MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1005

Modeling, building and evaluating an ontology for the automatic characterization of adverse drug effects during pharmacovigilance

Catherine Duclos[a], Lina F.Soualmia[a], Sonia Krivine[a], Anne Jamet[b], Agnes Lillo-Louët[b]

[a] *Laboratoire d'Informatique Médicale et Bioinformatique (LIM&Bio EA 3969), Université Paris 13, Bobigny, France*
[b] *Centre de Pharmacovigilance(CRPV), Hôpital Européen Georges Pompidou, Paris, France*

Abstract

Background: The characterization of spontaneous reported cases is fundamental for pharmacovigilance. This task is time consuming and its reproducibility is low. Objective: To develop a system founded on an ontology that automatically instantiates spontaneous reported cases as "known" adverse drug effects (ADE) only if the reported ADEs are described in drug compendia. Methods: A simple ontology of drugs and their related adverse effects represented in description logics was developed from a drug database. Manual evaluation was carried out on 378 spontaneous reported cases instantiated as "known ADE of a chemical class". The initial manual characterization was reviewed by a pharmacovigilance expert to validate the generated automatic characterization. Results: The ontology is composed by 57,704 concepts and 5 relations. It was successfully validated thanks to Pellet reasoner and it contains neither inconsistencies nor cycles. In this validation, 86% of the instantiated spontaneous reported cases effectively concerned notorious ADEs, whereas only 75% were initially identified manually as related to notorious ADEs. Conclusion: This system can assist characterization by applying a reasoning process similar to that used by experts in the search for ADEs.

Keywords:

Adverse drug reaction, Reporting systems, Iatrogenic disease/classification, Knowledge bases, Ontology, Description logics, Reasoning.

Introduction

Pharmacovigilance process has been developed at the local, national and international level to facilitate the collection and analysis of spontaneous reported adverse drug effects (ADE) in treated patients. This process aims to identify rare and dangerous adverse events likely to result in an unfavorable benefit risk ratio.

The nature of the relationship between an adverse event and a drug in a spontaneous reported case is defined by a pharmacovigilance expert on the basis of intrinsic and extrinsic imputability criteria. The intrinsic imputability is founded on chronological (the time interval between the drug intake and the occurrence of the adverse event) and semiological (the symptoms observed are compatible with this type of drug) criteria. Extrinsic imputability (characterization) is based on bibliographic references concerning existing knowledge about the adverse effects associated with the drug (well known effect described in drug compendia, predictable or previously reported side effects and side effects not described or not published).

When considering what is known about the possible adverse effects of the drug, the expert must consult several sources: (a) the drug monograph which reports all the side effects observed during development and post-marketing of drug, (b) a drug compendium reporting the side effects of active substances [1-2], (c) a specific side-effect compendium reporting side effects of active substances and classes of active substances [3], (d) published articles dealing with the reported side effects of the substance.

The automation of these tasks would help the expert considerably in his or her interactions with the available resources. It should also save time and result in more reproducible bibliographic documentation. One possibility for automation would be to develop a dedicated tool which automatically classifies a spontaneous reported case as a "case with known ADE" or as a "case with unknown ADE". To automate this task it is necessary to check the existence of known side effect at different levels : for the substance, for the pharmacological class of the substance and for the chemical class of the substance. For example, a patient takes ibuprofen and a gastric ulceration is reported. This spontaneous report can be characterized as a known ADE because non steroidal anti-inflammatory drugs (NSAID) are known to give gastric ulceration and ibuprofen is a NSAID. As illustrated, this characterization process should use subsumption reasoning.

Ontological reasoning based on formal representation languages such as description logics can be used for this type of classification. The use of description logics has several advantages including the possibility of using advanced inference services (satisfiability, subsumption, classification, consistency checking, instantiation and realization) [4].

An ontology is therefore required. This ontology must describe the drugs prescribed and their known related ADEs, but also the classes to which the drugs belong (chemical and, pharmacological properties, etc...) and the related ADEs of these classes.

The existing ontologies (Galen Drug ontology [5], VA NDF-RT [6], SNOMED CT [7]) do not completely satisfy these requirements: (a) ADEs of drug classes, or of drug are not always given, (b) these ontologies deal essentially with American or English pharmacopoeias but not French ones, (c) they are not necessary kept up to date, may contain information for obsolete drugs and lack information for new drugs.

Information about drugs can also be obtained from commercial drug databases which are regularly updated but have a formalism unsuitable for complex reasoning such as subsomption, classification or consistency checking [4].

In this work, we aimed to develop and use an ontology to describe drugs and their known ADEs, through the following steps: (a) modeling known ADEs, (b) representing the ontology with a formal language, (c) populating it with data from a French drug database, (d) instantiating it with reported cases from a pharmacovigilance center (CRPV), (e) evaluating the results obtained through a classification process.

Materials and Methods

Known ADEs model

The general ways a side effect can occur are explained using pharmaceutical and toxicological knowledge. The main needed concepts and relations are identified.

Modeling the known ADEs is guided by the objectives of the reasoning tasks:

- Classification task should provide (1) a hierarchy of drugs on the basis of their pharmacological and chemical properties (2) a hierarchy of "cases with known ADEs" on the basis of the side effect concerned and the drug involved.
- Instantiation task should provide the list of "cases with known ADEs" for which a "spontaneous reported case" is an instance.

Ontology building

We represent the ontology of drugs and their related ADEs using OWL-DL, a web Description Logics (DL) language [4]. DL structures the domain knowledge on two levels: a terminological level (TBox or ontology) containing the classes of domain objects (concepts) with their properties (roles), and an assertional level (ABox), containing individuals (instances).

The ontology was constructed in a sequential manner. The TOP ontology was first created and stabilized. We then added in the TBox subconcepts of TOP ontology using knowledge from a drug database. Finally, the instances were added in the ABox using a pharmacovigilance database.

TOP ontology

The TOP ontology was stabilized by (a) implementing the general concepts and relations previously identified in Protégé 4.0 ontology development environment[1] and (b) verifying the principles on which our model was based using fictive cases with known ADEs, reported cases and the Pellet reasoning engine [8].

Populating the ontology

We used the French drug database Thesorimed[2] to populate the drug iatrogeny ontology because :

- This drug database is highly structured,
- It includes side effect descriptions at various levels (chemical, pharmacological, active substance),
- It indexes side effects with normalized terms and will soon include a MedDRA (Medical Dictionnary of Regulatory Affairs) indexation of side effects. MedDRA is a terminology commonly used to describe side effects in spontaneous reported cases [9].

The pertinent tables and fields of this drug database were identified. They constitute a subset of the database dealing particularly with the side effects of drugs. Thesorimed side-effect terms were converted into MedDRA terms by computerized string matching and manual matching. A specific script was written to convert the drug database subset into subconcepts of the TOP ontology. The OWL-DL file obtained was concatenated to the TOP ontology OWL-DL file.

Instantiating the ontology

Instances of the ontology are the spontaneous reported cases included in the CRPV pharmacovigilance database. Side effects are expressed using MedDRA terminology, and drugs are expressed as active substances. A script generating the OWL file of these instances was concatenated to existing OWL files. The methodology is presented in Figure 1.

Classification, Instantiation

The ontology OWL file was imported in Protégé 4.0 ontology development environment, and the reasoning engine Pellet was then used for both classification and instantiation.

Classification computes the sub-concept relations between every named concept to create the complete hierarchy. It can be used to answer queries such as getting all or only the direct sub-concepts of a concept.

Instantiation finds the most specific concepts that an instance belongs to (direct types for each instance). It can only be performed after classification since direct types are defined with respect to a concept hierarchy. Using the classification hierarchy, it is then possible to get all the types for that instance.

[1] http://protege.standford.edu
[2] http://www.giesips.org

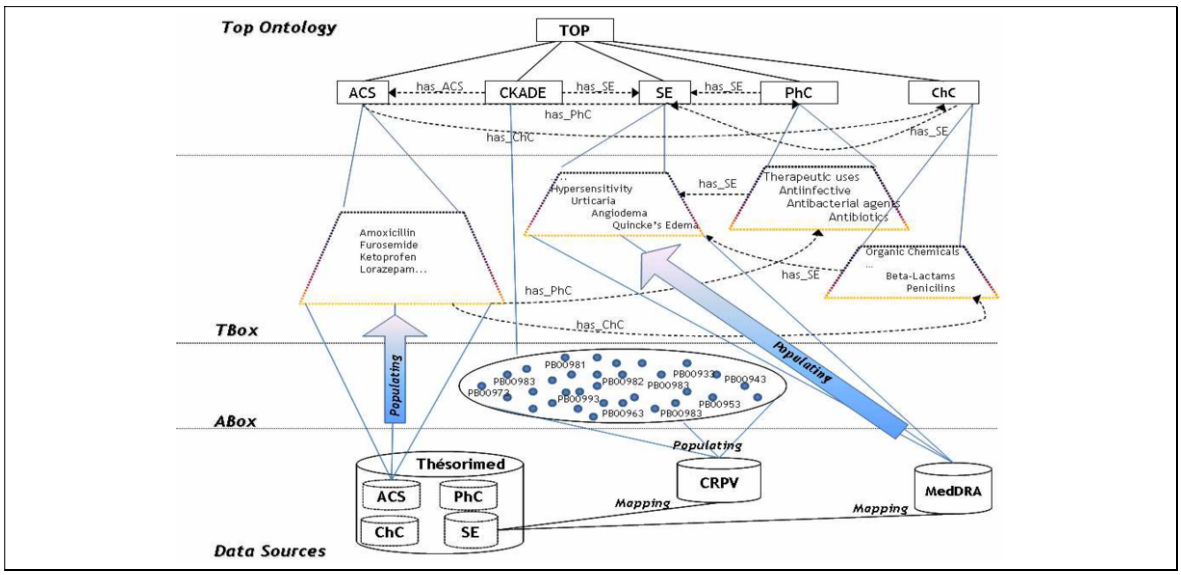

Figure 1 - Methodology of construction of ontology of drugs and their related ADEs. TOP ontology is composed by the concepts: ACS (Active Substance), CKADE (Case with known ADE), SE (Side Effect), PhC(Pharmacological Class) and ChC(Chemical Class).

Ontology evaluation

The logical consistency of the whole ontology (TOP, subconcepts, hierarchies, relations and instances) was analyzed by the reasoning engine. The conceptual validity of the ontology was evaluated by comparing the extrinsic imputability criteria of instantiated reported cases. If cases are instantiated, there is already bibliographic knowledge relating to them. Their extrinsic imputability criteria should therefore be 3 (well known effect described in drug compendium) or 2 (published side effect).

The list of instantiated cases was analysed by a pharmacovigilance expert that was asked to review the characterization of reported cases according to the information given by the "case with known ADE" concepts. The expert was asked whether the case should be considered "notorious" and had to select one of the following options for each reported case: she agrees and the initial criterion should be 3, she does not agree and the initial criterion is true, she partly agrees because she has other information explaining the initial criterion.

This evaluation was first performed on a preliminary set of "case with known ADEs" concepts dealing with chemical classes of drugs associated with side effects (*e.g. oxicam causing diarrhea*).

Results

Known ADEs model

Domain description

An adverse drug effect may be linked to:

- An active substance in the drug (*e.g. amoxicillin may cause tooth discoloration*),
- An auxiliary substance present in the drug with no therapeutic properties (*e.g. the aspartame present in oral formulation of amoxicillin may be dangerous for patients with phenylketonurias*),
- A chemical class of a drugs (*e.g. amidine penicillin may cause nausea*),
- A pharmacological class of a drugs (*e.g. antitussives may cause insomnia*).

An adverse effect may be due to:

- A generic interaction with an organ (non specific action),
- A molecular interaction (enzymatic induction, molecular competition, etc),
- An intrinsic toxicity of the drug (*e.g. ototoxicity of aminosides*).

An adverse effect may occur in various contexts:

- Patient context : physiological, pathological, genetics, allergy,
- Dose context : daily dose, frequency, duration of treatment,
- Administration context: form, route, flow,
- Exposure context: cumulative toxicity,
- Co-administration context (drug interactions).

Model of known ADEs

Modeling focused on the concept of "*case with known ADE*" relating to one or two drugs, one side effect and various con-

texts (patient, exposure, dose or administration contexts). The *case with known ADE* is a kind of *case*.

The drug involved may be a clinical drug (CD) or one of its components (active substance (ACS), auxiliary substance (AS)) or a substance belonging to a particular pharmacological class (PC), chemical class (CC) or interaction class (CI).

Figure 2 shows the model to describe known ADEs.

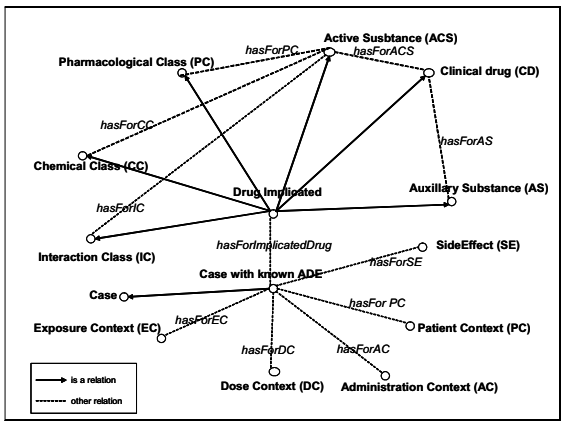

Figure 2- Model of known ADEs

Ontology Description

For the first version of the ontology, we limited the concept "*drug implicated*" to the: "*active substance*", "*pharmacological class*" or "*chemical class*" of the drug concerned. Context concepts had not been included in the ontology at this point.

TBox

Concepts relating to the drug implicated

Active substances are primitive concepts (n=4,481) described by two relations:

- *hasForPC* relating the active substance to its pharmacological class
- *hasForCC* relating the active substance to its chemical class

We created two defined concepts (*owl:Equivalent Class) - ActiveSubstanceHavingPC* (n=844) and *ActiveSubstanceHavingCC* (n=386) - making it possible to link *pharmacological* or *chemical class* to the "*drug implicated*".

Described case concepts

A "*case with known ADE*" is a defined concept based on 2 relations (*owl: ObjectProperty*):

- *hasForImplicatedDrug* relating the case to the drug implicated,
- *hasForSE* relating the case to a side effect

The *cases with known ADE* can be divided in 11,697 concepts of *PC causing side effect*, 5,198 concepts of *CC causing side effect*, and 32,685 concepts of *ACS causing side effect*.

Hierarchically organized concepts

Pharmacological class, chemical class and *side effects* are primitive concepts and are organized hierarchically (*owl:SubClassOf*). There are 844 concepts in the pharmacological hierarchy, 386 in the chemical one and 1,183 in the side effect one.

ABox

We consider all the "*spontaneous reported cases*" to be instances of the concept "*Case*".

Each "*spontaneous reported case*" is described with one or many instances of *drug implicated* and one or many instances of *side effects*. There are 2,555 instances in ABox.

Ontology evaluation

Neither inconsistent classes nor inconsistent instances have been inferred using our modeling principles.

For the preliminary evaluation, we considered only a subset of reported case instances of *"cases with known ADE"* defined as having a side effect related to a particular chemical class (378 of the 1,694 spontaneous reported cases in the CRPV database, corresponding to 565 side effect/substance pairs).

The initial extrinsic criterion was greater than 2 for 75% of the instantiated reported cases. For the others, the initial criterion was less than 2, indicating that the side effect concerned had not previously been described for this drug.

After review by the pharmacovigilance expert,

- 86% of the instantiated reported cases had a criterion greater then two (maintenance of the high initial criterion or increase in the criterion from an initial low value).
- 3% of the instantiated reported cases still had an extrinsic criterion below 2 after review (*e.g.* the reported case "clonazepam causing dyspepsia" is an instantiation of the described case "benzodiazepine causing dyspepsia", but the pharmacovigilance expert did not consider it to be notorious that benzodiazepine gives dyspepsia).
- 11% of the instantiated reported cases could have been given a criterion greater than 2, but the expert maintained the initial criterion because (a) other drugs were involved in the reported case for which the observed side effect had never been described (n= 44), (b) there were other side effects in the reported case that had never previously been described for the drug implicated (n=15), (c) the side effect was merely a sign that the drug was not effective (n=3) (*e.g. pain and ketoprofen*).

One of the benefits of this experiment was that the pharmacovigilance expert obtained new knowledge about drug properties. Indeed, some drugs were found to be classified exclusively on the basis of pharmacological principles and never on the basis of chemical principles (for example amprenavir is a sulfamide but is systematically described as an antiretroviral drug).

Discussion

The ontology we have developed is based on the pharmacovigilance observation and reasoning for the characterization of spontaneous reported cases. Based on these observations, we have repurposed the knowledge contained in a drug database to make it possible automation of the ADEs searching.

This method of ontology building also facilitates the maintenance of the ontology as the drug database is regularly updated and scripts automatically convert some parts of it into OWL-DL.

The automatic instantiation process efficiently identified spontaneous reported case of notorious bibliographically documented effects. Nevertheless, further evaluations with the whole ontology are required to quantify the specificity and accuracy of the instantiation process (true described reported cases, true not described cases, false described reported cases, false not described cases).

The evaluation of the ontology by the pharmacovigilance expert indicated several ways in which the ontology could be improved. New sub-concepts of cases with known ADEs could be created totally or partially matching to the reported cases in term of the number of drug/side effect pairs for the reported case automatically instantiated as cases with known ADEs.

The hierarchies included in the ontology are those provided by the drug database. They suffer from a lack of structure (no polyhierarchy and a very flat hierarchy). We are investigating ways to dissect concepts in these hierarchies to facilitate their reclassification.

In the near future, we will integrate the side effect ontology developed by another team of the VigiTermes project [10] into our system. This should greatly extend reasoning concerning the classification of side effects.

Other drug classification systems are currently integrated into the ontology (*e.g*: MeSH), which can thus be used to generate clusters of reported cases on the basis of common drug clusters. It will be used for the signal detection complementary to the approach developed by Henegar *et al* [11].

As our resource gives the knowledge about a drug and its known side effects, it could also be used to filter results of ADEs detection issued from Electronic Heath Records using NLP tools and data mining methods [12]. Among detected ADEs, known ADEs (True Positives) should be then automatically identified.

Finally, all the documentation services should be implemented in a web service.

Acknowledgments

This research was supported by the VigiTermes project funded by the French National Research Agency (ANR-07-TECSAN-026-04).

References

[1] Martindale: The Complete Drug Reference (36th Edition) 2009. London: Pharmaceutical Press

[2] The Merck Index – An Encyclopedia of Chemicals, Drugs, and Biologicals (14th Edition). 2006. Merck Publication

[3] Meyler's Side Effect of Drugs: The International Encyclopedia of Adverse Drug Reactions and Interactions (15th Edition). 2006. Elsevier Science Ltd

[4] Baader F, Calvanese D, McGuinness DL, Nardi D and Pater-Schneider PL. The Description Logic Handbook: Theory, Implementation and Applications, 2nd Ed. 2007. Cambridge University Press

[5] Solomon WD, Wroe CJ, Rector AL, Rogers JE A reference terminology for drugs. Proc AMIA Symp 1999, 152-55

[6] Carter JS, Brown SH, Bauer BA et al. Categorical information in pharmaceutical terminologies. Proc AMIA Symp 2006, 116-20

[7] Kim JM, Frosdick P. Description of a drug hierarchy in a concept-based reference terminology. Proc AMIA Symp 2001: 314-18

[8] Sirin E, Parsia B, Cuenca-Grau B, Kalyanpur A, and Katz Y. Pellet a Practical Owl DL Reasoner. Web Semantics 2007;5(2):51-53

[9] Brown EG, Wood L, Wood S. The Medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999; 20(2): 109-17

[10] Bousquet C, Trombert B, Kumar A, Rodrigues JM. Semantic categories and relation for modeling adverse drug reactions towards a categorical structure of pharmacovigilance. AMIA Annu Sympo Proc. 2008; 61-5

[11] Henegar C, Bousquet C, Lillo-Le Louët A, Degoulet P, Jaulent MC. Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. Comput Biol Med. 2006; 36(7-8): 748-67

[12] Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and Electronic Health Records: a feasibility study. JAMIA. 2009; 16(3): 328 - 337

Address for correspondence

Catherine Duclos (catherine.duclos@avc.aphp.fr),
LIM&Bio UFR SMBH, 74 rue Marcel Cachin,
93017 Bobigny cedex, France