Medical Informatics in a United and Healthy Europe
K.-P. Adlassnig et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-044-5-517

A Knowledge Management Platform for Documentation of Case Reports in Pharmacovigilance

Florence AMARDEILH [a,1], Cédric BOUSQUET [b], Sylvie GUILLEMIN-LANNE [c], Mathilde WISS-THEBAULT [c], Laetitia GUILLOT [d], Denis DELAMARRE [d], Agnès LILLO-LE LOUET [e], Anita BURGUN [d]

[a] *Mondeca, Paris, France*
[b] *Département de Santé Publique et Information Médicale, Université de Saint Etienne, France*
[c] *Temis, Paris, France*
[d] *INSERM U936, Université Rennes 1, France*
[e] *Centre Régional de Pharmacovigilance, Hôpital Européen Georges Pompidou – Assistance Publique-Hôpitaux de Paris, France*

Abstract. Most countries have developed information systems to report drug adverse effects. However, as in other domains where systematic reviews are needed, there is little guidance on how systematic documentation of drug adverse effects should be performed. The objective of the VigiTermes project is to develop a platform to improve documentation of pharmacovigilance case reports for the pharmaceutical industry and regulatory authorities. In order to improve systematic reviews of adverse drug reactions, we developed a prototype that first reproduces and standardizes search strategies, then extracts information from the Medline abstracts which were retrieved and annotates them. The platform aims at providing transparent access and analysis tools to pharmacovigilance experts investigating relevance of safety signals related to drugs. The platform's architecture consists in the integration of two vendor tools ITM® and Luxid® and one academic web service for knowledge extraction from medical literature. Whereas a manual search performed by a pharmacovigilance expert retrieved 578 publications, the system proposed a list of 229 publications thus decreasing time required for review by 60%. Recall was 70% and additional developments are required in order to improve exhaustivity.

Keywords. systems integration, information storage and retrieval, databases, drug adverse effects, text mining, semantic annotation, knowledge bases

1. Introduction

The information about drug safety is based mainly on spontaneous reporting systems by health care professionals to a central agency and/or to the pharmaceutical companies [1]. Other sources of information are of interest to improve our knowledge of adverse drug reaction. The medical literature is one of them, and its analysis is mandatory for

[1] Corresponding Author: Florence Amardeilh, Mondeca, 3 cité Nollez, 75018 Paris, France; E-mail: florence.amardeilh@mondeca.com.

the pharmaceutical companies, as noted in the recent Guidelines on Pharmacovigilance for medicinal products for human use. "*The Marketing Authorisation Holder is therefore expected to maintain awareness of possible publications by accessing a widely used systematic literature review and reference database (e.g., Medline, Excerpta Medica or Embase) no less frequently than once a week"* [2].
Currently, with widespread use of internet, information retrieval, extraction and analysis of individual cases reports, or of other information (clinical studies, epidemiological studies, press releases, …), is a challenge for pharmacovigilants [3]. Several studies showed that researchers reported difficulties when searching electronic resources for information on adverse effects [4–6]. Among others, these difficulties are due to poor indexing, to the wide variety of articles that may be potentially useful, to variation in their qualities, as well as the lack of tools to perform systematic search. Moreover, the field of pharmacovigilance is extensive as all drugs and all adverse events must be under surveillance. The question and/or the answer about safety may be broad or narrow in scope. For example, a review with a broad scope might be "*which adverse effects are associated with antidepressant therapy*" and a narrow one being "*the risk of suicide or suicidal behavior in teens taking a serotonin reuptake inhibitor*". Adverse Drug Reaction (ADR) reviews may be appropriate targets for clustering techniques and require complex literature search strategies, as well as information extraction driven by domain knowledge. Therefore, it is desirable to develop a literature search strategy based on key elements of the pharmacovigilance field, to automate data mining in electronic databases, to have a better information retrieval and to assist the pharmacovigilance specialist visualizing and mining the results. Automation is one of the very important requirements, as it is needed to ease the knowledge acquisition bottleneck, particularly for annotating large collections of legacy documents [7].

In this article, we present for the first time the design of a new platform that automatically performs searches and identifies information on adverse drug reactions from the medical literature, as part of the VigiTermes project (http://vigitermes.univ-rennes1.fr).

2. Material and Methods

The architecture of the system is based on an integrated platform composed of both existing and specifically developed web services. When a new case is reported, the workflow consists of three main steps (i) to build and refine a normalised query to retrieve pertinent abstracts from Medline, (ii) to perform domain-oriented annotation on Medline abstracts and (iii) to enrich the knowledge base and the documentation of the case report. The steps are encapsulated in a UIMA [11] (Unstructured Information Management Architecture) infrastructure, providing a rich text mining standardized environment for manipulating documents and annotations, and easing external components integration, being the ones presented in this article or others. Also, the UIMA infrastructure has been enriched to comply with the Semantic Web standards, so that the outputs can be easily accessed, shared and exploited by other applications.

Pubmed Querying Web Service
We reused and adapted the algorithm developed by Garcelon et al. for querying Medline [8]. The main entry is a pair of Adverse Effect (AE) – Active Ingredient (AI).

Additional parameters may be added for filtering purposes, such as gender, age class, publication type and period. The web service relies on a UMLS web service to normalize the entries [9] and the MEDLINE web service to retrieve the corresponding abstracts [10].

Luxid® Annotation Factory
Luxid® Annotation Factory (LAF) is a software component dedicated to information extraction: It takes as input original documents and enriches them with semantic annotations (categories, entities and relationships). All analysis steps are packaged in linguistics components that build a so-called Skill Cartridge™, which is plugged to the extraction server. It can be imagined as a cascade of extraction modules that condense the textual data into meaning. The analysis is based on finite state technology. A specific Skill Cartridge™ dedicated to pharmacovigilance has been defined, extracting medical entities (Diseases, Treatments, Diagnostic Methods, Symptoms, Patients, Clinical trials) as well as relationships (Diagnosis, Therapy, Drug use such as doses, route, frequency and duration).

Intelligent Topic Manager
The Intelligent Topic Manager® (ITM) [11] is an ontology-based solution for Taxonomy, Thesaurus, Controlled vocabulary and Knowledge base Management based on the Semantic Web standards (RDF, OWL, SKOS). It allows access to distributed knowledge, management of large terminologies and collaborative and customized access for end users to legacy content through semantic portals. ITM uses MedDRA or the WHO-ART terminologies to consolidate and enrich the annotations and exploit them to automatically create new instances in the knowledge base such as the detected relations between adverse effects and drugs. This is operated by the ITM's Content Augmentation Manager (CA Manager), an open-source tool that is used to bridge the gap between information extraction tools and semantic repositories and that extends the UIMA framework with the Semantic Web standards.

3. Result

The VigiTermes solution for the semantic annotation of pharmacovigilance-related articles is a semantic portal that performs: 1) automated extraction and annotation of Medline abstracts with regard to a pharmacovigilance ontology and standard medical terminologies, and 2) case reports documentation thanks to the knowledge instances detected from the produced annotations. The semantic portal provides a user interface from which the pharmacovigilance expert submits a query with at least one AE and one AI as reported in the case. The platform then calls the web service that translates and enriches the user query in order to obtain a list of relevant PMIDs from Pubmed. This list of PMIDs is then sent to the CA Manager (see Figure 1) that transmits the abstract content of each PMID to Luxid® which extracts the relevant information and annotates it. This enriched content is then controlled by the CA Manager with respect to the medical terminologies and the knowledge base instances already stored in ITM. It consolidates each annotation and new knowledge instance thanks to a set of heuristic rules detailed in [12] before sending back the results to the semantic portal. The visualization step takes advantage of the produced annotations by highlighting the relevant concepts and instances identified in the abstracts.

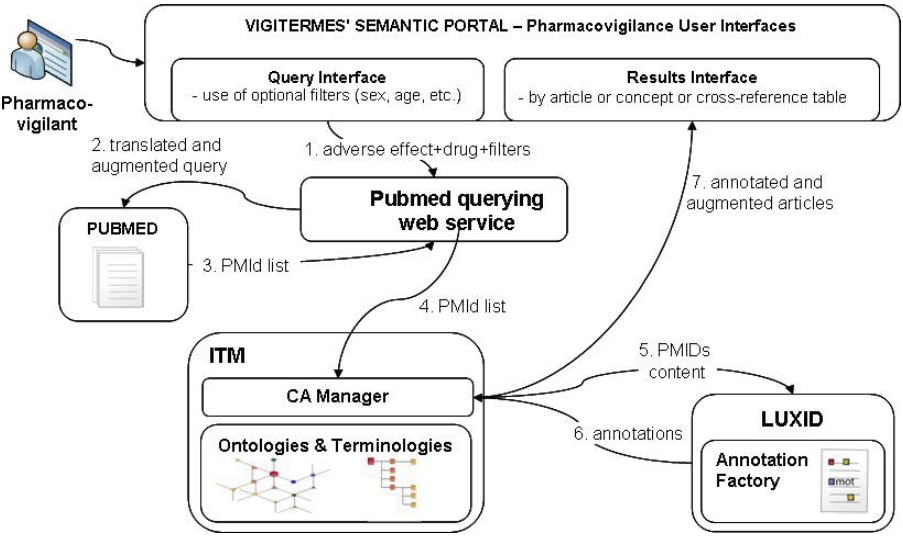

Figure 1. Case report documentation workflow

As of today, the pharmacovigilance regional centre of the European Georges Pompidou hospital experiments the VigiTermes semantic portal for functional validation. Use case is the documentation process for new ADR-AI case reports. Preliminary results have been obtained for the Pubmed Query web service. The gold standard was a Medline search performed by a pharmacovigilant expert who recorded his queries and identified the relevant publications based on 15 reported ADR cases. These relevant references represented 69 publications selected from the 578 retrieved with the Medline search (12%). The Pubmed Query component was then applied to the same set of 15 cases and the number of identified publications was 229. Within this set, 48 publications were belonging to the 69 relevant publications selected by the pharmacovigilance expert. Recall was 70% and precision 21%. Eleven cases of 15 have a recall of 100%.

4. Discussion and Conclusion

The medical literature is still a major source of information to evaluate drug safety: It has a valuable role in providing early warning alerts and can strengthen a signal detected by other means. Whereas pharmaceutical companies must organise a systematic survey only for the drugs they commercialize, regulatory authorities have to screen all the drugs available on the market. In any case, the amount of published literature on ADRs is overwhelming, so a standardized and automated tool would ensure that a minimum surveillance could be performed for each drug. The national French VigiTermes research project was funded to implement new methods for signal detection and analysis in pharmacovigilance. We presented the integration platform built to support the pharmacovigilance teams from industry and regulatory authorities by standardizing literature search on ADRs. The system works as a filter and it is able to reduce the number of abstracts to be reviewed by 60%. Additional developments are required in order to improve recall. The technical choices are based on Semantic Web

technologies, including ontology development and Web services, and the new standards recently developed in Natural Language Processing (UIMA).

Knowledge-based approaches have already been developed for screening drugs and predicting drug safety at the pre-clinical stage whereas this knowledge management platform is dedicated to the documentation of spontaneous case reports after market authorization. The development perspectives of our project will aim at adding other pharmacovigilance-related data sources (such as Martindale, Meyler's Side effects of Drugs, other databases like Micromedex) in order to complete data obtained from Pubmed. The presented platform can also be improved by developing tailored user interfaces for the pharmacovigilance end user. Moreover, to tackle some actual limitations of the platform, ontologies like SNOMED CT, by offering the ability to relate concepts together, will be most valuable in order to build clusters of MedDRA terms that express similar medical conditions. The PharmARTS tool developed by another partner of the VigiTermes project implements mappings between SNOMED CT and MedDRA/WHO-ART and a web-based graphical interface in order to select adverse reaction terms corresponding to a formal definition expressed thanks to SNOMED CT concepts [13]. This would help to complement the current search strategy by adding subsumption mechanisms in the algorithm. Finally the platform should accommodate cross-platform integration with proprietary spontaneous reporting system or other signal detection tools.

Acknowledgment. The VigiTermes project is funded by the French Agence Nationale pour la Recherche (ANR) ANR-07-TECSAN-026.

References

[1] Bright, R.A., Nelson, R.C. (2002) Automated support for pharmacovigilance: A proposed system. *Pharmacoepidemiology and Drug Safety* 11(2):121–125.
[2] Hopewell, S., Wolfenden, L., Clarke, M. (2008) Reporting of adverse events in systematic reviews can be improved: Survey results. *Journal of Clinical Epidemiology* 61(6):597–602.
[3] Cobert, B., Silvey, J. (1999) The internet and drug safety. *Drug Safety* 20(2):95–107.
[4] Loke, Y.K., Price, D., Herxheimer, A. (2007) Cochrane adverse effects methods group. Systematic reviews of adverse effects: Framework for a structured approach. *BMC Medical Research Methodology* 7:32.
[5] Golder, S., Loke, Y., McIntosh, H.M. (2008) Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. *Journal of Clinical Epidemiology* 61(5):440–448.
[6] Uren, V., Cimiano, P., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F. (2006) Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web* 4(1):14–26.
[7] Garcelon, N., Mougin, F., Bousquet, C., Burgun, A. (2006) Evidence in pharmacovigilance: extracting adverse drug reactions articles from MEDLINE to link them to case databases. *Studies in Health Technology and Informatics* 124:528–533.
[8] http://eutils.ncbi.nlm.nih.gov/entrez/query/static/efetch_help.html.
[9] http://kswebp1.nlm.nih.gov/DocPortlet/html/dGuide/webservices/metaops/find/find2.html.
[10] http://www.alphaworks.ibm.com/tech/uima.
[11] Amardeilh, F., Francart, T. (2004) A semantic web portal with HLT capabilities. In *Actes du colloque Veille Stratégique Scientifique et Technologique (VSST2004)*, Toulouse, France, Vol. 2, 481–492.
[12] Amardeilh, F. (2008) Semantic annotation & ontology population. In Cardoso, J., Lytras, M.D. (Eds.) *Semantic Web Engineering in the Knowledge Society*, Idea Group Publishing, 135–160.
[13] Alecu, I., Bousquet, C., Degoulet, P., Jaulent, M.C. (2007) PharmARTS: Terminology web services for drug safety data coding and retrieval. *Studies in Health Technology and Informatics* 129(Pt 1):699–704.