Knowledge representation of drug using ontology alignment and mapping techniques

Herlina Jayadianti [a,1,*], Alisya Amalia Putri Hasanah [a,2], Yuli Fauziah [a,3], Shoffan Saifullah [b,4]

[a] Department of Information System, Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. Babarsari 2 Yogyakarta, 55281, Indonesia
[b] Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. Babarsari 2 Yogyakarta, 55281, Indonesia
[1] herlina.jayadianti@upnyk.ac.id*; [2] alisyaamalia@gmail.com; [3] yuli.fauziah@upnyk.ac.id; [4] shoffans@upnyk.ac.id
* Corresponding Author

ARTICLE INFO

ABSTRACT

Drug searches are still based on drug names and brands, making it difficult for patients to come looking for a cure by saying that they feel sick. Likewise, when looking for drugs and information about their content to avoid overdose errors when changing drugs when drugs are supposed to be unavailable. Based on the issues raised, a study was conducted on applying semantic web ontology to search for drugs that can appear based on patients' names, compositions, or complaints of diseases. Protégé 5.5 serves to represent drug information based on knowledge. The application uses Netbeans with Jena API as a library and creates data and drug information on the semantic web. Drug search also uses similar in-formation meaning based on user knowledge. By representing knowledge on the search for drug and disease information with semantic web ontology technology, it can meet the purpose of research, namely to improve drug and disease information search following the user's wishes.

1. Introduction

The semantic web [1], [2] uses technology that allows computers to know the meaning of information based on metadata. Metadata can provide an understanding of essential information [3] for computers with search results that can be detailed and precise. The semantic web manages metadata using semantic data modeling techniques such as OWL (Web Ontology Language) and RDF (resource description framework) [4]–[8]. Information search has used the semantic web based on meaning, not just word, syntax, or tagging. Likewise, the data of drugs in pharmacies have the same name but different manufacturers or have other names but contain the same composition [9].

The Regulation of the Minister of Health of the Republic of Indonesia (Number 9 of 2017) states that pharmacies are a means and place for pharmaceutical services for pharmacists to practice. Using semantic web ontology technology in drug searches in pharmacies [10] is more

profitable than using the web with a regular database because it is still based on syntax. Search results depend on words typed, not based on knowledge.

In contrast, the search results will display drug information with different names but the same composition and similarities as a semantic web application [11], [12]. Another problem is that the search can only be based on the name of the drug, composition, or manufacturer, whereas when a patient or customer comes, they often ask for the drug based on the disease they are suffering from, not the name of the drug. Based on the problems, research has been conducted on applying semantic web ontology to look for drugs [13]–[15]. The application can search and show the name, composition, or complaints suffered, and the use of protégés to represent drug information based on knowledge and create data or information about web-based semantic drugs. Previous research was conducted by Afifa [16], establishing a semantic web searching for cures and diseases. Drug and disease data is integrated from 3 websites— drug data obtained from CSV format. The data is then converted to RDF ontology using CSV2RDF, which is most suitable for the Python platform [16]. Similar research has used the semantic web in an integrated information retrieval application, with a knowledge base of drug product information registered with BPOM, which aims to build an ontology-based knowledge model. Drug product data and information storage with an ontology model make searching easy [17]. In addition, other research references are the application of the semantic web in drug product search applications according to BPOM standards [17] and the implementation of web scraping on semantic-based ontology webs for drug and disease data [16].

2. Method

2.1. Fundamental

Ontology [18]–[21] in artificial intelligence has two definitions, namely ontology as a vocabulary (vocabulary) representation and ontology as a core of knowledge or body of knowledge [21], [22]. The ontology allows knowledge that plays a role in connecting related resources. Everyday examples that can be applied using semantic web ontology are like the sentence "I like pie," so some people with different perspectives and knowledge or routines can produce different meanings or meanings. The pie is a cake or sweet dessert with various decorations for the general public and food lovers. For people who struggle with data and statistics daily, the pie is a round curve with its distribution, and people relate it for distribution. Based on technology, especially smartphones, someone thinks that pie is the name of the android nine series. It is an example that everyone has different knowledge.

RDF (Resource Description Framework) is a method for modeling information created by the W3C [18]–[20], [23] with the (primary) idea of creating a web resource statement using subject, predicate, and objects (after this referred to as N-Triple) [5], [23]–[25]. The subject in question is the subject of the resource to be described. Meanwhile, the predicate describes the behavior (characteristics) of the source and the subject and object relationship. RDF is an extensible way to represent web source information flexibly. It represents several components such as personal information, social networks, and metadata about digital artifacts such as music and images and integrates various information sources. The RDF standard query language offered developers and end-users a way to write and use query results across a specific range of information. In particular, RDF can publish and link data on the web [4]–[6], [22]. The RDF schema uses the notion of classes to define categories in classifying resources. The relationship between an instance and its class uses the type property. RDF schema can be used as a reference for creating class and subclass hierarchies and properties and subsets. Domains and constraints define a particular triple object's subject and object types [24]. XML (Extensible Markup Language) [21], [25] is a design language that makes it easy to send documents using the web. Unlike the Hypertext Markup Language (HTML), XML has custom tags for the user. XML can describe the class of data objects (XML documents) and the behavior of computer

processing programs. XML is an application profile or a limited form of the Standard Generalized Markup Language (SGML), with a corresponding document [25]. Ontology Web Language (OWL) is one of the languages used to access documents on certain content.

OWL explicitly represents the meaning of terms in the vocabulary and their relationship, called ontology. OWL has more facilities for expressing meaning and semantics, such as XML, RDF, and RDFS. Thus, OWL surpasses these languages in its ability to represent machine-interpretable content on the web. OWL has been designed to meet the requirements of RDF, RDFS, and XML schemas [25]. Protégé [18] supports creating and editing one or more ontologies in a single workspace via a fully customizable user interface. Visualization tools allow for interactive navigation of ontology relationships. Available refactor operations include joining ontology, moving axioms between ontologies, renaming multiple entities, etc.

2.2. Research Design and Methodology

This research uses a quantitative research methodology in its design and preparation. The quantitative method was chosen because it was the most appropriate for data based on sample data used and processed back into the system. It was supported by some additional knowledge from brief questions and answers with relevant pharmacy officers and the processes carried out during the research process. According to Sugiyono, the quantitative research methodology steps are as follows [26]. The stages in the data collection process include literature study, observation, and interviews [27]. The data that has been collected is then analyzed and processed to obtain information. This study used the waterfall method with five steps: specification, conceptualization, formalization, implementation, and implementation [28]. Specifications relate to system development (why the system was created, its purpose, and its users) [29]. Conceptualization represents an intermediary with text or diagrams. Formalization transforms 'conceptual models' from domain experts to formal or semi-computing models, implemented by an ontology language (implementation).

Maintenance can be used for corrections and upgrades. Direct observations determine the pattern when a person buys drugs at a pharmacy. This activity is carried out for approximately two weeks for 4 hours daily. Patients can also confirm their complaints by visiting or consulting with a doctor on duty while visiting the doctor's office for routine checkups. Interviews were conducted with related people at the research site, namely interviews with two pharmacists, three assistant pharmacists, and one general practitioner. Information from interviews is used in developing the system.

3. Results and Discussion

This section describes the problem we solved with several troubleshooting approaches. People have special knowledge about it. Based on that fact, they may have different views, use different datasets, and have different perceptions. People can have different perceptions (Perception of drugs_1, Perception of drugs_2, and Perceptions_N). Perception is used as a database in the form of data and stored in separate storage. Repositories (db1, db2, and dbN) contain different data, concepts, terms, and semantics. It depends on the group that sees reality (policymakers) and the people who create and store data (users who use technology). Computer users play an essential role in controlling and changing the terminology and semantics of data.
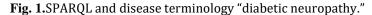
Everyone uses technology to search for data. It is hard for different groups to get the same results. Problems arise when people use data from other groups to share, collaborate, and find more effective solutions—the generation of ontologies from relational sources of drugs. We use ontology alignment and ontology matching techniques. The main contributions of the line of work introduced in this paper are the techniques that aim to improve the steps of creating the relational-to-ontology mapping. This study presents solutions based on different knowledge about the same reality (drugs) based on different perceptions. It uses a mechanism that works

with a standard set of concepts, common terms, semantics, a common language, and a standard set of queries. Users can still use their ideas, terms, and perceptions to input system queries.

This study attempts to provide similar answers (outputs) based on the standard layer between different systems and users based on these solutions. Ontology Alignment creates a new ontology from two or more ontologies overlapping the common parts. The source ontology domain is different from the resulting ontology domain but is related. The standard part is a common word recognized and used with the same meaning by a different user. *Example: Diabetic, Diabetic neuropathy, diabetic retinopathy, Diabetes mellitus, diabetes.* Semantic Web Search Engines such as google accept queries in a format that varies from one tool to another [30]. The concomitant diseases, complaints, the name of the drug, and the manufacturer will also be different in the user's mind, all arranged in one unit of knowledge.
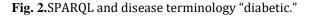
Example: Dizziness, Faintness, Giddiness, Unsteadiness, Vertigo, Lightheadedness, Wooziness. When a person with diabetes wants to buy a drug, he will say the symptoms are felt using many terms. *Example*: *frequent urination, urination.* Other symptoms of Diabetes, Examples: *Tired, overworked, drained, exhausted, sleepy, and exasperated.* Wide selection of Diabetes mellitus medications from doctors, Examples: *Metformin (biguanide), Sulfonylurea, Meglitinide, Thiazolidinediones (glitazone), and Inhibitor DPP-4 (gliptin).* Prefix below is one example of knowledge management with the same disease name but written in different terminology. Users and pharmacists can use both terminologies following the common term to search for a drug. Example: *Diabetic and Diabetic Neuropathy* (Fig. 1 and Fig. 2).


```
SPARQL query:
prefix :<http://www.semanticweb.org/asus/ontologies/2020/11/search.owl#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix xsd: <http://www.w3.org/2001/xmlschema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
select distinct name_of_the_disease ?name_of_the_drug ?equal
where {?obat :healed ?disease.
?disease :healed ?drug.
?drug :drugname ?drugname.
?disease : name_of_the_disease ?name_of_the_drug.
?penyakit owl:sameas ?equal.
filter (regex(?diseasename, "diabetic neuropathy"))}
```

Fig. 1.SPARQL and disease terminology "diabetic neuropathy."

SPARQL is an RDF (semantic) query language used for databases. This query can manipulate (fetching and processing) data stored in the Resource Description Framework (RDF) format. SPARQL [3] has "key-value" data that allows the user to query something specific or data with the RDF specification of the W3C. Thus, the entire database is a threefold set of "subject-predicate-object." It is analogous to the NoSQL database use of "key-value document."


```
SPARQL query:
prefix rdf: <http://www.semanticweb.org/asus/ontologies/2020/11/search.owl#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/xmlschema#>
select ? name_of_the_disease ?name_of_the_drug where {
?disease rdf:healed ?drug.
? disease owl:sameas ?equaldisease.
? disease rdf:diseasename ?name_of_the_disease.
?drug rdf:heal ?disease.
?drug rdf:drugname ?name_of_the_drug.
filter (regex(?name_of_the_disease, "diabetic"))}
```

Fig. 2.SPARQL and disease terminology "diabetic."

Dizziness and frequent urination are certainly not the only symptoms that occur only in people with diabetes. Other diseases may have symptoms such as diabetes; of course, the connectedness of knowledge is indispensable in searching for the right results. In this study, we have successfully created synonyms of terminology traceability for pharmaceuticals by using ontology alignment. The data we tested are shown in Table 1.

Table 1. Synonymous of disease

Terminology	Searching		
	Synonymous	*Tag*	*Knowledge*
Diabetic	Diabetic neuropathy, diabetic retinopathy, Diabetes mellitus, Diabetes, gula, kencing manis, Diabetes, sakit gula	-	√
renal lithiasis	Litiasis renalis, nephrolithiasis, urolitiasis. Nefrolitiasis, sakit ginjal, batu ginjal	-	√
Influenza	Flu, contagion, salesma, epidemic, curse, pilek, pandemic, outbreak, Radang Selaput Lendir, influensa	-	√
Heart attack	cardiovascular disease, serangan jantung, jantung, gagal jantung, heart failure, congestive heart failure, cardiac arrest, acute myocardial infarction	-	√
Malaria	Ague, fever and ague, jungle fever	-	√
Abdominal Pain	Gastric, gastrocolic, stomachic, sakit perut, melilit, mulas	-	√

Table 1 has synonyms in mentioning the name of the disease. There are five people in the pharmacy, and they will likely use different terminology to say the exact disease name. Table 1 is spoken in two languages, English and Bahasa Indonesia. Table 2 shows synonyms of symptoms of the disease.

Table 2. Synonyms of symptoms of the disease

Terminology	Searching		
	Synonymous	*Tag*	*Knowledge*
Dizzy	Dizziness, Faintness, Giddiness, Unsteadiness, Vertigo, Lightheadedness, Wooziness, pusing, pening	-	√
Fever	Delirium, turmoil, demam, panas, warm, red	-	√
Shortness of breath	Asthma, sesak nafas, manggah, sesak	-	√
Tired	Tiredness, overworked, drained, exhausted, sleepy, exasperated.	-	√
Frequent urination	Frequent urination, urination, sering buang air, pipis.	-	√

Not only frequent urination (Table 2), the symptoms of diabetes include increased thirst, tiredness, fatigue, drastic weight loss, itching around the genitals, slow healing wounds, and blurred vision. The same symptoms can appear in different diseases. We connect this knowledge by using relationships to form a complex understanding of diseases and symptoms, in contrast to the expert system where the symptoms must be the same. Someone will use the semantic web and start searching: "How I Fought Neuropathy." The system will search for the connectedness of the symptoms already present in this ontology. Knowledge in this ontology includes Diabetes symptoms: Random blood sugar (RBS): Measurement of sugar levels using a random blood sample and fasting blood sugar (FBS): Measurement of blood glucose levels by fasting for at least 8 hours. Post-prandial blood sugar (PPBS): Measurement of blood glucose levels after 2 hours of food intake. Glycated hemoglobin test: measuring average blood glucose over 2-3 months. Oral glucose tolerance test (OGTT): measurement of blood glucose levels before and 2 hours after consuming a glucose solution. Urine test: detects the excess amount of glucose excreted by the body (Fig. 3).

Fig. 3. The data property of "diabetic neuropathy" symptoms

Drugs are not yet available to cure diabetes. The first step in controlling blood sugar levels is a healthy diet and regular exercise. Available medications are used to manage specific conditions, which work differently. Some are used to increase insulin production, while some can increase insulin resistance. Implementing an information system is usually only an information provider, while an ontology can search based on knowledge.

Table 3. Synonyms of symptoms of the drug

Terminology	Searching		
	Synonymous	*Tag*	*Knowledge*
Drugs for Diabetic	Metformin (biguanide), Sulfonylurea, Meglitinide, Thiazolidinediones (glitazone), and Inhibitor DPP-4 (gliptin). Sulphonylureas.Thiazolidinediones	-	√

Table 3 shows the information of searching, such as Biguanide: helps in sensitizing body tissues to insulin. Sulphonylureas: increase insulin production in the body. Alpha-glucosidase inhibitors: Delay the absorption of carbohydrates. Thiazolidinediones: Increases sensitivity to insulin. It has many side effects, so it takes a prescription to get it. DPP-4 inhibitors: Help lower blood sugar levels without side effects. SGLT2 Inhibitors: Prevents the absorption of excess glucose in the kidneys. In effect, the urine will excrete excessive levels of sugar. In addition, recurrent urinary tract infections increase urination. Insulin injection is helpful for efficiently managing sugar levels. In its use, it is necessary to prescribe if the sugar level is long or very high, and there are many types of insulin. With ontology alignment, symptoms, diagnoses, and treatments are connected in a complex relationship and knowledge of medicine and disease. This system displays not only information but knowledge.

4. Conclusion

In this study, we used ontology alignment to solve differences in terminology when consumers buy drugs in pharmacies. The drug is referred to as the same disease, often using different terms and naming: Diabetes and diabetics. Through ontology, alignment between systems allows sharing of conceptualizations, terminology, and meanings. Ontology alignment is the process of determining the correspondence between concepts. New ontologies are obtained from the commonly overlapping sections of two or more ontologies. Using a standard set of terms (general ontology) is one way to make the system better understand to process information. It is expected that this research will provide additional knowledge for developing drug search engines to accommodate differences in knowledge from consumers. By representing knowledge about drug and disease information based on semantic web ontology technology, it can meet the purpose of research, which is to facilitate the search for drug and disease information following the user's wishes. The use of semantic web ontology technology in search of drug and disease information can improve the effectiveness of the search. The

significance of a search depends on how it represents data and information in knowledge representation.

Acknowledgment

References

[1] S. Staab and R. Studer, "Handbook on Ontologies," 2009, doi: 10.1007/978-3-540-92673-3.

[2] K. Kozaki, Y. Hayashi, M. Sasajima, S. Tarumi, and R. Mizoguchi, "Understanding Semantic Web Applications," *Semant. Web*, pp. 524–539, doi: 10.1007/978-3-540-89704-0_36.

[3] A. Kumari and J. Thakur, "Semantic Web Search Engines : A Comparative Survey," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 107–115, Jan. 2019, doi: 10.32628/CSEIT195115.

[4] "SPARQL Query Results XML Format (Second Edition)." https://www.w3.org/TR/rdf-sparql-XMLres/ (accessed Apr. 03, 2022).

[5] "OWL Web Ontology Language Reference." https://www.w3.org/TR/owl-ref/ (accessed Apr. 03, 2022).

[6] J. Euzenat and P. Shvaiko, "Ontology Matching," 2013, doi: 10.1007/978-3-642-38721-0.

[7] M. Horridge, "A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.3," Mar. 2011, Accessed: Apr. 03, 2022. [Online]. Available: http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf.

[8] A. Rashad Khalifa, "E-Learning and Semantic Web," *Int. J. Intell. Inf. Syst.*, vol. 4, no. 5, p. 84, 2015, doi: 10.11648/j.ijiis.20150405.11.

[9] Meyliana, Surjandy, E. Fernando, C. Cassandra, and Marjuki, "Medicine Information Record Based on Blockchain Technology," *2021 2nd Int. Conf. Innov. Creat. Inf. Technol.*, pp. 169–173, Sep. 2021, doi: 10.1109/ICITech50181.2021.9590133.

[10] G. Barisevičius *et al.*, "Supporting Digital Healthcare Services Using Semantic Web Technologies," pp. 291–306, 2018, doi: 10.1007/978-3-030-00668-6_18.

[11] C. V. S. Avila, W. Franco, A. D. P. Venceslau, T. V. Rolim, V. M. P. Vidal, and V. M. Pequeno, "MediBot: An Ontology-Based Chatbot to Retrieve Drug Information and Compare its Prices," *J. Inf. DATA Manag.*, vol. 12, no. 2, pp. 202–220, 2021, doi: 10.5753/jidm.2021.2148.

[12] M. Asfand-E-Yar and R. Ali, "Semantic Integration of Heterogeneous Databases of Same Domain Using Ontology," *IEEE Access*, vol. 8, pp. 77903–77919, 2020, doi: 10.1109/ACCESS.2020.2988685.

[13] S. Kanza and J. G. Frey, "A new wave of innovation in Semantic web tools for drug discovery," *Expert Opin. Drug Discov.*, vol. 14, no. 5, pp. 433–444, May 2019, doi: 10.1080/17460441.2019.1586880.

[14] M. C. Silva, D. Faria, and C. Pesquita, "Matching Multiple Ontologies to Build a Knowledge Graph for Personalized Medicine," *, al. Semant. Web. ESWC 2022. Lect. Notes Comput. Sci.*, vol. 13261, pp. 461–477, 2022, doi: 10.1007/978-3-031-06981-9_27.

[15] T. Thomas and R. Mervin, "Intelligent Agent System Using Medicine Ontology," *Semant. Web Eff. Healthc.*, pp. 139–157, Nov. 2021, doi: 10.1002/9781119764175.ch6.

[16] R. M. Afifa, "Implementasi Web Scraping Pada Web Semantik Berbasis Ontologi Untuk Data Obat dan Penyakit," *Repos. Institusi Univ. Sumatera Utara*, 2018, Accessed: Apr. 03, 2022.

[Online]. Available: https://repositori.usu.ac.id/handle/123456789/5531.

[17] R. Ekaningtyas, M. A. I. Pakereng, and Hendry, "Aplikasi Pencarian Produk Obat-obatan yang Beredar Sesuai Standar BPOM Berbasis Web Semantik," *Repos. Univ. Kristen Satya Wacana*, 2014, Accessed: Apr. 03, 2022. [Online]. Available: https://repository.uksw.edu/bitstream/123456789/12191/1/T1_672009196_Judul.pdf .

[18] "OWL Web Ontology Language Overview." https://www.w3.org/TR/owl-features/ (accessed Apr. 03, 2022).

[19] N. F. Noy and D. L. Mcguinness, "Ontology development 101: A guide to creating your first ontology," *Tech. Rep. KSL-01-05 SMI-2001-0880, Stanford Knowl. Syst. Lab. Stanford Med. Informatics*, 2001, Accessed: Apr. 03, 2022. [Online]. Available: https://corais.org/sites/default/files/ontology_development_101_aguide_to_creating_your_first_ontology.pdf.

[20] "OWL - Semantic Web Standards." https://www.w3.org/OWL/ (accessed Apr. 03, 2022).

[21] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications - Definition of Ontology as Specification," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, Feb. 1993, Accessed: Apr. 03, 2022. [Online]. Available: https://tomgruber.org/writing/ontolingua-kaj-1993.

[22] "Extensible Markup Language (XML) 1.0 (Fifth Edition)." https://www.w3.org/TR/REC-xml/ (accessed Apr. 03, 2022).

[23] "Protégé." https://protege.stanford.edu/products.php (accessed Apr. 03, 2022).

[24] "RDF 1.1 Primer." https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/ (accessed Apr. 03, 2022).

[25] "XML Schema Part 1: Structures Second Edition." https://www.w3.org/TR/2004/REC-xmlschema-1-20041028/ (accessed Apr. 03, 2022).

[26] Sugiyono, "Metode Penelitian Kuantitatif, Kualitatif dan R&D," pp. 1–334, 2016. Available: Google Scholar.

[27] G. Ataro, "Methods, methodological challenges and lesson learned from phenomenological study about OSCE experience: Overview of paradigm-driven qualitative approach in medical education," *Ann. Med. Surg.*, vol. 49, pp. 19–23, Jan. 2020, doi: 10.1016/j.amsu.2019.11.013.

[28] S. Suhirman, A. T. Hidayat, W. A. Saputra, and S. Saifullah, "Website-Based E-Pharmacy Application Development to Improve Sales Services Using Waterfall Method," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 2, pp. 114–129, Oct. 2021, doi: 10.25008/ijadis.v2i2.1226.

[29] A. T. Hidayat, A. M. D. Dewantara, and S. Saifullah, "The Development of Website on Management Information System for E-commerce and Services," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 380–386, Oct. 2020, doi: 10.32736/sisfokom.v9i3.992.

[30] S. Jain, K. R. Seeja, and R. Jindal, "A fuzzy ontology framework in information retrieval using semantic query expansion," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100009, Apr. 2021, doi: 10.1016/j.jjimei.2021.100009.