



Servidor de terminologias para a identificação e padronização de conceitos médicos a partir de anotações clínicas

Allan Ferreira¹, Nelson J. Oliveira Miranda², Joice B. Machado Marques³, Reinaldo P. Peres⁴, Fabrício A. Gualdani¹, Dilvan A. Moreira², Leonardo C. Botega¹

1 Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, SP

2 Universidade de São Paulo, São Carlos, SP.

3 Sofya Tecnologia LTDA.

4 Universidade Estadual Paulista Júlio de Mesquita Filho, Bauru, SP

*allan.ferreira1983@unesp.br, nelson.miranda@usp.br, joicemachado@1sti.com.br, reinaldo.peres@unesp.br
fabrício.gualdani@unesp.br, dilvan@icmc.usp.br, leonardo.botega@unesp.br*

Resumo. Os sistemas de apoio à decisão clínica podem melhorar o serviço de diagnóstico do paciente. No entanto, muitas vezes os dados dos pacientes estão distribuídos em diferentes sistemas, com cópias desconectadas e sem sincronização. Uma maneira de unificar o acesso a esses dados, possibilitando o auxílio à tomada de decisão clínica, é a categorização das informações dos pacientes. Nesse sentido, uma abordagem utilizada para enriquecer e categorizar um conjunto de dados é a construção de bases de conhecimento estruturadas usando conceitos de ontologias e terminologias, chamadas de servidores de terminologias. Esses grandes dicionários usam identificadores para mapear termos médicos, muitas vezes em línguas diferentes, aos conceitos que eles representam. Este trabalho tem como objetivo desenvolver um servidor de terminologias para classificar e categorizar termos clínicos, de maneira automatizada, levando em conta a sua grafia na língua portuguesa e no jargão médico brasileiro, com diferentes vocábulos regionais. Ele vai facilitar também a redução da redundância de dados e outros erros. A partir do uso do servidor de terminologias, os conceitos poderão ser consultados e atualizados de maneira automática. Este é um trabalho em andamento. O presente artigo apresenta sua proposta, arquitetura e os resultados obtidos inicialmente.

Palavras-chave: Sistema de Registros Eletrônicos de Saúde; Terminologias; Sistemas de apoio à decisão clínica.

1. INTRODUÇÃO

Nas últimas décadas, há uma ênfase na importância de estabelecer registros eletrônicos de saúde nas práticas clínicas visando aumentar a segurança do paciente, reduzir erros médicos, melhorar a eficiência e reduzir custos (1). Nesse sentido, os sistemas de apoio à decisão clínica colaboram diretamente para melhorar a qualidade do serviço de diagnóstico do paciente e a minimização de erros, assim como o aumento da produtividade dos profissionais. Dados de pacientes estão geralmente distribuídos em diferentes fontes e sistemas, muitas vezes repetidamente, sem conexões e automatizações para auxiliar as decisões clínicas, seja por meio de recomendações ou de atualizações automáticas (2). O conhecimento do especialista é fundamental para definir a representação dos dados do paciente que entram no sistema através de diferentes fontes. Uma forma de unificar o acesso

aos dados, promover uma governança mais eficiente e possibilitar o auxílio à tomada de decisão clínica é a categorização das informações dos pacientes. Essas categorizações dependem de diferentes informações, tais como: diagnóstico principal que levou à internação; diagnóstico atual das condições clínicas; cirurgias e intervenções. Uma abordagem utilizada para enriquecer e categorizar o conjunto de dados do paciente é a construção de bases de conhecimento estruturadas usando conceitos de ontologias e terminologias, representados por identificadores. Esses conceitos são descritos usando termos da área médica em inglês. Na maioria dos casos, não há tradução para a língua portuguesa e, muito menos, uma adaptação para o jargão médico brasileiro, incluindo seus diferentes vocábulos regionais. Para resolver esse problema, servidores de terminologia podem ser empregados. Serviços de terminologia compõem-se de uma representação conceitual do



conhecimento médico, possibilitando relações entre esses conceitos.

Este trabalho tem como objetivo desenvolver um servidor de terminologias para: classificar e categorizar termos clínicos de maneira automatizada; colaborar para a redução de erros; e otimizar a rotina do profissional, por meio da codificação dos termos clínicos nos sistemas de apoio à decisão. A partir do uso do servidor de terminologias os dados poderão ser consultados e atualizados, permitindo também a inserção de sinônimos em várias línguas para termos e definições de conceitos. A proposta, a arquitetura e os primeiros resultados que foram obtidos são apresentados nas próximas seções.

2. MATERIAIS E MÉTODOS

Os servidores de terminologia auxiliam na representação dos conceitos a fim de evitar ambiguidade dos termos. Essa ambiguidade pode ser criada em virtude da multiplicidade de sentidos de uma palavra ou de palavras que apresentam a mesma estrutura fonológica, mas apresentam significados divergentes dependendo do contexto em que estão inseridas.

As categorizações dos conceitos podem ocorrer por meio da utilização de padrões internacionais como as classificações CID (Classificação Internacional de Doenças), LOINC (Logical Observation Identifiers Names and Codes); terminologias como a SNOMED-CT, além dos padrões de saúde nacionais como o SIGTAP (Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS) e TUSS (Terminologia Unificada da Saúde Complementar). O servidor de terminologias foi desenvolvido com base no sistema de gerenciamento de terminologia de código aberto chamado Open Concept Lab - OCL e faz parte de uma aplicação para codificar automaticamente as doenças e os procedimentos usados em uma instituição de saúde. O servidor de terminologias proporcionará a consulta, atualização e inserção de sinônimos para os conceitos identificados.

Atualmente, o servidor contém os vocabulários CID, TUSS, SIGTAP e o conjunto GPS (Global Patient Set) da SNOMED-CT. A partir do reconhecimento das entidades nomeadas na etapa

de anotação, a aplicação identifica os códigos referentes às doenças, procedimentos e sintomas.

Vale ressaltar que, outros vocabulários podem ser facilmente adicionados ao servidor de acordo com o domínio da aplicação. A figura 1 ilustra o esquema de serviços associados ao servidor.

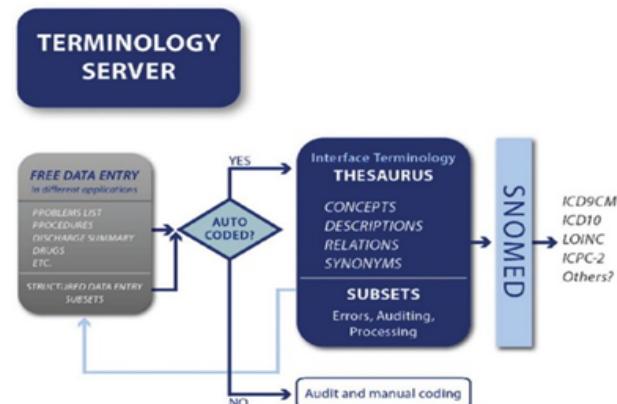


Figura 1. Esquema de serviços do servidor de terminologia

Na próxima seção são apresentadas as etapas de desenvolvimento do servidor de terminologias e seus respectivos resultados.

3. RESULTADOS

Para a implementação do servidor de terminologias, foi utilizado o padrão de projeto (*Design Pattern*) chamado *Proxy*. Ele oferece acesso às funcionalidades de um determinado recurso de maneira controlada, desacoplada e com possibilidade de implementação extra, quando necessário.

A aplicação foi desenvolvida com base na arquitetura REST para proporcionar maior abrangência de interoperabilidade com diversos clientes, representando o componente *Proxy*. Sua utilização é baseada no framework .NET 6, que tem acesso à uma instância concreta do OCL, disponibilizando um catálogo de APIs nativas que permitem a produção, manutenção e recuperação das informações em seu banco de dados. O componente *Client* representa a instância do Doccano, que é a ferramenta de anotação utilizada neste trabalho, e também do *Consumer*, software que enviará as solicitações de reconhecimento de entidade para o proxy.



A Figura 2 ilustra o fluxo de dados associado ao servidor de terminologias.

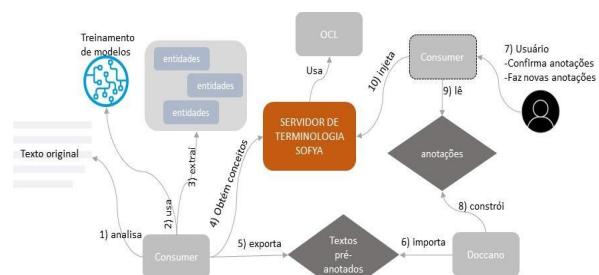


Figura 2. Fluxo de dados do servidor de terminologias

O fluxo de dados da Figura 2 tem como objetivo codificar automaticamente as entidades médicas identificadas, por meio do uso do servidor de terminologias. Para validar uma proposta, é necessário obedecer ao seguinte percurso: um usuário envia um conjunto de entidades nomeadas ao servidor e tem como retorno uma lista de conceitos pertencentes aos vocabulários padronizados. Os resultados são enviados para a plataforma de anotação. O usuário acessa a ferramenta de anotação e visualiza as sugestões de conceitos padronizados e reconhecidos pelo servidor, onde pode confirmar, excluir ou adicionar os conceitos. O resultado dessa validação do usuário é inserido novamente na base de conhecimento do OCL por meio do proxy. O Proxy será utilizado para melhoria de inferência em suas próximas requisições. Um usuário também pode requisitar novas associações de vocabulário ao servidor, para utilizar como insumo para re-treinamento do seu algoritmo de machine learning.

Considerando a Figura 2, no passo 1 o software consumidor recebe os textos clínicos provenientes dos registros eletrônicos do hospital. Nos passos 2 e 3, o software utiliza técnicas de NLP (Natural language processing) e Machine Learning para conseguir reconhecer e extrair as entidades nomeadas presentes no texto, ainda sem a definição de vocabulários, mas com uma semântica bem definida. Por exemplo, as entidades gripe e diabetes são classificadas como doenças; a entidade curativa é classificada como tratamento; febre e náusea são classificadas como sintomas.

No passo 4, o consumidor envia uma requisição para o servidor, com as entidades reconhecidas.

O proxy, por sua vez, acessa a base de dados do OCL e faz uma busca por termos principais e sinônimos. Por meio do uso do servidor de terminologias, o jargão médico brasileiro pode ser mapeado e associado às classificações nacionais e internacionais como a SNOMED-CT que possui um vocabulário clínico padronizado e que possibilita a interoperabilidade. O resultado dessa busca é formatado e devolvido ao consumidor, com detalhes dos conceitos, como código e vocabulário pertencente. Como é apresentado na Figura 3, a busca pelo termo “gripe” possui diferentes codificações com suas respectivas terminologias, com base nas classificações SIGTAP e CID.

```
{
  "inputTerm": "gripe",
  "results": [
    {
      "vocabulary": "SIGTAP",
      "code": "0303140151",
      "nameAtVocabulary": "TRATAMENTO DE PNEUMONIAS OU INFLUENZA",
      "conceptClass": "Procedure"
    },
    {
      "vocabulary": "CID",
      "code": "J09",
      "nameAtVocabulary": "Influenza [gripe] devida a vírus ident",
      "conceptClass": "Diagnosis"
    }
  ]
}
```

Figura 3. Resultado de busca de termo pelo proxy

Nos passos 5 e 6, o consumidor envia os resultados obtidos através do seu algoritmo de NLP para a plataforma de anotação. Ela importa os dados pré-formatados e os torna disponíveis com as devidas anotações, para os usuários da plataforma iniciarem suas avaliações.

No passo 7, o usuário acessa a plataforma de anotação e realiza operações como: aprovar a associação dos termos obtidos através das inferências realizadas pelos passos anteriores e é capaz também de realizar novas associações no texto, de acordo com o conhecimento de especialistas de domínio. É importante ressaltar, que, nesse passo, o modelo arquitetural recebe um grande enriquecimento informacional pois consegue aproveitar o conhecimento dos especialistas de domínio.

No passo 8, são utilizadas 2 técnicas para extrair as anotações realizadas: (i) exportação nativa da plataforma, em formato de arquivo JSON, da qual são apontadas as devidas associações dos termos com os vocabulários; (ii) exportação através de leitura direta do banco de dados, tornando possível o acesso direto a qualquer



dado em qualquer tabela da instância da ferramenta de anotação. Nos passos 9 e 10, um consumidor lê os resultados obtidos e insere novamente na base do OCL para validar o conceito. O servidor realiza todo o trabalho de desempacotamento e indexação dos termos adequados relacionados a cada vocabulário.

Nessa etapa é importante destacar a atuação do proxy como um agente que orquestra e simplifica toda a gestão de dados do servidor de terminologias. Esse comportamento é demonstrado na inserção de novos termos. Por exemplo, novos termos, como “terçol” e “pedras no rim”, podem ser adicionados e associados aos respectivos códigos CIDS de forma simplificada, sem conhecimento técnico dos detalhes de implementação do OCL. O passo 10 possibilita ao consumidor inicial realizar uma busca por um catálogo de termos associados a uma determinada classe de entidades, para que possa re-treinar o modelo de reconhecimento de entidades.

4. DISCUSSÃO

Até o momento, o servidor de terminologias possibilitou a otimização do trabalho dos especialistas na validação dos termos classificados e possibilitou o incremento de sinônimos utilizados na língua portuguesa. A utilização dos vocabulários já estabelecidos na prática médica e referenciados nos sistemas de saúde poderá facilitar a extração das classes de conceitos mais importantes, ao considerar as notas clínicas do cuidado do paciente. Há expectativa de suporte à tomada de decisão clínica por meio de uma padronização dos termos e vocabulários médicos utilizando o servidor de terminologias. O OCL é alimentado por terminologias clínicas padronizadas internacionais, como a CID-10 e SNOMED-CT, possibilitando assim um melhor aproveitamento da estrutura de

dados construída facilitando a interoperabilidade entre diferentes sistemas.

5. CONCLUSÃO

O objetivo deste trabalho foi introduzir o desenvolvimento do servidor de terminologias para classificar e categorizar termos clínicos, auxiliando nos sistemas de suporte à decisão. Foram utilizadas as categorizações CID-10, SNOMED-CT, TUSS e SIGTAP para criar um vocabulário padronizado e em consenso com as regulamentações do Ministério da Saúde. Nos próximos passos, a codificação automática dos prontuários poderá ser potencializada e enriquecida pelos serviços de terminologia. Este trabalho está em andamento e deve contemplar também a geração de recomendações a partir do uso de modelos de treinamento e ontologias de domínio, tendo como vocabulário as bases de conhecimento que estão sendo construídas no servidor de terminologias.

REFERÊNCIAS

1. Cramer K-A, Maher L, Van Dam P, Prior S. Personal electronic healthcare records: What influences consumers to engage with their clinical data online? A literature review. *Health Information Management Journal*. 2020 Jan 10.
2. Rajput AM, Triep K, Endrich O, Rajput AM. Semi-Automated Approach to Map Clinical Concepts to SNOMED CT Terms by Using Terminology Server. *Studies in health technology and informatics [Internet]*. 2022 May 16.