PAPER • OPEN ACCESS

Chatbot Implementation with Semantic Technology for Drugs Information Searching System

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

Chatbot Implementation with Semantic Technology for Drugs Information Searching System

A Amalia*[1], P Y C Sipahutar[2], E Elviwani[3], and F Purnamasari[4]

[1,2,3] Department of Computer Science, Universitas Sumatera Utara, Jl. Alumni No. 9-A ,Medan 20155- Indonesia
[4] Department of Information Technology, Universitas Sumatera Utara, Jl. Alumni No. 9-A, Medan 20155- Indonesia

Email: *amalia@usu.ac.id

Abstract. This research aims to design medicine information based on the ontology scheme and to display the relationship between information through chatbot as search media. The diversity of information does not guarantee to gather a relevant search result with user's needs. A conventional search engine would be substantially beneficial in information searching yet could not able to comprehend meaning of found information and its relationship with other information. However a machine could understand the relationship and the meaning of information through the implementation of semantic technology, which is ontology. Ontology is implemented to represent semantic metadata information. The information designed with the ontology model would be very assistive in searching for data which is relevant to users' need. The result of implementing ontology as a knowledge base in this research able to display the information of a disease and its relationship with medicine used as the medication for such a particular disease. The evaluation of our system shows that 90% of queries from users generate a valid response from the chatbot. Therefore, this research could be used as a reference in search of medical information and describe its relationship with a particular disease or related to medicine.

1. Introduction

Drug information is one of information that is mostly sought by society. Internet development has given convenience to the community to search for drugs and disease information. As stated on the prior research done by the Pew Research Center's Internet & American Life Project in 2013, tell us that 72% of internet users search information about health. According to the research, 77% of them uses a conventional search engine, and rest uses health website such as WebMD or Wikipedia. At certain times a conventional search engine not provide user needs because a conventional search engine do not understand a local language that is used by particular society. For example, word is used to define for Myopia (near sightedness), in Bahasa Indonesia is called "rabun ayam". Accordingly, semantic web technology able to increase efficiency of searching for health information. In this research, a chatbot is developed for searching information not only by keyword but also by user queries that impact to more accurate information result. A semantic web is a system which let the machine to comprehend and respond to complex human queries based on definition. This comprehension needs information which is structured semantically, such as Resource Description Framework (RDF) and ontology. The RDF is commonly used to represent data and its relationship. The

RDF is an important parts of semantic web especially to solve this problem. However, the RDF about drugs information in Bahasa Indonesia has not available to public. Therefore, in this research uses RDF which is done by [1]. The resulting RDF is cover information about drugs, disease, and a relationship between them. In this research, we also use ontology as a knowledge base in developing a chatbot. The ontology is used to describe semantic metadata and their relationship. Ontology data will be used as a knowledge base in developing chatbot as drug information searching media. The contribution of our study is semantic web technology is combined with Natural Language Processing (NLP) tasks. Our system will recognize and understand user queries by using sentences in Bahasa Indonesia. There are some pre-processing NLP such as case folding, tokenization, phrase detection, stemming, and filtering. This paper is divided into five sections. Section 2 describes related works about the semantic web, Section 3 describes the methodology of this study, Section 4 presents the result and discussion, and Section 5 concludes this study.

2. Related Works

Some of the previous works in the semantic web field are already conducted by [1][2][3][4]. Research in [1] describes how to crawl and scrape information about drugs in the tropical disease domain. They developed the information about drugs of tropical disease on RDF format. In our study, we utilize this RDF file to build a chatbot system based on the semantic web. Several studies has conducted to construct a semantic relationship using an RDF and used Bahasa Indonesia ontologies [2][4]. This study focuses on different domain; the tourist attraction field [2], and drug information [4]. The study by [4], using the semantic web to relate drug information on BPOM website. BPOM is an institution that maintains drug and food in Indonesia. In our study we use data from BPOM and website drugs.com to generate the RDF of drug information. In our study, we were also using other drug information on drugs.com. The drugs.com is an online pharmacy Encyclopedia that provides drug information from many health organizations. The example of health organizations in drugs.com is Cerner Multum, Micromedex from Truven Health Analytics, Wolters Kluwer Health, U.S. Food and Drug Administration (FDA), A.D.A.M., Stedmans, AHFS, Harvard Health Publications, Mayoclinic, North American Compendiums, and HealthDay. The other differences, in our study, we implemented a chatbot search system. Another previous research in the semantic web is [5],[6],[7],[9] most of these studies describe the survey of semantic web technologies.

3. Research Method

Ontology-based semantic web implementation in this research is consists of a few steps. The first step is retrieve the RDF dataset on tropical diseases drug. The second is designing data with an ontology scheme based on the RDF document. Then, the result of data design is used in searching system using a chatbot. The general architecture is explained in figure 1.
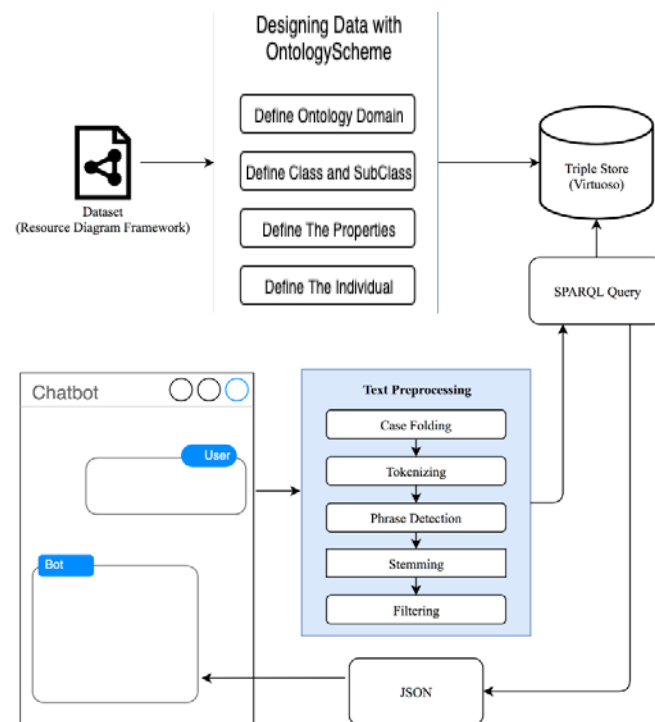
Figure 1. General Architecture

3.1 Data Collection

In this research, we use dataset in RDF format, the data consists of 28 tropical diseases, 13337 from BPOM data of drugs, and data from 211 drugs.com. Data is retrieved from trusted source and suitable to use in this research, BPOM and Drugs.com. By using this data, ontology form is developed.

3.2 Designing Data with Ontology Scheme

This step of process is core of this research. Ontology is a knowledge which describes an entity and relation between entities. Ontology has extensive vocabulary in describing things compared with RDF. RDF is a standard data modeling that is formed by W3C to represent information. Therefore it can be comprehended by computer and can be processed in many ways. According to Taye [8], RDF is a language that provides standard metadata for resources on the web. RDF can represent data and do knowledge exchange on the internet. RDF can be developed to be comprehended by computers to facilitate interoperability. In other words, RDF is a framework to represent metadata and to describe semantics of information that could be accessed by machines.
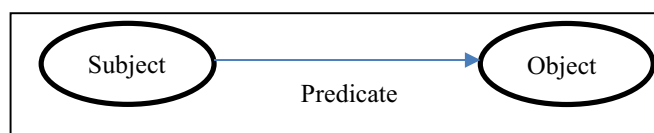

Figure 2. RDF Components

RDF has three components namely subject, predicate, and object. These components also known as triples or statements, as showed in Figure 2. RDF documents are converted to the ontology document. Ontology is a philosophical term to study a thing [8]. Meanwhile, ontology is a knowledge to explain about entities in the world and their relationship. In other areas, such as software engineering and artificial intelligence, ontology is defined as formal specifications of common conceptualization. An ontology consists of four main components, which are concepts, instances, relations, and axioms [8]. Ontology design has several steps:

- Determining domain of ontology

The field of this ontology is medical which has information about tropical diseases and drug related to tropical diseases.

- Defining Class, subclass and their hierarchy

Defining Class, subclass, and the hierarchy use a top-down approach, which starts with a very general concept and continues with a more specific idea.
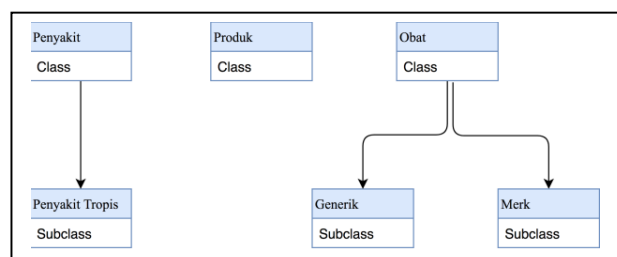

Figure 3. Ontology Class dan Sub-Class Hierarchy

In figure 3, the most general class in this research is "penyakit" (in English: disease), "obat" (in english: drug) and "produk" (in English: product). The Penyakit Class consists of subclass namely "penyakit tropis" (in English: disease tropic). The Obat Class has two subclasses which are "Generic" and "Merk". The Produk Class does not have any subclasses.

- Defining Property

A property describes the relationship between elements in the ontology. There are two types of properties, object property and datatype property. Object property connects two individuals, while datatype properties connect the individual with its datatype value such as string, integer, DateTime. The ontology properties as described in Table 1.

Table 1. Ontology Properties

Property	Function of Property	Type of Property
memilikiPengobatan (in English: Treatment)	Defining the relation between diseases and drugs	Object Property
obatPenyakit (in English: Treatment of)	Defining the relation between drugs and related diseases	Object Property
produkobat (in English:	Defining the relation between drugs and their products	Object Property
no_registrasi, nama_produk, bentuk_sediaan, komposisi, kemasan, pendaftar	Relating string value of registration number, product name, form, composition, packaging and registrant to related drugs.	Datatype Property
tanggal_terbit	Relating dateTime value to drugs' date of issue	Datatype Property

- Defining Individuals and individuals' properties

Individuals in the ontology are the essential element in the hierarchy, which represent an element of a class.
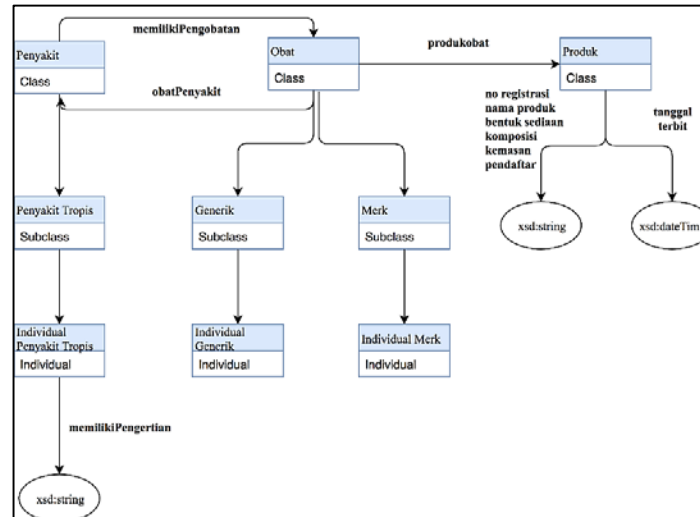

Figure 4. Individuals Ontology

- Text Preprocessing

In this study, chatbot is used as an application user interface that will retrieve sentence as a user's input. Some preprocessing is performed such as case folding, tokenizing, phrase detection, stemming, and filtering. Case folding is a process to convert all letters in lowercase. Tokenizing is a process to split a sentence into word by word or token [6]. Phrase detection is a method to find compound words. Stemming, a means to search the root of the word. Filtering is a process to retrieve essential words. Filtering process uses stop words that were built according to RDF document to make chatbot understand relevant keywords.

- SPARQL Query

SPARQL is a recursive acronym of SPARQL Protocol and RDF Query Language. SPARQL is used to do queries and to manipulate RDF Graph.

- Virtuoso

Openlink virtuoso is a scalable multi cross-platform server, which uses relational data management, graph and document with the web application server and web services functional platform. Virtuoso supports SPARQL to do the RDF query, which is stored in Virtuoso Triple Store. The query result of SPARQL will be in JavaScript Object Notation (JSON) form and will be parsed by system as a search result.

Figure 5. SPARQL Query of JSON Result

4. Result and Discussion

The chatbot will respond to user's query based on the drug's information in ontology on the RDF. RDF document consists of drugs and disease information. Figure 6 is figure out of visualization of ontology in this study, which is combined from two retrieved RDF documents. The RDF consists of the relation of triples or the relation of subjects, predicate, and object about drugs and disease information in RDF components. The graph RDF of this information can be seen in Figure 7 and Figure 8. Those RDF will be developed with ontology so that they are connected, as in Figure 9. The ontology can retrieve the hierarchy relation based on ontology class and subclass hierarchy.
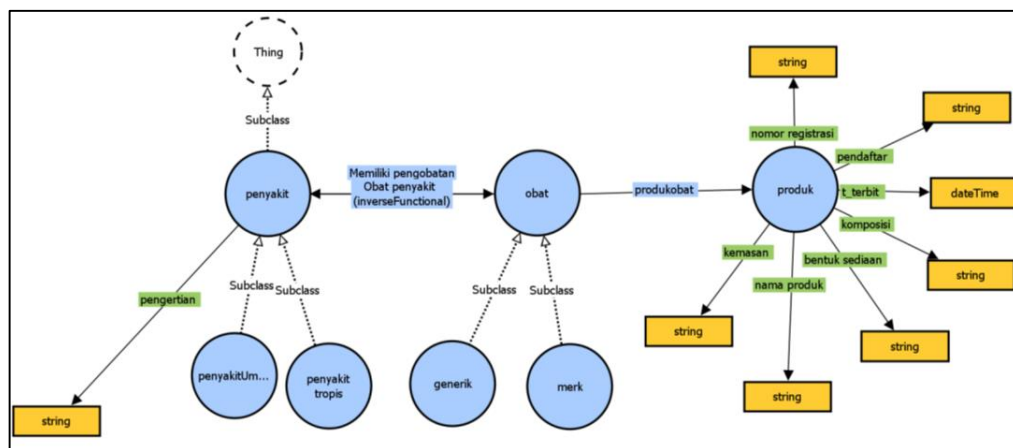

Figure 6. Ontology Visualization
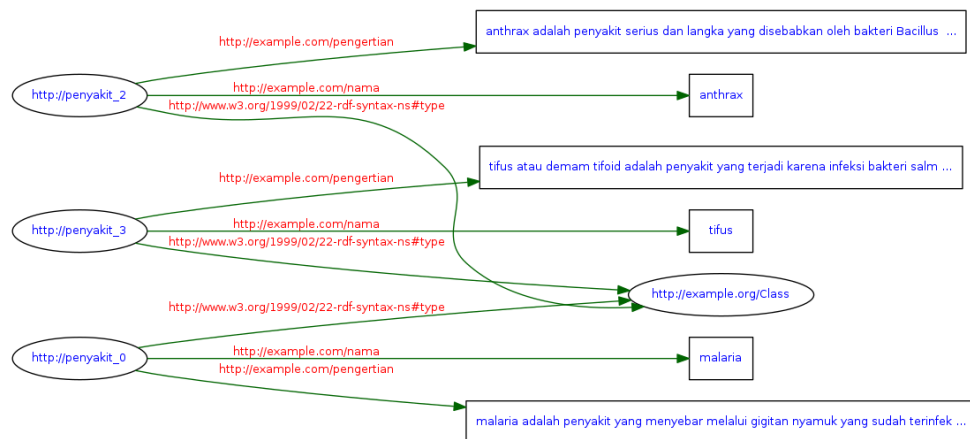
The RDF graph is shown in figure 7 and figure 8.

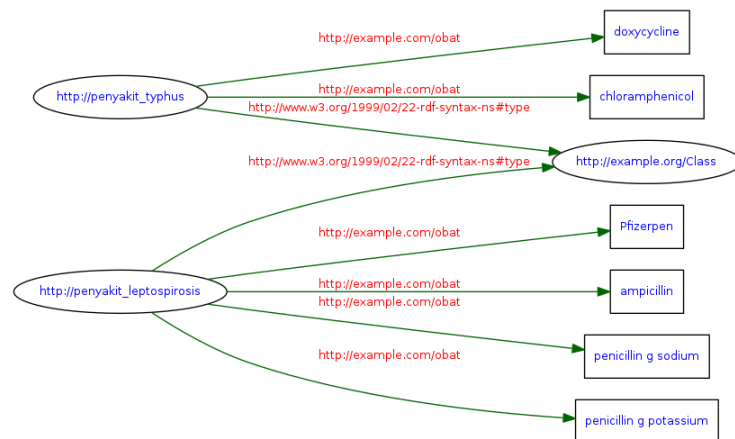**Figure 7**. Graph RDF Disease Data

Figure 8. Graph of Drugs Data
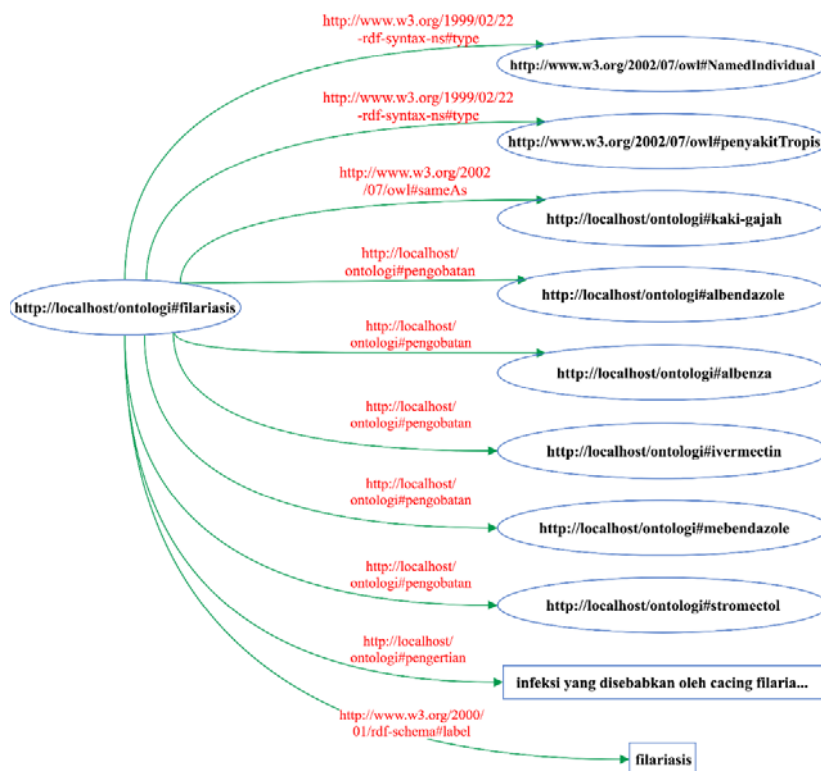
Figure 9. Graph Data of Drugs and Disease Information

The input queries from user is a sentence in Bahasa Indonesia. In this research, there are several pre-processing NLP steps to retrieve only keyword on RDF like case folding, tokenizing, phrase detection, stemming, and filtering. Keyword is an ontology property afterward send into Virtuoso to do the RDF query. Figure 10 is shown the result of the RDF document combination using ontology to describe semantic data from RDF. For example, to describe "filariasis" is similar to "kaki gajah" , in the vocabulary term of owl "sameAs" is used. Filariasis is classified as is a tropical disease class.

Figure 10. Chatbot User Interface

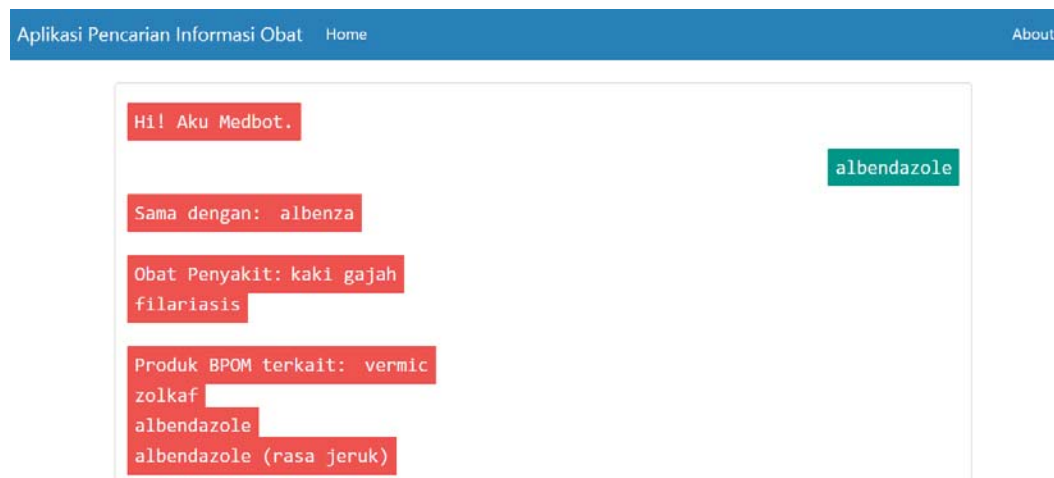Figure 10 shows the result of searching process on chatbot about drug information, where all RDF data is connected. Chatbot evaluation is done by inputting some examples of queries. The examples of input query and responses of the system for evaluation process can be seen in table 2. The result consist of two responses are VALID and NOT VALID. The VALID means the system retrieves a correct response or answer based on the user's question whereas NOT VALID means the system retrieves an incorrect answer.

Table 2 Response Test of Disease Questions

No.	User input	Chatbot response	Result
1	apa itu malaria? (in English: What is Malaria?)	Definition of Malaria: *penyakit yang menyebar melalui gigitan nyamuk yang sudah terinfeksi parasit. Infeksi malaria bisa terjadi hanya dengan satu gigitan nyamuk. Jika tidak ditangani dengan benar, penyakit ini bisa menyebabkan kematian.* *Pengobatan: aralen* *aralen phosphate* *artemether/lumefantrine* *atovaquone* *atovaquone/proguanil* *cardioquin* *oraxyl* *vibra-tabs* *vibramycin*	VALID
2	apa pengobatan demam kuning ? (in English: what is the treatment of yellow jack?)	*Pengobatan: cimetidine* *famotidine* *nizatidine* *ranitidine* *yellow-fever-vaccine*	VALID
3	apa obat untuk demam (in English: How to treat fever)	*Pengobatan: paracetamol* *acetaminophen* *actiprofen* *cetafen*	VALID

		ibuprofen	
4	coba jelaskan tentang demam berdarah (in English: what is demam berdarah?)	Definition of Demam Berdarah*: penyakit yang dapat membuat suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai sakit kepala, nyeri sendi, otot, dan tulang, serta nyeri di bagian belakang mata. Penyakit demam dengue disebabkan oleh virus dengue yang penyebarannya terjadi melalui gigitan nyamuk Aedes aegypti dan Aedes albopictus* *Same as: dengue fever*	VALID
5	penjelasan tentang filariasis (in English: explanation about filarisiasis)	Definition of filarisiasis: *infeksi yang disebabkan oleh cacing filaria. Penyakit ini dapat menyerang hewan maupun manusia. Parasit filaria memiliki ratusan jenis, tapi hanya 8 spesies yang dapat menyebabkan infeksi pada manusia.* Same As: *kaki gajah* *Pengobatan: albendazole* albenza ivermectin mebendazole stromectol	VALID
6	penyakit dengan pengobatan paracetamol (in English: disease with paracetamol treatment)	Treatment for fever	VALID
7	sakit rubella itu apa ya? (in English what is Rubella?)	Definition: *infeksi virus yang ditandai dengan ruam merah pada kulit. Rubella umumnya menyerang anak-anak dan remaja. Menurut data WHO, pada tahun 2016 di Indonesia terdapat lebih dari 800 kasus rubella yang sudah terkonfirmasi melalui pemeriksaan laboratorium.* Pengobatan: BayGam GamaSTAN S/D immune globulin intramuscular	VALID
8	apa obat tipus ? (in English: what is the medice of Typhus?)	Treatment: chloramphenicol doxycycline	VALID
9	tipus sama dengan apa? (in English: what is the synonym of typhus?)	Same As: typhoid typhus demam tifoid tifoid	VALID
10	botulisme itu apa ya? (in English: what is botulisme?)	Definition: *kondisi keracunan serius yang disebabkan oleh racun yang dihasilkan bakteri Clostridium botulinum. Kondisi ini cukup langka, tapi racun yang dihasilkan oleh bakteri ini dikenal sebagai salah satu racun yang paling berbahaya dan mematikan.* Same As: botulism Pengobatan: babybig	VALID

	botulism antitoxin
	botulism immune globulin
	penicillin g potassium
	penicillin g sodium
	pfizerpen

Table 3 Response Test of Drugs Information Questions

No.	User input	Chatbot response	Result
1	jelaskan tentang vibramycin ! (in English: what is vibramycin)	Same As: brodspec panmycin sumycin tetracap Obat Penyakit: brucellosis rickettsial infection Inhalation Bacillus Anthracis lyme disease malaria Produk BPOM terkait: soltrali itracycline ovabiotic ovacyclin	VALID
2	apa produk dari obat albenza ? (in English: what is products of albenza drug?)	ObatPenyakit: kaki gajah filariasis	NOT VALID
3	obat ranitidine untuk penyakit apa? (in English: What for is ranitidine?)	ObatPenyakit: yellow fever	VALID
4	paracetamol untuk apa (in English: what is the purpose of paracetamol?)	Same As: paracetamol cetafen tylenol Treatment of: demam headache Produk BPOM terkait: paracetamol panadol	VALID
5	clindamycin untuk penyakit apa (in English: clindamycin is for..?)	Same As: cleocin-pediatric cleocin phospate clindamycin ObatPenyakit: group a streptococcus infections group b streptococcus infections malaria staphylococcal Produk BPOM terkait: biodasi ethida clinium libroda clinatic	VALID
6	obat herpes dan campak (the medicine of herpes and rubeola)	No Answer Found	NOT VALID
7	omeprazole untuk sakit apa (in English: what treatment is using omeprazole ?)	ObatPenyakit: helicobacter pylori infection	VALID

8	acyclovir untuk pengobatan apa ? (in English: acyclovir for what treatment?)	Treatment of: herpes simplex varicella zoster	VALID
9	gentamicin sama dengan obat apa (in English: gentamicin is the same as what medicine?)	Same As: garamycin	VALID
10	doxycycline produknya apa (in English: what doxycycline products)	*Nomor Registrasi*: dkl0033401101a1 *Tanggal Terbit:* 2014-11-10 *Nama Produk*: soltrali *Bentuk Sediaan: kapsul* 500 mg *Komposisi:* tetracycline hydrochloride *Kemasan*: dus, 10 strip @ 10 capsul *Pendaftar*: solas langgeng sejagtera - indonesia - - *Nomor Registrasi*: dkl0033401101a1 *Tanggal Terbit*: 2014-11-10 Nama Produk: soltrali *Bentuk Sediaa*n: kapsul 500 mg *Komposisi:* tetracycline hydrochloride *Kemasan*: dus, 10 strip @ 10 capsul *Pendaftar*: solas langgeng sejahtera - indonesia - -	VALID

Based on testing of 20 questions or queries, we obtained 18 queries give VALID response to user's query. The accuracy of the chatbot system can be described as follow:

$$\text{Percentage of success} = \frac{\text{success test responses}}{\text{total test}} \times 100\% = \frac{18}{20} \times 100\% = 90\% \tag{1}$$

According to this experiment, we found that the system could detect two keywords from users' input, if there are more than two keywords, percentage of success will be lower. The hardest part in this research is to choose the correct keyword and its SPARQL query counterpart.

5. Conclusion

This study utilizes the RDF of drug information on the tropical disease from the previous research study. The methodology of this study is to retrieve the RDF dataset on tropical diseases drugs from valid resources like BOPM and drugs.com. The queries from users are transformed or pre-processed into the keyword. Furthermore, the system will retrieve the most related triples from the RDF. The evaluation of our system shows that 90% of queries from users generate a valid response from the chatbot. According to this, a semantic web approach can be one solution in building a search semantic engine. This system could be used as a reference to gain information about drugs and tropical diseases in Bahasa Indonesia (Indonesian language).

References

[1]   Amalia A, Afifa R M, and Herriyance H 2018 Resource Description Framework Generation for Tropical Disease Using Web Scraping, in *Int Conf on Communication, Networks and Satellite, Comnetsat* p. 44–48.
[2]   Amalia A, Herryance H and Angriani R 2019 North Sumatera tourism spots searching system based on semantic web, *J. of Physics: Conf. S.* p. 12099.

[3] Hellmund T, Hertweck P, Hilbring D, Moβgraber J, Alexandrakis G, Pouli P and Siatou A 2018 Introducing the HERACLES Ontology-Semantics for Cultural Heritage Management, *Heritage* ,**1** p 377-391.

[4] Ekaningtyas R 2014 Aplikasi Pencarian Produk Obat-obatan yang Beredar Sesuai Standar BPOM Berbasis Web Semantik Universitas Kristen Satya Wacana.

[5] Xu Z, Wei X, Liu Y, Mei L, Hu C, Raymond Choo K, and Zhu Y 2016 Building the search pattern of web users using conceptual semantic space model, *Int J. of Web and Grid Services*

[6] Uppal N, Rana S, and Sharma 2019 A Semantic Web Mining and Semantic Search Engine: A Review.

[7] Kumari A and Thakur J 2019 Semantic Web Search Engines : A Comparative Survey," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **5** 1 p 2456–3307

[8] Taye M M 2011 Web-Based Ontology Languages and its Based Description Logics **11** 11 ISSN: 2078-7952

[9] Rahma, Osmalina & Rahmatillah, Akif. (2019). Drowsiness Analysis Using Common Spatial Pattern and Extreme Learning Machine Based on Electroencephalogram Signal. Journal of Medical Signals & Sensors. 9. 130. 10.4103/jmss.JMSS_54_18.