNOMED combinados com a *UMLS*, criando um dicionário de palavras. A segunda etapa relata a criação de regras utilizando o PLN, com o objetivo de buscar similaridades entre o termo a ser mapeado e seu correspondente. A terceira etapa é responsável por agregar e sugerir dois ou mais termos possíveis como resultado para o mapeamento final. Através deste método, 80,58% dos termos puderam ser mapeados, porém com diferentes níveis de precisão. A causa destes diferentes níveis de precisão é proveniente da variação dos padrões utilizados entre os dois tipos de terminologias.

Ruch et al. (2008a) desenvolveram uma ferramenta que procura terminologias e apoia a categorização automática de conteúdos textuais utilizando o SNOMED. Os classificadores utilizados foram o SNOMED CT, *UMLS* e *MeSH*. Foi desenvolvido a partir de duas estratégias complementares de recuperação de informação. O primeiro utiliza uma classificação que envolve técnicas de espaço vetorial, enquanto o segundo classificador utiliza uma variação regular do dicionário de dados. Foi utilizada uma amostra retirada de resumos do MEDLINE (Sistema disponível em <http://eagl.unige.ch/SNOCat>).

Nystrom et al. (2007) desenvolveram uma ferramenta semiautomática integrada para a criação de arquétipos, utilizando a terminologia SNOMED CT.

Métodos léxicos e semânticos são aplicados para sugerir o mapeamento automático. Métodos de visualização de informação também são utilizados para auxiliar o usuário na exploração e seleção dos termos mapeados.

Meystre e Haug (2005) utilizaram o PLN no intuito de completar as informações em uma lista de problemas com diagnósticos descritos em narrativas clínicas, em que o diagnóstico pode estar já em seu formato do CID-9 ou escrito de outra forma.

A primeira etapa foi a identificação dos diagnósticos por meio do PLN. A regra para a identificação dos diagnósticos foi formada por palavras-chave observadas na lista de problemas, geram um XML (*eXtensible Markup Language*). Segundo os autores, para que esta etapa funcione corretamente, é recomendado que o profissional tenha cuidado com o formato da escrita. São utilizadas expressões regulares para realizar a detecção de sentença, excluir caracteres especiais e construir regras para relacionar diagnósticos.

O segundo passo foi a criação de uma rede Bayesiana para saber as probabilidades de um termo ser um diagnóstico contido no CID-9. Termos das narrativas são correlacionados como diagnósticos quando se parecem semanticamente.

Foram analisadas 3000 narrativas clínicas, procurando 80 diagnósticos cardiovasculares contidos no CID-9 (5% de todos os listados da CID-9), em que o método conseguiu identificar um correspondente do termo encontrado na lista de problemas para a CID-9 com uma taxa de acerto de 64%.

Naghavi et al. (2010) relatam dados referentes às causas de morte (CoD), que variam entre os países e são limitados por três problemas: mudanças na Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde; a utilização de listas tabuladas perdem detalhes importantes sobre as causas da morte e o número de mortes atribuídas a causas que não podem ou não devem ser consideradas causas de morte, pois muitas são inseridas no chamado código de lixo (GCs).

O objetivo é melhorar a utilidade dos dados nacionais de causa de morte para a análise da saúde pública. Cinquenta e seis causas são organizadas hierarquicamente para abranger todas as mortes. Cada causa foi mapeada da CID-6 para CID-10 e, quando possível, também foi mapeada para a GCS 1-5. Em cada revisão da CID, os GCs foram identificados. Esses GCs são redistribuídos e

identificados com base nas práticas de certificação e/ou fisiopatologia. Foram utilizados modelos estatísticos e algoritmos especialistas para redistribuir GCs, com o objetivo de mapear os códigos para cada grupo de idade e sexo.

Ruch et al. (2008b) relatam a criação de um sistema para auxiliar a atribuição da CID (Classificação Internacional de Doenças). Utilizaram a combinação de um conjunto de métodos de aprendizado de máquina e dados para gerar um único texto, que é categorizado automaticamente e retorna uma lista ordenada de códigos da CID.

Lowe, Huang e Regula (2009) relatam o mapeamento de informações de diagnósticos em relatórios da patologia para o SNOMED-CT. Utilizando relatórios da patologia, um sistema denominado *ChartIndex Medical Language Processing* foi adaptado para poder identificar e, automaticamente, mapear sintagmas e diagnósticos da patologia para SNOMED CT.

3 METODOLOGIA

Esta pesquisa é um estudo não experimental com característica exploratória descritiva. A partir dos resultados obtidos com a aplicação do método proposto para a identificação dos diagnósticos, também foi feita uma análise quantitativa, traduzindo opiniões e informações em números.

O trabalho está dividido em seis etapas principais:

1. Análise de narrativas clínicas.
2. Processamento das narrativas clínicas.
3. Definição das regras para extrair diagnósticos.
4. Correlação de diagnósticos da cardiologia para a CID-10.
5. Análise dos diagnósticos encontrados.
6. Desenvolvimento de uma ferramenta para integração ao PEP.

Da etapa de análise das narrativas clínicas até a de definição das regras para extrair diagnósticos (etapas 1 a 3), existe um ciclo recursivo, pois, após a definição das regras para extração, era verificado se todos os diagnósticos contidos nas narrativas analisadas durante esta etapa da pesquisa tinham sido identificados. Caso algum diagnóstico não tinha sido considerado, retornava-se à análise da narrativa e era feita uma tentativa de definir uma regra que contemplasse a situação em questão. Durante este processo, também era analisado se algum termo que não era diagnóstico tinha sido considerado, o que é denominado falso-diagnóstico. As regras foram definidas tentando maximizar o número de acertos e evitar os falsos diagnósticos. Na Figura 4, este ciclo é apresentado.

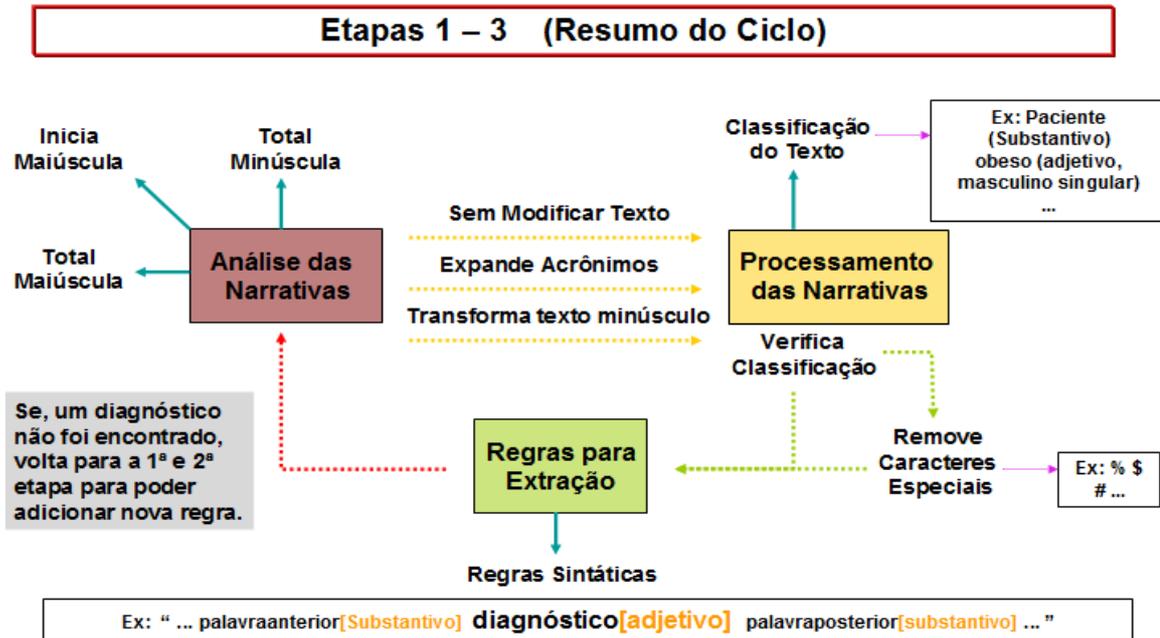


Figura 4 – Resumo do ciclo

Esta pesquisa foi realizada no Laboratório de Informática em Saúde (LAIS) do PPGTS da Pontifícia Universidade Católica do Paraná - PUCPR.

Esta pesquisa foi aprovada pelo Comitê de Ética (CEP) da PUCPR, tendo o parecer consubstanciado nº 0004422/10, para a participação de médicos na análise dos diagnósticos identificados (Apêndice A). A base de dados contendo as narrativas clínicas foi cedida pelo Hospital de Clínicas de Porto Alegre – HCPA ao grupo de pesquisa de recuperação de informações em saúde da PUCPR para o desenvolvimento de projetos, incluindo esta dissertação, com protocolo da Comissão Nacional de Ética em Pesquisa (CONEP) 0375.0.84.000-10.

Foram analisados 5.617 registros de pacientes que tiveram alta do serviço de cardiologia entre 2002 e 2007, e mais 383 registros referentes aos pacientes internados no HCPA que receberam alta durante o mês de julho de 2007, totalizando 6.000 registros.

3.1 ANÁLISE DE NARRATIVAS CLÍNICAS

As narrativas clínicas contidas nos sumários de alta foram elaboradas de três formas, considerando-se o tipo de letras:

- a) todo o texto em letras maiúsculas;
- b) todo o texto em letras minúsculas;
- c) os inícios de frases em letras maiúsculas e o restante do texto em minúsculas.

Além disso, as abreviaturas ou acrônimos, escritas em letras maiúsculas estão presentes em praticamente todos os sumários. Nas narrativas elaboradas totalmente em letras maiúsculas, existe uma maior dificuldade na identificação dos acrônimos. O texto do Quadro 3 apresenta dois diagnósticos abreviados (HAS – Hipertensão Arterial Sistêmica e DM2 – Diabetes Mellitus tipo 2), inseridos no meio de um texto com todas as letras maiúsculas. Os acrônimos, além de dificultar a identificação dos termos, podem alterar a classificação semântica.

PACIENTE HAS, DM2, CARDIOPATIA ISQUEMICA. INTERNA PARA ACTP+STENT CD. PROCEDIMENTO SEM INTERCORRÊNCIAS E COM SUCESSO ANGIOGRÁFICO. ALTA E AO SEU MÉDICO PARA SEGUIMENTO.
--

Quadro 3 – Texto em maiúscula e com acrônimos

Fonte: Sumários de Alta HCPA

Desta forma, para preparar a base de dados contendo os sumários, foi necessário fazer uma análise das narrativas clínicas para compreender o formato das mesmas, e poder identificar de que forma e em quais posições das narrativas os diagnósticos geralmente aparecem. Algumas vezes, os diagnósticos são escritos de forma abreviada, com acrônimo.

3.1.1 Formato de preenchimento

A narrativa clínica do sumário de alta foi escrita sem um padrão. Porém, existe um formato de texto ideal para trabalhar com o processamento de linguagem, que foi considerado quando não existem erros gramaticais. O diagnóstico é escrito por extenso, as frases dos textos são iniciadas em letras maiúsculas e o restante do texto é escrito em minúsculas. Um exemplo deste tipo de escrita é apresentado no Quadro 4.

Masculino, 55 anos, hipertenso, com história prévia de angioplastia mais implante de endoprótese em artéria coronária direita há dois anos, interna para nova angioplastia e implante de endoprótese. Procedimento não realizado por problemas apresentados no equipamento da sala 1 da hemodinâmica com previsão de necessidade de uma semana para conserto.

Quadro 4 – Exemplo da escrita ideal

Fonte: Sumários de Alta HCPA

Porém, mesmo utilizando-se a escrita ideal, geralmente existem mais de uma forma de escrever o diagnóstico. No exemplo apresentado no Quadro 4, o diagnóstico “Hipertenso” se refere também aos termos: “Hipertensa”, “Hipertensão” e “Hipertensivo (a)”.

Em algumas pesquisas, são utilizadas somente narrativas elaboradas em escrita ideal. No projeto MorphoSaurus (SCHULZ; HAHN, 2000; PACHECO et al., 2008), que está sendo desenvolvido na PUCPR e que utiliza esta mesma base de dados, são analisados somente os sumários elaborados utilizando a escrita ideal. Esta restrição facilita a identificação dos termos, evitando uma série de validações e correções. Neste trabalho, optou-se por utilizar as narrativas clínicas elaboradas, utilizando tanto letras maiúsculas quando minúsculas, sendo que nenhum padrão foi estabelecido.

3.1.2 Análise dos acrônimos

Os textos dos sumários do HCPA foram elaborados utilizando-se diversos acrônimos. O tratamento de acrônimos é essencial para que o corretor gramatical combinado ao PLN possa classificar sintaticamente os termos de uma frase corretamente.

Neste trabalho foi utilizada uma lista contendo 785 acrônimos, seus significados e o número de vezes que este aparece na base de dados do HCPA utilizada. Esta lista foi elaborada pelo grupo de pesquisa do MorphoSaurus, baseando-se nos sumários de alta HCPA (PACHECO, 2009).

Os acrônimos não se referem somente aos diagnósticos (por exemplo: “vO” significa via oral, “rx – raio x” e “cp - comprimido”). A lista contém alguns acrônimos

que receberam mais de um significado. Nestes casos a preferência foi para a expansão do acrônimo que indicou um diagnóstico ou aquele que apareceu com mais frequência. A lista de acrônimos utilizada é apresentada no Anexo A. A Figura 5 mostra um exemplo de como são expandidos dois acrônimos: ICC e HAS.

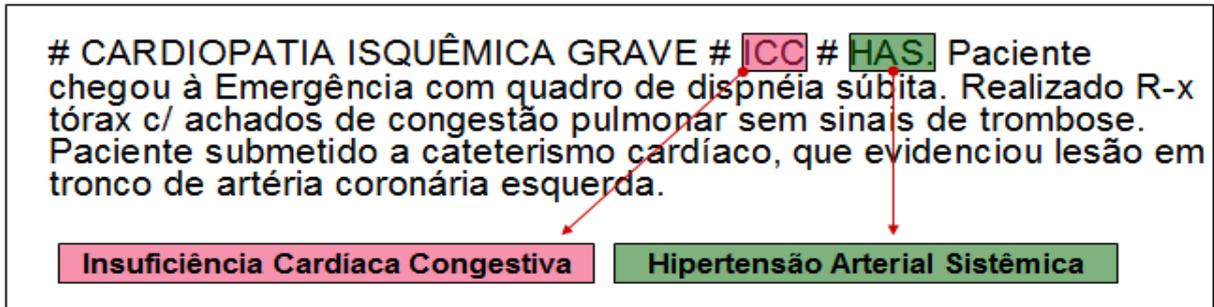


Figura 5 - Expansão de acrônimos

Para o algoritmo identificar um acrônimo na narrativa clínica, foram considerados: palavras com tamanho maior ou igual a dois caracteres até palavras com até quatro caracteres.

3.2 PROCESSAMENTO DAS NARRATIVAS CLÍNICAS

Para o processamento dos textos das narrativas clínicas, foi utilizado o Corretor Gramatical acoplável ao OpenOffice.org – CoGrOO. Esta ferramenta se mostrou adequada ao contexto deste projeto, uma vez que trabalha com textos em português, é *open source* e possui documentação completa, além de ser de fácil utilização.

A escolha pelo CoGrOO também se justifica pelo fato que outras ferramentas encontradas na literatura estão preparadas para o inglês, como é o caso do GATE e que acaba classificando as palavras de uma maneira com que se dificulta a criação de regras para a extração.

É importante ressaltar que o CoGrOO possui embutidos conceitos de Processamento em Linguagem Natural através de outra ferramenta *Open Source*, o OpenNLP, que é responsável por detectar sentenças, *tokens* e *pos-tagging*, que nada mais é do que uma etiqueta de classificação para cada palavra. A ferramenta

reúne, analisa, detecta entidades, ou seja, contém todo o conceito referente a PLN. São realizadas as 5 etapas apresentadas na Figura 6, sendo que, neste caso, a análise pragmática é representada pela etiquetagem (classificação) dos termos.

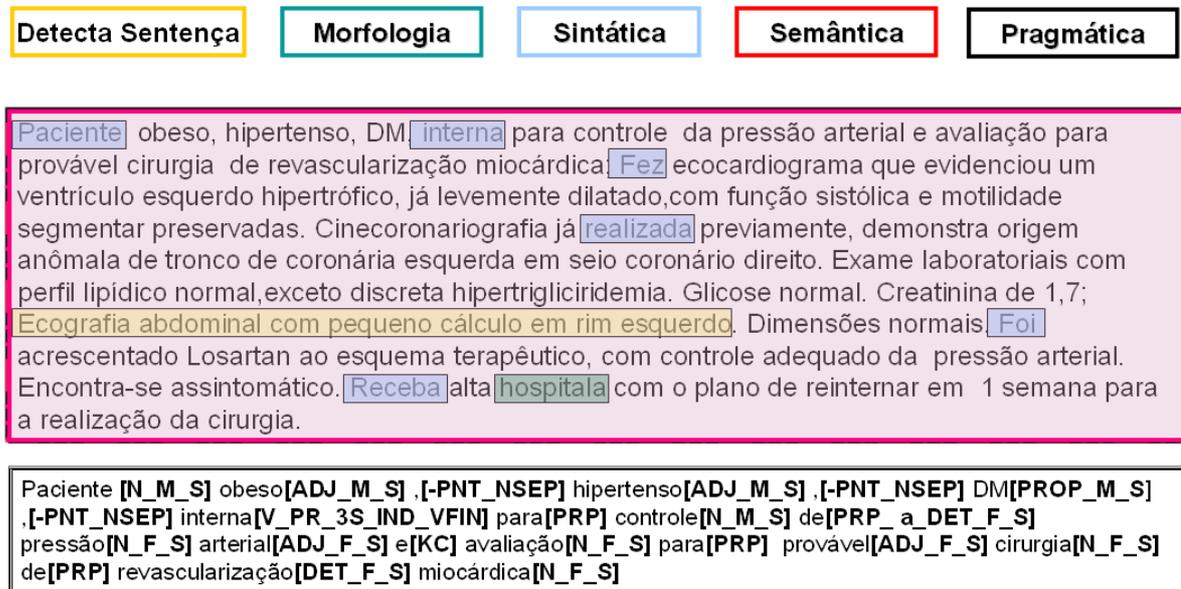


Figura 6 – Funcionamento interno CoGrOO

Embora o CoGrOO realize as cinco etapas – detecção de sentença, morfologia, sintática, semântica e pragmática, para a identificação dos diagnósticos, neste trabalho foi necessário utilizar somente as três primeiras etapas, até a função de análise sintática.

O *corpus* utilizado no CoGrOO é baseado em textos jornalísticos. Os resultados dos primeiros testes com os processamentos das narrativas clínicas revelaram a necessidade da utilização de um *corpus* com textos da área de saúde para obter uma classificação mais correta dos termos. O CoGrOO, através da ferramenta OpenNLP nele embutida, já havia sido utilizado na preparação de um *corpus* anotado, baseado em 10% da base de dados dos sumários de alta do HCPA, que foi elaborado pelo grupo do MorphoSaurus. A anotação deste *corpus* se refere à classificação sintática dos termos. Após a classificação realizada pelo CoGrOO, especialistas médicos verificaram manualmente se os termos tinham sido classificados corretamente e fizeram correções, quando necessário. A anotação feita pelos especialistas é que caracteriza o *corpus* clínico anotado. Um exemplo de parte do *corpus* utilizado nesta pesquisa é apresentado no Anexo B. Este *corpus*

anotado foi integrado ao CoGrOO para o processamento das demais narrativas clínicas dos sumários de alta. Para a identificação dos diagnósticos, foram elaboradas regras de extração, baseando-se nos resultados da fase de análise sintática. A metodologia utilizada para a definição destas regras é apresentada na subseção 3.3.

O processamento das narrativas foi realizado primeiramente com 80 sumários, nos quais foram verificados manualmente se os resultados do processamento no CoGrOO e a aplicação das Regras para Extração tinham identificado todos os diagnósticos contidos nestes 80 sumários. Casos alguns dos diagnósticos não fossem identificados, ou termos que não são diagnósticos fossem identificados como tais eram realizados ajustes nas Regras de Extração, objetivando maximizar a identificação dos diagnósticos (verdadeiros positivos) e minimizar a seleção de termos não relacionados aos diagnósticos (falsos-positivos). Isto foi realizado manualmente.

Antes de processar as narrativas, notou-se a necessidade de realizar um tratamento prévio. Este tratamento consistiu em forçar a narrativa a ser totalmente em minúscula, como é demonstrado no item B da Figura 7. Trabalhar com textos totalmente em minúsculas foi uma necessidade, pois o CoGrOO não conseguiu classificar todos os termos de uma frase quando o texto se apresentou em maiúsculas. Para evitar este problema, todos os textos foram transformados em letra minúsculas, utilizando uma função do Java chamada *toLowerCase*. A seguir, devem ser eliminados do texto pontuações e caracteres especiais, tais como “#” e “\$”, conforme apresentado no item C da Figura 7.

A) Narrativa para ser analisada

PACIENTE OBESO HIPERTENSO, DM, INTERNA PARA CONTROLE DE PRESSÃO ARTERIAL E AVALIAÇÃO PARA PROVÁVEL CIRURGIA DE REVASCULARIZAÇÃO MIOCÁRDICA.

(CoGrOO não consegue classificar todas as palavras quando totalmente em maiúsculo)

B) Texto passa a ser totalmente em minúsculas

paciente obeso hipertenso, dm, interna para controle de pressão arterial e avaliação para provável cirurgia de revascularização miocárdica.

C) Remover pontuações e caracteres especiais antes da classificação: , . |? / \# \$ etc.

paciente obeso hipertenso dm interna para controle de pressão arterial e avaliação para provável cirurgia de revascularização miocárdica

Neste caso, removeu . (ponto) e removeu , (vírgulas)

Figura 7 – Preparação da narrativa

No Quadro 5, são apresentados exemplos dos tipos de narrativas que foram testados. Eles estão separados por tipo de escrita, casos e classificadas, de acordo com a etiquetagem realizada pelo CoGrOO. A classificação é representada entre colchetes (por exemplo: [N_F_S] – Substantivo, Feminino e Singular).

Caso	Frase Exemplo	Resultado <i>TagSet</i>
“Escrita Ideal”	O paciente tem hipertensão.	O[DET_M_S] paciente[N_M_S] tem[V_PR_3S_IND_VFIN] hipertensão [N_F_S].[PNT_ABS]
“Escrita Ideal”	Paciente hipertenso , internou com quadro de dor anginosa.	Paciente[N_M_S] hipertenso [ADJ_M_S] ,[PNT_NSEP] internou[V_PS_3S_IND_VFIN] com[PRP] quadro[N_M_S] de[PRP] dor[N_F_S] anginosa[ADV em[PRP] .[PNT_ABS]
Escrito totalmente em maiúsculas	O PACIENTE TEM HIPERTENSÃO.	O[DET_M_S] PACIENTE[PROP_M_S] .[PNT_ABS]
“Escrita Ideal” com mais de um Diagnóstico	O paciente tem hipertensão e é diabético.	O [DET_M_S] paciente[N_M_S] tem[V_PR_3S_IND_VFIN] hipertensão [N_F_S] e[KC] é[V_PR_3S_IND_VFIN] diabético [ADJ_M_S]

Quadro 5 – Sumários de Alta x CoGrOO

O significado de cada etiqueta definida pelo CoGrOO é apresentado no Quadro 6.

[N_M_S]	Substantivo (<i>Noun</i>), masculino, singular.
[N_F_S]	Substantivo, feminino, singular.
[N_M_P]	Substantivo, masculino, plural.
[ADJ_M_S]	Adjetivo, masculino, singular.
[ADJ_F_S]	Adjetivo, feminino, singular.
[KC]	Conjunção Coordenada.
[PRP]	Preposição.
[V_PS_3S_IND_VFIN]	Verbo, perfeito simples, 3a pessoal do singular, indicativo, finito.
[PNT_ABS]	Ponto final.
[PNT_NSEP]	Vírgula.

Quadro 6 – Significado das etiquetas utilizadas

Nos casos classificados como “escrita ideal”, no Quadro 5, foi possível notar que não existe um padrão de classificação para o diagnóstico. No primeiro caso, este foi classificado como substantivo e, no segundo, como adjetivo.

O caso “escrito totalmente em maiúsculas” representa a dificuldade que o CoGrOO tem para classificar textos totalmente em maiúsculo, pois ele não consegue fazer a classificação até o final da frase.

3.3 DEFINIÇÃO DAS REGRAS PARA EXTRAIR DIAGNÓSTICOS

A metodologia para definir regras para extrair diagnósticos foi elaborada por um procedimento experimental e é baseada no modelo de regras sintáticas proposto por Bui et al. (2010), embora os autores deste método tenham utilizado estas regras em outro contexto, para extração da relação entre medicamentos e o HIV em resumos de artigos.

Como não foi encontrada a definição exata de diagnóstico na literatura, adotou-se a opção de que um diagnóstico é aquela palavra que tem correspondente morfológico na CID-10, salvo as exceções que médicos indicaram durante as

entrevistas, como, por exemplo, dislipidemia, que possui o termo lipidemia na CID-10 ((E78) Distúrbios do metabolismo de lipoproteínas e outras Lipidemias).

Em um primeiro momento, regras foram criadas, sem expandir os acrônimos. Esta opção levou a uma grande variação de opções de regras em que não foi possível encontrar um padrão para extrair os diagnósticos, necessitando, em alguns casos de até 4 regras para possibilitar a identificação de um único diagnóstico, enaltecendo, assim, a necessidade de tratá-los (Figura 6 – Etapa D).

Com os acrônimos expandidos, foi possível realizar o processamento das narrativas e analisar as classificações dos termos (Figura 6 – Etapa E, análise pragmática).

Com as palavras classificadas através do processamento e do CoGrOO, termos referentes aos diagnósticos, como hipertensão e diabetes, foram analisados, assim como seus antecessores e posteriores (Figura 8 – Etapa F). Hipertensão recebeu mais de uma regra, pois pode estar escrito de diferentes formas como hipertensão (Substantivo, Feminino, Singular – [N_F_S]), hipertensa (Adjetivo, Feminino, Singular – [ADJ_F_S]) e hipertensiva (Adjetivo, Feminino, Singular – [ADJ_F_S]).

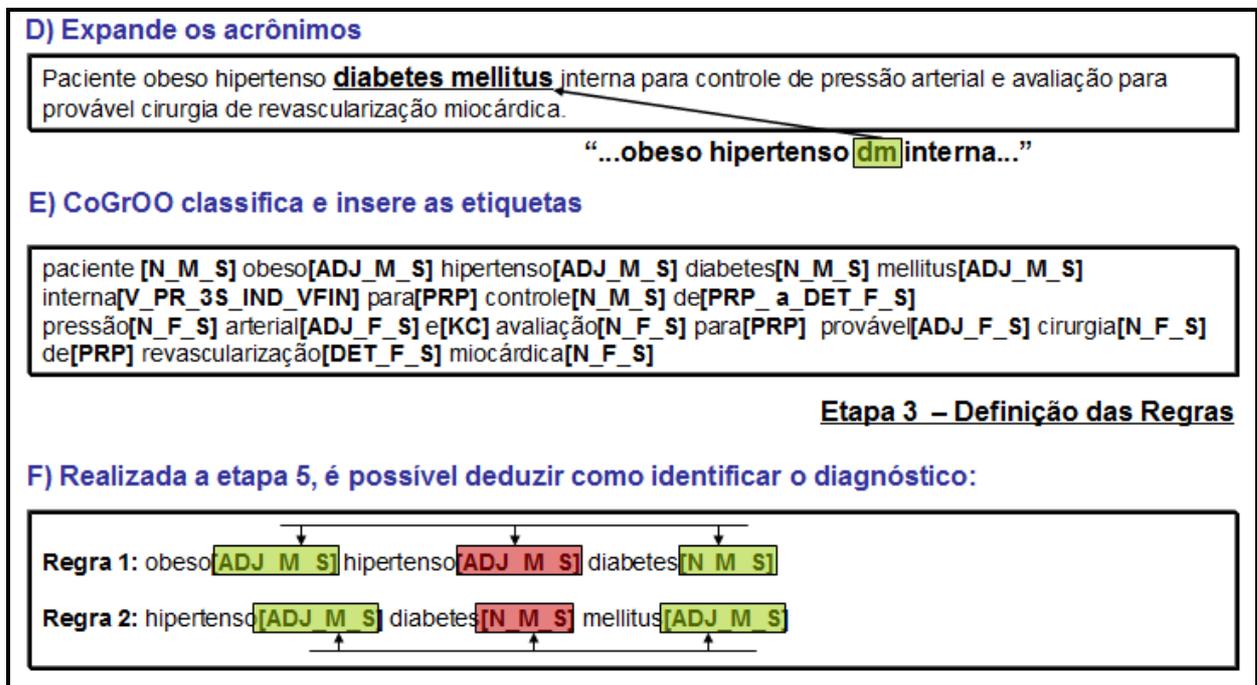


Figura 8 – Identificação de regras

Inicialmente, foram elaboradas 8 regras, para extrair os diagnósticos, que são apresentadas no Quadro 7. Estas regras foram testadas em 50 sumários. Após uma avaliação manual, observou-se uma precisão de 91,18% no acerto da extração de diagnósticos.

Regra	Descrição	Antecessor	Diagnóstico	Posterior
1	Diagnóstico adjetivo, masculino e singular. Antecessor nominal, masculino e singular. Posterior nominal, masculino e singular.	Paciente N_M_S	Hipertenso ADJ_M_S	Portador N_M_S
2	Diagnóstico adjetivo, masculino e singular. Antecessor nominal, masculino e singular. Posterior conjunção coordenativa.	Paciente N_M_S	Hipertenso ADJ_M_S	E KC
3	Diagnóstico nominal, feminino e singular. Antecessor nominal, masculino e singular. Posterior adjetivo, feminino e singular.	Paciente N_M_S	Cardiopata N_F_S	Sintomática ADJ_F_S
4	Diagnóstico adjetivo, masculino e singular. Antecessor nominal, masculino e singular. Posterior adjetivo, masculino e singular.	Paciente N_M_S	Hipertenso ADJ_M_S	Portador ADJ_M_S
5	Diagnóstico nominal, feminino e singular. Antecessor preposição. Posterior adjetivo, feminino, plural.	Com PRP	Angina N_F_S	Pectoris ADJ_F_P
6	Diagnóstico nominal, feminino e singular. Antecessor preposição. Posterior adjetivo, feminino, singular.	Com PRP	Angina N_F_S	Estável ADJ_F_S
7	Diagnóstico nominal, feminino e singular. Antecessor nominal, masculino e singular. Posterior adjetivo, masculino e singular.	Paciente N_M_S	Cardiopatia N_F_S	Isquêmico ADJ_M_S
8	Diagnóstico adjetivo, masculino e singular. Antecessor nominal, masculino e singular. Posterior adjetivo, feminino e singular.	Paciente N_M_S	Hipertensa ADJ_F_S	Tabagista ADJ_F_S

Quadro 7 – Regras iniciais para extração de diagnósticos

Com a utilização destas 8 regras, foram encontrados 62 diagnósticos. Alguns destes sumários continham mais de um diagnóstico, sendo que 33 narrativas não continham nenhum diagnóstico descrito. Nestes casos, nos sumários são descritos somente ações ou resultados de procedimentos (por exemplo, “realizou angioplastia com sucesso”).

Os diagnósticos não identificados representaram 8,3%, sendo que quase metade (4%) desta não identificação foi causada por erros de ortografia ou digitação, que impossibilitaram classificar a palavra como um diagnóstico.

Utilizando o mesmo processo descrito anteriormente, foram elaboradas mais 24 regras, com o objetivo de tentar contemplar os diagnósticos que não foram identificados, utilizando-se somente as 8 regras iniciais, resultando em 32 regras.

Um novo teste com os mesmos 50 sumários, aplicando-se as novas regras, foi realizado para avaliar se as novas inclusões permitiriam a identificação de todos os diagnósticos contidos nestes sumários. Na Figura 9, é apresentado um esquema de como estes testes foram realizados. Toda vez que uma nova regra era inserida ou modificada, ou era aplicado nos textos um dos itens de 1 a 4, era necessário reavaliar todos os 50 sumários, pois, qualquer uma destas alterações poderia adicionar novos falsos diagnósticos ou deixar de identificar um termo como diagnóstico.

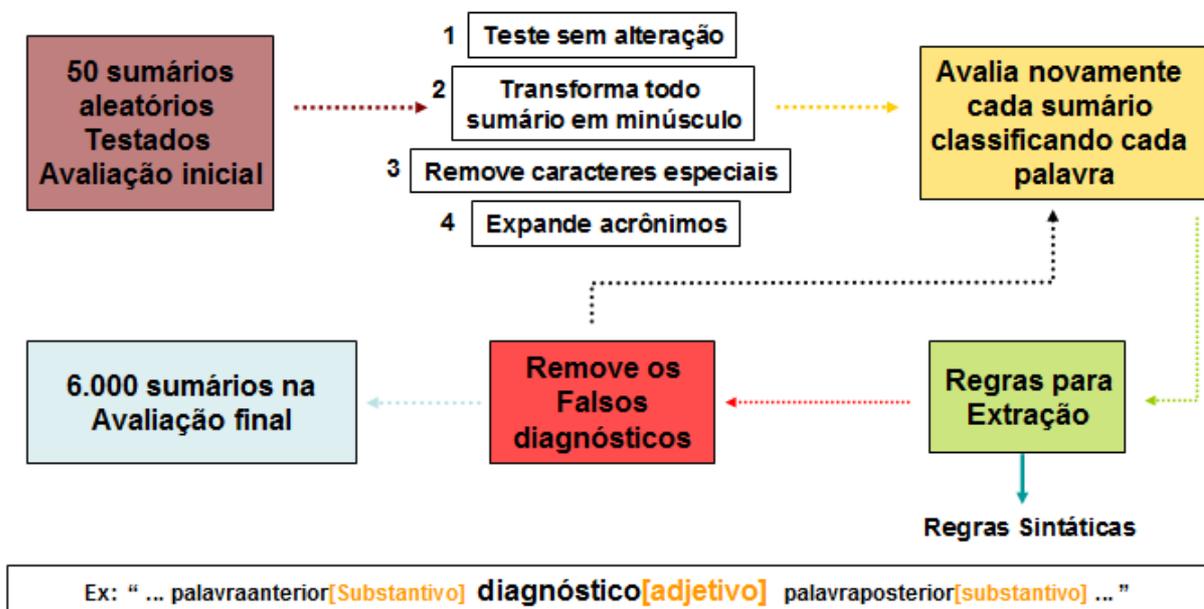


Figura 9 – Ciclo para definição de regras para identificar diagnósticos

Após esta segunda avaliação, todas as 6.000 narrativas clínicas foram processadas com todas as regras criadas. Ao final deste processamento, percebeu-se a necessidade de remover caracteres especiais (apresentado na Etapa C da Figura 7), pois estes alteravam a classificação das palavras próximas, prejudicando a exatidão do método utilizado para identificar os diagnósticos.

3.4 CORRELAÇÃO DE DIAGNÓSTICOS DA CARDIOLOGIA PARA O CID-10

A correlação ou mapeamento dos diagnósticos identificados foi realizado através da semelhança morfológica entre a classificação de um termo do sumário como sendo um diagnóstico com a descrição de um diagnóstico na CID-10.

Cada diagnóstico recebe um grupo e uma categoria, podendo haver mais de uma opção de CID-10 para um mesmo diagnóstico, por exemplo: Hipertensão em um sumário poderá ser classificada em CID-10 como hipertensão essencial, renovascular ou secundária, recebendo códigos CID-10 diferentes.

A Figura 10 mostra algumas das opções de mapeamento para o diagnóstico “hipertensão” e a associação com outras formas de escrevê-lo (“hipertenso”, “hipertensa”).

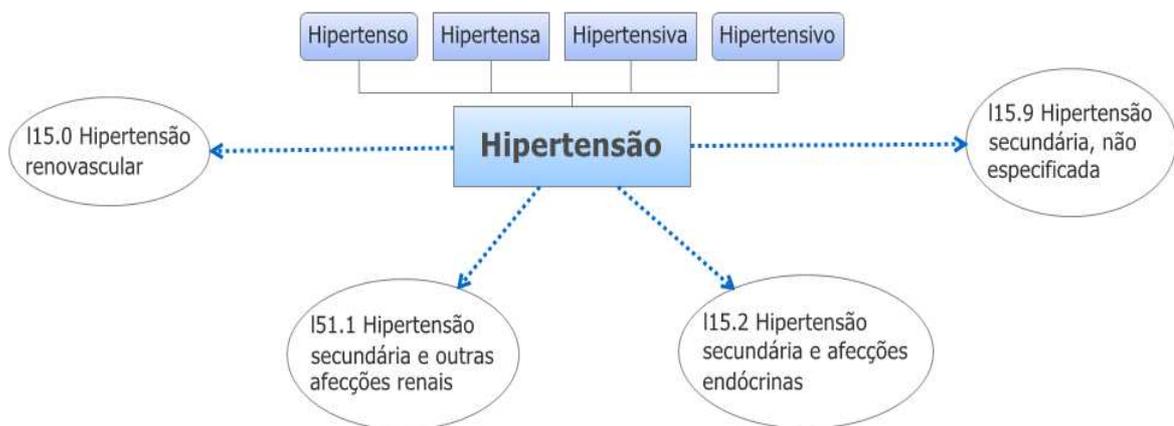


Figura 10 – Correlação para hipertensão

Para alguns diagnósticos que não possuem um correspondente morfológico na descrição da CID-10, o mapeamento foi definido com o auxílio de especialistas. Um exemplo disto é o diagnóstico dislipidemia, que é correlacionado com o código E78 - Distúrbios do metabolismo de lipoproteínas e outras lipidemias.

O grupo e categoria são sugeridos em ordem alfabética, porém, quando a sugestão se torna a mais comum – “genérica” – e se encaixa em vários diagnósticos, foi colocada na última posição da lista de diagnósticos CID-10 apresentados para o mapeamento. Por exemplo, para o diagnóstico Pericardite, o grupo poderia ser I30-I52 (Outras formas de doença do coração), porém existem

relações que podem ser mais específicas como: I48 “Flutter” e fibrilação atrial ou I44 (Bloqueio atrioventricular e do ramo esquerdo). Nesta caso, a opção “Outras formas de doença do coração” é colocada na última posição da lista de diagnósticos CID-10.

Caso não exista nenhuma categoria ou grupo, a ferramenta NOTTUS, implementada para a utilização do método de extração proposto nesta pesquisa, permite que o próprio usuário preencha manualmente o diagnóstico mais adequado.

3.5 ANÁLISE DOS DIAGNÓSTICOS ENCONTRADOS

Para analisar as regras de identificação de diagnósticos e suas correlações, foram escolhidos dez sumários de alta com características e resultados diferentes, representados por vários modelos de preenchimento: respeitando a escrita, texto totalmente em maiúsculo, presença ou não de acrônimos, entre outros exemplos.

A escolha dos dez sumários foi feita aleatoriamente, porém foram descartados desde o início os sumários que apresentavam frases curtas e que descreviam apenas um procedimento (por exemplo, retirado da base de dados: “internou para implante de *stent*”).

Após a seleção dos sumários, o método proposto foi aplicado em cada um deles. A seguir, foi elaborado um questionário, apresentado no Anexo C, contendo o texto dos 10 sumários, os diagnósticos encontrados e as correlações para a CID-10 de cada um dos textos. O objetivo deste questionário foi possibilitar que os médicos que analisaram os resultados do método pudessem anotar se todos os diagnósticos foram encontrados e indicar se faltava algum diagnóstico, além de verificar se os CIDs relacionados estavam corretos.

A função dos médicos foi comparar se os diagnósticos contidos em uma narrativa clínica também foram encontrados pelo método. Para isso, cada médico analisou o sumário de alta, encontrou os diagnósticos na narrativa clínica e os comparou com os diagnósticos apresentados pela ferramenta que foi desenvolvida, NOTTUS. Para a correlação dos diagnósticos, foi necessária uma verificação para cada diagnóstico encontrado, analisando se a ferramenta indicou o grupo e a categoria correta.

3.6 DESENVOLVIMENTO DE UMA FERRAMENTA PARA INTEGRAÇÃO AO PEP

Nesta fase, foi desenvolvida uma ferramenta denominada NOTTUS, que inclui o método proposto nesta pesquisa. Assim como o CoGrOO, o NOTTUS é baseado na linguagem de programação Java. Desta forma, em PEPs que foram codificados em Java, basta informar uma relação no banco de dados entre o número do sumário e o próprio sumário de alta, possibilitando 100% de integração com a ferramenta. Para sistemas codificados em outras linguagens de programação, a ferramenta também funciona e consegue importar informações da base, porém será um executável à parte. Ao invés do usuário utilizar a nova ferramenta através do PEP, deverá utilizar em paralelo.

O desenvolvimento da ferramenta foi realizado em etapas, que consistem em desenvolver e avaliar resultados, reformulando regras, caso algum diagnóstico não fosse encontrado.

No processamento das narrativas clínicas foram observadas as classificações dos diagnósticos como hipertensão (hipertensa, hipertenso, hipertensiva, hipertensivo), diabetes (diabético, diabética), cardiopatia (cardiopata) e angina, nas quais foi possível identificaram algumas semelhanças: palavras e classificações similares que antecediam ou sucediam um diagnóstico.

Os Quadros 8, 9 e 10 apresentam exemplos da necessidade de regras diferentes para diagnósticos iguais. Os textos de exemplo dos quadros contêm cardiopata e, para o modo flexivo, cardiopatia. Estes textos são partes reais retiradas dos sumários de alta e do resultado do processamento das narrativas na ferramenta NOTTUS. Existem outros diagnósticos nestes textos, porém o objetivo dos quadros é apresentar as diferentes classificações encontradas para cardiopata e cardiopatia, além do resultado do funcionamento do algoritmo que executa o método de identificação dos diagnósticos e do NOTTUS.

Entrada: Paciente cardiopata isquêmico submetido previamente a estudo eletrofisiológico

Classificação: paciente_N_M_S cardiopata_N_F_S isquêmico_ADJ_M_S
submetido_V_PCP_M_S previamente_ADV_a_PRP_estudo_N_M_S_
eletrofisiológico_ADJ_M_S_

Regra: _N_M_S_ _N_F_S_ _ADJ_M_S_ onde: _N_M_S_ (Termo Antecessor – Substantivo Masculino singular) _N_F_S_ (Diagnóstico – Substantivo Feminino Singular) _ADJ_M_S_ (Termo Posterior – Adjetivo Masculino Singular)

Como o algoritmo processou: **1)** procurou a classificação _N_F_S_ que foi definido previamente como possível diagnóstico. **2)** encontrou _N_F_S_ na segunda palavra classificada. **3)** parou. **4)** verificou a classificação do termo anterior e encontrou _N_M_S_. **5)** A classificação da palavra anterior faz parte da regra, verifique a palavra posterior ao diagnóstico. **6)** Encontrou _ADJ_M_S_, faz parte da regra. **7)** Defina a palavra que recebeu a classificação _N_F_S_ com a classificação da palavra antecessor _N_M_S_ e a palavra posterior classificada _ADJ_M_S_ como diagnóstico. **8)** Continuou a procurar classificação de termos que possam ser diagnóstico, comparando sempre a classificação do termo antecessor e do termo posterior.

Quadro 8 – Classificação diagnóstico cardiopata isquêmico

Entrada: pct has cardiopatia isquêmica angina.

Classificação: Paciente_N_M_S_ hipertensão_N_F_S_ arterial_ADJ_F_S_ **sistêmica_ADJ_F_S_ cardiopatia_N_F_S_ isquêmica_ADJ_F_S_ angina_N_F_S_ hipertensão**

Regra e processamento do algoritmo idem ao Quadro 10.

Quadro 9 – Classificação diagnóstico cardiopatia isquêmica

Entrada: Paciente diabetica hipertensa cardiopata isquemica interna para realização de stent

Classificação: paciente_N_F_S_ diabetica_ADJ_F_S_ **hipertensa_ADJ_F_S_ cardiopata_N_F_S_ isquemica_ADJ_F_S_ interna_ADJ_F_S_ para_PRP_ realização_N_F_S_ de_PRP_ stent_N_M_S_**

Regra: _ADJ_F_S_ _N_F_S_ _ADJ_F_S_ onde: _ADJ_F_S_ (Termo Antecessor – Adjetivo feminino singular) _N_F_S_ (Diagnóstico – Substantivo Feminino Singular) _ADJ_F_S_ (Termo Posterior – Adjetivo Feminino Singular)

Como o algoritmo processou: **1)** procurou a classificação _N_F_S_ que foi definido previamente como possível diagnóstico. **2)** encontrou _N_F_S_ na segunda palavra classificada. **3)** parou. **4)** verificou a classificação do termo anterior e encontrou _ADJ_F_S_. **5)** A classificação da palavra anterior faz parte da regra, verifique a palavra posterior ao diagnóstico. **6)** Encontrou _ADJ_F_S_, faz parte da regra. **7)** Defina a palavra que recebeu a classificação _N_F_S_ com a classificação da palavra antecessor _N_M_S_ e a palavra posterior classificada _ADJ_M_S_ como diagnóstico. **8)** Continuou a procurar classificação de termos que possam ser diagnóstico, comparando sempre a classificação do termo antecessor e do termo posterior.

Quadro 10 – Classificação diagnóstico cardiopata isquêmica

É possível notar, no Quadro 9, que o sistema já se antecipa e, antes da classificação, expande os acrônimos “pct” e “has”, permitindo que o CoGrOO faça a classificação mais adequada para os termos.

Quando um termo antecessor ou posterior não está na regra, a palavra central que o algoritmo analisa é automaticamente descartada, não sendo considerado como um diagnóstico.

Os falsos diagnósticos são termos identificados como diagnósticos, mas que não são relacionados. O Quadro 11 apresenta alguns exemplos destes falsos diagnósticos. É possível notar que alguns destes termos poderiam ser corrigidos, utilizando-se um corretor ortográfico.

Ñão, cirúrgico, qualit , relmínimo, rapamicina, arteriog, nível, diagnóstico, devido, planejamento, bilateral, externa, empírica, cessação, oclusão, ticlopidina, recanalização, bilateralmente, revisão, claudicação, otização, plano

Quadro 11 – Falsos diagnósticos

A ferramenta NOTTUS foi simplificada o máximo possível para ser utilizada pelo usuário. Para encontrar diagnósticos em um determinado sumário de alta, basta escolher o número do sumário (ou digitar um novo) e clicar em um botão “procurar diagnóstico”. Então, o sistema automaticamente verifica e sugere os possíveis grupos e categorias CID-10 para que o médico confirmar o diagnóstico.

Caso a ferramenta não encontre a opção exata, existe a opção de inserir uma nova sugestão de diagnóstico, em que o médico digita as informações de acordo com o padrão CID-10.

Os administradores da ferramenta podem inserir novas regras em um modelo prático contido em um arquivo de texto, tornando a ferramenta flexível, além de também ter acesso a uma lista de palavras que excluem falsos diagnósticos dos resultados. Porém, é importante ressaltar que é necessário cuidado para a inclusão de novas regras e que isto deve ser avaliado pelos especialistas médicos. Cada nova regra criada para identificação de novos diagnósticos aumenta a possibilidade do aparecimento de novos falsos diagnósticos.

4 RESULTADOS

O principal resultado desta pesquisa é o método para extração dos diagnósticos. Além dele, as regras para extrair o diagnóstico, o mapeamento para a CID-10, a análise realizada pelos médicos e a ferramenta NOTTUS compõem os resultados.

4.1 MÉTODO PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS

O método elaborado é composto de 5 etapas:

1. Transformação do texto do sumário de alta totalmente em letras minúsculas.
2. Remoção de caracteres especiais e pontuação.
3. Expansão dos acrônimos.
4. Classificação dos termos (etiquetas), utilizando o CoGrOO com um corpus da área de saúde.
5. Aplicação das regras de extração.

O método é composto por uma última etapa, que é a Correlação dos Diagnósticos para a CID-10.

4.2 REGRAS PARA EXTRAIR DIAGNÓSTICOS

No Quadro 12, são apresentadas as 32 regras construídas para extração dos diagnósticos dos sumários de alta. As regras estão em negrito. Os termos antecessores e sucessores ao diagnóstico numa frase também são apresentados com suas classificações, com o intuito de exemplificar o método elaborado. A

definição das etiquetas utilizadas para classificar os termos foi apresentada no Quadro 6 (subseção 3.2).

Antecessor	Diagnóstico	Sucessor
1) paciente_N_M_S	hipertenso_ADJ_M_S	cardiopata_N_M_S
2) paciente_N_M_S	hipertenso_ADJ_M_S	e_KC
3) diabético_N_M_S	cardiopata_ADJ_M_S	isquêmico_ADJ_M_S
4) paciente_N_M_S	cardiopata_N_F_S	isquêmico_ADJ_M_S
5) paciente_N_M_S	cardiopata_N_F_S	isquêmica_ADJ_F_S
6) paciente_N_M_S	hipertenso_ADJ_M_S	com_PRP
7) paciente_N_M_S	hipertenso_ADJ_M_S	cardiopata_N_F_S
8) paciente_N_F_S	hipertensa_ADJ_F_S	diabética_ADJ_F_S
9) paciente_N_F_S	hipertensa_ADJ_F_S	e_KC
10) paciente_N_F_S	diabetico_N_M_S	dislipidêmico_ADJ_M_S
11) paciente_N_F_S	hipertensa_ADJ_F_S	diabética_ADJ_F_S
12) anos_N_M_P	dislipidêmico_ADJ_M_S	diabético_N_M_S
13) anos_N_M_P	hipertenso_ADJ_M_S	cardiopata_N_F_S
14) com_PRP	angina_N_F_S	há_PRP
15) de_PRP	hipertensão_N_F_S	renovascular_ADJ_F_S
16) de_PRP	endocardite_N_F_S	em_PRP
17) de_PRP	aneurisma_N_M_S	dissecante_ADJ_M_S
18) com_PRP	angina_N_F_S	classe_N_F_S
19) com_PRP	reestenose_N_F_S	intra-stent_N_M_S
20) isquemico_ADJ_M_S	diabetico_N_M_S	com_PRP
21) hipertenso_ADJ_M_S	cardiopata_N_M_S	isquêmico_ADJ_M_S
22) hipertenso_ADJ_M_S	cardiopata_N_F_S	isquêmico_ADJ_M_S
23) diabética_ADJ_F_S	hipertensa_ADJ_F_S	dislipidêmica_ADJ_F_S
24) hipertensa_ADJ_F_S	dislipidêmica_ADJ_F_S	cardiopata_N_F_S
25) diabética_ADJ_F_S	dislipidêmica_N_F_S	internou_V_PS_3S_IND_VFIN

26) diabética_N_F_S	hipertensa_ADJ_F_S	internou_V_PS_3S_IND_VFIN
27) apresenta_V_PR_3S_IND_VFIN	estenose_N_F_S	crítica_ADJ_F_S
28) hipertensa_ADJ_F_S	cardiopata_N_F_S	isquêmica_ADJ_F_S
29) e_KC	estenose_N_F_S	de_PRP
30) e_KC	dislipidêmica_ADJ_F_S	de_PRP
31) anos_N_M_P	diabética_N_F_S	hipertensa_ADJ_F_S
32) paciente_N_M_S	cardiopata_N_F_S	isquêmica_ADJ_F_S

Quadro 12 – Regras para extração de diagnósticos

Com a aplicação das 32 regras, foi possível identificar 52.379 termos como diagnósticos, incluindo repetições. Por exemplo, o termo hipertensão foi encontrado 20 vezes, sendo que todas estas ocorrências foram consideradas. Apenas 2,5% destes termos correspondiam a falsos diagnósticos.

É possível observar, no Quadro 12, que as regras para identificar o diagnóstico são caracterizadas sempre por um substantivo (18 vezes) ou adjetivo (14 vezes) no singular. No caso dos substantivos, 14 são do gênero feminino, e 4 do gênero masculino. Nos adjetivos, 7 são do gênero masculino e 7 do gênero feminino.

4.2.1 Aplicação das regras

A aplicação das regras deve seguir a ordem sequencial de 1 a 32, conforme a relação definida no Quadro 12, pois existem regras que identificam um diagnóstico composto e, caso a ordem não seja seguida, estes não serão encontrados. Após o médico inserir a narrativa clínica a ser analisada e ter sido realizado o PLN para a etiquetagem dos termos, cada regra varre toda a narrativa em busca do termo principal que é o diagnóstico. Por exemplo: baseando-se na regra 1, todos os termos classificados como adjetivo terão seus antecessores e sucessores verificados. Se estes representarem a classificação do termo como um Substantivo-Masculino-Singular, o termo é considerado um diagnóstico. Nas próximas regras, os termos que já foram definidos como diagnósticos não são mais analisados.

O Quadro 13 apresenta um exemplo de diagnóstico composto: cardiopatia isquêmica.

<p>1) Sumário analisado</p> <p>PCT CARDIOPATIA ISQUEMICA, DM2, ACTP + STENT CD. INTERNA P/ ACTP + STENT. PROCEDIMENTO REALIZADO SEM INTERCORRENCIAS E COM SUCESSO ANGIOGRÁFICO. RECEBE ALTA EM USO DAS MEDICAÇÕES ABAIXO E COM RETORNO AO SEU MÉDICO PARA SEGUIMENTO.</p> <p>2) Classificação dos diagnósticos e complemento do diagnóstico</p> <p>2.1) pct_N_M_S_ cardiopatia_N_F_S_ isquemica_ADJ_F_S_</p> <p>2.2) cardiopatia_N_F_S_ isquemica_ADJ_F_S_ diabetes_N_M_S_</p> <p>2.3) isquemica_ADJ_F_S_ diabetes_N_M_S_ melitus_ADJ_M_P_</p> <p>3) Diagnósticos identificados:</p> <p>Cardiopatia isquêmica, Diabetes</p> <p>Conclusão:</p> <p>Se a regra demonstrada no item 2.2) _N_F_S_ _ADJ_F_S_ _N_M_S_ estiver no documento de texto antes da regra 2.1) _N_M_S_ _N_F_S_ _ADJ_F_S_ o resultado da extração seria Isquêmica Cardiopatia.</p>

Quadro 13 – Importância da ordem das regras para extrair diagnóstico

4.3 CORRELAÇÃO DE DIAGNÓSTICOS

A correlação dos diagnósticos com a CID-10 foi feita através da semelhança morfológica entre as palavras. Os médicos entrevistados sugeriram na análise que diagnósticos contidos na CID-10, como, por exemplo “outra formas de doença do coração”, fossem ordenados por último, para exigir que o especialista procure e relacione melhor o diagnóstico encontrado pela ferramenta. A sugestão foi aceita e implementada.

A correlação feita por semelhança morfológica pode não ser a melhor forma para indicar a classificação mais precisa dos diagnósticos encontrados, porém,

trata-se de uma segunda contribuição da pesquisa. Essa contribuição pode ser otimizada utilizando-se algoritmos de aprendizagem, para que conforme a utilização do NOTTUS, a sugestão da CID-10 seja cada vez mais próxima do ideal.

Quando a correlação não foi encontrada na CID-10, sugestões de especialistas foram aceitas, como no caso de dislipidemia - (E78) Distúrbios do metabolismo de lipoproteínas e outras lipidemias.

4.4 ANÁLISE PELOS ESPECIALISTAS DOS DIAGNÓSTICOS IDENTIFICADOS

No Quadro 14, é apresentado um resumo das sugestões e modificações sugeridas pelos três médicos durante a análise dos 10 sumários de alta com características diferentes. Também foram feitos comentários em relação à utilização e resultados obtidos com o NOTTUS.

Três tipos de diagnósticos identificados pela ferramenta não foram considerados pelos médicos: cardiopata, estenose e trombose - o último, por ser uma negação (ausência de trombose), e os dois primeiros por representarem um termo muito geral, podendo até não significar um diagnóstico.

Os médicos também identificaram dois diagnósticos que não foram considerados pela ferramenta: pressão arterial elevada, que deveria ter sido considerada como diagnóstico de hipertensão arterial (Anexo C, Caso 2), e abscesso, para o qual não foi criada uma regra especial e que deve ser adicionado. No caso do diagnóstico “abscesso” apresentado no questionário, a ferramenta não identificou como diagnóstico pelo fato do *corpus* não ter feito a correção de uma palavra anterior ao diagnóstico, invalidando a regra.

Origem	Descrição	Sugestão	Correção
Identificação de Diagnósticos	Cardiopata não é um diagnóstico ou é um diagnóstico sindrômico, ou seja, pertence a um grande grupo.	Cardiopata deverá ser considerado diagnóstico somente quando acompanhado da palavra isquêmica. Ex: Cardiopata isquêmico, cardiopatia isquêmica.	Realizada uma validação em que é verificado se a palavra posterior ao cardiopata é isquêmica.
Identificação de Diagnósticos	Deverá ser especificado o local da estenose.	Incluir uma verificação para identificar se é uma artéria do coração ou renal, pois somente estenose é um termo muito amplo.	A remoção de uma das regras da metodologia para extração de diagnósticos resolveu a sugestão.
Identificação de Diagnósticos	Não identificou abscesso.	Abscesso é um diagnóstico e precisa ser identificado.	Não havia uma regra para identificação de abscesso. A correção para a sugestão foi criar uma nova regra para identificar o diagnóstico.
Identificação de Diagnósticos	Identificou um falso-positivo como diagnóstico.	Frase: “sem sinais de trombose”, onde trombose foi identificada como diagnóstico.	A ferramenta não foi preparada para todo tipo de negação de diagnósticos, no momento, sem correção.
Mapeamento CID-10	Colocar em última posição a CID-10 que pode receber mais de um tipo de diagnóstico.	CID-10 que representa, por exemplo: “[I30-I52] - Outras formas de doença do coração”, pode se encaixar em mais de um diagnóstico. Para evitar a generalização, a sugestão é oferecer outras opções antes.	A sugestão foi corrigida durante as entrevistas com os médicos.
Comentários Adicionais do Especialista 1	Análise semântica de termos no sumário.	A análise semântica poderia relacionar termos como pressão arterial elevada para ser indicada como hipertensão.	A sugestão ficou adotada para ser realizada na continuação do projeto, em trabalhos futuros.
Comentários Adicionais do Especialista 2	Análise semântica de termos no sumário.	A análise semântica poderia relacionar procedimentos para identificar diagnósticos. Assim como pressão arterial elevada ser indicada como hipertensão, Ecodopler relacionar com estenose e hemoglobina glicada com diabetes.	A sugestão ficou para ser realizada na continuação do projeto, em trabalhos futuros.
Comentários Adicionais do Especialista 3	A exclusão de resultados de exames pode ajudar a evitar os falsos diagnósticos.	A sugestão foi relacionar termos de exames junto com termos de negação antes da verificação de diagnóstico.	A negação de diagnósticos em narrativas clínicas deverá ser realizada em trabalhos futuros utilizando a análise semântica.

Quadro 14 – Sugestões e correções da validação com médicos

4.5 NOTTUS – NOVO SISTEMA PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS

A nova ferramenta permite extrair diagnósticos de narrativas clínicas, possibilitando ao médico estruturar ou descartar a sugestão, além de poder verificar um prontuário que já tenha sido processado pela ferramenta. Com o número do prontuário, é possível auxiliar na integração do NOTTUS em prontuários eletrônicos e sistemas de informação em saúde.

A Figura 11 apresenta a tela que o médico utiliza para encontrar os diagnósticos, utilizando-se um exemplo de sumário de alta escrito totalmente em maiúsculas e com a presença de acrônimos. Para realizar a identificação dos diagnósticos e a correlação com a CID-10, o primeiro passo foi o preenchimento do campo “número do sumário” – neste caso, o sumário já existia na base de dados do sistema. Este campo deverá ser conectado com a base de dados em que o sistema NOTTUS deverá atuar.

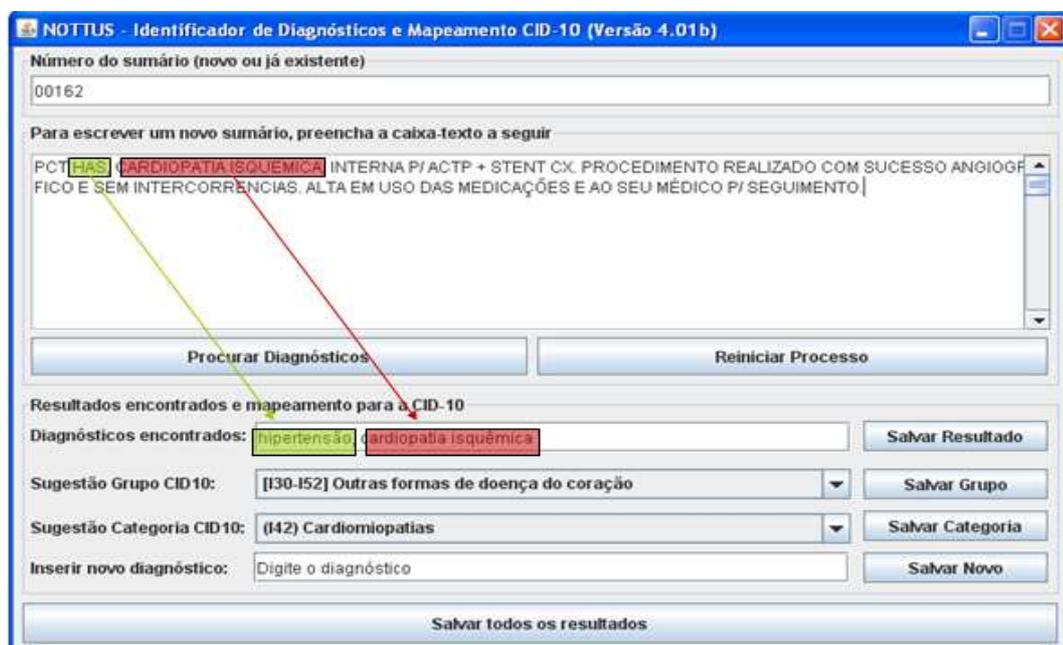


Figura 11 – NOTTUS, funcionamento

Após a conexão do campo da base de dados referente ao número do sumário, deve ser associado o campo que corresponde ao sumário de alta existente na “caixa-texto” do NOTTUS. Associando estes dois campos, as correlações são feitas automaticamente. Basta enviar as informações dos campos para o botão

“Procurar Diagnóstico” e a ferramenta está integrada, de forma a identificar diagnósticos em narrativas clínicas. Os botões do tipo “salvar” são opções para guardar os resultados de diagnósticos encontrados, grupo e categoria CID-10 ou um novo diagnóstico. Estes podem ser novos campos na base de dados do prontuário.

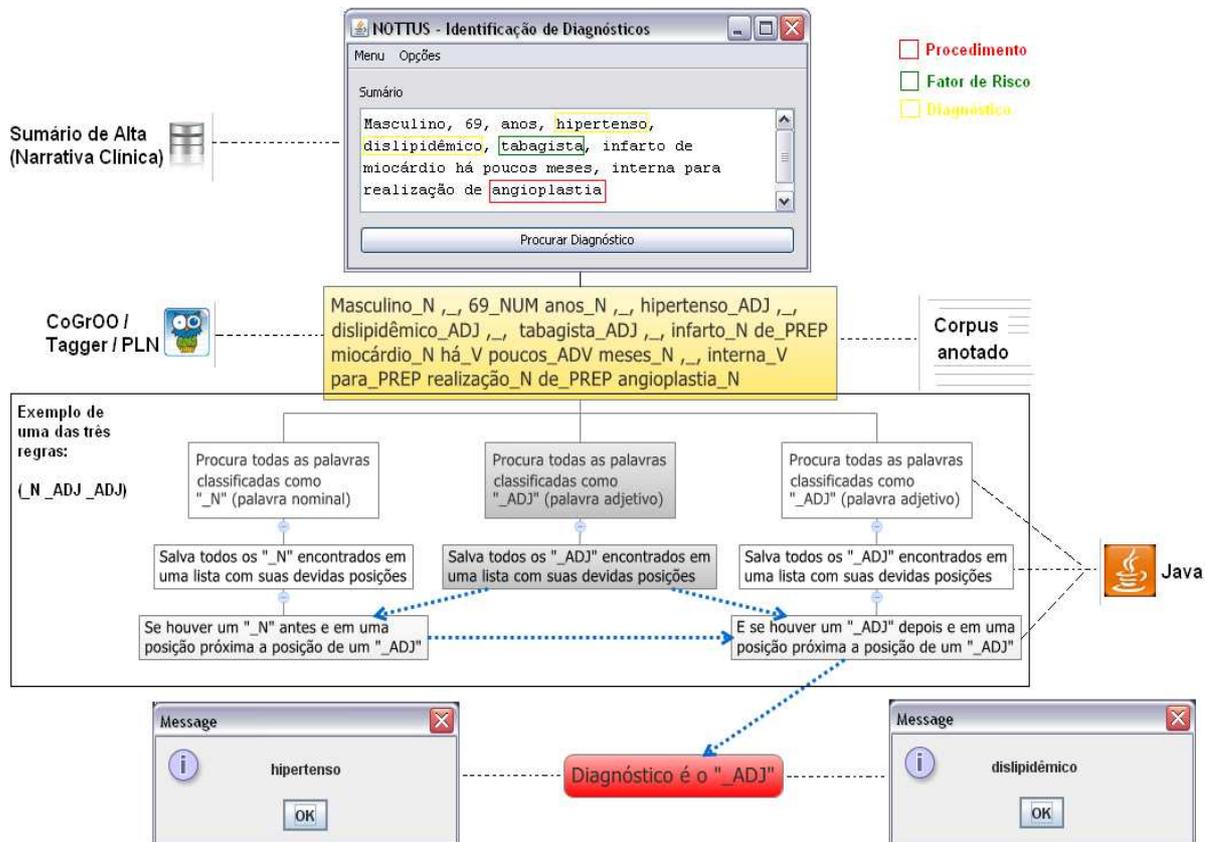


Figura 12 – Esquema da ferramenta para a extração de diagnósticos clínicos, extrair diagnósticos

O esquema de funcionamento interno de todo processamento do texto do sumário de alta, até a extração, identificação dos diagnósticos realizadas pelo NOTTUS, é mostrado com um exemplo na Figura 12. A partir do momento que o sumário é inserido na área de texto (tela NOTTUS) e o usuário aperta o botão “procurar diagnósticos”, o sumário é interpretado pelo sistema e cada palavra recebe uma classificação (quadro em amarelo). Esta classificação é realizada pelo CoGrOO, baseando-se no *corpus* clínico anotado descrito na metodologia. Após a classificação da narrativa, um conjunto de regras inseridas dentro de um documento de texto é acionado (Figura 13). Neste exemplo, uma das regras utilizadas para encontrar o diagnóstico hipertenso e dislipidêmico tenta localizar no texto os termos classificados como “_ADJ” (adjetivo). A seguir, é verificada se a classificação da

palavra anterior é um “_N”, ou seja, um substantivo, e se o termo posterior é um “_ADJ”. Se a resposta desta comparação for verdadeira, então a palavra que correspondente ao primeiro adjetivo encontrado é exibida como diagnóstico.

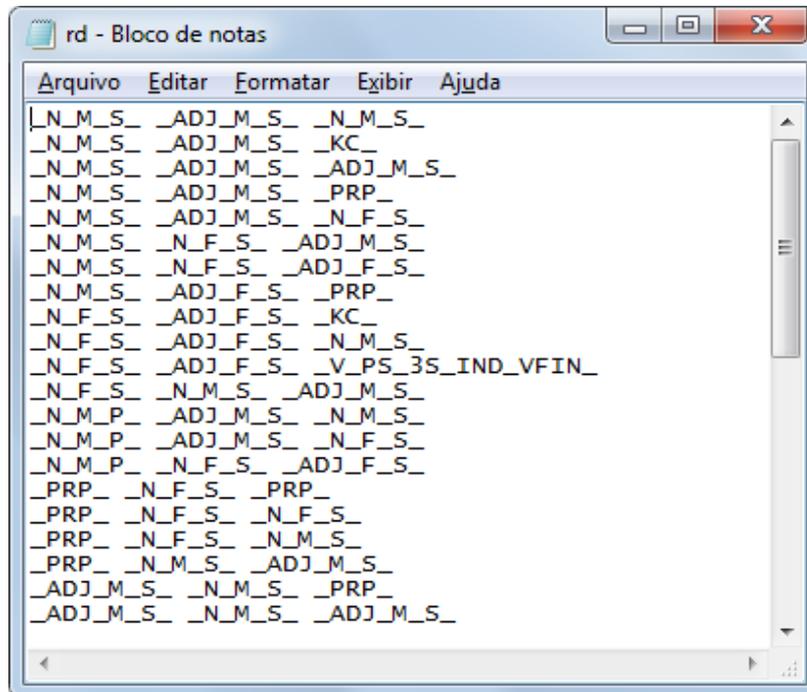


Figura 13 – Arquivo com regras para extração de diagnóstico

5 DISCUSSÃO

5.1 MÉTODO PARA IDENTIFICAÇÃO DE DIAGNÓSTICOS EM NARRATIVAS CLÍNICAS

A primeira dificuldade para o desenvolvimento do método para identificação de diagnósticos contidos em narrativas clínicas foi a falta de definição do que é um diagnóstico e como ele é caracterizado. Atualmente, não existe na literatura um consenso sobre o que é um diagnóstico. Este é considerado como um termo generalizado e que depende muito da área em que o profissional de saúde atua (LACQUANITI et al., 2010; SELLEVOL; STENSETH, 2010).

Um outro ponto relevante na especificação do método é a estrutura dos textos e a forma de escrita. A base de dados do Hospital de Clínicas de Porto Alegre apresenta narrativas clínicas sem um formato de escrita e com informações que dificultam a sua extração, por não serem padronizadas e estarem registradas de diferentes maneiras (CHU, 2002).

Trabalhos como *Lessons Extracting Diseases from Discharge Summaries*, de Long (2007), *Using a natural language processing system to extract and code family history data from admission reports*, de Friedlin e McDonald (2006), e outros citados na revisão da literatura relatam extração de informações baseando-se em cenários ideais. São textos elaborados com estrutura no quais foram identificados dados bem específicos. Porém, esta não é a realidade dos PEPs que são preenchidos com informações clínicas, na maioria das vezes sem uma estrutura definida. Além disso, a maioria das narrativas clínicas, incluindo os sumários de alta do HCPA, contém acrônimos.

O tratamento dos acrônimos é uma característica essencial para o sucesso do método. Acrônimos presentes nas narrativas precisaram ser expandidos. Nos acrônimos com dois ou mais significados, optou-se por utilizar a descrição que preferencialmente representa um diagnóstico. Se nenhuma das opções refere-se a um diagnóstico, foi escolhida aquela que apresenta maior frequência na base de dados dos sumários. Alguns acrônimos não foram expandidos, por exemplo, o “na”.

Além de “na” ter o significado de sódio, pode ser somente a preposição mais artigo “na”. Porém, exceções deste tipo não são frequentes nos textos.

Considerando-se que, nos textos de alguns dos sumários do HCPA, o diagnóstico é a primeira palavra, foi necessário incluir o termo “paciente” em todos os inícios dos textos. Durante os testes, foi descoberto que, se inserida a palavra “paciente” no início do texto, o diagnóstico é encontrado, mesmo que este esteja em forma de acrônimo. Este resultado pode ser visualizado no Quadro 15. Nas situações em que a palavra “paciente” não era necessária, a sua inserção não interferiu no resultado da identificação dos diagnósticos.

Sumário	# CARDIOPATIA ISQUÊMICA GRAVE # ICC - FE=41% # HAS # EAP NA BAIXA # ECG: BRE Paciente chegou à Emergência com quadro de dispnéia súbita. Realizado R-x tórax c/ achados de congestão pulmonar e ECO Doppler MiS sem sinais de trombose. Paciente apresentou boa evolução e resposta clínica às medidas para edema pulmonar agudo.
Como o programa analisou	paciente cardiopatia isquêmica grave icc fe 41 has eap na baixa ecg: bre paciente chegou à emergência com quadro de dispnéia súbita realizado r x tórax c achados de congestão pulmonar e eco doppler mis sem sinais de trombose paciente apresentou boa evolução e resposta clínica às medidas para edema pulmonar agudo
O que encontrou como diagnóstico	Cardiopatia isquêmica, trombose, insuficiência cardíaca.

Quadro 15 – Tratamento das regras em sumários iniciando com diagnósticos

A opção por não associar palavras, ou criar listas, para a identificação de diagnósticos parte do modelo proposto por Bui et al. (2010), que utiliza regras. Também se deve ao fato que as narrativas clínicas analisadas para este trabalho possuem formato e padrões diferentes entre si, o que dificulta a associação de palavras. Porém, as regras não são restritas aos sumários de alta. Existe a possibilidade de estendê-las para a extração de diagnósticos em outros tipos de narrativas clínicas.

A utilização do gênero e grau na elaboração das regras de extração foi fundamental para limitar a quantidade de falsos diagnósticos. Em um trabalho realizado por Pacheco (2009), foram elaboradas regras a partir da classificação de termos do SNOMED – *Clinical Finds*, em que uma regra é bastante semelhante à definida neste estudo: `_ADJ _ADJ _ADJ`.

Para esta regra específica, foram encontradas 19.922 frases indicando que deveria ter um diagnóstico em cada uma delas, sendo que o sumário trabalhado contém apenas 5.617 registros. No total, estes números revelam que, nos 5.617 sumários de alta, poderiam haver 187.518 diagnósticos.

A inclusão do *corpus* clínico anotado é fundamental para a correta identificação dos diagnósticos. O CoGrOO possui um *corpus* dentro do contexto jornalístico. Testes sem a utilização do *corpus* preparado revelaram a necessidade de um maior número de regras, sendo que a metodologia de identificação através do método sintático pode não ser o mais indicada. Porém, como foi utilizado apenas 10% da base dos sumários para criar o *corpus* anotado, na análise dos resultados foram encontradas algumas situações que indicam a necessidade de aumentar a porcentagem de narrativas corrigidas no *corpus*. Um exemplo é o caso do termo “perivesicular”, que foi classificado como verbo, fez com que o diagnóstico “abscesso” não fosse identificado.

A lista com falsos diagnósticos poderia ser reduzida, utilizando-se um corretor ortográfico que auxiliasse no preenchimento do sumário. Palavras como “emsegunda” apareceram como diagnóstico.

5.2 MAPEAMENTO DOS DIAGNÓSTICOS PARA A CID-10

Os diagnósticos, após serem identificados pelo método, foram mapeados para a CID-10, conforme a semelhança morfológica. O mapeamento foi possível e atende ao objetivo de relacionar possíveis indicações entre o diagnóstico e opções da CID-10. Porém, existem melhorias que podem ser feitas para a classificação e ordem de apresentação das opções de codificação CID-10. Por exemplo, utilizar técnicas de aprendizagem baseadas em casos, nas quais o sistema armazena as opções mais utilizadas no mapeamento e as utiliza nas análises posteriores.

Porém, é preciso ter cuidado para não induzir o médico ao erro e escolher a opção mais conveniente sempre, como é o caso de diagnósticos considerados genéricos. Atualmente, no NOTTUS, a opção mais genérica aparece na última posição da lista de códigos, incentivando o profissional a escolher um diagnóstico mais específico.

5.3 ANÁLISE PELOS ESPECIALISTAS DOS DIAGNÓSTICOS IDENTIFICADOS

O método de Bui et al. (2010), utilizado neste trabalho, permitiu a identificação de diagnósticos em narrativas clínicas. Trata-se de um método exaustivo por tentativas que consiste em analisar a classificação de termos de uma frase, identificar o termo que se deseja identificar (neste caso diagnóstico) e a classificação do termo antecessor e posterior.

A cada regra adicionada, o método identifica também (se não na mesma narrativa, na seguinte), resultados indesejados que precisam ser removidos. A cada remoção, é preciso reavaliar todo o conteúdo anteriormente analisado para se obter a perfeição do método, pois, dependendo da palavra removida, pode inutilizar a identificação de outros termos ou diagnóstico.

Melhorias no método de Bui et al. (2010) foram necessárias para a identificação de diagnósticos, como a remoção de caracteres especiais, expansão de acrônimos e a modificação forçada dos textos para trabalhar somente em minúsculos, este último por necessidade do CoGrOO.

Os diagnósticos podem ser encontrados em outros tipos de narrativas. Porém, para utilizar o método desenvolvido, pode ser necessário inserir ou remover regras, ou seja, é necessário uma pré-avaliação antes da utilização em um novo contexto diferente do sumário de alta.

5.4 NOTTUS

Considerando que a ferramenta NOTTUS é semiautomática, o ideal é ter uma alta sensibilidade. Por isso, foram utilizadas apenas regras com três etiquetadores. Testes envolvendo apenas dois etiquetadores do exemplo: `_ADJ _ADJ` (primeiro formato de regra a ser considerado envolvendo PLN e alta sensibilidade, neste caso) mostraram semelhança nos verdadeiro-resultados, porém, um número muito grande de falsos resultados.

Neste trabalho, não foram utilizadas regras para eliminar a negação de diagnósticos, embora, em alguns casos, a ferramenta já consiga ignorar a negação devido às regras estabelecidas sintaticamente.

A preferência é apenas indicar a negação do diagnóstico, em que algumas palavras que antecedem o diagnóstico podem ser grafadas no texto, como, por exemplo: “suspeita”, “nega” “não”. Casos da negação do diagnóstico estão representados no Quadro 16.

Os médicos consultados consideraram que, para utilizar a ferramenta em outras narrativas clínicas, deverá ser realizado um trabalho futuro para o tratamento da negação de diagnósticos. Segundo informações, em alguns documentos é recomendada a prática do “nega diagnóstico”, em que o especialista revela a negação de diagnósticos em documentos.

Exemplo	Ferramenta indicou como diagnóstico? Qual?	Regra
1) ... estudo eletrofisiológico com indução de taquicardia ventricular.	Sim, taquicardia.	_PRP__N_F_S__ADJ_F_S_
2) ... sem recorrência da arritmia após instituição da amiodarona.	Não	Não se aplica, desconsiderou o diagnóstico negado.
3) ... investigação para hipertensão renovascular negativa.	Sim, hipertensão renovascular.	1) PRP__N_F_S__ADJ_F_S_ 2)_N_F_S__ADJ_F_S_ _ADJ_F_S_
4) ... não apresentou mais episódios de taquicardia durante o tratamento.	Sim, taquicardia.	_PRP__N_F_S__PRP_
5) ... dor torácica não sugestiva de angina, mas de forte intensidade.	Não	Não se aplica, desconsiderou o diagnóstico negado.
6) ... não apresentou mais sintomas de angina, evoluindo bem.	Não	Não se aplica, desconsiderou o diagnóstico negado.
7) ... sem recorrência da arritmia após instituição da amiodarona	Não	Não se aplica, desconsiderou o diagnóstico negado.

Quadro 16 – Negação de diagnósticos

Diagnósticos que recebem a palavra “nega” são excluídos diretamente pelas regras criadas pela ferramenta, mas, como visto no Quadro 16, a ferramenta,

através destas regras, não consegue avaliar todo tipo de negação de diagnósticos, necessitando de um novo tipo de tratamento.

Na versão inicial da ferramenta NOTTUS, foi desenvolvida uma função para o usuário remover os falsos diagnósticos dos diagnósticos encontrados. Essa função foi desabilitada após os testes finais, pois pode induzir os usuários a construir regras que podem excluir radicais que fazem parte de um diagnóstico, inutilizando o método e, conseqüentemente, a ferramenta. Alguns exemplos de erros encontrados utilizando o mecanismo de remoção de falsos diagnósticos estão relatados no Quadro 17.

Palavra encontrada nos resultados	Palavra removida	Exemplo de erro ocorrido devido à palavra removida
Dishipertensão	Dis	Quando encontra o diagnóstico dislipidemia, mostra resultado como lipidemia.
Arterialda	Arterialda	Remove toda palavra com arterial. Excluir resultados de hipertensão arterial, por exemplo.
Arterialda	Da	Exclui vários resultados de diagnósticos, entre um exemplo: hemorroida – em que mostra apenas hemorrói.

Quadro 17 – Problemas revelados por *Stop Words*

5.5 TRABALHOS FUTUROS

O método desenvolvido para identificação de diagnósticos permite várias pesquisas para o desenvolvimento da ferramenta, de acordo com cada especialidade médica. Cada área de atuação (hospitalar, saúde da família, clínico, administrativo, entre outros) pode utilizar o método e especificá-lo para um determinado uso. Para isso, é necessário adequar as regras de extração para o contexto a ser aplicado.

O manejo fácil das regras permite que o administrador ou desenvolvedor do sistema faça uma análise e altere o contexto do sistema. A grande vantagem é que o mesmo método poderá ser utilizado para verificar a eficiência do método para

identificação de outras informações (por exemplo: sintomas, doenças associadas, fatores de risco, entre outros).

Para mudar de domínio de aplicação, pode ser necessária a construção de um novo *corpus* ou da ampliação do *corpus* já adicionado à ferramenta. Quanto mais frases e termos forem incorporados, maior probabilidade no acerto da identificação dos termos.

A implementação de um corretor ortográfico (o CoGrOO é um corretor gramatical) dentro da ferramenta NOTTUS pode melhorar a eficiência da ferramenta.

A análise semântica para correlacionar procedimentos e/ou sintomas para diagnósticos é uma contribuição que pode aprimorar a ferramenta. Assim como melhorar o tratamento da negação de diagnósticos, que, por meio sintático, não apresentou ser um método eficaz.

A metodologia utilizada neste trabalho, proposta por Bui et al. (2010), pode ser reutilizada para o desenvolvimento do método para encontrar as negações dos diagnósticos, em que sintaticamente, deve-se utilizar as regras para identificação de diagnósticos e, em uma lista, sugerir as palavras que possam ser a negação do diagnóstico, tais como: “nega”, “negação”, “suspeita”, “não”, “indução”, entre outras.

6 CONCLUSÃO

Os diagnósticos contidos em sumários de alta elaborados em formato de texto livre foram identificados, incluindo a correlação destes para a CID-10. Para isto, foi elaborado um método para identificar diagnósticos. O método desenvolvido está direcionado para a cardiologia, pois foram utilizados principalmente textos de sumários de alta do serviço de cardiologia do HCPA. Porém, os resultados obtidos nos experimentos com sumários de outras áreas deste hospital mostraram que o método possibilita também a identificação de diagnósticos mais relacionados a outras especialidades clínicas.

A correlação dos diagnósticos identificados, baseando-se na morfologia das descrições da CID-10, contempla a necessidade de mapeamento entre estes e os códigos desta classificação.

Em relação às análises realizadas pelos médicos, os especialistas validaram a identificação dos diagnósticos e sua correlação com a CID-10. O resultado da pesquisa foi satisfatório e recebeu contribuições importantes para a sua continuação.

NOTTUS é a nova ferramenta que integra o CoGrOO e permite identificar diagnósticos em narrativas clínicas. O uso da ferramenta é bastante intuitivo e permite integração com outras bases de dados de prontuários eletrônicos. As regras elaboradas que estão embutidas dentro do sistema permitem a fácil adaptação para ser utilizada em outros contextos.

A inserção de um corretor ortográfico e o aumento no número de narrativas corrigidas dentro do *corpus* pode melhorar a qualidade da extração dos diagnósticos. Porém, o método elaborado utilizando o *corpus* clínico anotado se apresentou bem adequado para a identificação dos diagnósticos no contexto utilizado.

Esta é uma pesquisa que pode ser facilmente incorporada ao dia a dia clínico, durante o preenchimento e utilização dos sistemas de PEP.

REFERÊNCIAS

- Alecu I, Bousquet C, Jaulent M. A case report: using SNOMED CT for grouping adverse drug reactions terms. *BMC Medical Information Decision Making*. 2008 Oct.; 27(8 Suppl 1):S4.
- Barnicke LT, Wetson J, Collobert RM, et al. Large scale application of neural network based semantic role labeling for automated relation extraction from biomedical texts. *PLoS One*. 2009 July;4(7):e6393.
- Bui QC, Nualláin BO, Boucher CA, Sloom PM. Extracting causal relations on HIV drug resistance from literature. *BMC Bioinformatics*. 2010 Feb.;11(101):11-23.
- Chanlekha H, Kawazoe A, Collier N. A framework for enhancing spatial and temporal granularity in report-based health surveillance systems. *BMC Medical Informatics and Decision Making*. 2010 Jan.;10(1):1-15.
- Charnovsck R, Nascimento Filho JR, Machado GB, Nicoleit ER. Registro Eletrônico para Acompanhamento Médico de Pacientes em uma UTI . In: IX Congresso Brasileiro de Informática em Saúde; 2004; Ribeirão Preto. Disponível em: <<http://www.sbis.org.br/cbis9/arquivos/650.pdf>>.
- Chu SS. Information retrieval and health/clinical management. *Yearbook of Medical Informatics*; 2002:271-275.
- DATASUS. Banco de dados do Sistema Único de Saúde. CID-10 - Classificação Internacional de Doenças. [acesso em 20 out. 2009]. Disponível em: <<http://www.datasus.gov.br/cid10/v2008/cid10.htm>>.
- CoGrOO. Corretor gramatical acoplável ao OpenOffice.org. [acesso em 13 ago. 2009]. Disponível em: <<http://cogroo.sourceforge.net>>.
- Farkas R, Szarvas G. Automatic construction of rule-based ICD-9-CM coding systems. *BMC Bioinformatics*. 2008;9(Suppl 3):S10.
- Friedlin J, McDonald CJ. Using a natural language processing system to extract and code family history data from admission reports. *AMIA Annu Symp Proc*. 2006; 2006: 925. PMID: PMC1839517.
- Groen PC, Tavanapong W, Oh JH, Wong J. Manual annotation of colonoscopy videos: a first step towards automation. *AMIA Annu Symp Proc*. 2006; 2006: 901. PMID: PMC1839295.
- Grossman E, Cardoso MHCA. As narrativas em Medicina: contribuições à prática clínica e ao ensino médico. *Revista Brasileira de Educação Médica*. 2006;30(1):6-14.
- IHTSDO. International Health Terminology Standards Development Organisation. [acesso em 15 nov. 2009]. Disponível em: <<http://www.ihtsdo.org/snomed-ct>>.

Kling J. Diagnosis or drug? Will pharmaceutical companies or diagnostics manufacturers earn more from personalized medicine? *Science and Society*. 2007;8(10):903-906.

Kluck MM, Guimarães JR. Sumário eletrônico de alta: garantindo a continuidade da assistência ao paciente através da informação. *Revista de Informática Pública*. 1999 dez.;1(2):123-137.

Lacquaniti A, Bolignano D, Donato V, et al. Alterations of lipid metabolism in chronic nephropathies: mechanisms, diagnosis and treatment. *Kidney Blood Press Res*. 2010 Mar.;33(2):100-110.

Lapelle NR, Luckmann R, Simpsons EH, et al. Identifying strategies to improve access to credible and relevant information for public health professionals: a qualitative study. *BMC Public Health*. 2006;6:89.

Long W. Lessons extracting diseases from discharge summaries. *AMIA Annu Symp Proc*. 2007; 2007: 478–482. Published online 2007. PMID: PMC2655845.

Lowe HJ, Huang Y, Regula DP. Using a statistical natural language Parser augmented with the UMLS specialist lexicon to assign SNOMED CT codes to anatomic sites and pathologic diagnoses in full text pathology reports. *AMIA Annu Symp Proc*. 2009; 2009: 386–390. Published online 2009 November 14. PMID: PMC2815395.

Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *Journal American Medical Information Association*. 2005 July-Aug.;12(4):448-57.

Meystre SM, Haug PJ. Comparing natural language processing tools to extract medical problems from narrative text. *AMIA Annu Symp Proc*. 2005; 2005: 525–529. PMID: PMC1560561.

Naghavi M, Makela S, Foreman K, et al. Algorithms for enhancing public health utility of national causes-of-death data. *Population Health Metrics*. 2010;8(9):1-14.

Naryshkin S, Schultz BL. Assessment of quality of data provided on Pap test requisitions: implications for quality of care and patient safety. *Cytojournal*. 2009 July;6:11.

Nystrom M, Merkel M, Petersson H, et al. Creating a medical dictionary using word alignment: the influence of sources and resources. *BMC Medical Informatics and Decision Making*. 2007 Nov.;7:37.

Pacheco E, Cancian P, Nohama P, Schulz S. MorphoSaurus: a cross-language information retrieval system. In: Blobel B, Pharow P, Zvarova J, Lopez D, editors. *EHealth: combining health telematics, telemedicine, biomedical engineering and bioinformatics to the edge*. CeHR International Conference Results 2007. Berlin: Aka; 2008. p. 111-116.

Pacheco EJ. MorphoMap: mapeamento automático de narrativas clínicas para uma terminologia médica. [Tese]. Curitiba: Universidade Tecnologia Federal do Paraná;

2009. 154 p. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial.

Pakhomov S, Pedersen T, Chute CG. Abbreviation and acronym disambiguation in clinical discourse. *AMIA Annu Symp Proc.* 2005; 2005: 589–593. PMID: PMC1560669.

Ribeiro MB. Propostas de aplicações para tratamento linguístico computacional do Esperanto. [Artigo apresentado à disciplina Fatores Humanos em Sistemas de Informação]. Instituto de Computação da Universidade Estadual de Campinas (IC/UNICAMP); 2001. [acesso em 05 fev. 2009]. Disponível em: <<http://www.aleph.com.br/kce/ftp/artigo-mbr01.pdf>>.

Rich E, Knight K. *Inteligência artificial*. 2^a. ed. São Paulo: Makron Books; 1994.

Rosier A, Burgun A, Mabo P. Using regular expressions to extract information on pacemaker implantation procedures from clinical reports. *AMIA Annu Symp Proc.* 2008; 2008: 81–85. Published online 2008. PMID: PMC2656039.

Ruch P, Gobeill J, Lovis C, Geissbühler A. Automatic medical encoding with SNOMED categories. *BMC Medical Informatics and Decision Making.* 2008 Oct.;8 Suppl 1:S6. PMID: PMC258279.

Ruch P, Gobeill J, Tbahriti I, Geissbühler A. From episodes of care to diagnosis codes: automatic text categorization for medico-economic encoding. *AMIA Annu Symp Proc.* 2008; 2008: 636–640. Published online 2008. PMID: PMC2655971.

Russel S, Norvig P. *Artificial intelligence: a modern approach*. 3rd ed. Local: Prentice Hall; 2009.

Schank RC, Riesbeck CK. *Inside computer understanding*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers; 1981.

Schulz S, Hahn U. Morpheme-based cross-language indexing for medical document retrieval. *International Journal of Medical Informatics.* 2000;58:87-89. [acesso em 05 fev. 2009]. Disponível em: <http://www.ijmijournal.com/article/S1386-5056%2800%2900078-2/abstract>.

Sellevoid OF, Stenseth R. Non-cardiac surgery in patients with cardiac disease. *Tidsskr Nor Laegeforen.* 2010 Mar.;130(6):623-627.

Shortliffe EH, Barnett GO. *Medical data: their acquisition, storage, and use*. 2^a. ed. Nova Iorque: Springer; 2001.

Sohn S, Comeau DC, Kim W, et al. Abbreviation definition identification based on automatic precision estimates. *BMC Bioinformatics.* 2008 Sept.;9:402.

South BR, Shen S, Jones M, et al. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics.* 2009 Sept.;10(Suppl. 9):S12.

Sun JY, Sun YA. System for automated lexical mapping. *Journal American Medical Informatics Association*. 2006 May-June;13(3):334-43.

Tange HJ, Hasman A, Robbé PFV, et al. Medical narratives in electronic medical records. *International Journal of Medical Informatics*. 1997;46:7-29.

Uzuner Ö, Sibanda TC, Luo Y, et al. A de-identifier for medical discharge summaries. *Artificial Intelligence in Medicine*. 2008;42:13-35.

VISL. Visual Interactive Syntax Learning. [acesso em 29 out. 2009]. Disponível em: <<http://beta.visl.sdu.dk/visl/pt/info/portsymbol.html>>.

Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports. *AMIA Annu Symp Proc*. 2008; 2008: 783–787. Published online 2008. PMID: PMC2656103.

Wang Y, Patrick J, Miller G, et al. A computational linguistics motivated mapping of ICPC-2 PLUS to SNOMED CT. *BMC Medical Informatics and Decision Making*. 2006 Oct.;8 Suppl 1:S5.

Wojciechowski J. Um sistema computacional para verificação do nível de compreensão de textos em linguagem natural. [Tese]. Curitiba: Pontifícia Universidade Católica do Paraná; 2003. 119p. Programa de Pós-Graduação em Informática Aplicada (PPGIA).

WHO. World Health Organization. International Classification of Diseases (ICD). [acesso em 1 dez. 2009]. Disponível em: <<http://www.who.int/classifications/icd/en/>>.

Xu H, Stetson PD, Friedman C. A study of abbreviations in clinical notes. *AMIA Annu Symp Proc*. 2007; 2007: 821–825. Published online 2007. PMID: PMC2655910.

Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decision Making*. 2006 July;26(6):30.

Zhou G, Shen D, Zhang J, et al. Recognition of protein/gene names from text using an ensemble of classifiers. *BMC Medical Informatics*. 2005;6(Suppl 1):S7.

APÊNDICES

APÊNDICE A – APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
Núcleo de Bioética
Comitê de Ética em Pesquisa
Ciência com Consciência

PARECER CONSUBSTANCIADO DE PROTOCOLO DE PESQUISA

Parecer Nº **0004422/10** Protocolo CEP Nº 5854
 Título do projeto **IDENTIFICAÇÃO DE DIAGNÓSTICOS CONTIDOS EM NARRATIVAS CLÍNICAS E MAPEAMENTO PARA A CLASSIFICAÇÃO INTERNACIONAL DE DOENÇAS** Grupo III
 Protocolo CONEP **0375.0.84.000-10** Pesquisador responsável **Claudia Maria Cabral Moro Barra** Versão **001**
 Instituição **Laboratório de Informática em Saúde – PUCPR/CCBS - Curitiba**

Objetivos

O projeto tem como objetivo principal "Identificar diagnósticos contidos em sumários de alta hospitalar elaborados em formato de texto livre propondo correlações com a CID-10", e como objetivos secundários:

1. Propor um método para a identificação de diagnósticos relacionados à cardiologia presentes nos sumários de alta.
2. Validar os diagnósticos identificados.
3. Mapear os diagnósticos para o CID-10.
4. Implementar um protótipo do sistema para identificação de diagnósticos codificando-os em CID-10.

Comentários e considerações

O projeto pretende desenvolver uma ferramenta (sistema) que possibilite coletar diagnósticos de narrativas clínicas e indicar um diagnóstico estruturado segundo a terminologia da Classificação Internacional de Doenças (CID-10). Para tanto os pesquisadores utilizarão uma base de dados contendo um conjunto de sumários de alta (descaracterizados) que foi solicitada ao Hospital de Clínicas de Porto Alegre - HCPA através do projeto CONEP FR – 126681. O parecer positivo do comitê de ética do HCPA permitiu a partir de 2007 o acesso a um corpus composto pela totalidade dos sumários de alta de 5 anos da área de cardiologia e de 1 mês dos sumários de alta referente a todos os atendimentos hospitalares, que são utilizados pelo grupo de pesquisa de Recuperação de Informações em Saúde do PPGTS em projetos de extração de informações destes sumários. Este conjunto é composto por 8.332 documentos, sendo esta a base que foi utilizada para os experimentos da dissertação relacionada à este projeto. Assim, os pesquisadores entendem que não é utilizada uma base de dados real que necessite o TCUD. Três médicos deverão validar os diagnósticos encontrados e deverão responder a um questionário referente à qualidade dos resultados identificados pelo sistema. O critério de inclusão é "atuar na área de cardiologia" e os critérios de exclusão são:

1. Profissionais que não estão habituados com elaboração de sumário de alta na área de cardiologia.
2. Profissionais que não estão acostumados em codificar diagnósticos utilizando a Classificação Internacional de Doenças (CID-10).

A metodologia consiste nas seguintes realizações:

- Análise de narrativas clínicas.
- Adaptação de um analisador gramatical.
- Definição das regras para extrair diagnósticos.
- Correlação de diagnósticos da cardiologia para o CID-10.
- Desenvolvimento de um novo sistema.
- Validação pelos cardiologistas.

Os dados serão tratados e analisados estatisticamente. Não são identificados riscos aos participantes, e é apontado como benefício é apontado o resultado da ferramenta a ser desenvolvida, que auxiliará no preenchimento e na identificação de informações do Prontuário Eletrônico do Paciente - PEP



Termo de consentimento livre e esclarecido e/ou Termo de compromisso para uso de dados.

O TCLE está de acordo com a resolução deste Comitê.

Conclusões

O projeto atende às recomendações deste Comitê.

Devido ao exposto, o Comitê de Ética em Pesquisa da PUCPR, de acordo com as exigências das Resoluções Nacionais 196/96 e demais relacionadas a pesquisas envolvendo seres humanos, em reunião realizada no dia: **03/11/2010**, manifesta-se por considerar o projeto **Aprovado**.

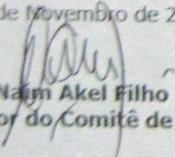
Situação Aprovado

Lembramos aos senhores pesquisadores que, no cumprimento da Resolução 196/96, o Comitê de Ética em Pesquisa (CEP) deverá receber relatórios anuais sobre o andamento do estudo, bem como a qualquer tempo e a critério do pesquisador nos casos de relevância, além do envio dos relatos de eventos adversos, para conhecimento deste Comitê. Salientamos ainda, a necessidade de relatório completo ao final do estudo.

Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP-PUCPR de forma clara e sucinta, identificando a parte do protocolo a ser modificado e as suas justificativas.

Se a pesquisa, ou parte dela for realizada em outras instituições, cabe ao pesquisador não iniciá-la antes de receber a autorização formal para a sua realização. O documento que autoriza o início da pesquisa deve ser carimbado e assinado pelo responsável da instituição e deve ser mantido em poder do pesquisador responsável, podendo ser requerido por este CEP em qualquer tempo.

Curitiba, 03 de Novembro de 2010.


Prof. MSc. Naim Akel Filho
Coordenador do Comitê de Ética em Pesquisa
PUC PR



ANEXOS

ANEXO A – LISTA DE ACRÔNIMOS

Acrônimo	Ocorrências	Significado Principal
A2	3	amostra 2
A2RV	2	alteração de repolarização ventricular
AA	20	após almoço
AA2	2	Aminoácidos
AAA	6	aneurisma da aorta abdominal
AAS	966	ácido acetil salicílico
AB	6	AB
ABD	8	detecção automática de fronteira
AC	71	ausculta cardíaca
ACD	161	artéria coronária direita
ACE	2	artéria coronária esquerda
ACFA	97	fibrilação atrial crônica
ACM	5	artéria cerebral média
ACO	20	anticoagulante oral
ACTH	8	hormônio adrenocorticotrófico
ACTP	810	angioplastia coronária transluminal percutânea
ACTPs	5	angioplastias coronarianas transluminais percutâneas
ACV	4	aparelho cardiovascular
ACX	68	artéria coronariana circunflexa
ACx	26	artéria circunflexa A
AD	12	átrio direito
ADA	98	artéria descendente anterior
ADAE	41	átrio direito átrio esquerdo
ADC	3	coeficiente de difusão aparente
ADS	2	Amniocentese descompressiva seriada
AE	221	átrio esquerdo
AESP	5	atividade elétrica sem pulso
AI	2	angina instável
AIG	9	adequado para a idade gestacional
AIH	7	Autorização de Internação Hospitalar
AINE	14	antiinflamatório não-esteroidal
AINES	9	antiinflamatórios não-esteroidais
AINEs	2	antiinflamatórios não-esteroidais
AIRV	5	alterações inespecíficas da repolarização ventricular
AIT	17	ataque isquêmico transitório
AITs	6	ataques isquêmicos transitórios
AJ	22	antes do jantar
AMBU	6	Ambulatório
AMgCx	3	ramo marginal da artéria circunflexa
AMO	8	alteração de medula óssea
AMPI	4	Ampicilina
ANCA	3	anticorpo anticitoplasma de neutrófilo
ANGIO	2	Angiografia
ANTI	5	Anti
AP	73	ausculta pulmonar
APGAR	178	Apgar
ARA	2	antagonistas dos receptores da angiotensina

ART	2	Artéria
ARV	12	anti-retro viral
ARV's	3	Antiretrovirais
ATB	90	Antibiótico
ATC	24	angioplastia transluminal coronária
AV	39	átrio ventricular
AVC	176	acidente vascular cerebral
AVCi	5	acidente vascular cerebral isquêmico
AVCs	12	acidentes vasculares cerebrais
AVE	26	acidente vascular encefálico
AVEi	5	acidente vascular encefálico isquêmico
AVEs	2	acidente vascular encefálico
AVF	2	ante-verso-flexão
AZT	5	Zidovudina
B1	2	B1
B12	11	B12
B2	25	ausculta cardíaca
B3	2	terceira bulha
B4	2	quarta bulha
B6	5	B6
BAAR	39	bacilo álcool ácido resistente
BAV	48	bloqueio atrioventricular
BAVT	28	bloqueio atrioventricular total
BB	8	beta-bloqueador
BBloq	2	beta-bloqueadores
BC	2	bloco cirúrgico
BCF	13	batimentos cardíacos fetais
BCG	2	bacilo de Calmette-Guérin
BCP	92	broncopneumonia
BCPs	2	broncopneumonias
BD	34	morte cerebral
BDAS	2	bloqueios divisionais ântero-superiores
BEG	7	bom estado geral
BFM	3	Berlin-Frankfurt-Münster
BI	8	Duas
BIA	8	balão intra-aórtico
BID	104	duas vezes ao dia
BIPAP	2	pressão positiva em vias aéreas com dois níveis
BiPAP	2	pressão positiva em vias aéreas com dois níveis
BMO	4	biópsia de medula óssea
BNF	2	bulhas normofonéticos
BQLT	6	Bronquiolite
BQLTE	3	Bronquiolite
BQTE	5	Bronquite
BR	15	bloqueio de ramo
BRD	11	bloqueio de ramo direito
BRE	51	bloqueio de ramo esquerdo
BT	39	bilirrubina total
BT:41	2	bilirrubina total
BX	8	Biópsia
BZD	2	benzodiazepínicos
C1	2	cesariana 1
C3	10	Complemento 3
C4	8	Complemento 4
CA	40	Câncer
Ca2	2	Cálcio

CAPD	3	diálise peritoneal ambulatorial crônica
CAPS	9	Centro de Atendimento Psicoprofissionalizante
CAT	358	avaliação crítica tópica
CAt	2	avaliação crítica tópica
CaT	2	avaliação crítica tópica
CBZ	7	carbamazepina
CC	3	cardiopatias congênitas
CCA	12	cardiopatias congênitas em adultos
CCT	5	carcinoma de células transicionais
CD	437	coronária direita
CD34	2	grupo de diferenciação 34
CD4	23	grupo de diferenciação 4
CDI	54	cardioversor-desfibrilador implantável
CE	15	coronária esquerda
CEA	12	antígeno carcinoembrionário
CEC	9	circulação extracorpórea
CEN	8	catéter endonasal
CF	28	colo femoral
CHAD	46	concentrado de hemácias adulto
CHAd	2	concentrado de hemácias adulto
CHC	3	carcinoma hepatocelular
CHILD	7	Child
CI	96	cardiopatia isquêmica
CIA	7	comunicação interatrial
CIC	2	cirurgia cardíaca
CID	2	Classificação Internacional de Doenças
CIE	4	carótida interna esquerda
CIG	10	Cigarro
CIP	2	carcinoma incidental da próstata
CIPED	9	Centro de Investigação de Doenças Pediátricas
CIV	12	comunicação interventricular
CK	43	creatinoquinase
CKMB	8	fração MB da creatinofosfoquinase
CK-MB	21	fração MB da creatinofosfoquinase
CM	4	Centímetro
cm2	5	centímetro quadrado
CMV	23	citomegalovírus
CN	14	catéter nasal
CO	35	monóxido de carbono
CO2	6	dióxido de carbono
CP	35	Comprimido
CPAP	31	pressão positiva contínua em vias aéreas
CPAPn	4	pressão positiva contínua em vias aéreas nasal
CPER	10	colangiopancreatografia endoscópica retrógrada
CPRE	14	colangiopancreatografia retrógrada endoscópica
CPT	3	capacidade pulmonar total
CR	4	Creatinina
CREAT	3	Creatinina
CRM	246	cardio-ressonância magnética
CT	9	centro de tratamento
CTC	6	Corticóide
CTI	316	Centro de Terapia Intensiva
CTi	2	centro de terapia intensiva
CTICC	3	Centro de Tratamento Intensivo Clínico-cirúrgico
CV	13	carga viral
CVE	2	cardioversão elétrica

CVF	18	capacidade vital forçada
CVM	2	contração voluntária máxima
CX	160	artéria circunflexa
D14	2	dia 14
D3	7	dia 3
D4	2	dia 4
DA	420	artéria descendente anterior
DA/Dg	2	descendente anterior / primeira diagonal
DAE	432	dimensão do átrio esquerdo
DAOP	3	doença arterial obstrutiva periférica
DBP	2	diâmetro biparietal
DBPOC	3	doença broncopulmonar obstrutiva crônica
DC	3	doença celíaca
DCE	32	depuração de creatinina endógena
DD	44	diâmetro diastólico
DDD	18	DDD
DDI	3	DDI
DDVE	33	diâmetro diastólico do ventrículo esquerdo
DG	9	ramo diagonal
Dg1	2	primeira diagonal 1
Dg-DA	2	primeira diagonal / descendente anterior
DHEA	2	deidroepiandrosterona
DI	8	Dois
DIP	8	Dipiridamol
DIPI	10	Dipiridamol
DIU	2	dispositivo intra-uterino
DM	272	diabetes melitus
DM1	5	diabetes melitus tipo 1
DM2	92	diabetes melitus tipo 2
Dm2	2	diabete melitus tipo 2
DM-2	2	diabete melitus tipo 2
DMG	4	diabetes melitus gestacional
DMII	2	diabete melitus tipo 2
DMO	3	densidade mineral óssea
DMSA	4	ácido dimercaptosuccínico
DN	15	data do nascimento
DNA	7	ácido desoxirribonucléico
DP	61	doença periapical
DPCD	9	diálise peritoneal continuada
DP-CD	2	diagonal posterior - coronária direita
DPN	6	dispnéia paroxística noturna
DPOC	144	doença pulmonar obstrutiva crônica
DPP	2	Descolamento prematuro de placenta
DPT	2	espessamento peritoneal difuso
DR	6	Doutor
DRA	4	Doutora
DRC	3	doença renal crônica
DRGE	12	doença do refluxo gastresofágico
DS	44	diâmetro sistólico
DSVE	29	diâmetro sistólico do ventrículo esquerdo
DTS	2	dose total semanal
DV	4	densitovolumetria
DVO	28	distúrbio ventilatório obstrutivo
DVP	4	derivação ventrículo peritoneal
DVR	2	distúrbio ventilatório restritivo
DX	3	Diagnóstico

EA	11	emergência ambulatorial
EAP	27	edema agudo de pulmão
EBV	5	Epstein-Barr vírus
ECA	15	enzima conversora de angotensina
ECG	552	eletrocardiograma
ECG's	2	Ecografias
ECGs	19	eletrocardiogramas
ECO	133	ecocardiograma
ECT	2	eletroconvulsoterapia
E-D	2	esquerda-direita
EDA	49	endoscopia digestiva alta
EDSS	2	escala ampliada do estado de incapacidade
EEF	15	escala de expressões faciais
EEG	23	eletroencefalograma
EF	3	exame físico
EFV	2	Efavirenz
EFZ	3	Efavirenz
EGD	39	esofagogastroduodenoscopia
EGDA	5	anastomose esofagogastroduodenal
EHCPA	35	serviço de emergências do Hospital de Clínicas de Porto Alegre
ELLA	3	endoprótese arterial de perna esquerda
EME	2	Emergência
EMEP	2	emergência pediátrica
EMG	4	Emergência
ENMG	7	eletroneuromiografia
EP	2	epitélio pigmentado
EPED	35	especialidades pediátricas
EPF	3	exame parasitológico de fezes
EPO	3	Eritropoetina
EQ	3	esquema quimioterápico
EQU	68	exame qualitativo de urina
ESBL	3	produtoras de beta-lactamase com espectro estendido
ESQ	7	Esquerdo
ESSV	7	extra-sístoles supraventriculares
ESV	9	extra-sístoles ventriculares
EV	126	Endovenosa
EX	10	Ex
FA	87	fibrilação atrial
FAF	5	arma de fogo
FAN	11	fator anti-nuclear
FAV	11	Favorável
FBC	7	freqüência do batimento ciliar
FC	121	função cardíaca
FCmáx	2	função cardíaca máxima
FE	588	fração de ejeção
FEJ	2	fração de ejeção
FEj	2	fração de ejeção
FEM	2	Feminino
FEV	2	Fevereiro
FEVE	4	fração de ejeção do ventrículo esquerdo
FH	2	formação do hipocampo
FID	2	fossa ilíaca direita
FiO2	4	fração de oxigênio inspirado
FK	4	Trakolimus
FNB	15	Fenobarbital
FO	110	ferida operatória

FOP	3 forame oval patente
FR	25 frequência respiratória
FS	7 frequência sinusal
FSH	3 hormônio folículo estimulante
FSV	2 fundo de saco vaginal
FV	23 fibrilação ventricular
g/dia	2 grama/dia
G3	3 gestação 3
G4	2 gestação 4
G6PD	2 glicose-6-fosfato dehidrogenase
GA	4 Gasometria
GGT	12 gama GT
GH	6 hormônio do crescimento
GI	2 gastro intestinal
GNMP	2 glomerulonefrite membrano-proliferativa
GRAFT	2 Graft
GRAM	2 Gram
GT	2 glutariltransferase
GTS	2 Gotas
h/h	3 de hora em hora
HAP	16 hipertensão arterial pulmonar
HAS	731 hipertensão arterial sistêmica
HAS,e	2 hipertensão arterial sistêmica
HB	13 Hemoglobina
HBAE	2 hemibloqueio anterior esquerdo
HBP	2 hiperplasia benigna da próstata
HBs	2 vírus da hepatite B
HBsAg	35 antígeno de superfície do vírus da hepatite B
HbsAg	22 antígeno de superfície da hepatite B
HBSAG	12 antígeno de superfície para hepatite B
HBV	5 vírus da hepatite B
HCG	5 hormônio da gonadotrofina coriônica
HCO3	4 Bicarbonato
HCPA	772 Hospital de Clínicas de Porto Alegre
HCSA	2 Hospital da Criança Santo Antônio
HCTZ	112 hidroclortiazida
HCV	76 vírus da hepatite C
HD	31 Hemodiálise
HDA	14 hemorragia digestiva alta
hdl	6 lipoproteínas de alta densidade
HELLP	3 anemia hemolítica, níveis elevados de enzimas hepáticas e contagem baixa de plaquetas
HF	12 história familiar
HGT	18 hemoglicoteste
HGTs	17 hemoglicotestes
HIC	2 hemorragia intracraniana
HIV	80 vírus da imunodeficiência humana
HM	2 hipertermia maligna
HMC	26 Hemocultura
HMCs	2 Hemoculturas
HMG	71 Hemograma
HMP	3 história médica pregressa
HMV	5 Hospital Moinhos de Vento
HNF	2 heparina não-fracionada
HNSC	11 Hospital Nossa Senhora da Conceição
HP	8 Helicobacter pylori

HPB	10	hiperplasia prostática benigna
HPIV	4	hemorragia peri-intraventricular
HPS	44	Hospital de Pronto Socorro
HS	4	Horas
HSL	5	Hospital São Lucas
HT	7	Hematócrito
Ht/Hb	2	hematócrito/hemoglobina
HTC	2	Hematócrito
HTLV	4	vírus T-linfotrópico humano
HVE	19	hipertrofia ventricular esquerda
HX	2	Histórico
IA	2	infarto agudo
IAM	470	infarto agudo do miocárdio
IAMs	12	infarto agudo do miocárdio
IAM's	2	infartos agudos do miocárdio
IAo	13	insuficiência aórtica
IAO	3	insuficiência aórtica
IC	130	insuficiência cardíaca
ICC	632	insuficiência cardíaca congestiva
ICE	6	insuficiência congestiva esquerda
ICP	3	intervenção coronária percutânea
ICT	27	índice cardiotorácico
IECA	29	inibidor da enzima conversora de angiotensina
iECA	6	inibidores da enzima conversora de angiotensina
I-ECA	12	inibidor da enzima conversora de angiotensina
i-ECA	6	inibidores da enzima conversora de angiotensina
IF	2	forma indeterminada
IFI	6	imunofluorescência indireta
IFN	2	interferon alfa-recombinante
IG	194	imunoglobulina
IgG	73	imunoglobulina G
IGG	50	imunoglobulina G
IgM	45	imunoglobulina M
IGM	41	imunoglobulina M
IGO	24	índice do grau de obesidade
IGP	3	idade gestacional no parto
ILA	18	índice de líquido amniótico
IM	82	insuficiência mitral
IMC	5	índice de massa corporal
IMi	9	infarto do miocárdio
IMT	2	calibre intermediário da carótida
IN	3	Intranasais
INR	109	razão normalizada internacional
IOT	2	intubação orotraqueal
IPC	2	índice de potencial de contaminação
IR	65	insuficiência renal
IRA	30	insuficiência renal aguda
IRC	189	insuficiência renal crônica
ISHAK	2	Ishak
ISQ	4	Isquêmico
ISRS	2	inibidor seletivo da recaptção de serotonina
IT	49	transição interna
ITB	31	índice tornozelo/braço
ITC	2	insuficiência tricúspide
ITr	7	insuficiência tricúspide
ITU	93	infecção do trato urinário

ITUs	7	infecções do trato urinário
IVa	2	veia interventricular anterior
IVAS	9	infecções das vias aéreas superiores
IVC	12	insuficiência venosa crônica
IVUS	10	ultra-sonografia intravascular
JUP	3	junção uretero-piélica
K2	2	Potássio
KG	6	Quilograma
KTTP	4	tempo de tromboplastina parcialmente ativada
L1	3	lombar 1
L2-L3	2	lombar 2 - lombar 3
L3	2	lombar 3
L4	2	lombar 4
L4-L5	2	lombar 4 - lombar 5
L5-S1	2	lombar 5 - sacro 1
LBA	6	lavado brônquico alveolar
LCR	14	Licor
LDH	20	desidrogenase láctica
LDL	6	lipoproteínas de baixa densidade
LE	2	laparotomia exploradora
LFN	2	Linfonodos
LH	5	hormônio luteinizante
LI	2	liquido intersticial
LID	8	lobo inferior direito
LIE	5	lobo inferior esquerdo
LLA	2	leucemia linfocítica aguda
LLC	2	leucemia linfocítica crônica
LM	3	lobo médio
LMA	4	leucemia mielóide aguda
LMC	2	leucemia mielóide crônica
LN	2	Linfonodo
LNH	2	linfoma não-Hodgkin
LOC	4	lúcido, orientado e consciente
LQR	2	Líquor
LSD	13	lobo superior direito
LSE	4	lobo superior esquerdo
LT	7	leucograma total
LUTS	2	sintomas do trato urinário inferior
LV	3	leite de vaca
M0	6	medula óssea
MA	4	meningites assépticas
MAC	4	ambulatório de cardiologia
MAM	5	artéria mamária
MAP	11	pressão média de vias aéreas
mcg/d	6	microgramas/decilitro
mCi	8	Milicurie
MED	5	Médio
MEI	43	medicina interna
MELD	2	modelo para doença hepática terminal
METS	19	equivalente metabólico
METSs	5	equivalentes metabólicos
MF	2	microorganismos filamentosos
MG	42	Miligrama
mg/d	320	miligrama/decilitro
mg/dl	8	miligrama/decilitro
Mg1	2	primeira marginal

MgCx	17	ramo marginal da artéria circunflexa
MgCX	12	ramo moarginal da artéria circunflexa
MGCX	5	ramo marginal da artéria circunflexa
MGCx	2	ramo marginal da artéria circunflexa
Mg-Cx	14	ramo marginal da artéria circunflexa
MGLIS	2	artéria coronária marginal localizada intrastent
MI	3	membro inferior
MI's	3	membros inferiores
MIBI	3	metoxi-isobutil-isonitrila
MID	58	membro inferior direito
MIE	73	membro inferior esquerdo
mL	9	Millilitro
MM	20	mieloma múltiplo
MM-DA	4	mamária - descendente anterior
MMG	2	Mamografia
mmHg	79	milímetro de mercúrio
MMII	17	membros inferiores
MMIIs	4	membros inferiores
MMSS	2	membros superiores
mnmHg	2	milímetros de mercúrio
MO	4	medula óssea
MP	95	marca-passo
MRSA	3	Staphylococcus aureus resistente à metilina
MSD	21	membro superior direito
MSE	27	membro superior esquerdo
MsIs	32	membros inferiores
MsIS	3	membros inferiores
MSS	2	membros superiores
MSs	2	membros superiores
MSSA	2	Staphylococcus aureus sensível à metilina
MsSs	2	membros superiores
MT	4	metiltestosterona
MTD	4	dose máxima tolerada
MTX	16	Metotrexato
MTZ	3	Mirtazapina
MV	8	murmúrio vesicular
mVE	2	massa ventricular esquerda
NA	78	Sódio
Na3	2	Potássio
NAC	2	neuropatia autonômica cardiovascular
NAN	2	leite NAN
NBZ	10	Nebulização
NC	3	nervo craniano
ND	2	nefropatia diabética
NEG	11	Negativo
NEO	4	Neonatal
NEURO	4	Neurologia
NOV	2	Novembro
NPH	120	protamina neutra Hagedorn
NPO	20	nada por via oral
NPS	4	Nitroprussiato de sódio
NPT	24	nutrição parenteral total
NR	178	não-reagente
NS	15	não significativo
NTG	15	Nitroglicerina
NYHA	2	Associação do Coração de Nova York

O2	126	Oxigênio
o2	5	Oxigênio
OAA	5	obstrução arterial aguda
OBS	3	Observação
OD	3	olho direito
OE	3	olho esquerdo
OEA	2	otoemisiones acústicas
OFT	2	Oftalmologia
OK	18	Ok
OMA	11	otite média aguda
ORL	4	otorrinolaringologia
OUT	2	Outubro
P/C	5	polifenol/carboidrato
P4	2	parto 4
PA	102	pressão arterial
PAAF	5	punção aspirativa por agulha fina
PAAf	2	punção aspirativa por agulha fina
PAC	2	pneumonia adquirida na comunidade
PAM	17	pressão arterial média
PaO2	4	pressão arterial de oxigênio
PAS	6	pressão arterial sistêmica
PASP	2	pressão sistólica da artéria pulmonar
PAVM	2	pneumonia associada à ventilação mecânica
PBE	14	peritonite bacteriana espontânea
PBF	19	perfil biofísico fetal
PC	43	parada cardíaca
PCO2	3	pressão de dióxido de carbono
pCO2	2	pressão de dióxido de carbono
PCP	3	pressão capilar pulmonar
PCR	136	parada cárdio-respiratória
PCR _s	5	paradas cardio-respiratórias
PCT	3	Paciente
PEG	9	pré-eclâmpsia grave
PFE	2	peso fetal estimado
PFH	6	prova de função hepática
pH	4	potencial hidrogeniônico
PICC	2	catéter central de inserção periférica
PIG	10	pequeno para idade gestacional
PL	22	punção lombar
PLAQ	4	Plaquetas
PMAP	3	pressão média da artéria pulmonar
PMT	16	Prematuro
PN	14	Pneumonia
PNA	5	pielonefrite aguda
PNE	3	portador de necessidades especiais
PNI	2	psiconeuroimunologia
PNM	2	Pneumonia
PNP	2	Polineuropatia
PNTx	3	Pneumotórax
PO	17	pós-operatório
PO2	2	pressão parcial do oxigênio
POA	3	Porto Alegre
POP	2	Procedimento operacional padrão
PP	72	parto prematuro
PPD	3	derivado protéico purificado
PPL	16	pressão pleural

PPNL	4	Propranolol
PROX	15	Próximo
PS	82	pronto socorro
PSA	22	antígeno específico da próstata
PSAP	142	pressão sistólica da artéria pulmonar
PSG	2	polissonografia
PS-MG	2	Pronto Socorro de Minas Gerais
PSP	6	punção suprapúbica
PT	6	Perímetro
PTH	4	Paratormônio
PTU	7	Propiltiouracil
PUC	16	Pontifícia Universidade Católica
PUCRS	2	Pontifícia Universidade Católica do Rio Grande do Sul
PV	3	parto vaginal
PVC	4	pressão venosa central
PVM	2	prolapso da valva mitral
QD	7	quantidade diária
QI	6	quociente de inteligência
QID	3	quadrante inferior direito
QMT	11	Quimioterapia
QN	7	quando necessário
QT	38	Quimioterapia
R1	9	residente 1
R2	7	residente 2
R3	2	residente 3
RA	4	rêmora atrial
RBV	3	Ribavirina
RC	3	risco cardiovascular
RCP	3	reanimação cardiopulmonar
RCR	2	Ressuscitação cardio-respiratória
RCT	2	rastreamento corporal total com radioiodo
RD	3	retinopatia diabética
RDNPM	2	retardo do desenvolvimento neuropsicomotor
RDP	2	retinopatia diabética proliferativa
RDT	16	Radioterapia
RDW	4	amplitude de distribuição eritrocitária
RE	2	retículo esquerdo
REAG	3	Reagente
REED	8	radiograma de esôfago, estômago e duodeno
RG	9	regime geral
RGE	3	refluxo gastresofágico
RH	4	rifampicina e isoniazida
RHA	3	ruídos hidroaéreos
RHJ	2	refluxo hepato-jugular
RHZ	22	rifampicina, isoniazida e pirazinamida
RM	10	ressonância magnética
RMN	6	ressonância magnética
RN	212	recém-nato
RN1	56	recém-nato 1
RN2	3	recém-nato 2
RNI	8	razão normalizada internacional
RNM	58	ressonância nuclear magnética
RPD	3	retinopatia diabética proliferativa
RR	2	risco relativo
RS	7	ritmo sinusal
RT	2	Radioterapia

RTU	28	ressecção transuretral
RTX	2	neurotoxina resiniferatoxina
RUB	2	Rubéola
RV	23	remodelamento ventricular
RX	159	raio X
R-x	5	raio X
R-X	3	raio X
r-x	2	raio X
RxABD	4	raio X abdominal
RxT	9	raio X de tórax
RXT	7	raio X de tórax
RxTx	13	raio X de tórax
RXTX	3	raio X de tórax
S/N	3	se necessário
s/n	2	se necessário
s/pp	7	sistólico / parede posterior
S1	2	sacro 1
SAD	2	sobrecarga atrial direita
SAE	2	sobrecarga atrial esquerda
SAMU	13	Serviço de Atendimento Médico de Urgência
satO2	2	saturação de oxigênio
SC	41	Subcutâneo
SCA	100	síndrome coronariana aguda
SCD	2	seio coronário distal
SF	32	soro fisiológico
SFA	4	sofrimento fetal agudo
SG	6	assobrevida global
SIADH	5	síndrome de secreção inapropriada de hormônio anti-diurético
SIDA	16	síndrome da imunodeficiência adquirida
SK	5	estreptoquinase
SL	97	Sublingual
SM	4	silagem de milho
SMC	3	serviço médico de cirurgia
SMD	2	síndrome mielodisplásica
SMO	3	serviço médico de oncologia
SN	16	sistema nervoso
SNC	7	sistema nervoso central
SNE	19	sonda nasoenteral
SNG	12	sonda nasogástrica
SO	7	sala de observação
SOG	2	sonda orogástrica
SOP	5	síndrome do ovário policístico
SP	2	sala de politraumatizados
SPA	4	Serviço de Pronto Atendimento
SPECT	2	spect cardíaco
SpO2	4	saturação percutânea de oxigênio
S-PP	11	sistólico - parede posterior
SR	3	Senhor
SSST	34	sem supradesnível do segmento ST
SST	15	supradesnível do segmento ST
STENT	204	Stent
STK	8	streptoquinase
SULFA	2	sulfametoxazol
SUS	5	Sistema Único de Saúde
SV	65	supraventricular
SVA	3	sonda uretral plástica

SVD	18	sobrecarga ventricular direita
SVE	7	sobrecarga ventricular esquerda
SVs	6	sinais vitais
T:0,7	2	troponina 0
T:2,1	2	troponina 2
T1	6	tempo 1
T12	2	torácica 12
T2	8	tempo 2
T4	21	Tiroxina
T4L	2	tetraiodotironina
TA	36	artéria transversa
TAB	4	transtorno afetivo bipolar
TAP	4	tempo de atividade de protrombina
TARV	4	terapia antiretroviral
TB	26	Tuberculose
TBC	12	Tuberculose
TC	412	tomografia computadorizada
TCC	10	tomografia computadorizada cardiovascular
TCE	85	tronco de coronária esquerda
TCEC	3	tempo de circulação extracorpórea
TE	25	teste ergométrico
TEP	38	tromboembolismo pulmonar
TET	17	tubo endotraqueal
TG	4	Triglicerídeos
TGI	5	trato gastrointestinal
TGO	20	Transaminase glutâmica oxalacética
TGP	15	Transaminase glutâmico-pirúvica
THB	2	transtorno de humor bipolar
TID	89	dilatação isquêmica transitória
TIFF	3	Tiffeneau
TIG	15	imunoglobulina antitetânica humana
TIG5	2	taxa de infusão de glicose
TILT	4	teste de inclinação ortostática
TIMI	181	medida do fluxo coronário e microvascular
TISQ	2	tempo de isquemia
TJV	2	transfusão intravascular
TMO	5	transplante de medula óssea
TMP	2	Trimetoprim
TnT	2	Troponina
TOT	3	tubo orotraqueal
TOXO	42	Toxoplasmose
TP	72	trabalho de parto
TPO	5	tireoperoxidase
TPP	8	trabalho de parto prematuro
TRAM	2	retalho miocutâneo transverso abdominal
TS	90	tipo sanguíneo
TSC	2	tiragem subcostal
TSE	2	spin-eco turbo
TSH	56	hormônio tireo-estimulante
TSH:5	2	hormônio tireo-estimulante 5
TSM	10	tipo sanguíneo da mãe
TSRN	3	tipo sanguíneo do recém-nato
TT	22	Transtorácico
TTG	2	teste de tolerância a glicose
TTO	30	Tratamento
TU	2	trato urinário

TV	39 taquicardias ventriculares
TVNS	5 taquicardia ventricular não sustentada
TVP	37 trombose venosa profunda
TVS	2 taquicardias ventriculares monomórficas sustentadas
TX	11 Tórax
TxH	2 transplante hepático
UBS	3 Unidade Básica de Saúde
UCA	4 cultura de urina
UCC	2 Unidade de Cardiopatias Congênitas
UCI	4 unidade de cuidados intermediários
UFC	2 unidades formadoras de colônia
UI	92 unidades internacionais
UR	2 Uréia
URC	4 Urocultura
URO	2 Urocultura
US	4 ultra-sonografia
UTI	66 Unidade de Terapia Intensiva
UTIN	4 Unidade de Terapia Intensiva Neonatal
UTIP	21 Unidade de Terapia Intensiva Pediátrica
VA	3 vias aéreas
VA-AD	2 valvula anterior do átrio direito
VAD	3 vincristina, adriplastina e dexametasona
VANCO	2 Vancomicina
VAS	2 vias aéreas superiores
VB	3 vesícula biliar
VCI	11 veia cava inferior
VCM	11 volume corpuscular médio
VCR	3 Vincristina
VD	131 ventrículo direito
VD/AD	7 ventrículo direito / átrio direito
VD>AD	5 ventrículo direito > átrio direito
VD17	2 ventrículo direito 17
VD28	4 ventrículo direito 28
VD-AD	9 ventrículo direito - átrio direito
VDRL	100 laboratório de pesquisa de doenças venéreas
VE	507 ventrículo esquerdo
VED	3 diâmetro diastólico
VEd	2 diástole do ventrículo esquerdo
VEDF	3 ventrículo esquerdo diástole final
VEF	9 volume expiratório forçado
VEF1	29 volume expiratório forçado 1
VES	3 diâmetro sistólico
VEs	2 sístole do ventrículo esquerdo
VESF	3 ventrículo esquerdo sístole final
VLP	10 videolaparoscopia
VM	54 ventilação mecânica
VMA	2 ácido vanilmandélico
VO	718 via oral
vO	2 via oral
VO2	12 volume de oxigênio
Vo2	2 volume de oxigênio
VP	32 Vasopressina
VPA	8 valproato de sódio
VPP	7 ventilação com pressão positiva
VR	2 via retal
VSG	16 velocidade de sedimentação globular

ANEXO B – EXEMPLO DO CORPUS UTILIZADO PARA CLASSIFICAÇÃO SINTÁTICA

Cardiopatia_N isquêmica_ADJ com_PREP disfunção_N ventricular-_ADJ Ecocardiografia_N de_PREP 1_NUM 3.09.2005_NUM :_: Classe_N Funcional_ADJ 1_NUM em_PREP a_ART alta_N hospitalar_ADJ infarto_N agudo_ADJ do_PREP miocárdio_N anterior_ADJ extenso_ADJ em_PREP Janeiro_N de_PREP 2005_NUM -_- tratamento_N conservador_ADJ pois_CJ chegou_V com_PREP >12_NUM horas_N ._. Paciente_N internou_V por_PREP apresentar_V piora_ADJ funcional_ADJ caracterizada_PCP por_PREP dispnéia_N a_PREP os_N moderados_ADJ. Pcte_N interna_V eletivamente_ADV para_PREP fechamento_N de_PREP colostomia_N ._. Durante_PREP internação_N apresentou_V episódios_N repetidos_ADJ de_PREP obstrução_N intestinal_ADJ ,,, prejudicando_V o_ART preparo_N para_PREP cirurgia_N ._. Durante_PREP procedimento_N houve_V contaminação_N grosseira_ADJ de_PREP o_ART campo_N cirúrgico_ADJ ,,, sendo_V optado_PCP por_PREP a_ART realização_N de_PREP nova_ADJ colostomia_N terminal_ADJ ._. Paciente_N com_PREP os_ART seguintes_ADJ diagnósticos_N :_: Ex-tabagista_N Enzimas_N cardíacas_ADJ sem_PREP. Recebe_V alta_N com_PREP tratamento_N para_PREP Hipertensão_N arterial_ADJ sistêmica_ADJ com_PREP plano_N de_PREP seguir_V acompanhamento_N clínico_ADJ com_PREP seu_PRN médico_N ._. Diversas_ADJ internações_N por_PREP descompensação_N de_PREP insuficiência_N cardíaca_ADJ congestiva_ADJ e_CJ doença_N pulmonar_ADJ obstrutiva_ADJ crônica_ADJ desde_PREP 2004_NUM ._. Paciente_N vem_V a_PREP a_ART emergência_N com_PREP queixa_N de_PREP dispnéia_N a_PREP os_ART pequenos_ADJ esforços_N ((sem_PREP tosse_N ou_CJ expectoração_N)) e_CJ dor_N importante+_N aumento_N de_PREP volume_N em_PREP joelho_N esquerdo_ADJ ._.

ANEXO C – QUESTIONÁRIO UTILIZADO PARA VALIDAÇÃO DA FERRAMENTA

O foco deste trabalho é identificar os diagnósticos dentro de uma narrativa clínica. A correlação feita com a Classificação Internacional da Doença – CID-10 é uma sugestão que pode ser aprimorada nessa ferramenta de apoio.

Nos casos apresentados a seguir, confirme se os diagnósticos apresentados estão corretos. Caso não, no espaço indicado ao lado, por gentileza preencha com os diagnósticos faltantes.

Na segunda fase, verifique a correlação com a CID-10. Caso não esteja correto, por gentileza indique as correções.

Caso 1) Paciente cardiopata isquêmico, submetido previamente a estudo eletrofisiológico com indução de taquicardia ventricular, foi submetido a implante de cardioversor-desfibrilador com sucesso.

Resultados da ferramenta:

Diagnósticos: cardiopata, taquicardia.

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

CID Cardiopata

Grupo(s): [I30-I52] Outras formas de doença do coração

Categoria(s): (I42) Cardiomiopatias / (I43) Cardiomiopatia em doenças classificadas em outra parte / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas

CID Taquicardia

Grupos(s): [R00-R09] Sintomas e sinais relativos ao aparelho circulatório e respiratório / [I30-I52] Outras formas de doença do coração

Categoria(s): (R00) Anormalidades do batimento cardíaco / (I47) Taquicardia paroxística

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 2) Paciente internou apresentando pressão arterial elevada. Durante a internação foi aumentada a dose de betabloqueador, introduzido diurético e ansiolítico, obtendo-se controle adequado da pressão arterial. Função renal normal. Ecodoppler de artérias renais com estenose de 70 % na artéria renal direita (não se afastando a possibilidade de acotovelamento da mesma) e estenose de 50 a 70 % na artéria renal esquerda. Em vista do fácil controle da pressão arterial, da paciente encontrar-se bem clinicamente e com função renal normal, opta-se pelo tratamento conservador, devendo fazer revisões periódicas ambulatoriais e ser reavaliada pelo seu nefrologista.

Resultados da ferramenta:

Diagnósticos: estenose

Sim ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Não ()		
Sim ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?
Não ()		

CID Estenose

Grupo(s): [I60-I69] Doenças cerebrovasculares, [N30-N39] Outras doenças do aparelho urinário / [Q38-Q45] Outras malformações congênitas do aparelho digestivo

Categoria(s): (I65) Oclusão e estenose de artérias pré-cerebrais que não resultam em infarto cerebral / (I66) Oclusão e estenose de artérias cerebrais que não resultam em infarto cerebral / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas / (N35) Estenose da uretra / (Q41) Ausência, atresia e estenose congênita do intestino delgado / (Q42) Ausência, atresia e estenose congênita do cólon

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 3) Paciente hipertensa e portadora de angina estável, encaminhada para angioplastia. Realizada ACTP com implante de Stent em DAE, com sucesso angiográfico e sem intercorrências.

Resultados da ferramenta: hipertensa, angina.

Sim ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Não ()		
Sim ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?
Não ()		

CID Hipertensa

Grupo(s): [I10-I15] Doenças hipertensivas / [O10-O16] Edema, proteinúria e transtornos hipertensivos na gravidez, no parto e no puerpério

Categoria(s): (I10) Hipertensão essencial (primária) / (I11) Doença cardíaca hipertensiva / (I12) Doença renal hipertensiva / (I13) Doença cardíaca e renal

hipertensiva / (I15) Hipertensão secundária / (O10) Hipertensão pré-existente complicando a gravidez, o parto e o puerpério / (O11) Distúrbio hipertensivo pré-existente com proteinúria superposta / (O12) Edema e proteinúria gestacionais [induzidos pela gravidez], sem hipertensão / (O13) Hipertensão gestacional [induzida pela gravidez] sem proteinúria significativa / (O14) Hipertensão gestacional [induzida pela gravidez] com proteinúria significativa / (O16) Hipertensão materna não especificada

CID Angina

Grupo(s): [I20-I25] Doenças isquêmicas do coração

Categoria(s): (I20) Angina pectoris

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 4) PACIENTE 74 ANOS, DISLIPIDÊMICO, DIABÉTICO, CARDIOPATA ISQUÊMICO , CAT FEITO NO HOSPITAL DA ULBRA COM LESÃO EM ADA (ÓSTIO E SEGMENTO MÉDIO), CD COM LESÃO EM SEGMENTO PROXIMAL , INTERNA PARA REALIZAÇÃO DE ACTP + IMPLANTE DE STENT EM CD. PROCEDIMENTO SEM INTERCORRÊNCIAS E COM BOM RESULTADO ANGIOGRÁFICO.

Resultados da ferramenta: cardiopata, dislipidêmico, diabético.

Sim ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Não ()		
Sim ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?
Não ()		

CID Cardiopata**Grupo(s):** [I30-I52] Outras formas de doença do coração**Categoria(s):** (I42) Cardiomiopatias / (I43) Cardiomiopatia em doenças classificadas em outra parte / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas**CID Dislipidêmico****Grupo(s):** [E70-E90] Distúrbios metabólicos**Categoria(s):** (E78) Distúrbios do metabolismo de lipoproteínas e outras lipidemias**CID Diabético****Grupo(s):** [E10-E14] Diabetes mellitus / O20-O29 Outros transtornos maternos relacionados predominantemente com a gravidez**Categoria(s):** (E10) Diabetes mellitus insulino-dependente / (E11) Diabetes mellitus não-insulino-dependente / (E12) Diabetes mellitus relacionado com a desnutrição / (E13) Outros tipos especificados de diabetes mellitus / (E14) Diabetes mellitus não especificado / O24 Diabetes mellitus na gravidez

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 5) Paciente hipertenso, com história de infartos prévios e ICC, transferido de sua cidade por ter apresentado IAM complicado por angina pós-IAM. Realizou cateterismo cardíaco que mostrou oclusões em DAE e CD, apresentando, no dia seguinte ao CAT, quadro de angina instável, com necessidade de internação em CTI. Submetido a Cirurgia de revascularização miocárdica em 19/09/02 (mamária para DAE e safena para DPCD), tendo apresentado no 3o dia de pós-operatório quadro de fibrilação atrial com alta resposta ventricular com repercussão hemodinâmica (necessidade de cardioversão elétrica com urgência, uso de amiodarona EV e entubação oro-traqueal, sem recorrência da arritmia após

instituição da amiodarona. Apresentou também quadro de anemia sintomática, sendo transfundido com 2 unidades de concentrado de hemáceas, e perda de função renal. Após, evoluiu bem no período pós-operatório, sem outras complicações. Recebe alta em bom estado, com boa cicatrização da ferida operatória, sem sinais de descompensação de ICC, em ritmo sinusal, com plano de retorno ao médico assistente para manter acompanhamento.

Resultados da ferramenta: hipertenso, angina, anemia, insuficiência cardíaca, arritmia.

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

CID Hipertenso

Grupo(s): [I10-I15] Doenças hipertensivas / [O10-O16] Edema, proteinúria e transtornos hipertensivos na gravidez, no parto e no puerpério

Categoria(s): (I10) Hipertensão essencial (primária) / (I11) Doença cardíaca hipertensiva / (I12) Doença renal hipertensiva / (I13) Doença cardíaca e renal hipertensiva / (I15) Hipertensão secundária / (O10) Hipertensão pré-existente complicando a gravidez, o parto e o puerpério / (O11) Distúrbio hipertensivo pré-existente com proteinúria superposta / (O12) Edema e proteinúria gestacionais [induzidos pela gravidez], sem hipertensão / (O13) Hipertensão gestacional [induzida pela gravidez] sem proteinúria significativa / (O14) Hipertensão gestacional [induzida pela gravidez] com proteinúria significativa / (O16) Hipertensão materna não especificada

CID Angina

Grupo(s): [I20-I25] Doenças isquêmicas do coração

Categoria(s): (I20) Angina pectoris

CID Anemia

Grupos(s): D50-D53 Anemias nutricionais / D55-D59 Anemias hemolíticas / D60-D64 Anemias aplásticas e outras anemias

Categoria(s): D50 Anemia por deficiência de ferro / D51 Anemia por deficiência de vitamina B12 / D52 Anemia por deficiência de folato / D53 Outras anemias nutricionais / D55 Anemia devida a transtornos enzimáticos / D56 Talassemia / D57 Transtornos falciformes / D58 Outras anemias hemolíticas hereditárias / D59 Anemia hemolítica adquirida / D60 Aplasia pura da série vermelha, adquirida [eritroblastopenia] / D61 Outras anemias aplásticas / D62 Anemia aguda pós-hemorrágica / D63* Anemia em doenças crônicas classificadas em outra parte / D64 Outras anemias

CID Insuficiência Cardíaca

Grupo(s): [J95-J99] Outras doenças do aparelho respiratório / [K70-K77] Doenças do fígado / [N17-N19] Insuficiência renal / [I30-I52] Outras formas de doença do coração /

Categoria(s): (I50) Insuficiência cardíaca / (J96) Insuficiência respiratória não classificada de outra parte / (K72) Insuficiência hepática não classificada em outra parte / (N17) Insuficiência renal aguda / (N18) Insuficiência renal crônica / (N19) Insuficiência renal não especificada

CID Arritmia

Grupo(s): [I30-I52] Outras formas de doença do coração

Categoria(s): (I49) Outras arritmias cardíacas

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 6) PACIENTE DIABETICA, HIPERTENSA, CARDIOPATA ISQUEMICA INTERNA PARA REALIZAÇÃO DE ACTP+STENT EM ADAE. PROCEDIMENTO

REALIZADO COM SUCESSO ANGIOGRAFICO. PACIENTE APRESENTOU ELEVÇÃO DA PA LOGO APÓS O PROCEDIMENTO, A QUAL APÓS FOI NORMALIZADA E SE MANTEVE ESTAVEL.RECEBE ALTA EM BOAS CONDIÇÕES COM ORIENTAÇÃO PARA MANTER MEDICAÇÕES PREVIAMENTE EM USO E TICLOPIDINA 250MG 12/12 H POR 30 DIAS.

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

Resultados da ferramenta: diabética, hipertensa, cardiopata isquêmica.

CID Diabética

Grupo(s): [E10-E14] Diabetes mellitus / O20-O29 Outros transtornos maternos relacionados predominantemente com a gravidez

Categoria(s): (E10) Diabetes mellitus insulino-dependente / (E11) Diabetes mellitus não-insulino-dependente / (E12) Diabetes mellitus relacionado com a desnutrição / (E13) Outros tipos especificados de diabetes mellitus / (E14) Diabetes mellitus não especificado / O24 Diabetes mellitus na gravidez

CID Cardiopata Isquêmica

Grupo(s): [I30-I52] Outras formas de doença do coração

Categoria(s): (I42) Cardiomiopatias / (I43) Cardiomiopatia em doenças classificadas em outra parte / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas

CID Hipertensa

Grupo(s): [I10-I15] Doenças hipertensivas / [O10-O16] Edema, proteinúria e transtornos hipertensivos na gravidez, no parto e no puerpério

Categoria(s): (I10) Hipertensão essencial (primária) / (I11) Doença cardíaca hipertensiva / (I12) Doença renal hipertensiva / (I13) Doença cardíaca e renal hipertensiva / (I15) Hipertensão secundária / (O10) Hipertensão pré-existente complicando a gravidez, o parto e o puerpério / (O11) Distúrbio hipertensivo pré-existente com proteinúria superposta / (O12) Edema e proteinúria gestacionais [induzidos pela gravidez], sem hipertensão / (O13) Hipertensão gestacional [induzida pela gravidez] sem proteinúria significativa / (O14) Hipertensão gestacional [induzida pela gravidez] com proteinúria significativa / (O16) Hipertensão materna não especificada

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 7) Paciente com história de dislipidemia, tabagismo e etilismo admitido em 03/08/02 por quadro de IAM em parede antero-septal. Realizada ACTP primária na 5ª hora do início da dor, com colocação de stent em segmento proximal da DAE. Boa evolução clínica, tendo apresentado quadro de pericardite pós-IAM não-complicada manejada com AINE. Recebe alta clinicamente estável com retorno para o ambulatório de Cardiologia do HCPA. Medicamentos em uso na alta: - AAS 100 mg VO 1x/dia; - Ticlopidina 250 mg VO 12/12 horas por 30 dias; - Sinvastatina 20 mg VO 1x/dia; - Captopril 25 mg VO 8/8 horas; - Propranolol 40 mg VO 8/8 horas; - Indometacina 50 mg VO 8/8 horas por 5 dias.

Resultados da ferramenta: pericardite, dislipidemia.

Sim ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Não ()		
Sim ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?
Não ()		

CID Pericardite**Grupo(s):** [I30-I52] Outras formas de doença do coração**Categoria(s):** (I30) Pericardite aguda / (I32) Pericardite em doenças classificadas em outra parte**CID Dislipidemia****Grupo(s):** [E70-E90] Distúrbios metabólicos**Categoria(s):** (E78) Distúrbios do metabolismo de lipoproteínas e outras lipidemias

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 8) Paciente hipertensa, diabética, dislipidêmica, internou com quadro de angina instável, estabilizada com tratamento clínico, realizando cateterismo cardíaco que mostrou lesão grave em DAE proximal e lesão suboclusiva em CD (pequena, não-dominante). Transferida ao HCPA para angioplastia, desenvolveu quadro de colecistite aguda alitiásica com pequeno abscesso peri-vesícula; tratou com ampicilina por 32 dias, com melhora clínica, sendo submetida, após, a colecistectomia convencional com exploração de vias biliares. Apresentou boa evolução pós-operatória, com suspensão do ATB no sétimo P.O. Realizada angioplastia com implante de Stent em DAE após a cirurgia, com sucesso angiográfico e sem complicações (procedimento realizado apenas após a colecistectomia pela necessidade de uso de anti-agregantes plaquetários). Identificados nódulos pulmonares, com plano de encaminhamento para pneumologista para acompanhamento.

Resultados da ferramenta: hipertensa, angina, colecistite, diabética, dislipidêmica.

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

CID Hipertensa

Grupo(s): [I10-I15] Doenças hipertensivas / [O10-O16] Edema, proteinúria e transtornos hipertensivos na gravidez, no parto e no puerpério

Categoria(s): (I10) Hipertensão essencial (primária) / (I11) Doença cardíaca hipertensiva / (I12) Doença renal hipertensiva / (I13) Doença cardíaca e renal hipertensiva / (I15) Hipertensão secundária / (O10) Hipertensão pré-existente complicando a gravidez, o parto e o puerpério / (O11) Distúrbio hipertensivo pré-existente com proteinúria superposta / (O12) Edema e proteinúria gestacionais [induzidos pela gravidez], sem hipertensão / (O13) Hipertensão gestacional [induzida pela gravidez] sem proteinúria significativa / (O14) Hipertensão gestacional [induzida pela gravidez] com proteinúria significativa / (O16) Hipertensão materna não especificada

CID Angina

Grupo(s): [I20-I25] Doenças isquêmicas do coração

Categoria(s): (I20) Angina pectoris

CID Colecistite

Grupo(s): [K80-K87] Transtornos da vesícula biliar, das vias biliares e do pâncreas

Categoria(s): (K81) Colecistite

CID Diabética

Grupo(s): [E10-E14] Diabetes mellitus / O20-O29 Outros transtornos maternos relacionados predominantemente com a gravidez

Categoria(s): (E10) Diabetes mellitus insulino-dependente / (E11) Diabetes mellitus não-insulino-dependente / (E12) Diabetes mellitus relacionado com a desnutrição / (E13) Outros tipos especificados de diabetes mellitus / (E14) Diabetes mellitus não especificado / O24 Diabetes mellitus na gravidez

CID Dislipidêmica

Grupo(s): [E70-E90] Distúrbios metabólicos

Categoria(s): (E78) Distúrbios do metabolismo de lipoproteínas e outras lipidemias

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 9) PACIENTE DIABÉTICA , HIPERTENSA, DISLIPIDÊMICA , CARDIOPATA ISQUÊMICA DE LONGA DATA , COM ACTP + STENT EM 1996 NA ART.DECENDENTE ANTERIOR, APRESENTANDO ANGINA DE REPOUSO NAS ÚLTIMAS SEMANAS. REALIZOU CATETERISMO CARDÍACO QUE MOSTROU DA COM REESTENOSE INTRA-STENT ,1' Dg COM LESÃO GRAVE ENVOLVENDO ÓSTIO, CX COM IRREGULARIDADES DIFUSAS E CD OCLUÍDA. VENTRICULOGRAFIA COM ACINESIA ANTERIOR E DISFUNÇÃO LEVE DE VE. EVOLUIU SATISFATORIAMENTE DURANTE A INTERNAÇÃO, SINTOMAS CONTROLADOS. MEDICAÇÕES PRESCRITAS NA ALTA HOSPITALAR: #METOPROLOL 100 MG DE 12/12 H; #ENALAPRIL 10 MG ÀS 8 H; #AAS 100 MG NO ALMOÇO; #ISSORBIDA 40 MG 3X/DIA; #SINVASTATINA 20 MG/NOITE; #TIROXINA 325 MCG 1X/DIA; #GLIBENCLAMIDA 10MG AC E AJ; #CARBAMAZEPINA 200MG:2CP/MANHÃ E 3 CP/NOITE; #LEVOMEPRIMAZINA 25MG:1 CP/MANHÃ E 2 CP/NOITE.

Resultados da ferramenta: diabética, estenose, hipertensa, dislipidêmica, cardiopata.

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

CID Diabética

Grupo(s): [E10-E14] Diabetes mellitus / O20-O29 Outros transtornos maternos relacionados predominantemente com a gravidez

Categoria(s): (E10) Diabetes mellitus insulino-dependente / (E11) Diabetes mellitus não-insulino-dependente / (E12) Diabetes mellitus relacionado com a desnutrição / (E13) Outros tipos especificados de diabetes mellitus / (E14) Diabetes mellitus não especificado / O24 Diabetes mellitus na gravidez

CID Estenose

Grupo(s): [I60-I69] Doenças cerebrovasculares, [N30-N39] Outras doenças do aparelho urinário / [Q38-Q45] Outras malformações congênitas do aparelho digestivo

Categoria(s): (I65) Oclusão e estenose de artérias pré-cerebrais que não resultam em infarto cerebral / (I66) Oclusão e estenose de artérias cerebrais que não resultam em infarto cerebral / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas / (N35) Estenose da uretra / (Q41) Ausência, atresia e estenose congênita do intestino delgado / (Q42) Ausência, atresia e estenose congênita do cólon

CID Hipertensa

Grupo(s): [I10-I15] Doenças hipertensivas / [O10-O16] Edema, proteinúria e transtornos hipertensivos na gravidez, no parto e no puerpério

Categoria(s): (I10) Hipertensão essencial (primária) / (I11) Doença cardíaca hipertensiva / (I12) Doença renal hipertensiva / (I13) Doença cardíaca e renal hipertensiva / (I15) Hipertensão secundária / (O10) Hipertensão pré-existente

complicando a gravidez, o parto e o puerpério / (O11) Distúrbio hipertensivo pré-existente com proteinúria superposta / (O12) Edema e proteinúria gestacionais [induzidos pela gravidez], sem hipertensão / (O13) Hipertensão gestacional [induzida pela gravidez] sem proteinúria significativa / (O14) Hipertensão gestacional [induzida pela gravidez] com proteinúria significativa / (O16) Hipertensão materna não especificada

CID Cardiopata

Grupo(s): [I30-I52] Outras formas de doença do coração

Categoria(s): (I42) Cardiomiopatias / (I43) Cardiomiopatia em doenças classificadas em outra parte / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas

CID Dislipidêmica

Grupo(s): [E70-E90] Distúrbios metabólicos

Categoria(s): (E78) Distúrbios do metabolismo de lipoproteínas e outras lipidemias

Sim ()	A correlação com o grupo e categoria CID-10 está correta? .	Indique as correções:
Não ()		

Caso 10) # CARDIOPATIA ISQUÊMICA GRAVE # ICC - FE=41% # HAS # EAP NA BAIXA # ECG: BRE Paciente chegou à Emergência com quadro de dispnéia súbita. Realizado R-x tórax c/ achados de congestão pulmonar e ECO Doppler MiS sem sinais de trombose. Paciente submetido a cateterismo cardíaco, que evidenciou lesão em tronco de artéria coronária esquerda. A lesão apresentada pelo paciente é indicação de cirurgia de revascularização do miocárdio; entretanto, o procedimento cirúrgico foi suspenso por decisão da família (filha) .Após conversa com a equipe, o paciente referiu interesse em realizar o procedimento, e sua filha deixou a decisão a critério do paciente. Plano- Retorno ao ambulatório de cirurgia cardíaca para marcação de cirurgia eletiva. - Usar medicamentos prescritos.

Resultados da ferramenta: cardiopatia isquêmica, trombose

Sim () Não ()	Estes resultados realmente são considerados diagnósticos?	Indique os termos que não são diagnósticos:
Sim () Não ()	A ferramenta encontrou todos os diagnósticos?	Qual (s) está faltando?

CID Cardiopatia Isquêmica

Grupo(s): [I30-I52] Outras formas de doença do coração

Categoria(s): (I42) Cardiomiopatias / (I43) Cardiomiopatia em doenças classificadas em outra parte / (I51) Complicações de cardiopatias e doenças cardíacas mal definidas

CID Trombose

Grupo(s): I70-I79 Doenças das artérias, das arteríolas e dos capilares / I80-I89 Doenças das veias, dos vasos linfáticos e dos gânglios linfáticos, não classificadas em outra parte

Categoria(s): I74 Embolia e trombose arteriais / I81 Trombose da veia porta / I82 Outra embolia e trombose venosas

Sim () Não ()	A correlação com o grupo e categoria CID-10 está correta?	Indique as correções:
--------------------	--	-----------------------