Modeling and Acquisition of Drug-Drug Interaction Knowledge

Frédéric Mille, Patrice Degoulet, Marie-Christine Jaulet

Abstract

Objectives: The effectiveness of computerized clinical decision support systems (CDSS) depends on the quality of the knowledge they refer to. In this article, we are interested in the acquisition, modeling and representation of the knowledge embedded in the “national reference framework of drug-drug interaction” published by the French Health Products Safety Agency.

Methods: A model of drug-drug interactions has been designed using bottom-up and top-down approaches. This model is the basis for the design of an XML format to represent and extract information on drug interactions from the reference framework.

Results: A specific tool has been developed to extract the information from a corpus of 1053 drug monographs using a methodology similar to the one used by the GEM-Cutter tool to extract information from clinical guidelines. Strategies to integrate the XML files produced into CDSSs are discussed.

Discussion-conclusion: Modeling and acquisition of drug-drug interaction knowledge from a corpus of drug monographs is a potential approach to foster the development of CDSS and improve their specificity.

Keywords: drug interactions, knowledge modeling, XML

Introduction

Technology–based interventions have been recommended for reducing the likelihood of medication errors. Computerized physician order entry (CPOE) has been cited as one of the most effective ways to avoid medication errors[1-3] caused by misinterpretation of handwritten orders, incorrect doses, wrong dose forms and inappropriate administration times. The next step is to integrate effective Clinical Decision Support Systems (CDSS) into CPOE systems.

CDSS are information systems designed to improve clinical decision making, for example detecting drug-drug interactions (DDIs) during the order entry process. Indeed, DDI detection systems were among the first CDSS developed. When a potential problem is identified, these systems can provide real-time alerts, allowing the clinician to make appropriate changes before the prescription is finalized.

The effectiveness of CDSS depends on the quality of the knowledge base which they use. Obviously, for detection of DDIs, any CDSS requires knowledge of these DDIs. The main issue addressed here is the acquisition, modeling and representation of the knowledge in this field. In France, the main source of knowledge for DDIs is the “national reference framework of the drug-drug interaction” published by French Health Products Safety Agency “Agence Française de Sécurité Sanitaire des Produits de Santé” (AFSSAPS) [4]. DDIs are described in free text in these monographs and are not directly accessible by a computerized system.

A simplest way to represent knowledge in monographs would be to build manually two-column tables describing the interactions between pairs of drugs. This approach is difficult to maintain as knowledge progresses. The most advanced approach would be automatic free text understanding by Natural Language Processing (NLP) techniques. These techniques could automatically identify concepts in DDI monographs and semantic relations between these concepts. However, the NLP approach is complex and beyond the scope of this work. Between these two extremes, an affordable and plausible approach is to propose tools to structure the knowledge contained in the monographs so as to make this knowledge available.

The aim of the study is to acquire DDI knowledge in a structured knowledge base that could be used by a computerized system. The paper presents a method and a tool for building a knowledge resource usable in the context of prescription systems.

Our work is based on the hypothesis that DDI knowledge formulation has similarities with clinical guidelines knowledge formulation. Indeed, DDI monographs contain recommendations and risk factors that can be considered in the same way as guidelines. Clinical guidelines have become an important medium for the standardization and dissemination of medical knowledge. The “document-centric approach” has been introduced to facilitate the use of guideline knowledge. In the document-centric approach, the original text of the guidelines is systematically marked-up with respect to the model and kept in the form of a structured document.
The best known instance of "document-centric approach" is the Guideline Elements Model (GEM) which is an XML (Extensible Markup Language) framework based on a hierarchy of concepts describing guideline contents [5].

The encoding of a clinical guidelines using the GEM framework consists of structuring the textual document using the set of XML markups provided by the framework.

The encoding of a monograph consists of structuring the textual document using the set of XML markups. This can be a complex process, as it requires in-depth analysis of the text content. Similarly, for clinical guidelines, substantial variation is observed in the encoding by different users [6]. This would suggest, as concerns our study, that the complexity of manual analysis can affect the structure of the encoded monographs.

To tackle these problems and support the process of manual marking up of monographs, we developed a specific interface, inspired by the GEM-Cutter application [5].

We propose to use the same methodology to structure the information within monographs. The first step was to identify the relevant information in the monographs so as to build an XML schema related to DDI information. This schema is implemented in a specific environment to acquire descriptions of DDIs directly from text. The second part of our work was to use and evaluate this tool to develop a knowledge base concerning DDI information from 1000 monographs. The practical result is the knowledge base itself. We present the advantages and limitations of our approach and consider the exploitation of the resulting knowledge base.

Materials and methods

Methods

We used a two-step approach to identify the information elements contained in the "national reference framework of the drug-drug interaction" published by AFSSAPS. First, we identified the main concepts of DDI. Second, we used natural language processing tools to analyze the content of DDI texts.

Then we built a XML schema of DDI information, which was evaluated for its ability to represent the initial text information and maintain its meaning.

Finally, we built specific software to write XML files.

Top-down approach knowledge identification

The identification of large classes relied on the manual study of structure in the monographs. It was done based on the reference frame of AFSSAPS.

Bottom-up approach knowledge identification

We applied a data-driven knowledge approach to identify the concepts of the domain present in the monographs. We used a syntax analyzer module based on the hypothesis of similar dependencies between terms which have similar meanings. Then, a pharmacist reviewed the different terms on the basis of shared syntactical contexts and exploited all the network data to cluster and organize the terms.

The lexicon built with this software made possible a second manual step of semantic analysis registering the principal semantic structures contained in the DDI texts, and their frequencies.

Establishing a XML schema

The creation of an XML language to describe monographs requires conceptual modeling of DDIs. We built an XML schema of DDI knowledge from existing knowledge and terminological analysis. This hierarchy of concepts constitutes the skeleton of the XML schema.

Development of an XML Editor

We built X-DIE (XML Drug Interaction Editor), an XML editor with added functionalities to facilitate the editing of XML. X-DIE is a graphic editor which hides the code in the background and presents the content to the user in a more readable format, similar to the version which must be ultimately posted. The interface is based on a multiple windowing system that displays the original monograph, a XML browser and the corresponding XML-encoded text.

Material

The French drug database Theriaque® developed by the Centre National Hospitalier d'Information sur le Medicament (CNHIM) is responsible for the dissemination of independent information about all drugs available in France [7]. The database contains complete information about the drugs: the pharmaco-therapeutic group, the active component, the excipient, the commercial presentation, indications, contra-indications and AFSSAPS reference concerning DDIs. The set of all 1053 monographs of DDIs available in the Theriaque® database constitutes the initial corpus.

The syntax analyser module is SYNTEx, a natural language processing tool that is widely used to build ontologies from texts. SYNTEx performs a syntactic analysis of the sentences of the corpus, and yields a network of the dependence of words and phrases [8, 9].

The PROTEGE editor was used to build and browse the hierarchy of concepts.

The XML schema was built using an open-source XML file editor, JAXE.

The interface for the marking up of monographs was developed in JAVA with the API DOM (Application Programming interface Document Object Model) dedicated to the production and management of XML files.

Results

Knowledge acquisition results

Top-down approach

In the frame of reference of AFSSAPS, a DDI is defined by a pair of protagonists “A + B” who can be an active substance, indicated by their international non-proprietary names, or a therapeutic class, itself being the subject of
interactions “of class”. The wording of a DDI is structured in 4 paragraphs: 1) nature of the risk, 2) brief mechanism of action (if known), 3) action to be taken and 4) degree of constraint.

For the last type of paragraphs four degrees of constraint are possible: 1) Contra-indication: This is an absolute character and should not be transgressed, 2) Not advised: The association should generally be avoided, except if rigorous examination of the risk/benefit ratio suggests otherwise, and imposes close monitoring of the patient, 3) Cautioned: The association is possible if, in particular at the beginning of treatment, simple recommendations are respected making it possible to avoid undesirable DDI (adaptation of doses, reinforcement of monitoring, be it clinical, biological, ECG, or other), 4) Take into account: The action to be taken is generally summarized in case of a DDI being classified as a contra-indication or not advised, the action to be taken is generally summarized only as a constraint. Cautioned DDI are often associated with recommendations that are simple to implement to avoid the interaction (adaptation of doses, biological controls, etc.). The classification “take into account” is not associated with any practical recommendation because it announces the addition of adverse effects that can only be avoided by using other therapeutic strategies.

This first stage of the analysis generated various categories of knowledge found in the DDI monographs. These categories are considered to be the main concepts to be identified in DDI texts and are used to group terms in the following analysis.

Several classes are distinguished but there is no organization of knowledge in each class.

The level of organization of the knowledge was insufficient for application in CDSS.

**Bottom-up approach**

The initial corpus contained about 87,707 words. Once processed with SynTex, the corpus gave 9393 candidate terms appearing at least twice. The expert selected 2657 candidate terms: 1) 1349 noun phrases out of 3150, 2) 1071 nouns out of 3775, 3) 237 adjectives out of 336. These terms have been classified into 5 categories: 1) pharmacology and pharmacokinetics (931 terms), 2) galenical, active principle, dose and regimen (936 terms, this category includes the various characteristics of the associations causing DDIs), 3) physiopathology (282 terms, this group includes all diseases or symptoms secondary to DDIs), 4) physiology (188 terms), 5) others (320 terms).

These selected candidate terms were normalized within each category and this gave a final total of 888 concepts.

The combination of bottom-up and top-down approaches enabled us to build a DDI model with 6 main classes (figure 1): risk association, consequence, mechanism, limitation, precaution for use, risk factors.

**Figure 1 - DDI model**

The XML representation schema

The JAXE software allows an XML schema to be built, compatible with the W3C norm.

XML files are constructed as a hierarchy of 72 discrete tags with 6 major branches intended to capture the information in the 6 main sections of a monograph. The 6 branches are:

- **Risk association:** the sub-tree describes the drug association responsible for the DDI, and the clinical situations (pathologies, ages) in which the DDI occurs.
- **Mechanism:** the sub-tree describes the DDI mechanisms (pharmacology, pharmacokinetics, etc.).
- **Risk factors:** the sub-tree describes any physiology (age, pregnancy) and/or diseases that may increase the risk of adverse effects of a given DDI.
- **Consequences:** the sub-tree describes the diseases and symptoms resulting from the DDI.
- **Precaution for use:** the sub-tree describes the precautions to take to reduce the effects of the DDI (substitution, alternative drugs) and/or to reduce its consequences (clinical or biological exams).
- **Limitations:** the sub-tree describes any disease or condition where the benefit for the patient outweighs the risk of an adverse event due to the DDI.

Other types of elements are represented in the XML schema. They include: drug, pathology, biological examinations.

XML elements can have attributes in the start tag. Attributes provide additional information about elements. The attributes of our elements are found in published classifications particularly ATC (Anatomical Therapeutic Chemical), ICD10 (International Classification of Disease – tenth edition) and MedDRA (Medical Dictionary for Regulatory Activities). With these attributes, DDI detection system (DIS) will be able to perform various types of inference between DDI information and the clinical context of the patient. In addition, the XML schema includes...
elements to allow quantitative information (for example duration, dose, etc.) to be coded.

**The X-DIE environment for the encoding of monographs**

The X-DIE environment has been developed to encode monographs. The main interface is a three-windowing system (figure 2), inspired by the GEM-Cutter interface: 1 window A displays the DDI monograph in full text, 2 window B displays the hierarchy of the elements for the XML file under construction and 3 window C displays the information to provide for each element of the file.

![Figure 2 - X-DIE](image)

Our environment allows the user to look for drugs (active principles, therapeutic class and chemical class) for which there is a link, in the data base Theriaque®, with the displayed monograph. It also allows the user to search classifications including ICD10 and MedDra for pathologies similar to those cited in the visualized monograph. Other information is obtained by the user.

**Knowledge base**

One thousand and six of the total of 1053 monographs present in the Theriaque® database could be successfully entered; the coverage was thus 96%. The monographs which were not successfully encoded had at least one of the following two characteristics: 1 The monograph describes an DDI involving a drug which is no longer marketed; in these cases we chose not to enter the corresponding XML file. 2 The monograph describes a DDI whose mechanism is a modification of plasma protein binding. This mechanism was omitted by the expert.

**Discussion and conclusion**

We developed a DDI model capable of representing full DDI information and we developed the X-DIE environment to capture this structured data from full text monographs. We used both general knowledge about DDI contained in the “national reference framework of the drug-drug interaction” published by AFSSAPS, and knowledge extraction with NLP techniques from the same database.

General knowledge helped us to define the generic structure of the DDI model. This top-down approach, although performed manually, rapidly gave an organization of the main concepts, because the domain is delimited. To refine this generic structure, we extracted the knowledge present in the “national reference framework of the drug-drug interaction”. For this bottom-up approach, we combined natural language processing results using manual semantic analysis of the relevant DDI candidate terms specifying generic concepts identified by the top-down analysis. The bottom-up and top-down approaches are standard methods in knowledge engineering.

Their implementation in this way ensures that the final structure covers a significant proportion of the information contained in the DDI monographs. However, the percentage of coverage of all the monographs has not yet been established. Moreover, the use of monographs from only one knowledge source potentially imposes limitations: the interoperability of the XML files is reduced in theory because the source for the knowledge and the data to encode is the same. An improvement would be to design the model from different sources of DDI data.

It became clear, during the encoding process, that the monographs were poor in recommendations. A few XML markups were rarely used, for example “clinical examination”. We also noted that some recommendations are vague; biological examinations are cited under the generic term “biological monitoring” and clinic examinations are quoted under the generic term “clinical monitoring”. This was a surprising finding because monographs may contain: 1 monitoring options (blood tests, clinical and extra clinical examinations), 2 guidelines to manage DDIs (modification of the times or sequence of administration, routes of administration, compensation by administration of a third compound or substitution of depleted endogenous substances, adjusting doses, discontinuation as symptoms appear, and suggestions of non interacting alternatives), 3 factors increasing the risks of DDI (age, sex, genetics, predisposing diseases, and use of alcohol and tobacco).

This work clearly illustrated the paucity of the content of some of the monographs as concerns recommendations. Thus, our model could be used as a tool to enrich and standardize the DDI monographs.

X-DIE tool facilitates the encoding of monographs. However, the data capturing is still largely manual. The encoding process could be considerably simplified if X–DIE could incorporate automatic text processing functions, such as the identification of linguistic markers. Previous work by our group on guidelines demonstrates the value of the identification of linguistic markers [10]. Such extension will exploit the already built ontology to highlight pertinent information in the text.

Some studies have indicated that the weak specificity and irrelevance of alerts in CDSS lead to loss of confidence in
these systems, a phenomenon known as the "Cry wolf syndrome" [11, 12]. Perceived poor specificity of drug alerts may be a major obstacle to efficient utilization of information and may prevent such alerts contributing fully to improving safety. If there is a false alert, these systems substantially increase the time required to carry out a task, due to the disruption in workflow. High signal-to-noise ratios may also produce alert fatigue and result in physicians skipping past alerts without considering or even reading them. One major limitation of existing DDI detection systems is associated with the use of static knowledge, most often embedded in fixed mapping tables: they proceed by comparing between drugs according to a hard-coded interaction mapping table, without any consideration of the clinical context of the patient or of all the information contained in monographs. These systems are often highly inclusive, placing more emphasis on breadth of coverage than on clinical relevancy or severity of adverse events. In published studies, practitioners explained that CDSS should integrate context relevance information, guidelines or evidence-base medicine. In fact, the majority of DDI can be compensated by dose adjustment or prevented by a well-considered sequence of administration and represents therefore a manageable risk [13]. Physicians wish informative support on DDI, concerning management. A distinction between clinically relevant and negligible DDI is essential [14].

For all these reasons we suggest that the reference framework should be integrated in the form of XML files in CDSS. The current XML-Schema has to be augmented in order to include all information for each DDI in the reference framework: some instructions for laboratory testing to monitor side-effects, replacement of ordered drug with another drug, change in dose, or additional drug. It also contains clinical situations where the use of an association is acceptable in spite of the potential consequences of the DDI; in these clinical situations, it is not necessary to announce the DDI or post an alert. Indeed, in these situations, the clinician can deviate from recommendations for good clinical reasons and the benefits of the drug association outweigh the disadvantages of the potential DDI.

To construct a more accurate and evidence-based CDSS for DDI detection, knowledge from DDI databases as provided by AFSSAPS must be arranged and integrated, but it is also essential to construct algorithms that can exploit this knowledge base appropriately. The next steps include the development of algorithms able to use XML files. Also, the impact of this combination (algorithms and knowledge base) on CDSS specificity will have to be evaluated.

References


Address for correspondence
Dr. Frédéric Mille. UMR_S 872, 15 rue de l’école de médecine, 75006 Paris, France. frederic.mille@rdb.aphp.fr