Evaluation of an Ontological Resource for Pharmacovigilance

Marie-Christine JAULENT, Iulian ALECU
INSERM, UMR_S 872, eq.20 Paris, France; Paris Descartes University, Paris, France

Abstract. In this work, we present a methodology for evaluating an ontology designed in a previous study to describe adverse drug reactions. We evaluate it in term of its fitness for grouping cases in pharmacovigilance. We define as gold standard the Standardized MedDRA Queries (SMQs) developed manually to group terms representing similar medical conditions. We perform an automatic search in the ontology in order to retrieve concepts related to the medical conditions. An optimal query is built for each medical condition. The evaluation relies on the comparison between the terms in the SMQ and the terms subsumed by the query. The result is quantified by sensitivity and specificity. We applied this methodology for 24 SMQs and we obtain a mean sensitivity of 0.82. This work allows validating the semantic resource and provides, in perspective, tools to maintain the ontology while the knowledge is evolving.

Keywords. ontology, evaluation methodology, pharmacovigilance

1. Introduction

The prevention of drug-related adverse events (ADRs) is a major issue in public health. The aim of pharmacovigilance is to detect, evaluate and prevent ADRs. WHO-ART (World Health Organization – Adverse Reaction Terminology) is the terminology developed by the World Health Organization for the coherent and rational coding of ADRs. MedDRA (the Medical Dictionary for Regulatory Activities) was developed more recently by the Maintenance and Support Service Organization (MSSO) and is now recommended for the description of ADRs by the pharmaceutical industry and by regulatory authorities. All WHO-ART terms are present in the MedDRA terminology. The structure of relationships used to organize terms is particularly important to retrieve similar medical conditions and has a direct impact on the specificity and sensitivity of pharmacovigilance signal detection [1, 2]. We have previously shown that neither WHO-ART nor MedDRA allows similar clinical conditions to be clustered together due to the lack of polyhierarchy. For instance in MedDRA, the term “gastric ulcer hemorrhage” is linked to the “gastric ulcer and perforations” term but not to the “gastric and esophageal hemorrhage” term. This deficiency limits the detection, evaluation and prevention of ADRs [3]. We have developed an ontology describing ADRs and a web tool, PharmARTS for navigation in the ontology [4]. Our current
objective is to evaluate the existing ontology for the specific purpose of grouping relevant cases together.

Three criteria are usually distinguished for the evaluation of ontologies: 1) the relevance of the vocabulary describing the concepts, 2) the relevance of the “is-a” hierarchy and 3) the relevance of the semantic relations [6]. Possible errors are: a) the errors of addition, b) the errors of omission and c) the errors of relation [6]. In a recent review, Jeremy Rogers pointed out that few studies had dealt with methods for evaluating the suitability of the ontology for the task for which it was created in the medical field [7]. More recently, Cornet et al. review approaches for auditing medical terminological systems and propose a method based on detecting concepts with equivalent definitions [8]. One significant problem for ontology quality assurance is the lack of a gold standard against which to determine the correctness of the ontology and its suitability for the given purpose [7]. We present here a method of evaluation for the ADR ontology. We define a gold standard specific to the use of this ontology and propose a formalization of this method enabling it to be reused throughout the development of the ontology.

We will first consider the material with which we worked and the methodology adopted. We will then present the results obtained during the first use of this method.

2. Material and Methods

2.1. ADR Ontology

The resource is a network of concepts. It contains concepts from WHO-ART and concepts from SNOMED CT relating to drug safety. The concepts from WHO-ART are obtained by aligning WHO-ART terms with SNOMED CT concepts, using the synonymy link in the metathesaurus of UMLS. The links are associative relationships extracted from SNOMED CT (e.g., “bladder neoplasm” is associated with the localization “bladder structure”) and the relationship “is a” is used to indicate taxonomic relationships (e.g., “renal failure” is a “renal disease”) [4]. The ontology contains 7,357 concepts accounting for 1.6% of SNOMED CT (366,170 concepts). The resource is augmented with the MedDRA terminology. In the resource, the origin of a concept is identified by a prefix: W for WHO-ART, S for SNOMED CT and M for MedDRA.

2.2. MedDRA SMQs

Queries in pharmacovigilance are currently defined by experts. These queries, known as “SMQs” (standardized MedDRA queries), are developed manually by the MSSO to facilitate the grouping together of pharmacovigilance cases, according to a controlled method. As the entire WHO-ART terminology is included in MedDRA, some of the MedDRA terms included in SMQs are also WHO-ART terms. For example, the SMQ “acute pancreatitis” contains 40 preferred MedDRA coding terms (and more than 100 terms in total if we consider synonyms), including 7 terms from WHO-ART. Twenty-four SMQs, reduced to WHO-ART terms constitute the “gold standard” for the first use

\[ http://www.ihtsdo.org/ \]
\[ http://www.nlm.nih.gov/research/umls/ \]
of our method. In the following, to simplify the notation, we call SMQ, MedDRA SMQ reduced to WHO-ART terms.

2.3. Evaluation of the ADR Ontology

Our aim is to validate the coverage and structure of the ontology with respect to its proposed use. The methodological objective is twofold: 1) to demonstrate that it is possible, for a given SMQ, to construct a query for which all the concepts subsumed by the query match as closely as possible the terms supplied by the SMQ; 2) to develop an evaluation method sufficiently generic for automatic (or almost so) reappraisal whenever the resource is updated or experts create new SMQs. The evaluation of the result of the query with respect to the medical condition is based on the measurement of sensitivity/specificity\(^6\). The evaluation can be modeled as follows:

- Group of individuals: Collection of WHO-ART terms (more than 3,500)
- Definition of the disease or condition: WHO-ART term present in SMQ
- Definition of the test: A query composed of one or several concepts of the ontology. The result of the test is the set of concepts subsumed by at least one concept of the query. We say that the test is:
  - Positive, if the term is subsumed by the query (Q+)
  - Negative, if the term is not subsumed by the query (Q-)

Thus, for each SMQ, we can consider sensitivity and specificity to measure the capacity of the query to predict the terms of the SMQ (Table 1). In other words, sensitivity, or the probability of the term being subsumed by the query when it is in SMQ, is given by: \(a/(a+c)\).

<table>
<thead>
<tr>
<th>Test/SMQ</th>
<th>in SMQ</th>
<th>outside SMQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q+</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Q-</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

As the number of WHO-ART terms outside SMQ is always much larger than the number of terms inside, \(d/(d+b)\) is close to 1. Indeed, the number of terms within SMQ does not exceed 50, whereas there are 3,500 terms in total in WHO-ART.

For the evaluation of the ontology, it would be valuable to determine the test (query) minimizing \(b\) and \(c\). The numbers \(b\) and \(c\) refer to the terms that must be explained as errors in SMQ (such errors may occur) or as errors in the ontology. This should result in a high level of specificity for the tests carried out. For this reason, we do not consider the Q- terms outside SMQ (the true negatives) below. Instead, in the following, we only give the results for sensitivity.

2.4. Construction of the Query

The construction of the query is a two-steps process. During the first step, the method selects candidate concepts in the ontology to represent the medical condition under study or part of the medical condition. The selection is first done directly by correspondence when the label designating the SMQ corresponds to a concept in the

ontology or to a MedDRA term. For example, for the SMQ “acute pancreatitis” we identify two candidate concepts M: Pancreatitis_acute and S: Acute_pancreatitis. The selection is then enlarged to the ascendants of all candidate concepts (direct parents in the “is-a” hierarchy). When the SMQ label does not correspond to a concept of the ontology (for example, the SMQ “Neuroleptic malignant syndrome”), the candidate concepts are searched within the ascendants of the terms present in the SMQ. For each query (each candidate concept), we calculate the values of a, b and c (Table 2).

<table>
<thead>
<tr>
<th>Candidate concept</th>
<th>Selection</th>
<th>T*</th>
<th>a</th>
<th>c</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>M: Pancreatitis_acute</td>
<td>Correspondance</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>M: Acute_and_chronic_pancreatitis</td>
<td>Enlargement</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S: Acute_pancreatitis</td>
<td>Correspondance</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>S: Acute_inflammatory_disease</td>
<td>Enlargement</td>
<td>25</td>
<td>3</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>S: Acute_digestive_system</td>
<td>Enlargement</td>
<td>22</td>
<td>3</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>S: Pancreatitis</td>
<td>Enlargement</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

* T = number of WHO-ART concepts subsumed by the candidate concept

During the second step, the method builds the optimal query by choosing the best union of the candidate concepts, the one that maximizes a and minimizes both b and c. In the previous example, the final query is « M: Acute_and_chronic_pancreatitis » since there is no union that provides better values than sensitivity = 1, b=1 and c=0.

3. Results and Discussion

The methodology has been tested on 24 SMQs. Table 3 summarizes the results. One interesting aspect of the method is to distinguish different error types using the information provided by the b and c values. We identified the following types of errors:

- **b≠0.** The term is missing in the SMQ. For example, in the SMQ « angioedema », the term W :OEDEMA_CIRCUMORAL does not exist while it should.

- **b≠0.** There is a taxonomic error in the ontology. For example, for the “acute pancreatitis” SMQ, the query returns a wrong concept, W :PANCREATITIS_CHRONIC, due probably to the fact that there is a missing level in the hierarchy that explicit differences between acute and chronic.

- **c≠0.** Formal concept definitions are not complete. During the construction of the ontology, not all WHO-ART terms were matched with SNOMED CT synonyms. Some formal definitions are missing that can be pointed out here.

- **c≠0.** Semantic relations are missing in the ontology. For example, the relation “evoke” between an exam result and a diagnosis does not exist in the ontology yet so that concepts related to exam results cannot be subsumed by a query representing a medical condition.

<table>
<thead>
<tr>
<th>Table 3. Results for the 24 SMQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of WHO-ART terms in the SMQs</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>23.5</td>
</tr>
</tbody>
</table>
4. Conclusion and Perspectives

In this article, we present a methodology for the evaluation of an ADR ontology built to improve signal detection in pharmacovigilance. The method does not support the development of the ontology but allows identifying how it has to evolve by finding the corrections that have to be made (missing definitions or missing relations, wrong definitions or wrong relations). This work shows the relevancy of the approach to evaluate the ontology and the feasibility for automation. The study, led on 24 SMQs, shows satisfactory results in term of sensibility and encourages us to use the methodology at different steps during the construction or the evolution of the ontology. The methodology is also useful to measure if the ontology remains adequate when there is a new context (construction of new SMQs by the MSSO). Other tools must be envisaged to assist the resolution of the errors revealed by the method. In the near future, we want to collect manually the classification and corrections of errors done during the evaluation in order to extract, if possible, some rules that could help the correction process.

References