

MediBot: Um chatbot para consulta de riscos e informações sobre medicamentos

Caio Viktor S. Avila, Tulio Vidal Rolim,
José Wellington Franco da Silva, Vania Maria Ponte Vidal

¹Departamento de Computação – Universidade Federal do Ceará (UFC)
Fortaleza, CE – Brazil
caioviktor@alu.ufc.br, tulio.xcrtf@gmail.com
wellington@crateus.ufc.br, vvidal@lia.ufc.br

Abstract. *Self-medication without medical advice can lead to health problems through intoxication. The objective of this study is to present MediBot - a chatbot to consult information about medications and their risks. MediBot enables queries through natural language, transformed them into SPARQL queries over a Linked Data Mashup about data on medicines provided by ANVISA and Sider sources. Finally, MediBot presents itself as a timely tool in promoting access to information by the general public.*

Resumo. *A automedicação sem orientação médica pode vir a acarretar problemas de saúde por meio da intoxicação. O objetivo deste estudo consiste em apresentar MediBot - um chatbot para consulta de informações sobre medicamentos e seus riscos. O MediBot permite a realização de consultas através de linguagem natural, transformado-as em consultas SPARQL sobre um Linked Data Mashup envolvendo dados sobre medicamentos providos pelas fontes ANVISA e Sider. Por fim, MediBot apresenta-se como instrumento oportuno na promoção do acesso à informação por parte do público geral.*

1. Introdução

O uso de medicamentos para tratamento de doenças e enfermidades é percebido como uma prática válida para melhorias no que tange ao estado de saúde da população. Contudo, a ingestão inadequada de medicamentos pode acarretar a ocorrência de efeitos colaterais e intoxicação. Segundo a Organização Mundial da Saúde (OMS) [WHO et al. 2002], são definidos como eventos adversos à medicamentos quaisquer ocorrências médicas desfavoráveis que possam ocorrer durante a fase de tratamento medicamentoso.

No Brasil, entre os agentes causadores de intoxicação, como agrotóxicos, drogas ilícitas, rodenticidas, inseticidas e alimentos impróprios para consumo, os medicamentos ocupam a primeira posição [Corrêa et al. 2013]. Somente em 2016, houveram aproximadamente 56.937 casos registrados de intoxicação humana, dos quais, 226 evoluíram para óbito, sendo 20.527 gerados em decorrência do uso de medicamentos [SINITOX 2016].

Na *web*, existe uma grande variedade de dados sobre medicamentos que podem auxiliar na orientação de uma medicação correta, tais dados são provenientes de organizações governamentais, como agências reguladoras e portais de dados públicos, fabricantes, farmácias *on-line* e boletins eletrônicos. Contudo, a grande maioria destes

dados está em formato proprietário, como planilhas, em *backups* de bancos de dados relacionais ou disponíveis apenas por meio de páginas da *web*. Além disso, esses dados são isolados em silos de dados, sem conexão direta entre recursos, onde por vezes, esses *datasets* são representados com diferentes vocabulários, tendo um mesmo conceito do mundo real representado através de termos distintos. Um exemplo de informação que não pode ser recuperada diretamente é dada a seguir: “*Quais são os riscos do medicamento XX?*”.

Apesar de ser uma consulta simples, apresenta alguns desafios como: heterogeneidade das fontes (dados em *rdf*, *csv*, *xls* e etc) e a necessidade do conhecimento prévio dos usuários sobre tecnologias e domínio em questão. Esses problemas podem ser resolvidos usando tecnologias da Web Semântica [Shadbolt et al. 2006] e *Linked Data* [Bizer et al. 2011] que permitem uma integração semântica entre recursos em diferentes fontes, representando os dados em um vocabulário unificado através do uso de ontologias. A ontologia fornece um vocabulário uniforme, atuando como uma camada de abstração para acesso a dados. Além disso, permite sua publicação em um formato aberto não proprietário que permite a recuperação de informações por agentes computacionais e usuários humanos através de consultas *SPARQL*.

Este trabalho tem como objetivo demonstrar como *MediBot*, um chatbot baseado em ontologia, pode ser utilizado como ferramenta de acesso à dados sobre medicamentos e seus riscos.

O resto deste artigo é estruturado da seguinte forma: na seção 2 é apresentada a ferramenta *MediBot*. A seção 3 apresenta a arquitetura de *MediBot*. Na seção 4 é apresentada a demonstração proposta. Por fim, na seção 5 são apresentadas nossas considerações finais e trabalhos futuros.

2. *MediBot*

MediBot é um *chatbot* que pode ser utilizado tanto através do aplicativo móvel do *Telegram* e de sua interface *web*, quanto no ambiente *desktop*. *MediBot* foi implementado em JavaScript utilizando NodeJS. *MediBot* pode ser contactado via *Telegram* através do id @LDM_MediBot. Para mais informações, vídeos demonstrativos e exemplos acessar o link¹.

MediBot possui dois modos de funcionamento, sendo o primeiro o modo de **respostas rápidas** e o segundo o **interativo**. No modo de respostas rápidas, *MediBot* possui um conjunto de consultas *SPARQL* pré-definidas para as quais a entrada do usuário é mapeada. O mapeamento da entrada do usuário para o seu respectivo padrão de consulta é realizado através de uma abordagem baseada em expressões regulares. Durante o processo de avaliação da entrada, os termos-chave e os parâmetros de filtragem são recuperados. Os termos-chave ajudam a classificar em qual tipo de consulta a entrada deve ser mapeada, enquanto os parâmetros de filtragem são usados em cláusulas *FILTER* para restringir o resultado da consulta à intenção específica do usuário. Finalmente, a consulta *SPARQL* é construída e executada via *HTTP* no *Virtuoso endpoint*. Além disso, o processo de construção da resposta também usa padrões pré-definidos. A Figura 1 mostra o *workflow* realizado pelo *MediBot* durante este modo. Sete tipos de consultas foram definidos, conforme mostrado na Tabela 1.

¹<https://anonimoanim10.github.io/mediBot/>

Em complemento ao modo resposta rápida, *MediBot* possui um modo interativo. Enquanto o primeiro fornece acesso rápido e fácil às informações, o segundo fornece uma forma versátil de acesso ao conhecimento contido no LDM.

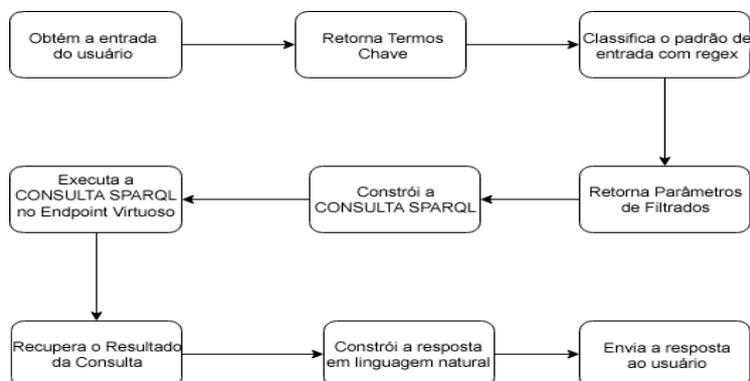


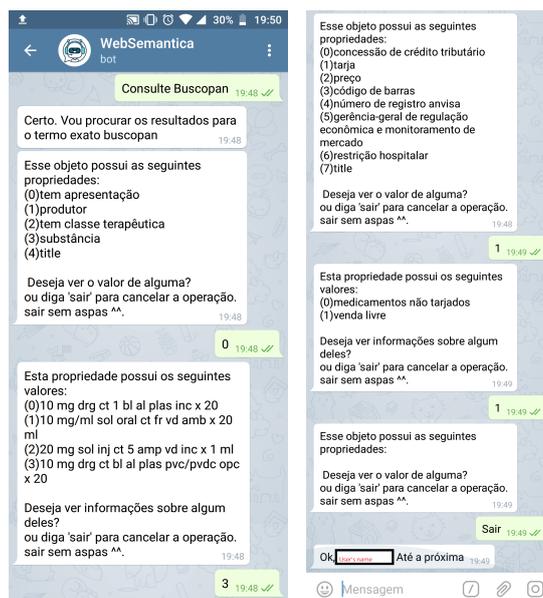
Figura 1. MediBot workflow.

Tabela 1. Tipos de consultas respondidas pelo *MediBot* no modo respostas rápidas.

Tipo de Consulta	Exemplo
Medicamentos com um princípio ativo	Quais são os medicamentos com a substância dipirona?
Definição de termos no domínio	Defina Classe Terapêutica
Informações sobre um certo medicamento	Fale sobre o medicamento aspirina
Riscos de um medicamento	Quais são os riscos de um medicamento reopro?
Lista de representações de um medicamento	Quais são as apresentações do medicamento reopro?
Informações sobre a apresentação do código de barra	Me informe sobre a apresentação do código de barras 7896382701801
Preço de uma apresentação com a taxa de ICMS em um Estado	Quais são os preços com taxa de ICMS da apresentação 7896382701801 no estado do Ceará?

Durante o modo interativo, dois tipos de tarefas podem ser executadas, **navegar** e **consultar**. Para a decisão de quais próximos passos devem ser seguidos durante a conversação em uma tarefa específica *MediBot* usa informações sobre interações passadas durante a mesma tarefa (**contexto**), o ponto atual da tarefa (**state**) e, nos casos em que o *chatbot* espera uma resposta do usuário, é recebida uma mensagem (**input**). A seguir são descritas as tarefas:

- Tarefa de Consulta:** Permite ao usuário consultar e visualizar os dados sobre instâncias contidas na base de conhecimento, mostrando valores de atributos de instâncias. A tarefa de consulta é recursivamente interativa. Ao receber uma consulta de um usuário, *MediBot* retorna uma lista de propriedades da instância, dando ao usuário a opção de selecionar uma destas para visualizar seus valores. Na Figura 2 é apresentado um exemplo de interação da tarefa de consulta através do Telegram.
- Tarefa de Navegação:** Permite ao usuário navegar recursivamente através dos termos existentes na base de conhecimento (*knowledge base*), mostrando definições e esquema da fonte. Quando o usuário inicia a tarefa de navegação sobre um termo, *MediBot* apresenta diferentes nomes, tipos e definições de termo, além de apresentar suas propriedades e permitir ao usuário selecionar uma destas para ser o novo pivô da tarefa, podendo navegar sobre os conceitos definidos na ontologia. Para iniciar esta tarefa basta digitar a palavra “Explore”.



(a) Tarefa Consulta

(b) Tarefa Consulta

Figura 2. Exemplos de Consulta no Telegram

3. Arquitetura

A arquitetura do *MediBot* é organizada em três camadas: *i)* interface de usuário; *ii)* servidor e *iii)* *conhecimento*. A Figura 3 apresenta a arquitetura do *MediBot* e seus componentes.

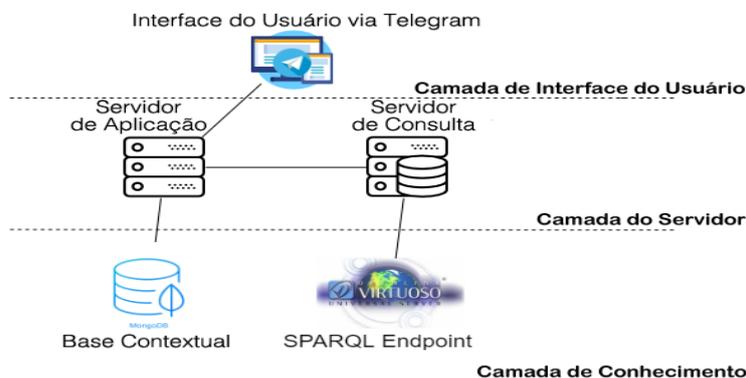


Figura 3. Arquitetura MediBot

Na primeira camada, a interface do usuário fornece um meio de interação com o usuário através do aplicativo de mensagens instantâneas *Telegram*. A utilização do *MediBot* pode ser feita através do aplicativo móvel do *Telegram*, ou, por seu aplicativo *desktop* no computador. A razão para escolher o *Telegram* como um canal para o *MediBot* se dá em razão dessa ferramenta já possuir uma grande base de usuários em conjunto com uma infraestrutura robusta, além de ter uma *API* fácil que permite a criação e uso de *chatbots*.

A segunda camada proposta é a camada do servidor. Esta camada é responsável por processar as requisições e prover respostas, sendo composta por dois componentes

principais. O primeiro componente desta camada é o servidor de aplicação (*application server*), responsável por receber e processar as requisições do usuário. O *application server* recebe a requisição do usuário e constrói a consulta SPARQL responsável por retornar a informação desejada. Ainda, tal componente é também responsável por enviar a consulta SPARQL para o servidor de consultas (*query server*) e emitir uma mensagem de confirmação de entrega ao usuário. Sendo o segundo componente, o *query server* é responsável por receber a consulta SPARQL e executá-la no SPARQL *endpoint*. Além disso, esse componente também é responsável por obter a resposta da consulta SPARQL e construir a mensagem de resposta ao usuário.

A terceira e última camada é a camada de conhecimento. Esta camada é responsável por armazenar o conhecimento necessário para interagir e responder o usuário. Esta camada também é composta por dois componentes:

- **Base Contextual:** Responsável por armazenar informações relevantes ao contexto, tais como informações pessoais sobre o usuário como seu nome, sobrenome, linguagem e outras informações referentes ao fluxo de diálogo. As informações sobre o contexto da interação são armazenadas em uma instância no banco de dados não-relacional orientado a objetos *mongoDB*, que permite um rápido armazenamento e recuperação de dados armazenados em *JSON*, permitindo o processamento direto por código *javascript* sem a necessidade de pré-processamento dos dados. Os links para as fontes originais podem ser encontrados em ².
- **Linked Data Mashup (LDM):** O processo de *Linked Data Mashup* foi realizado com os dados de medicamentos presentes nas bases da Agência Nacional de Vigilância Sanitária (ANVISA) e da base SIDER presente no projeto *BIO2RDF* [Belleau et al. 2008]. Da ANVISA, foram selecionados:
 - Preços de Medicamentos ao Consumidor (PMC);
 - Preços de Medicamentos ao Governo (PMG);
 - Riscos do medicamento em gestantes e lactantes (RGL);

PMC e PMG são encontrados nos formatos de arquivo *XLS* e *PDF*, sendo usadas neste trabalho as versões *XLS*. Ambos os *datasets* contêm informações sobre medicamentos alopáticos, como o nome do medicamento, o produtor, o código de barras, a classe terapêutica, a apresentação, o princípio ativo e os preços. O *dataset* RGL contém as categorias de risco de substâncias durante o período da gravidez e amamentação. Este *dataset* só está disponível no formato não estruturado *PDF*. Por sua vez, o *dataset* do *SIDER* disponibilizado pelo projeto *BIO2RDF* já está no formato *RDF*. O *dataset* *SIDER* contém dados sobre drogas, suas indicações, efeitos colaterais e diferentes rótulos. No entanto, o banco de dados só possui dados em inglês, não contendo informações sobre medicamentos brasileiros, tornando necessário traduzi-los para o português. Este *dataset* foi selecionado porque contém informações sobre os efeitos colaterais dos princípios ativos, tais informações são necessárias para informar os riscos de um medicamento.

Por fim, foi gerado um *Linked Data Mashup* (LDM) representando uma visão integrada das múltiplas fontes seguindo um mesmo vocabulário e método de armazenamento e acesso, além dos arquivos de mapeamento e da implementação OWL da ontologia, podendo ser acessados publicamente via *datahub* ³. O LDM

²<https://anonimoanim10.github.io/mediBot/#fontes>

³<https://datahub.io/linkeddatabashupeducacional/data-med/v/2>

resultante foi hospedado no *virtuoso triplestore*, que é acessado diretamente pelo aplicativo MediBot através do servidor de consulta.

4. Demonstração proposta

Para a demonstração da efetividade da ferramenta *MediBot* como uma ferramenta de apoio a recuperação de informação, propomos o seguinte cenário:

- Um usuário busca saber quais são os riscos de um determinado medicamento através do modo de respostas rápidas.
- Posteriormente o usuário busca por medicamentos que possuem o mesmo princípio ativo. Esta atividade pode ser realizada tanto através do modo rápido, quanto por meio do modo de consulta.
- Ao ver termos como tarja, tipo de medicamento, classe terapêutica, etc. O usuário seleciona a opção no qual deseja, levando-o a navegar sobre os termos da ontologia e posteriormente obtendo informações.

5. Considerações Finais

A automedicação sem instrução prévia é uma problemática que vem apresentando indicadores preocupantes no que tange a intoxicação. Neste trabalho foi apresentado *MediBot*, um chatbot para consulta de riscos e informações sobre medicamentos. Para tanto, *MediBot* apresenta-se como um instrumento interativo para auxílio no ato de recuperação de informações sobre medicamentos e seus riscos, possibilitando uma redução em relação às eventuais dúvidas do paciente.

Como trabalhos futuros, pretende-se disponibilizar *MediBot* para outras plataformas de mensagem instantânea, como *Facebook Messenger* e *Instagram Direct*, além de expandir o conjunto de padrões de consultas rápidas.

Referências

- Belleau, F. et al. (2008). Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics*, 41(5):706–716.
- Bizer, C. et al. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global.
- Corrêa, A. et al. (2013). Uma abordagem sobre o uso de medicamentos nos livros didáticos de biologia como estratégia de promoção de saúde. *Ciência & Saúde Coletiva*, 18:3071–3081.
- Shadbolt, N. et al. (2006). The semantic web revisited. *IEEE intelligent systems*, 21(3):96–101.
- SINITOX (2016). Sistema Nacional de Informações Toxicológicas registro de intoxicações no brasil. Acessado: 10-02-2019.
- WHO et al. (2002). Safety of medicines: a guide to detecting and reporting adverse drug reactions. Technical report, Geneva: World Health Organization.