

Using Chatbots to Compare Drug Prices Extracted in Real-Time from the WEB

Caio Viktor S. Avila
Federal University of Ceará
Fortaleza, Brazil
caioviktor@alu.ufc.br

Vania Ponte Vidal
Federal University of Ceará
Fortaleza, Brazil
vvidal@lia.ufc.br

Asley Caldas
Federal University of Ceará
Fortaleza, Brazil
asleylc@gmail.com

Tulio Vidal Rolim
Federal University of Ceará
Fortaleza, Brazil
tulio.xcrtf@gmail.com

Thays Lavor
Federal University of Ceará
Fortaleza, Brazil
thays.lavor@gmail.com

ABSTRACT

In this paper we present a chatbot for the medicines information consultation. The system presented aims to act as a single access point for the natural and simplified information retrieval of medicines, its prices and risks, democratizing the access to knowledge and promoting the conscious consumption of medicines. The chatbot has two modes of operation: Quick Response and Interactive modes. The first answers questions asked in natural language, while the second has three interactive tasks, namely Navigation, Query and Price Comparison. In this work, we focus on the presentation of the system architecture and the Price Comparison task. This task presents the best prices for medicines and their best possible substitutes extracted in real-time from web with the help of information obtained from a linked data mashup. The work is still under development, however in this paper we present the results of preliminary assessments and possible directions for future work.

CCS CONCEPTS

• Information systems → Question answering; Information extraction.

KEYWORDS

Drugs, ChatBots, Linked Data, Semantic Web

1 INTRODUÇÃO

Segundo a Organização Mundial de Saúde (OMS), o uso racional de medicamentos por pacientes dá-se quando estes recebem medicamentos apropriados para suas condições clínicas, em doses adequadas às suas necessidades individuais, por um período adequado e ao menor custo para si e para a comunidade[8]. Deste modo, o uso racional é uma prerrogativa básica para o uso efetivo de medicamentos. Contudo, de acordo com [1], as práticas de consumo da

população brasileira estão muito distantes de tal cenário, onde um primeiro passo para a promoção do uso racional de medicamentos passa pela educação e informação da população.

Deste modo, neste trabalho buscamos desenvolver um método que democratiza o acesso ao conhecimento acerca do domínio de medicamentos, empoderando o usuário com conhecimento de modo a garantir o uso racional de medicamentos pela população em geral.

Como solução para tal cenário, viemos a realizar a integração semântica de diversas fontes de dados disponíveis publicamente na web, gerando uma base de conhecimento (KB, *Knowledge Base*) sobre o domínio que permite a recuperação da informação através de um único ponto de acesso e com um vocabulário comum e acessível à usuários leigos. Como interface de consulta e acesso ao conhecimento, viemos a desenvolver um *chatbot* para um conhecido cliente de mensageiro instantâneo (*Instant Messenger*). O *chatbot* disponibiliza para o usuário o acesso às informações contidas no KB com o uso da linguagem natural, permitindo que usuários não técnicos realizem consultas e recuperem suas informações de interesse sem que estes precisem entender a estrutura dos dados e linguagens de consulta. O *chatbot* possui dois modos de operação o *modo de Respostas Rápidas* e o *Modo Interativo*. O primeiro responde perguntas feitas em linguagem natural, enquanto o segundo possui três atividades interativas, sendo elas *Navegação*, *Consulta* e *Comparação de Preços*.

Neste trabalho focaremos no ponto de "...e ao menor custo para si e para a comunidade." da definição do uso racional de medicamento. Segundo [6], a diferença de preços entre medicamentos pode ultrapassar 5.000% de uma farmácia para outra e entre diferentes versões de um mesmo medicamento. Assim, neste trabalho focamos na apresentação da atividade interativa de *Comparação de Preços*. Esta atividade realiza a comparação dos preços disponíveis para um medicamento e seus similares. Para tal, é efetuada a recuperação na web em tempo real dos preços ofertados para as diversas apresentações de um medicamento nas principais redes farmacêuticas do país. Além disso, como um diferencial, o sistema não busca apenas as ofertas relativas a um único medicamento, mas também aquelas de seus medicamentos similares. Para a localização de medicamentos similares, recorreremos ao uso das informações contidas em nosso KB.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebMedia '19, October 29–November 1, 2019, Rio de Janeiro, Brazil

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6763-9/19/10...\$15.00

<https://doi.org/10.1145/3323503.3360635>

2 TRABALHOS RELACIONADOS

O “MedChatBot” é um *chatbot* para estudantes de medicina baseado no AIML UMLS de código aberto para gerar respostas a consultas através da extração de conhecimento[7]. Contudo, “MedChatBot” não tem acesso aos preços de medicamentos ofertados em tempo real por revendedores brasileiros.

Alguns aplicativos para *smartphone* como o *Consulta Remédios*¹, *MediPreço*², etc. disponibilizam o serviço de comparador de preços de medicamentos. Contudo, tais serviços realizam apenas a comparação de preços de uma mesma apresentação de um medicamento para diferentes revendedores (farmácias e drogarias), não apresentando de maneira direta a comparação entre o preço dos medicamentos similares. Além disso, tais serviços necessitam que o usuário baixe e instale um *APP*, o que consome tempo em sua instalação e espaço no armazenamento do dispositivo, o que pode ser uma desvantagem quando o usuário necessita apenas realizar consultas rápidas e pontuais. Com o uso de um *chatbot*, tal dificuldade é contornada, pois o usuário precisa apenas adicionar e enviar mensagens ao *chatbot* através de um aplicativo de mensagens (cada vez mais comum na vida das pessoas), não precisando baixar ou instalar um novo aplicativo apenas para essa função.

3 ARQUITETURA DO SISTEMA

A arquitetura do sistema apresentada na Figura 1 é organizada em três camadas: *i*) interface de usuário; *ii*) servidores e *iii*) conhecimento.

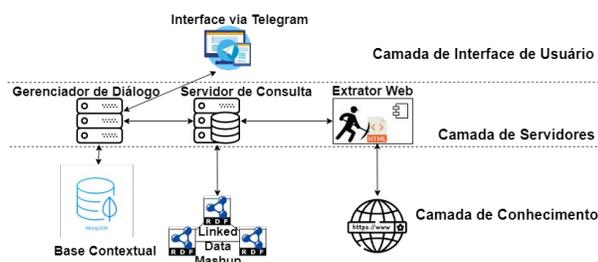


Figure 1: Arquitetura do sistema

Na primeira camada, a **Interface de Usuário** fornece um meio de interação com o usuário através do aplicativo de mensagens instantâneas *Telegram*. A razão para a escolha do *Telegram* se deu em razão dessa plataforma já possuir uma grande base de usuários em conjunto com uma infraestrutura robusta. Além disso, O *Telegram* disponibiliza uma *API* que permite a criação e o uso de *chatbots* de maneira fácil e prática.

A segunda camada proposta é a **Camada de Servidores**. Esta camada é responsável por processar as requisições e prover respostas, sendo composta por três componentes principais:

- **Gerenciador de Diálogo (GD):** É o módulo responsável por receber, processar e interpretar as requisições do usuário, controlando o fluxo do diálogo.
- **Servidor de Consulta (SC):** É responsável por receber uma solicitação de informação encaminhada pelo GD, podendo

consultar diretamente o KB ou encaminhar a solicitação ao extrator web. No primeiro caso, o SC cria consultas SPARQL que serão realizadas sobre o *endpoint* do KB;

- **Extrator Web:** É composto por *web scrapers* responsáveis por acessar páginas da Web e minerar destas as informações solicitadas pelo SC. Este processo é feito em tempo de execução, o que garante a atualização constante da informações retornadas.

A terceira e última camada é a **Camada de Conhecimento**. Esta camada é responsável por armazenar o conhecimento necessário para a interação e resolução de consultas. Esta camada também é composta por três componentes:

- **Base Contextual:** Responsável por armazenar informações relevantes ao contexto, tais como informações pessoais sobre o usuário como seu nome e outras informações referentes ao fluxo de diálogo, guardando o estado atual da conversa.
- **Linked Data Mashup (LDM):** O sistema utiliza um LDM como sua principal fonte de conhecimento. O LDM provê uma visão semanticamente integrada das diferentes fontes, disponibilizando-as através de um único método de acesso (consultas SPARQL), sendo representadas por um único vocabulário (com o uso de uma ontologia de aplicação), tendo as fontes armazenadas como uma única e em um mesmo formato (em um *triplestore* RDF) e mesclando informações sobre um mesmo objeto distribuídas pelas diferentes fontes (através da fusão de entidades). As fontes integradas foram sobre dados de medicamentos presentes nas bases da Agência Nacional de Vigilância Sanitária (ANVISA) e da fonte RDF *SIDER* presente no projeto *BIO2RDF* [4]. A fonte *SIDER* contém informações sobre os efeitos colaterais e indicações de medicamentos, já sendo disponibilizada como *Linked Open Data* [5]. Da ANVISA, foram selecionados os *datasets*:

- **Preços de Medicamentos ao Consumidor (PMC) e Preços de Medicamentos ao Governo (PMG)**³: Ambos os *datasets* contém informações sobre medicamentos alopáticos, como o nome do medicamento, o produtor, o princípio ativo e os preços máximos de comercialização permitidos. A única diferença entre PMC e PMG é que o primeiro possui os preços máximos de venda para o consumidor, enquanto o segundo contém os preços máximos de venda para agências governamentais;
- **Riscos do medicamento em gestantes e lactantes (RGL):** O *dataset* RGL contém as categorias de risco de substâncias durante os períodos de gravidez e amamentação. Este *dataset*⁴ só está disponível no formato não estruturado *PDF*.
- **Sites de preços de medicamentos:** Embora os *datasets* PMC e PMG possuam informações sobre os preços dos medicamentos, estes referem-se apenas aos preços máximos permitidos por lei, mas não os preços que são de fato aplicados pelos revendedores. Deste modo, os preços aplicados pelos revendedores são obtidos diretamente das informações publicadas na *Web* com o uso de extratores (*web scrapers*) que

³<http://portal.anvisa.gov.br/listas-de-precos>

⁴<http://portal.anvisa.gov.br/documents/33880/2561889/116.pdf/2292b730-2bd5-4acc-b378-10682b1fc344?version=1.0>

¹<https://play.google.com/store/apps/details?id=com.consultaremedios>

²<https://play.google.com/store/apps/details?id=br.com.illevel.medipreco>

recuperam os preços em tempo real, garantido a atualização e validade das ofertas. Atualmente, utilizamos o portal *Consulta Remédios*⁵ como fonte. O *Consulta Remédios* é um portal que agrega informações sobre medicamentos e seus preços fornecidos por diversas redes farmacêuticas, sendo atualizado em tempo real com as ofertas de medicamentos.

O sistema possui dois modos de operação, os modos de *Respostas Rápidas* e o *Modo Interativo*. No modo de respostas rápidas o sistema busca mapear a pergunta realizada em linguagem natural para um dos padrões de consulta definidos *à priori*, estes padrões são definidos com o uso de expressões regulares. O modo de respostas rápidas dá suporte apenas a interações de um único estágio, onde sempre há apenas uma única pergunta e uma única resposta.

O segundo modo, o modo interativo é performado no estilo de conversação, onde há uma sequência de perguntas e respostas relacionadas, onde as interações anteriores são consideradas durante a computação da interação atual. Este modo possui três atividades distintas: *Navegação*, *Consulta* e *Comparação de Preços*. As duas primeiras são apresentadas em mais detalhes em [2, 3], onde resumidamente a primeira atua como um dicionário eletrônico sobre os termos presentes na ontologia de aplicação, permitindo que o usuário navegue sobre seus conceitos; já a segunda atividade permite que o usuário consulte valores de propriedades de instâncias contidas no LDM.

4 ATIVIDADE DE COMPARAÇÃO DE PREÇOS

4.1 Workflow da interação com o chatbot

Nesta atividade o *chatbot* busca na *web* os diferentes preços ofertados para um mesmo medicamento em diferentes revendedores (Figura 2). Além disso, o sistema também busca os preços de medicamentos que possam substituir o original, mostrando aqueles com os menores preços como possíveis alternativas. O *workflow* da atividade de comparação de preços segue os seguintes passos:

- (1) O *chatbot* solicita ao usuário o nome do medicamento;
- (2) Usuário digita o nome completo ou parcial do medicamento;
- (3) O sistema busca no LDM medicamentos que contenha o nome solicitado pelo paciente em seus nomes e apresenta a lista de possibilidades ao usuário;
- (4) Usuário seleciona o medicamento de interesse M da lista;
- (5) O sistema busca no LDM medicamentos similares S_1, \dots, S_n à M . Os medicamentos similares são aqueles que possuem os mesmos princípios ativos e efeitos;
- (6) O sistema busca na *web* ofertas de preços para os medicamentos M, S_1, \dots, S_n . Para cada resultado de M encontrado, este é considerado uma apresentação $A_k M$, onde $1 \leq k \leq RM$, sendo RM o número de resultados encontrados para a busca usando M como palavra-chave. Além disso, para cada resultado de S_j encontrado, este é considerado uma apresentação $A_k S_j$, onde $1 \leq k \leq RS_j$, sendo RS_j o número de resultados encontrados para a busca usando S_j como palavra-chave e $1 \leq j \leq n$. Para cada resultado encontrado $(A_1 M, \dots, A_{RM} M, A_1 S_1, \dots, A_{RS_1} S_1, \dots, A_1 S_n, \dots, A_{RS_n} S_n)$ o sistema extrai as informações sobre nome do medicamento, descrição da apresentação, menor preço, nome do revendedor

com menor preço, site da oferta do revendedor com menor preço, descrição sobre o medicamento e a descrição sobre seu modo de uso;

- (7) Com as informações sobre os medicamentos extraídas, o sistema agrupa os medicamentos por apresentações. Cada grupo $G_{A_k M}$ é composto por um representante, sendo este $A_k M$, e as apresentações de medicamentos similares. $A_k S_j \in G_{A_k M}$, ou seja, uma apresentação $A_k S_j$ pertence ao grupo da apresentação $A_k M$, quando a distância de *levenshtein* entre a descrição da apresentação de $A_k S_j$ e da apresentação de $A_k M$ é menor ou igual a um *threshold* t , (*levenshtein*(*aprese*($A_k S_j$), *aprese*($A_k M$)) $\leq t$);
- (8) Para cada grupo $G_{A_k M}$ o sistema elege uma melhor alternativa $S_k M$, onde $\{S_k | A_k S_j \in G_{A_k M} \ \& \ \text{preco}(A_k S_j) = \text{precoMin}(G_{A_k M})\}$, ou seja, S_k é a apresentação de um medicamento similar com o menor preço dentro de $G_{A_k M}$;
- (9) Ao fim, o *chatbot* mostra a lista de apresentações $(G_{A_1 M}, \dots, G_{A_{RM} M})$ ordenadas em ordem crescente pelo preço do representante de cada grupo $A_k M$ e mostrando S_k caso $\text{preco}(S_k) < \text{preco}(A_k M)$.

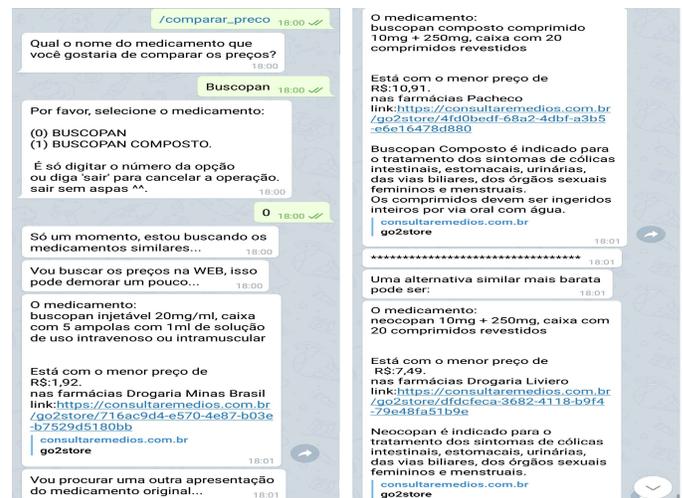


Figure 2: Demonstração da comparação de preços

4.2 Algoritmo de agrupamento e comparação

```

1 input: original, similars, t_name, t_presen
2 itens_original = filter(original.itens, original.name, t_name)
3 groups = Group_By_Presentation(itens, t_presen)
4 for similarGroup in similars:
5     for similar in similarGroup.itens:
6         if(distance(similar.name, similarGroup) <= t_name):
7             best_group = null
8             best_dist = infinity
9             for group in groups:
10                if(distance_group(group, similar, t_presen) < best_dist):
11                    best_group = group
12                    best_dist = distance_group(group, similar)
13            if(best_group != null):
14                best_group.add(similar, best_dist)
15 return groups

```

Algoritmo 1: Algoritmo de agrupamento de itens e comparação de preços.

No Algoritmo 1 é apresentado o algoritmo executado nos passos (7) e (8) da seção 4.1. A lógica por trás do algoritmo é a seguinte:

⁵<https://consultaremedios.com.br/>

O algoritmo recebe como entrada o conjunto de itens encontrados utilizando o nome do medicamento original como termo de busca (original), o conjunto de itens encontrados utilizando cada um dos nomes de medicamentos similares como termo de busca (similars), o *threshold* de tolerância para os nomes dos medicamentos (*t_name*) e o *threshold* de tolerância para as descrições das apresentações dos medicamentos (*t_presen*). Então, nas linhas 2 e 3 o algoritmo filtra possíveis ruídos dos medicamentos originais e agrupa-os por apresentações. Posteriormente, para cada possível medicamento similar (ruídos filtrados em 6) o algoritmo o classifica para um dos grupos originais, tendo como base a distância (atualmente *levenshtein*) da apresentação do medicamento para a apresentação do representante de cada grupo (linhas 9 - 14). Quando um medicamento é adicionado à um grupo (linha 14) é decidido se ele é o similar representante ou não, isto acontece quando o valor de relevância do item sendo avaliado é maior do que a do atual representante. a relevância de um item é calculada como: $R(I) = 1/(WP * \max(I.preco, 0.001)) + 1/(WD * \max(I.dist, 1))$, onde $WP + WD = 1$ e 0.001 é um hiper-parâmetro. Esta medida tenta ponderar a importância do menor preço e da menor distância para a relevância de um item no grupo.

5 AVALIAÇÕES PRELIMINARES

Table 1: Resultados da avaliação preliminar.

Medicamento	Encontrou	Menor Preço	Similar	Preço Similar	Diferença de Preço	Preço Máximo
RUSUCOR	1	32,65	1	23,46	9,19	48,51
SORINE SSC	1	15,37	0	-	-	35,33
RELAXMED	1	3,99	1	1,9	2,09	5,96
VICK VAPORUB	1	8,04	0	-	-	23,86
DROPROZINA	1	6,35	1	6,34	0,01	6,35
CORISTINA D	1	6,19	0	-	-	9,09
BENEGRIP	1	6,6	0	-	-	9,3
BUSCOPAN	1	10,91	1	7,49	3,42	14,4
MERTHIOLATE	1	13,28	1	5,07	8,21	17,92
RENALAPRIL	1	8,67	1	1,8	6,87	17,31
DORFLEX	1	3,99	0	-	-	15,02
XARELTO	1	66,9	0	-	-	66,9
SELOZOK	1	20,32	0	-	-	57,68
NEOSALDINA	1	1,52	0	-	-	29
TORSILAX	1	1,9	0	-	-	36,52
ARADOIS	1	43,84	1	28,88	14,96	135,56
GLIFAGE	1	15,69	1	4,69	11	29,5
ADDERA D3	1	18,4	1	15,46	2,94	114,04
ANTHELIOIS	0	-	-	-	-	-
BUSCOPAN COMPOSTO	1	10,91	0	-	-	14,98
MEDIAS TOTAIS	0,95	15,55368421	0,474	10,56555556	6,521111111	36,17
MEDIAS TOP	0,9	20,38555556	0,333	14,89888889	9,633333333	55,46666667
MEDIAS ALEATÓRIOS	1	11,205	0,6	7,67666667	4,965	18,803

Avaliações preliminares foram realizadas com o objetivo de avaliar a eficácia do sistema em encontrar os menores preços ofertados para uma apresentação de um medicamento e os menores preços para possíveis alternativas. Para este fim, realizamos o processo de consulta manual de 20 medicamentos, onde os 10 primeiros (RUSUCOR-RENALAPRIL) foram escolhidos aleatoriamente a partir do catálogo de opções e os 10 últimos foram escolhidos seguindo o top dos medicamentos mais vendidos⁶. O site *Consulta Remédios* foi utilizado como fonte para as comparações. Em nossos teste utilizamos os valores de $t_name = t_presen = 8$ e $WP = 0.9$, $WD = 0.1$, dando uma maior relevância ao preço.

Em nossos testes (com os resultados na Tabela 1) comprovamos que o preço de um medicamento pode variar muito entre os diferentes revendedores, onde a média do menor preço fica em média

⁶<https://guiadafarmacia.com.br/interfarma-faz-lista-dos-dez-medicamentos-mais-vendidos-no-brasil/>

R\$ 15,50, já a média de ofertas mais caras chega a R\$ 36,17. Além disso, em nossos testes comprovamos que o sistema é eficaz em encontrar os medicamentos buscados, tendo encontrado 90% dos medicamentos mais vendidos. Já para os medicamentos similares, encontramos uma alternativa em cerca de 33% dos medicamentos mais vendidos e 60% nos medicamentos aleatórios, o que indica que ainda temos bastante espaço para melhorias nesta área.

6 CONCLUSÃO

Neste trabalho apresentamos um *chatbot* para a consulta de informações sobre medicamentos. O sistema apresentado tem como objetivo atuar como um ponto de acesso único para a recuperação de informações, seus preços e riscos, sobre medicamento em uma maneira natural e simplificada, democratizando o acesso ao conhecimento e promovendo o consumo consciente de medicamentos.

O sistema utiliza como fonte de conhecimento um *Linked Data Mashup* (LDM) e informações extraídas da *web* em tempo real. O LDM integra semanticamente *datasets* heterogêneos de diferentes fontes publicadas na *Web*, permitindo uma visão homogênea e integrada dos dados. Além disso, o sistema recupera os preços ofertados de medicamentos nas principais redes farmacêuticas do país.

Neste trabalho focamos na apresentação da arquitetura do sistema e na atividade de comparação de preços, que apresenta os melhores preços para medicamentos e seus melhores possíveis substitutos extraídos em tempo real da *web*. Em nossos testes preliminares comprovamos a eficácia do sistema em encontrar os menores preços para as apresentações de um medicamento solicitado, alcançando uma taxa de sucesso de 90% para os medicamentos mais vendidos no país. No entanto, a taxa de recomendações de substitutos ficou em torno de apenas 33% para os mesmos medicamentos, indicando assim o caminho para possíveis melhorias, em particular no mecanismo de agrupamento de apresentações de medicamentos que leva em conta apenas aspectos sintáticos da descrição da apresentação, sendo sensível a variações em sua escrita.

REFERENCES

- [1] Daniela Silva de Aquino. 2008. Por que o uso racional de medicamentos deve ser uma prioridade? *Ciência & Saúde Coletiva* 13 (2008), 733–736.
- [2] Caio Viktor Avila, A. Calixto, Tulio Vidal Rolim, Wellington Franco, A. Venceslau, Vânia M. P. Vidal, Valéria M. Pequeno, and Francildo Felix De Moura. 2019. MediBot: An Ontology based Chatbot for Portuguese Speakers Drug's Users. In *Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 1: ICEIS*, INSTICC, SciTePress, 25–36. <https://doi.org/10.5220/0007656400250036>
- [3] Caio Viktor S Avila, Tulio Vidal Rolim, José Wellington Franco da Silva, and Vania Maria Ponte Vidal. 2019. MediBot: Um chatbot para consulta de riscos e informações sobre medicamentos. In *Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*. SBC, 1–6.
- [4] François Belleau et al. 2008. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics* 41, 5 (2008), 706–716.
- [5] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2011. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*. IGI Global, 205–227.
- [6] G1. 2016. Terra Remédios têm diferença de preços que pode passar de 5.000%. <http://glo.bo/29qiFag>. Accessed: 2019-07-18.
- [7] Hameedullah Kazi, BS Chowdhry, and Zeesha Memon. 2012. MedChatBot: An UMLS based Chatbot for Medical Students. *International Journal of Computer Applications* 55, 17 (2012).
- [8] World Health Organization et al. 2014. The rational use of drugs: report of the Conference of Experts. 1987. *Geneva: World Health Organization* (2014).