Building a Comprehensive Romanian Knowledge Base for Drug Administration

Bogdan Nicula[1], Mihai Dascalu[1], Maria-Dorinela Sirbu[1], Stefan Trăușan-Matu[1] and Alexandru Nuta[2]

[1]University Politehnica of Bucharest, 313 Splaiul Independentei, 060042, Bucharest, Romania
[2]All Business Management, Str. Vasile Conta 19, sc. A, et. 1, ap. 8, Bucharest, Romania

Abstract

Information on drug administration is obtained traditionally from doctors and pharmacists, as well as leaflets which provide in most cases cumbersome and hard-to-follow details. Thus, the need for medical knowledge bases emerges to provide access to concrete and well-structured information which can play an important role in informing patients. This paper introduces a Romanian medical knowledge base focused on drug-drug interactions, on representing relevant drug information, and on symptom-disease relations. The knowledge base was created by extracting and transforming information using Natural Language Processing techniques from both structured and unstructured sources, together with manual annotations. The resulting Romanian ontologies are aligned with larger, well-established, English medical ontologies. Our knowledge base supports queries regarding drugs (e.g., active ingredients, concentration, expiration date), drug-drug interaction, symptom-disease relations, as well as drug-symptom relations (e.g., searching for the drug that might be most useful for treating a given set of symptoms).

1 Introduction

The conventional way of accessing information regarding drug administration and storage strategies, recommendations and precautions, effects and side-effects is via medical leaflets. However, most of the times the text is too complex, too cluttered with information, that the leaflet ends up being ignored by the consumer. At the same time, when consumers must take multiple drugs as result of overlapping treatment schemes, it becomes increasingly difficult for them to keep in mind all mentioned precautions and counter-indications. This paper presents a knowledge base built upon medical leaflets and existing medical ontologies aimed at aiding Romanian consumers when adding new drugs to their treatment schemes.

The knowledge base consists of a set of two ontologies built from information extracted from websites of pharmaceutical producers and national agencies, which were combined with English medical ontologies. The first ontology is used to better structure the leaflet content and provide easy access to information concerning administration and storage strategies, together with possible side-effects. This ontology is aligned with a larger English-based ontology - DINTO (Herrero-Zazo et al., 2015) - in order to discover incompatibilities between drugs and to warn the user whether two administered drugs might interact one with another. The second ontology is focused on diseases. Translations for both diseases and symptoms were added for Romanian language, allowing customers to lookup possible explanations for their symptoms. Moreover, the gap between the two ontologies is filled in by indexing the description and recommendation texts for the drug leaflets, therefore enabling users to directly search what drugs might help them deal with certain symptoms.

These tools are not supposed to replace in any manner actual pharmacists, and an extra opinion from a pharmacist or a physician is always recommended when asking for a drug, given a set of symptoms. The aim of our system is to provide support and easy access to essential information, when no similar solutions are available for Romanian language.

Proceedings of Recent Advances in Natural Language Processing, pages 829–836,
Varna, Bulgaria, Sep 2–4, 2019.
https://doi.org/10.26615/978-954-452-056-4_096

The second section of this paper covers similar knowledge bases, as well as systems providing similar information for English language. The third section describes the data extraction and the architecture of our knowledge base. The last two sections focus on the current results, shortcomings and ways of further improving our knowledge base.

2 Related Work

2.1 Medical Ontologies

Multiple knowledge bases for English language exist, covering different medical areas of interest. Part of them were developed by authors of the Open Biological and Biomedical Ontology (OBO) Foundry (Smith et al., 2007) which focuses on collaborations and on defining a common set of principles for developing medical ontologies. The foundry's mission is to develop a set of interoperable ontologies that are well formed and scientifically accurate. These ontologies are built using semantic web technologies (Berners-Lee et al., 2001) and they are usually made available in OWL (McGuinness et al., 2004) format. Over 150 ontologies are currently listed on the OBO webpage (http://www.obofoundry.org/).

Out of all the ontologies developed by OBO members, the following knowledge bases are relevant for the functionalities presented in this paper:

- CHEBI (Chemical Entities of Biological Interest), containing information regarding a diverse set of chemical compounds relevant for biological interests (Hastings et al., 2015).
- DINTO (The Drug-Drug Interactions Ontology), covering information on how 2 active ingredients interact one with another; DINTO is integrated with CHEBI.
- DOID (Human Disease Ontology), a taxonomy of human diseases (Kibbe et al., 2014).
- SYMP (Symptom Ontology), including symptoms which may be indicative of a disease. SYMP was developed as part of the Gemina project (Schriml et al., 2009) and is integrated with DOID.

Besides OBO, other detailed medical ontologies have been created, such as FMA (The Foundational Model of Anatomy Ontology) (Rosse and Mejino Jr, 2003) developed by the University of Washington, which focuses on the structure of the human body. Generic ontologies are also available, such as DBpedia (Bizer et al., 2009) built by using data from Wikipedia, but they do not offer information relevant for the task at hand.

2.2 Medical Applications

Several applications offering drug-related information exist for English Language. They allow users to search a drug by name, illness or medical procedure, and offer the possibility of testing whether two drugs are incompatible due to harmful interactions between their active substances.

One of the most known portals making medical information available and easy to interpret for non-professional users is WebMD. WebMD offers two applications, Medscape and the WebMD app. Medscape is focused on the more practical aspects of healthcare, offering features such as identifying pills based on a set of physical features, computing the body mass index (BMI) or other relevant metrics based on user's input, and searching for nearby medical professionals and hospitals. The WebMD app is focused more on offering theoretical insights regarding drugs and diseases. It offers the possibility of searching a disease based on a list of symptoms, searching for remedies based on age, gender and severity of the symptoms, and it can notify the user when the US Food and Drug Administration (FDA) has published new information regarding a drug from the user's treatment profile.

Other applications, such as Drugs.com, offer similar features, but also take into account community feedback as a mechanism for informing users. The application facilitates communication within its user-base, allowing customers to find relevant and useful information from peers who underwent similar experiences. This application also differentiates itself from the rest by offering two different views based on the user's medical proficiency. Users with little medical knowledge are directed towards pages with simple and easy-to-grasp information, while experts are provided access to more complex content, which includes more scientific terms.

A specific sub-category of applications is focused on providing drug-administration assistants. One such example is CareZone. These systems allow users to register all drugs on their current treatment scheme, and offer the possibility of entering

and keeping track of different medical parameters, such as blood sugar levels. Apart from that, the application can also be used to set up reminders for administering drugs.

3 Method

3.1 Corpus

There is no established medical or biological ontology for Romanian language. However, public information is readily available in both structured and unstructured format. Medical leaflets are available either as .pdf files or integrated in web pages made available by both private drug producers (e.g., Biofarm) and by state authorities (e.g., The National Agency for Drugs and Medical Devices - ANMDM). The web page of ANMDM also contains structured information (e.g., active substances, concentrations, therapeutical role) for all approved drugs. Figure 1 contains an overview of extracted information from the considered data sources; specific details are provided in the follow-up subsections.

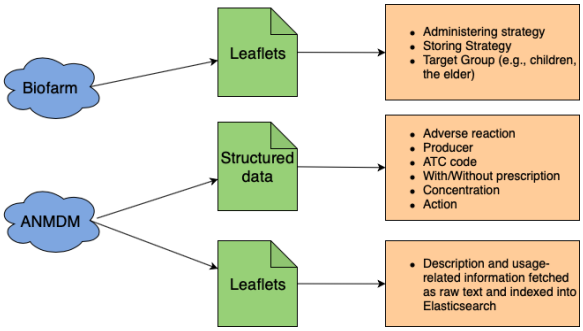


Figure 1: Considered data sources.

The functionalities supported by our knowledge base can be split into three different categories:

- Accessing drug-related information, including inferences of drug-drug interactions, given two or more drugs.
- Finding the most probable disease given a set of symptoms, or listing the set of symptoms that correspond to a certain disease.
- Finding the drug that is most likely to address a given symptom (e.g., flu medication in case of fever and coughing).

3.1.1 Information Regarding Drugs

Both ANMDM structured web information regarding drugs, as well as medical leaflets obtained from ANMDM and private drug producers were used to create a drugs ontology. A total of 220 leaflets from Biofarm (Figure 2) and 1138 leaflets from ANMDM were parsed (Figures 3-4), containing information on 15,093 drugs having 1330 different active substances.

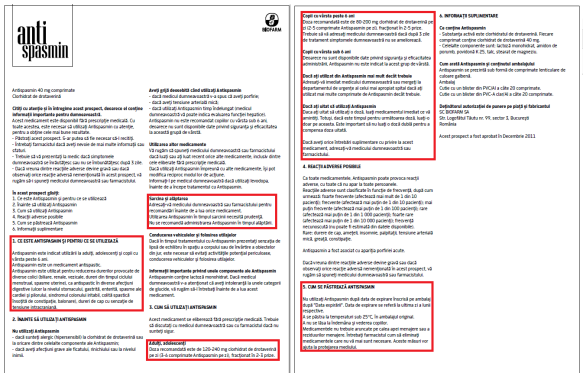
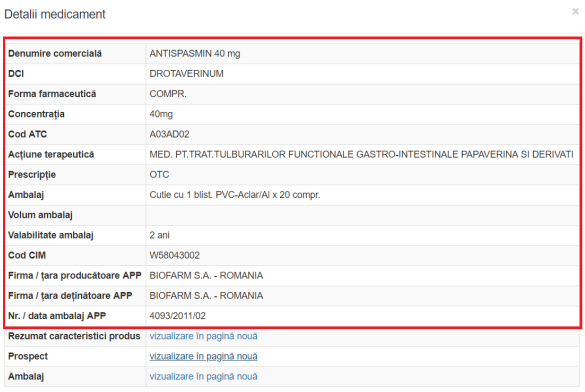


Figure 2: Content extracted from Biofarm leaflets.

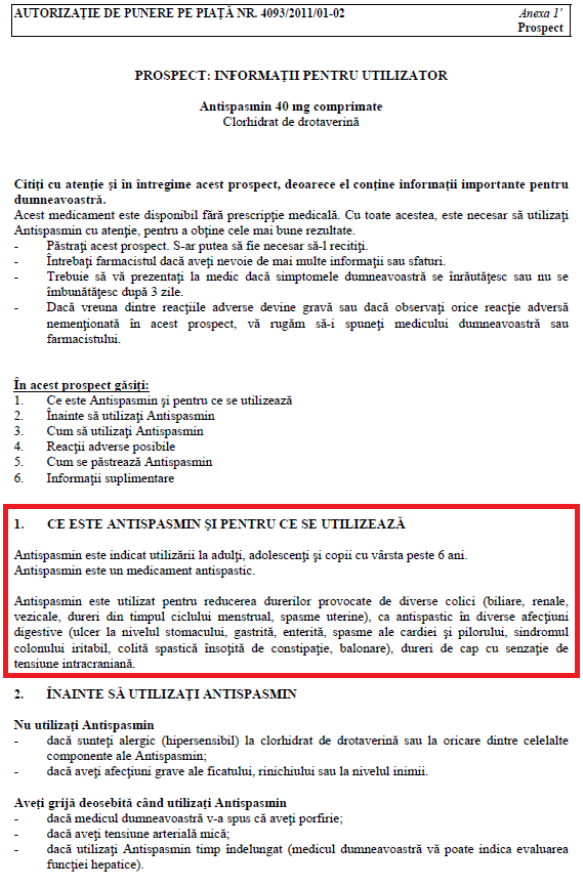

Detalii medicament

Denumire comercială	ANTISPASMIN 40 mg
DCI	DROTAVERINUM
Forma farmaceutică	COMPR.
Concentrația	40mg
Cod ATC	A03AD02
Acțiune terapeutică	MED. PT.TRAT.TULBURARILOR FUNCTIONALE GASTRO-INTESTINALE PAPAVERINA SI DERIVATI
Prescripție	OTC
Ambalaj	Cutie cu 1 blist. PVC-Aclar/Al x 20 compr.
Volum ambalaj	
Valabilitate ambalaj	2 ani
Cod CIM	W58043002
Firma / țara producătoare APP	BIOFARM S.A. - ROMANIA
Firma / țara deținătoare APP	BIOFARM S.A. - ROMANIA
Nr. / data ambalaj APP	4093/2011/02
Rezumat caracteristici produs	vizualizare în pagină nouă
Prospect	vizualizare în pagină nouă
Ambalaj	vizualizare în pagină nouă

Figure 3: Structured information scrapped from ANMDM.

The ontology was built from scratch and the attributes for each drug instance were either copied from the structured ANMDM web entry, or were extracted from the corresponding medical leaflets (such as the administering strategy). The extraction was done partially via a rule-based approach, but manual extraction was necessary for some of the more complex entries, when no reliable pattern could be identified.

As it can be seen in Figure 5, the ontology is centered on the Drug class, which represents a certain type of drug, but not an actual product that can be bought in stores. All instances of the Drug



AUTORIZAȚIE DE PUNERE PE PIAȚĂ NR. 4093/2011/01-02 *Anexa 1'* Prospect

PROSPECT: INFORMAȚII PENTRU UTILIZATOR

Antispasmin 40 mg comprimate
Clorhidrat de drotaverină

Citiți cu atenție și în întregime acest prospect, deoarece el conține informații importante pentru dumneavoastră.
Acest medicament este disponibil fără prescripție medicală. Cu toate acestea, este necesar să utilizați Antispasmin cu atenție, pentru a obține cele mai bune rezultate.
- Păstrați acest prospect. S-ar putea să fie necesar să-l recitiți.
- Întrebați farmacistul dacă aveți nevoie de mai multe informații sau sfaturi.
- Trebuie să vă prezentați la medic dacă simptomele dumneavoastră se înrăutățesc sau nu se îmbunătățesc după 3 zile.
- Dacă vreuna dintre reacțiile adverse devine gravă sau dacă observați orice reacție adversă nemenționată în acest prospect, vă rugăm să-i spuneți medicului dumneavoastră sau farmacistului.

În acest prospect găsiți:
1. Ce este Antispasmin și pentru ce se utilizează
2. Înainte să utilizați Antispasmin
3. Cum să utilizați Antispasmin
4. Reacții adverse posibile
5. Cum se păstrează Antispasmin
6. Informații suplimentare

1. CE ESTE ANTISPASMIN ȘI PENTRU CE SE UTILIZEAZĂ

Antispasmin este indicat utilizării la adulți, adolescenți și copii cu vârsta peste 6 ani.
Antispasmin este un medicament antispastic.

Antispasmin este utilizat pentru reducerea durerilor provocate de diverse colici (biliare, renale, vezicale, dureri din timpul ciclului menstrual, spasme uterine), ca antispastic în diverse afecțiuni digestive (ulcer la nivelul stomacului, gastrită, enterită, spasme ale cardiei și pilorului, sindromul colonului iritabil, colită spastică însoțită de constipație, balonare), dureri de cap cu senzație de tensiune intracraniană.

2. ÎNAINTE SĂ UTILIZAȚI ANTISPASMIN

Nu utilizați Antispasmin
- dacă sunteți alergic (hipersensibil) la clorhidrat de drotaverină sau la oricare dintre celelalte componente ale Antispasmin;
- dacă aveți afecțiuni grave ale ficatului, rinichiului sau la nivelul inimii.

Aveți grijă deosebită când utilizați Antispasmin
- dacă medicul dumneavoastră v-a spus că aveți porfirie;
- dacă aveți tensiune arterială mică;
- dacă utilizați Antispasmin timp îndelungat (medicul dumneavoastră vă poate indica evaluarea funcției hepatice).

Figure 4: Content extracted from ANMDM leaflets.

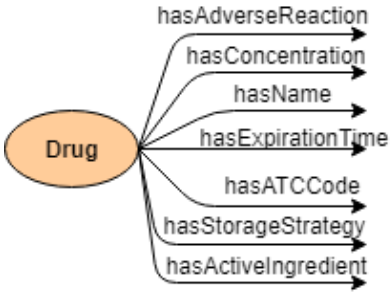


Figure 5: Romanian drug ontology - drug class simplified view.

class contain the same combination of active ingredients, but they may have different concentrations, different names, different expiration times, etc.

This ontology was aligned to its larger English counterpart, DINTO. The data from the two ontologies was merged at active substance level. Difficulties were encountered as, even though most active substances use conventional names for chemical entities derived from Latin, their names differ from Romanian to English (for instance, ”acidum ascorbicum” in the Romanian ontology is equivalent with ”ascorbic acid” in DINTO). This alignment was done in 2 phases. First, 500 active substances were merged because they either represented perfect matches, or they matched after applying a small set of conventional changes (e.g., removing the ”-um” prefix from the Romanian version). Second, the remaining 800 active substances were matched by analyzing the closest correspondent in the other ontology in terms of Levenshtein distance (Levenshtein, 1966), followed by a manual validation of the match.

3.1.2 Information Regarding Symptoms and Diseases

The English DOID and SYMP ontologies were used as a starting point (see Figure 6). Both the names of the symptoms and of the diseases were translated into Romanian via Google Translate, and then they were manually corrected in order to eliminate mistakes. Names containing polysemic words proved to be most difficult for the automated translation – for instance, a symptom describing a change in the pupil, was mistakenly interpreted as something related to a school student, not to the human eye.

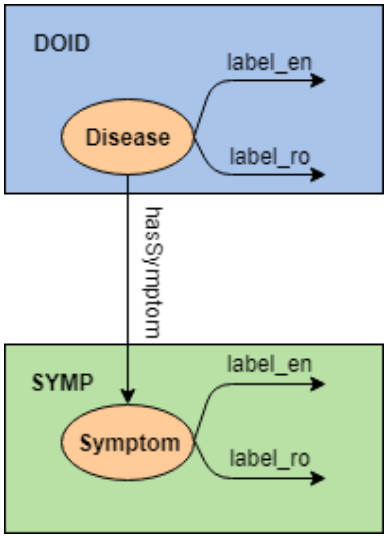


Figure 6: Romanian disease ontology - simplified view.

In total, names for over 900 symptoms and 10000 diseases were translated, allowing Romanian users to use the full benefits offered by the DOID ontology, without the need of English proficiency.

3.1.3 Connecting Symptoms to Drugs

There are no datasets containing information on what drugs should be administered for certain symptom in Romanian language. However, a usage section exists in each medical leaflet describing cases in which the drug should be taken (e.g., to numb pain, to reduce coughing, to lower body temperature and eliminate fever, etc.). Thus, the usage section for each of the 1138 leaflets was extracted from the ANMDM website and was indexed into Elasticsearch (Divya and Goyal, 2013), allowing users to find drugs with the usage description that most closely fits the set of provided symptoms.

3.2 Architecture and Processing Pipeline

The knowledge base can be hosted on a single server consisting of two different applications (see Figure 7). First, an instance of a Fuseki semantic repository server (Jena, 2019) was used to host the aforementioned ontologies: the Romanian drug ontology, DINTO, and a merge between DOID and SYMP called DOID-merged, to which we have added extra labels for Romanian. This repository allows users to query the ontologies via the SPARQL query language (Prud'hommeaux and Seaborne, 2008). Second, an Elasticsearch instance stores unstructured leaflets and can be queried in order to find drugs that are most likely helpful in relieving one or more symptoms.

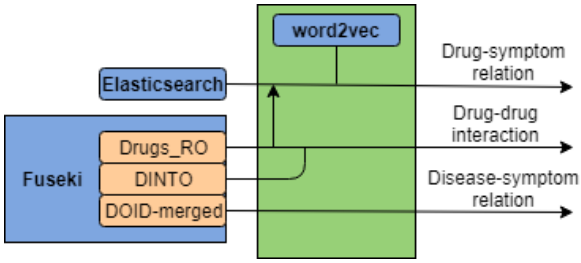


Figure 7: Knowledge base architecture.

On top of the two applications that act as information sources, a user interface allows users lacking experience on semantic web technologies to easily access the information. The user requests made at these endpoints are transformed into valid SPARQL and Elasticsearch queries. Furthermore, this interface can act as an autocorrect, by suggesting symptoms and drugs present in our dataset.

In the case of querying for the most helpful drugs given a set of symptoms, the relevance of the Elasticsearch information retrieval is improved by adding semantic information. For a given set of candidate texts from leaflets, each corresponding to a different drug, a word2vec (Mikolov et al., 2013) model trained on a 1-billion word Romanian corpus is used to compute the semantic similarity between it and the given set of symptoms. This is done by computing aggregate embeddings for the two text representations, and then computing cosine similarity. The Elasticsearch query score and the cosine similarity are both min-max normalized with regard to the set of candidate symptom-drug pairs extracted with Elasticsearch, their average is computed, and the candidate having the best average score is selected in the end.

Extra functionalities are added on top of the Fuseki repository and Elasticsearch server, such as suggesting possible matches when the string representing a searched drug/disease/symptom was not found. These suggestions are made by using a Levenshtein edit distance. In the case of symptoms, however, the same sensation can be expressed in multiple ways; thus, we rely on a very strict Levenshtein distance search to account for small typing errors. If this fails, a semantic search is used, looking for the symptom in the knowledge base for which the word2vec embedding is closest to the embedding of the input text.

4 Results

4.1 Drug Information and Drug-Drug Interactions

Our knowledge base offers access to structured information regarding the 15,093 drugs and 1,330 active substances. The information regarding drugs includes both numeric attributes (e.g., time until expiration), as well as text attributes (e.g., usage recommendations which were not standardized as format).

Users can search for the list of drugs with which any given one may interact because the Romanian drug ontology was aligned with DINTO. This search is done at active substance level. In case of drugs containing a combination of active substances, we consider that drugs A and B may interact if, for at least one active substance from A, there is at least one active substance from B with which it interacts. For example, if the user wants to check the interaction between *"OTOTIS"* which is based on a combination of two active ingredients (namely *"ciprofloxacinum"* and *"fluocinolonum"*) and *"ENAFILZIL"* which is based on

"sildenafilum", the knowledge base will search if either *"ciprofloxacinum"* or *"fluocinolonum"* interact with *"sildenafilum"*. As *"ciprofloxacinum"* interacts with *"sildenafilum"*, the system will conclude that the 2 drugs interact. If the same user wants to find the list of all the drugs which interact with *"OTOTIS"*, a list of all the drugs containing at least one active substance that interacts with either of *"ciprofloxacinum"* or *"fluocinolonum"* is generated and it will contain 35 entries.

4.2 Symptom-Disease Information

Our consolidated knowledge base contains 900 symptoms and 10,000 diseases from DOID and SYMP with names translated into Romanian language. The user can enter a list of symptoms to search for the disease that matches most symptoms. As mentioned before, if a symptom does not exist, two sequential attempts are performed to find its closest correspondent in our knowledge base. First, symptoms with a Levenshtein distance of 2 or less are searched. Second, if no result is found in the previous step, a word2vec embedding of the input string is computed, and the symptom from the knowledge base having the closest embedding to it is considered its equivalent.

For example, if a user searches for diseases that have *"mic de statură"* (eng, *"short stature"*) as a symptom, the results would be *"trichorhinophalangeal syndrome type II"*, *"Albright's hereditary osteodystrophy"*, *"spondyloepimetaphyseal dysplasia, strudwick type"* or *"Renpenning syndrome"*, as these are the only diseases linked to that symptom according to DOID and SYMP. If users make 1-2 typos when writing the symptom, the most relevant symptoms is suggested, and they can redo the search with the correct version. If they enter *"corp mic"* (eng, *"small body"*) or *"scund"* (eng, *"short"*), the first recommendation would fail, but the second one would suggest "short stature" as the most similar symptom based on a semantic similarity search.

4.3 Connecting Symptoms to Relevant Drugs

Considering a search for a combination of *"tuse, febră, durere de cap"* (eng, *"coughing, fever, headache"*), several types of analgesics, aspirin, paracetamol, and 2 types of cough syrup are recommended. The bottom results focus on yellow fever or other diseases that contain only a part of the symptoms specified as input.

In most simple use cases, the first entries are relevant. However, the symptom of a disease can be, in some cases, the effect of a drug that is targeted against a totally different disease. For instance, if the user searches symptoms related to diarrhea, such as *"scaun moale"* (eng, *"loose stool"*), some of the first drugs to be recommended are laxatives, but this would not be the wisest choice of medication. This happens because the effect of the medicine is, in some cases, mentioned alongside with the symptoms it should alleviate.

5 Conclusions

This paper presents a medical knowledge base for Romanian language focused on drugs. It is built using medical leaflets in Romanian and structured information regarding the drugs that was extracted from the ANMDM website. The knowledge base is also integrated with English ontologies in order to make more powerful inferences, such as searching for drug-drug interactions. The provided information can be structured into three main categories: drug related information, disease-symptom information, and drug-symptom information. To our knowing, this is the most comprehensive effort of building a knowledge base for Romanian drugs, their counter-indications, as well as potential relations to exhibited conditions.

The drug related information was extracted directly from official Romanian sources, thus it can be considered reliable. In order to keep the knowledge base up to date, the sources need to be crawled periodically in order to ensure that new information is always added, and that deprecated records are eliminated promptly. Information concerning drug-drug interactions is based on the DINTO ontology, which was last updated in 2016 and is still relevant. Nevertheless, we warn users that the information presented by our services is not a valid substitute for the opinion of a medical professional or a pharmacist. In the future, apart from drug-drug interactions, the knowledge base could also take into account pre-existing conditions or dietary choices which may interact with a certain treatment scheme. Part of this information is already available in DINTO, but it would need to be translated and integrated in our knowledge base.

The disease-symptom information is based on the DOID and SYMP ontologies. The name of the

diseases and symptoms were automatically translated using Google Translate, and then manually corrected, if necessary. The two ontologies are still actively maintained; thus, the knowledge base needs to refresh this information from time to time in order to get the latest version. As is the case with the previous category, this information is not exhaustive and cannot substitute the knowledge of a professional.

The drug-symptom information is based only on medical leaflets crawled from the ANMDM website, which needs to be updated every several weeks. In some cases, the drug-symptom queries are very effective, for instance when searching for drugs targeting flu-like symptoms, such as fever and coughing. In other cases, the queries mistake the symptoms that the drug should address, with the drug's effect. These types of mistakes cannot be avoided for the time being due to manner in which the information was indexed. If more structured information regarding the drug-symptom relation could be extracted from the leaflets, either manually or by using different NLP techniques, these outlier cases would be addressed.

Our knowledge base provides real aid for Romanian users requiring drug-related information, and no similar initiatives exist at national level. The system cannot substitute the knowledge of a professional, and there are still problems to be addressed, but it is still an easy-to-use and useful tool for informing a user on medical treatments. Further improvements will be explored, including the orientation towards a personal health assistant for drug administration, similar in some degree to Babylon Health AI (https://www.babylonhealth.com/ai).

6 Future Work

In the future, we aim to expand even further our knowledge base. This can be done by indexing medical leaflets from other drug producers, as well as extracting more complex information from leaflets - for instance, contraindications expressed as rules (e.g. do not take certain antibiotics, such as tetracycline, with milk, other dairy products, calcium supplements, or antacids). Furthermore, we aim to integrate our knowledge base with other information sources, such as the Unified Medical Language System (Bodenreider, 2004).

7 Acknowledgements

The work presented in this paper has been funded by the "Intelligent platform for drugs administration – PIAM", subsidiary contract no. 1267/22.01.2018, from the NETIO project ID: P_40_270, MySMIS Code: 105976.

References

Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284(5):34–43.

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia - A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web* 7(3):154–165. https://doi.org/10.1016/j.websem.2009.07.002.

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research* 32(suppl_1):D267–D270.

Manda Sai Divya and Shiv Kumar Goyal. 2013. Elasticsearch: An advanced and quick search technique to handle voluminous data. *Compusoft* 2(6):171.

Janna Hastings, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, and Christoph Steinbeck. 2015. Chebi in 2016: Improved services and an expanding collection of metabolites. *Nucleic acids research* 44(D1):D1214–D1219.

María Herrero-Zazo, Isabel Segura-Bedmar, Janna Hastings, and Paloma Martínez. 2015. Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms. *Journal of chemical information and modeling* 55(8):1698–1707.

Apache Jena. 2019. Apache jena fuseki documentation. http://jena.apache.org/documentation/fuseki2/.

Warren A Kibbe, Cesar Arze, Victor Felix, Elvira Mitraka, Evan Bolton, Gang Fu, Christopher J Mungall, Janos X Binder, James Malone, Drashtti Vasant, et al. 2014. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research* 43(D1):D1071–D1078.

Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*. volume 10, pages 707–710.

Deborah L McGuinness, Frank Van Harmelen, et al. 2004. Owl web ontology language overview. *W3C recommendation* 10(10):2004.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv* http://arxiv.org/abs/1301.3781.

Eric Prud'hommeaux and Andy Seaborne. 2008. SPARQL Query Language for RDF. *W3C Recommendation* 2009(January):1–106. http://www.w3.org/TR/rdf-sparql-query/.

Cornelius Rosse and José LV Mejino Jr. 2003. A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of biomedical informatics* 36(6):478–500.

Lynn M Schriml, Cesar Arze, Suvarna Nadendla, Anu Ganapathy, Victor Felix, Anup Mahurkar, Katherine Phillippy, Aaron Gussman, Sam Angiuoli, Elodie Ghedin, et al. 2009. Gemina, genomic metadata for infectious agents, a geospatial surveillance pathogen database. *Nucleic acids research* 38(suppl_1):D754–D764.

Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, et al. 2007. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25(11):1251.