Applied Medical Informatics
Original Research Vol. 40, No. 1-2 /2018, pp: 7-14

Extracting and Structuring Drug Information to Improve e-Prescription and Streamline Medical Treatment

Oana-Sorina LUPŞE* and Lăcrămioara STOICU-TIVADAR

Faculty of Automation and Computers, University Politehnica Timişoara, Blvd. Vasile Pârvan, no. 2, 300223, Timişoara, Romania.
E-mails: oana.lupse@aut.upt.ro; lacramioara.stoicu-tivadar@aut.upt.ro

*Author to whom correspondence should be addressed; Tel.: +4-0256-403228; Fax: +4-0256-403214

Received: January 2, 2018 /Accepted: May 11, 2018 / Published online: June 30, 2018

Abstract
Currently, physicians are using the patient electronic health record (EHR) to support their practice. The Romanian healthcare system switched to the electronic prescription starting with 2012. Physicians use the electronic medical record and health card to access patient data whenever available. To improve the medical act, we propose a tool supporting the prescription process, structuring and extracting important information from drug characteristics leaflets (prospectus). The application processes data extracted from around 3.000 medical prospectuses using several Romanian language Web sources. The drug leaflet data is structured on sections: therapeutic action, contraindications, mode of administration, adverse reactions, etc. A stemming algorithm has been applied to each section, extracting the root of the word for an easy search. The result is a text in an *.xml file. After structuring step, the application searches in the structured file the necessary information to prescribe the patient's medication as closely as possible related to patient state. The application suggests all the drugs matching the patient's disease and are not contraindicated, or enter in conflict with other diseases, treatments or allergies of the patient, and the physician may select the best solution for the given situation.

Keywords: Medical prescription; Prospectuses; Stemming; Electronic health record

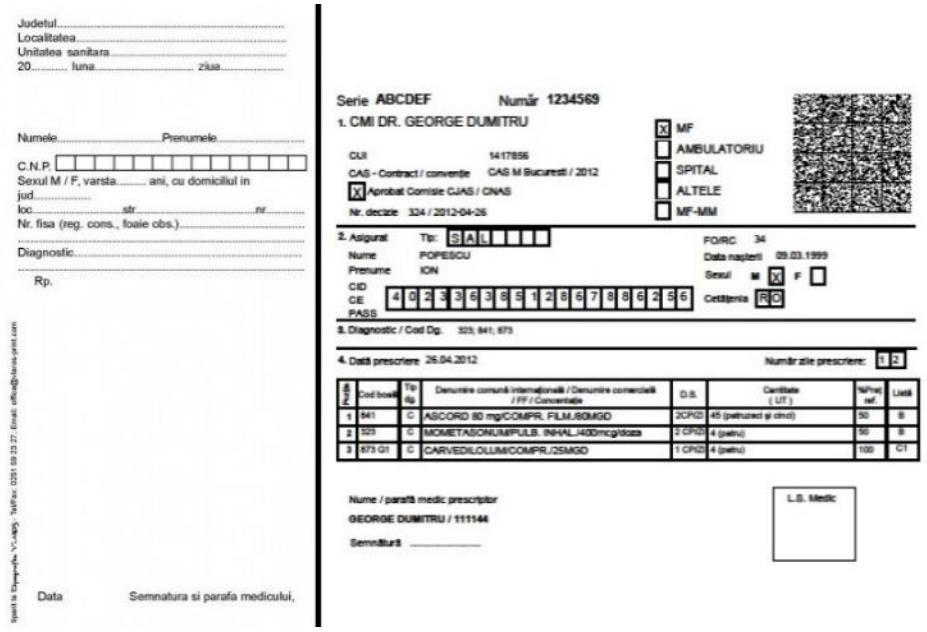
Introduction

The medical prescription process plays a very important role in treating patients and has the potential to generate health risks if the chosen drugs are not appropriate for the patient. When a physician chooses a medication he/she must consider many factors, some of them generating potential health hazard to the patient. Mistakes in prescribing medication are often related to poor information about the patient's history, current medication of the patient, allergies, contraindications and other particularities of the prescribed medicine. In Germany in 2006, 3900 cases of malpraxis were reported, with regards to errors in treatment prescription [1]. In Romania, electronical patient medical records are managed by SIUI (Unique Integrated Information System) application where physicians store patient medical data and that provides reports and statistics for CNAS (National Health Insurance Agency). Physicians and other healthcare providers input in this system patient's consultations, medical history, diseases and treatments. SIUI system has three modules: Health Card, e-Prescription and Electronic Health Record [2].

The Health Card allows physicians to access the medical information of the patient, Electronic Prescription is used to control drug prescription and monitor the drug consumption and the module Electronic Health Record is in a developing state [2].

Authorized physicians and patients connected from any device to Internet may check medical information as provided by the national Insurance House [3]: allergies and intolerances, chronic diseases, history of diseases/diagnoses (other than allergies and chronic), interventions and procedures, medical services, immunizations, and treatments.

The medical prescription in Romania has two formats: a hand-written prescription with basic information of the patient – name, demographics, diagnosis and medical recommendation (Figure 1.a), for which the patient is paying all the medication, and a form generated by a medical information system in the physician's office (Figure 1.b) containing patient name, demographics, patient ID, QR code, insurance status, diagnosis codes and active substance prescriptions with doses recommendation, useful for free or compensated medicines.



(a) (b)

Figure 1. Romanian medical prescription – template for hand written form (a) and for electronic use (b)

In Romania, the e-Prescription was implemented as a pilot project in 2012 and was launched and used by all healthcare providers (hospitals, practices, pharmacies) 1st of January 2013 [4].

The problems and needs in the field of medical prescriptions are the following:

- More and more discovered diseases requiring effective treatments;
- Discovery of new treatments/medications which doctors need to access;
- Malpraxis of physicians (prescription of wrong treatments/medication);
- Accessing all information regarding the medication and diseases in a prompt manner;

The information needed to assist the physician in deciding the appropriate treatment is taken from the Electronic Health Record of the patient and from the information provided by each drug manufacturer.

The most important sections of a medical prospectus are: Commercial name; DCI (International Joint Name); Pharmaceutical presentation; Weight; Active substance; Other ingredients (contained substances); Producer; What it treats; Contraindications: Allergies, Other diseases, Other medicines; Indications by age; Indications in pregnancy; Breastfeeding instructions; Driving alerts; Overdose; Adverse effects.

In conclusion, according to presented information, in the national healthcare system it would be a real help for the doctors to have the most detailed information about the medicines considering the progress made in the field of medical applications, especially for young practitioners and residents.

State-of-the-Art

The main source of errors in medicine is related to drugs prescriptions in spite of several resources relating the drug prospectuses being available online [5, 6, 7]. A quarter of the total mistakes in medical errors are produced by prescriptions and this has serious consequences to patient's health. 50% of these errors are related to medical prescription and 50% are caused by drug-drug interactions [8].

Several studies were conducted and different models to support physicians during the medical prescription process were reported in the scientific literature. Luna et al. [8] developed a decision support system to alert doctors with regards to the interactions between drugs as a tool to support the decrease of medical errors. Keil et al. [9] presented a drug management system associated to a smartwatch to allow physicians to manage the patient's drugs during treatment. With this system, doctors can control the patient's treatment and properly recommend them, reducing errors in drug administration.

To reduce risk of medication errors a module [10] that collects successful treatments for certain diseases starting from patient characteristics, saving and making them available on a cloud database for future suggestions of treatments for other patients with the same characteristics has been developed. The physicians authorized to access these data may use it to compare situations and select the best treatment for a certain situation.

Fung et al. [11] compared three knowledge databases in use in clinical practice for potential drug-drug interactions. The three knowledge databases (First DataBank, Micromedex, Multum) were mapped to RxNorm and compared from clinical and composition point of view and drug-drug interaction levels. The developed knowledge database covered 99.8-99.9% of alerts from the Office of the National Coordinator for Health Technology reference list of drug-drug interaction.

Deléger et al. presented a system that extract drugs and information about them from French clinical texts after a system initially designed for English language [12]. The system has a specialized lexicon and a set of extraction rules. The system was evaluated for 50 clinical texts and 86.7 % accuracy in French language was obtained, a higher level than in the original English system.

The goal of our study was to develop a tool for prescribing medication through structuring and extracting the necessary information from drug prospectuses available on Internet. The benefits consist in having all information about drugs from multiple sources and providing a structure where it is easier to add the information in medical applications.

Material and Method

Methods

The prospectuses have a different format and different section names (Figure 2), that makes them hard to analyze in a coherent manner.

Each prospectus is saved as an xml file (based on the Internet source) ensuring a better interoperability within the application. From several sources the prospectuses are collected directly on sections, by html tags, and for others they are unstructured, available only as text. We used the resulting structured file as a module for the clinical decision support system in medical applications for Romanian prescriptions. For the collected unstructured prospects, we created an algorithm to divide the content into specific sections. To prepare for implementation, we created a file with section names that may appear in each prospectus (therapeutic action, way of administration, pregnancy and breastfeeding, composition, dosage, etc.).

The steps required to implement the structuring algorithm on sections are:

- Create lists that save the start and end positions of each section name
- Look for section names in the prospectus text

- Sort the names of the sections found in the text to remove the duplicates (e.g., for the text prospectus *Therapeutic action*, the algorithm will interpret as sections *action* and also *therapeutic action* and will consider the longer text).
- Delete sections with the same starting item or sections that are contained in other sections
- Delete similar words in sections
- Extract the section name and the related text to the next section of the prospectus
- Create the xml file with the prospectus with sections

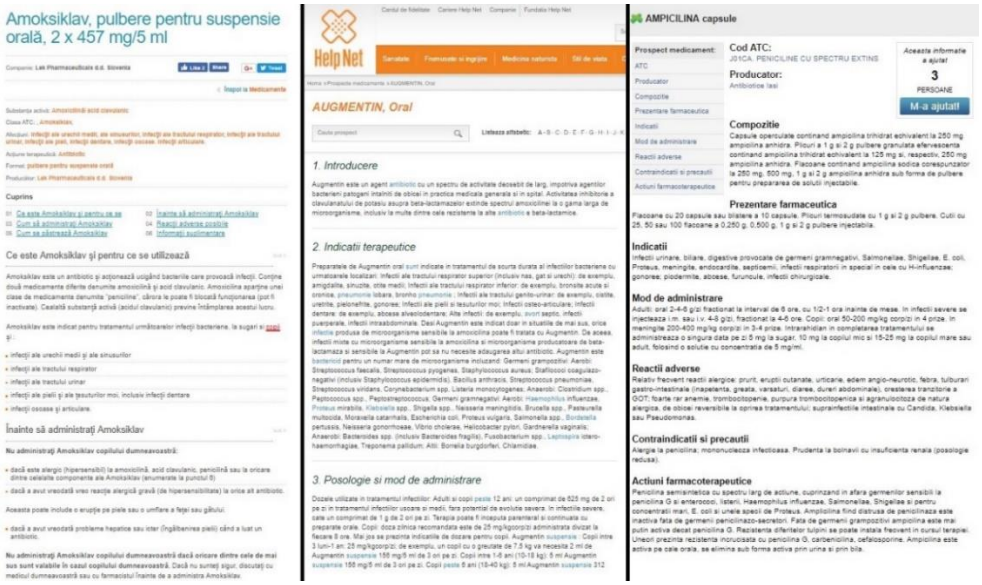


Figure 2. Examples for 3 on-line medical prospectuses

Figure 3 shows the files obtained from the collected prospectuses, formatted with tag sections and without sections.

The next step after completion of the splitting algorithm is to perform an uniformization of the information in the prospectuses. As can be seen in each prospectus, for the same type of section, different names may appear, e.g. for indications in the use of a drug, we may have one of the several section names: action, pharmaceutical action, pharmacodynamics action, pharmacological action, pharmaco-therapeutic action, therapeutic action, therapeutic action and indications, indications or therapeutic indications. To reach a common name for the same type of section, we use regular expressions and create an algorithm that performs a uniformization of the sections and creates an xml file with the same names for all the prospectuses related to the same section content.

The newly created file contains all prospectuses with homogeneous sections names. The next step in our application is to simplify the text that appears in different sections where the application will look for information needed to treat a certain disease. To simplify the text (removing link words and bringing words to a shorter form, extract the root of the word), we apply a stop word removal algorithm and a Romanian stemming algorithm. The stemming algorithm divides text into words and extracts its root by eliminating ending of words (e.g., Stemming in Romanian language: the words *tratamentul, tratamente, tratamentele* are transformed all in *tratament*). This stemming algorithm is applied to each section of the prospectus. There are medical terms for which, when we remove the ending of the word, the medical term does not have the same meaning or means nothing. To avoid this error, we created a file with medical terms for which word change makes no sense. For these words the stemming algorithm is not applied in order not to distort the correct medical terms. The sections that were considered so far were the therapeutic indications section, dosage and contraindications. These sections are considered as the most important in establishing the appropriate treatment. In the therapeutic indication section, the disease is searched for the needed treatment, and in the contraindications section the application looks for the patient's allergies, illnesses and treatments. The dosing section provides information about correct dosage based on age. This structure is helping the

physician find the best drug and to prescribe the correct dosage for the patient. This is particularly valuable in a pediatric practice setting.

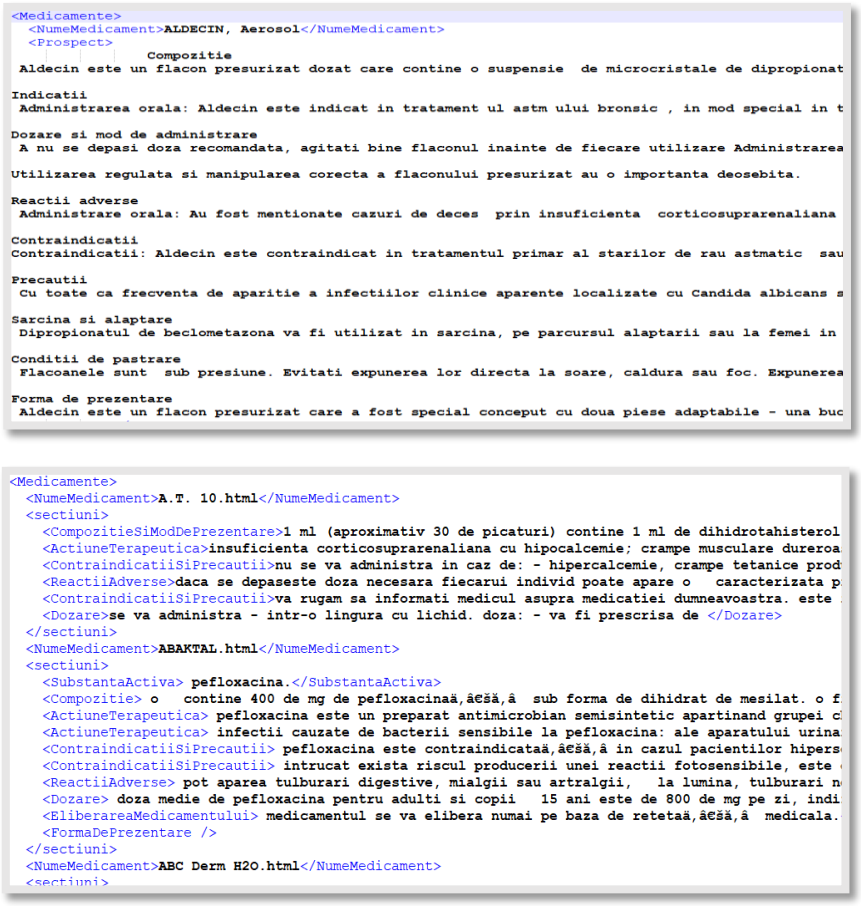


Figure 3. Unstructured (a) and structured (b) prospectuses

Figure 4 presents the resulting xml structured file.

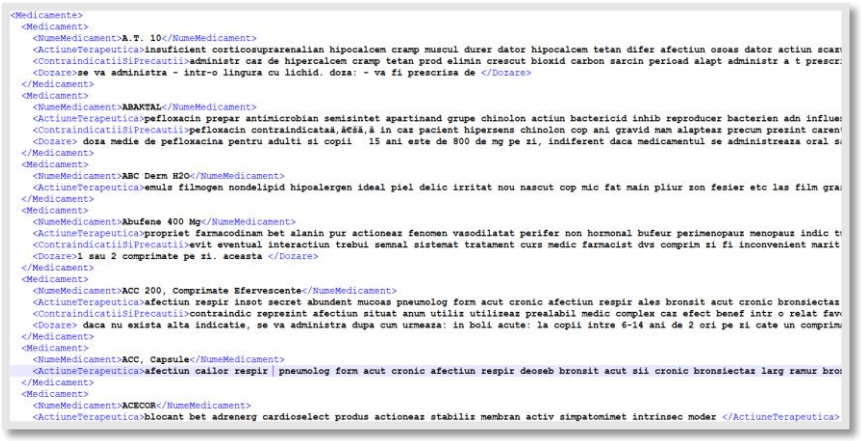


Figure 4. Structured prospectuses using the stemming algorithm

Figure 5 shows the flow of actions followed to structure the medical prospectuses.

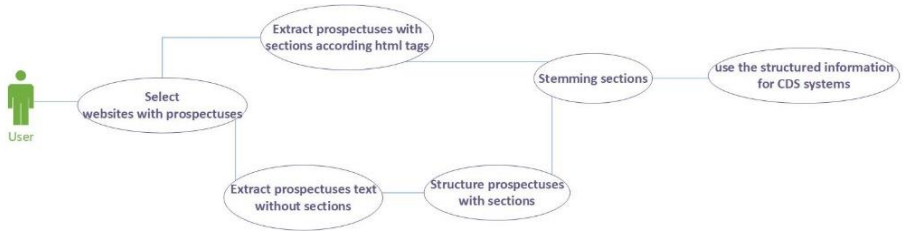


Figure 5. The flow of the actions made to structure the prospectuses

To test the effectiveness of the algorithms and structured prospectuses we created a simple application that introduces a diagnosis and contraindication for the patient condition and the application provides, based on the data, appropriate and inappropriate drugs. The application was tested using around 3.000 prospects with different formatting and section names. Thus, if the physician introduces a diagnosis, the application will look for the therapeutic action of drugs inclusive contraindication, and possible interactions with the drug. Once a drug is selected from the list of appropriate drugs, the dosage will be displayed in another box. The application is in development and will be improved for a more complex structure of the information provided by the physician (patient allergies, treatments, other diseases, age-related dosing, etc.). Figure 6 presents an interface of the structuring application. Figure 7 presents the information flow in the application.

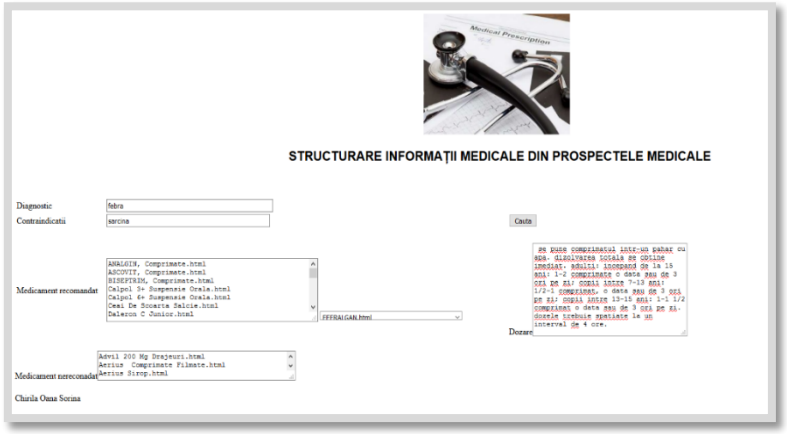


Figure 6. Test structured prospectuses – work in progress

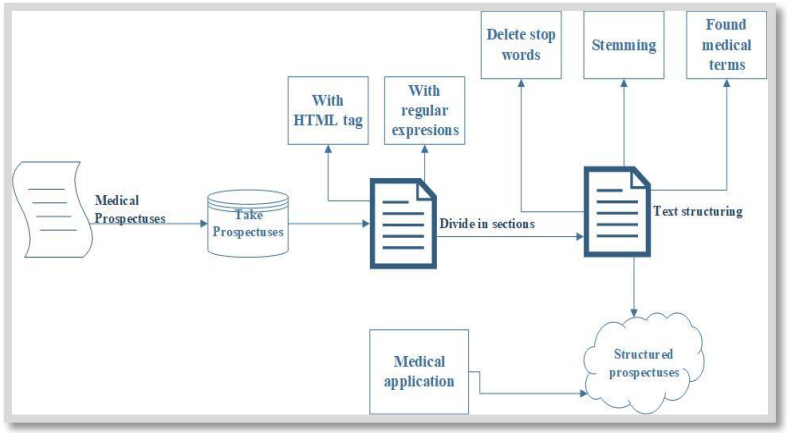


Figure 7. Information flow for structuring the medical prospect

The functions of the application are:

- Creating a management system for drugs prospectuses for easy treatment management
- Creating functions to compare, structure and retrieve useful information to establish the right treatment

- Creating functions to check for drug incompatibility with other drugs / diseases / other characteristics
- Storing information in the cloud, in xml or owl format, for data access and interoperability

Results and Discussion

The result of our work is an application supporting physicians in prescribing medication by structuring and extracting the necessary information from drug prospectuses. To this point, literature reports applications for drug-drug interaction alerts, as well as applications for drug management, with the end goal of controlling the patients' treatment. There are authors that use extraction of medication information from clinical texts. Before this work we developed a system that collects successful treatments for future use in other treatments for patients according to similar characteristics. In this article we present a method to extract drugs information from Romanian medical prospectuses to create databases for clinical decision support application in medical prescriptions. We did not use information from English databases and did not use English language tools because our prospectuses are in Romanian and medical terms are different from English medical terms. The application can be generalized for other languages prospectuses according to their sections. Using the prospectuses, we have all the needed information about drugs: indications, contraindications, dosage, adverse reactions, drug-drug interactions, etc. To find the best way to structure the prospectuses we developed two methods of extracting text. One method was to browse the html text and extract the sections with the path to the section node, and the second method was based on browsing the text and dividing it into sections using regular expressions and section names from the file previously created. We developed the methods for extracting data from online resources. After completing these steps, we concluded that the most effective method is to browse the text and split into sections of the text file because the html pages, even from the same source, can have different paths at the same node type, so we cannot access the desired information.

We choose for first step three sources with medical prospectus: Pagina Farmacistilor [5] that has 1983 prospectuses, HelpNet [6] that has 6084 prospectuses and CSID [7] that has 3879 prospectuses. For HelpNet we investigated both methods (extract the html sections and extract all prospectuses for future structuring). For the first method we extracted 2672 prospectuses and for the second method we extracted 3002 prospectuses. For the two other websites we could not use the first method because the sections of the drugs are different. With the second method we extracted from Pagina Farmaciştilor 1630 prospectuses and from CSID 3814 prospectuses. For the prospectuses that we could not extract we conclude that presented links with some errors, like null text or other problems.

To evaluate our method of extraction we use F-measure method [13] and we have the following results: for Pagina Farmaciştior the accuracy is 0.90, for HelpNet – the first method (with html tags) the accuracy is 0.61 and for the second method is 0.66, for CSID the accuracy is 0.88.

For a simpler processing, we chose the homogenization of the section names, so that after applying the algorithms a homogeneous structure of the prospectuses from all the collected sources has been reached.

Structuring medical prospectuses and using a specific application to support it results in helping doctors to choose the best treatments for their patients. Doctors are alerted to possible contraindications of medicines to certain diseases or other treatments. This kind of application reduces medical errors in prescribing medication and doctors have access to the latest products from which they can choose the right treatment.

The application is developed for the Romanian prospectuses, but in a future work we want to extend it to a multilingual one. The structured prospectuses can be used in medical applications to create clinical decision support for medical prescriptions and lead to reduced medical errors. To use the created database in a medical application, one can search the disease related to a drug with the ICD10 codes. Some problems can appear when some medical term is altered with the stemming algorithm. To solve this problem, we also work to build databases with the root of medical terms to use them in the stemming algorithm.

To improve the application, we will consider in the future to pass the structured information of prospectuses into ontologies. Using ontologies, relationships can be created between certain elements. Because of these relationships, an application will more easily detect the indicated or contra-indicated drugs of a patient. The goal is to create an ontology for medical prospectuses to be used globally. We will test the application on the prospectus in English after creating a file with English sections. For Romanian prospectus we intend to extract prospectus from diverse sources like The National Agency of Drugs and Medical Devices Nomenclature, where more than 30000 prospectuses can be found, but for this source we need to investigate more because all medical prospectuses are available for now in a PDF format. Another future development is structuring the dosing sections to display the correct dose according to age or disease.

List of abbreviations

SIUI – Unique Integrated Information System
e-Prescription – Electronic Prescription
XML – Extensible Markup Language
CNAS – National Health Insurance Agencies
DCI – International Joint Name
OWL – Web Ontology Language

References

1. Poşta Medicală. Available from URL: http://www.postamedicala.ro (accessed April 2018)
2. Rac-Albu EM, Ciobanu V, Rac-Albu M, Popescu N. Interoperability of Medical Data Through e-Health Service in Romania, 7th International Conference, IESS 2016 Bucharest, Romania, May 25–27, 2016, pp. 683-692
3. CNAS. Available from: http://www.des-cnas.ro/wps/portal/cnas/Dosarul-de-sanatate. (accessed April 2018)
4. Drăgănescu D, Lupuleasa D, Dumitrescu IB, Dinu Pârvu CE, Ciolan DF. Evidence On E-Prescribing Systems Worldwide. First Romanian Results. Farmacia 2013;61(2):353-360.
5. Prospectuses from HelpNet. Available from: URL: http://www.helpnet.ro/prospecte (accessed April 2018)
6. Prospectuses from CSID. Available from: URL: http://www.csid.ro/medicamente/ (accessed April 2018)
7. Prospectuses from Pagina Farmacistilor. Available from: URL: http://www.paginafarmacistilor.ro (accessed April 2018)
8. Luna D, Otero C, Risk M, Stanziola E, González Bernaldo De Quirós F. Impact of Participatory Design for Drug-Drug Interaction Alerts. A Comparison Study Between Two Interfaces. Stud Health Technol Inform. 2016;228:68-72.
9. Keil A, Gegier K, Pobiruchin M, Wiesner M. A Smartwatch-Driven Medication Management System Compliant to the German Medication Plan. Stud Health Technol Inform. 2016;228:185-189.
10. Lupse OS, Stoicu-Tivadar L, Golie C. Assisted prescription based on successful treatments. 2013 E-Health and Bioengineering Conference (EHB), Iasi, 2013, pp. 1-4. doi: 10.1109/EHB.2013.6707286
11. Fung KW, Kapusnik-Uner J, Cunningham J, Higby-Baker S, Bodenreider O. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support, J Am Med Inform Assoc. 2017;24(4):806-812.
12. Deléger L, Grouin C, Zweigenbaum P. Extracting medication information from French clinical texts. Stud Health Technol Inform. 2010;160(Pt 2):949-53.
13. Sasaki Y. The truth of the F-measure. University of Manchester, 26th October, 2007. Available from: http://www.cs.odu.edu/~mukka/cs795sum10dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf (accessed April 2018)