F. Lau et al. (Eds.)

doi:10.3233/978-1-61499-951-5-298

# Modeling Keyword Search Strategy: Analysis of Pharmacovigilance Specialists' Search of MedDRA Terms

Romaric MARCILLY[a,1], Laura DOUZE[a], Cedric BOUSQUET[b] & Sylvia PELAYO[a]
[a]*Univ. Lille, INSERM, CHU Lille, CIC-IT/Evalab 1403 - Centre d'Investigation Clinique, EA 2694, F-59000 Lille, France*
[b]*INSERM, U1142, LIMICS, F-75006; Sorbonne Universités, UPMC, Univ. Paris 06, Paris; Univ. Paris 13, F-93430, Villetaneuse, France*

**Abstract.** In the information retrieval task, searching and choosing keywords to form the query is crucial. The present study analyzes and describes the keywords' search strategy into a thesaurus in the field of pharmacovigilance. Two ergonomics experts shadowed 22 pharmacovigilance specialists during their daily work. They focus on the strategies for searching and choosing MedDRA terms to build pharmacovigilance queries. Interviews of four pharmacovigilance specialists completed the observations. Results highlight that, for unusual or complex searches, pharmacovigilance specialists proceed iteratively in three main phases: (i) preparation of a list of terms and of evaluation criteria, (ii) exploration of the MedDRA hierarchy and choice of a term, and (iii) evaluation of the results against the criteria. Overall, the search and the choice of keywords within a thesaurus shares similarity with the information retrieval task and is closely interwoven with the query building process. Based on the results, the paper proposes design specifications for new interfaces supporting the identification of MedDRA terms so that pharmacovigilance reports searches achieve a good level of expressiveness.

**Keywords.** Cognitive work analysis; pharmacovigilance; MedDRA; Information retrieval; Modelling;

## 1. Introduction

Information retrieval (IR) from a database is a complex cognitive work that has been the object of different kinds of modelling [1]: some models describe the IR behavior [2] while other, explicative models, represent the cognitive activities underpinning this behavior [3]. Despite their differing perspectives, both kinds share common features [1]. First, they all consider that the IR work comprises three main steps: (i) identifying the research question, (ii) expressing and performing the query, and (iii) analyzing, synthesizing and evaluating the results compared to the research question. Second, IR is an iterative process with loops between the above-mentioned steps.

For instance, Sutcliffe and Ennis [4] (Figure 1) highlight how the initial search question is turn into needs, then into a query entered in an IR system that provides in turn

[1] Corresponding author, Centre Hospitalier Universitaire de Lille, Institut Cœur-Poumon (ICP), Bd du Professeur Jules Leclercq, 3[ième] étage Aile Est, CS 70001, 59037 Lille Cedex, France ; e-mail : romaric.marcilly@univ-lille.fr

results that are evaluated by the users against their expectancies. This model considers various kinds of iterations: unsatisfactory results lead to reformulate the query or to rephrase the needs, while a failed search leads to changing the initial question. Iterations will stop as soon as the results are satisfactory. Two kinds of knowledge may impact the search strategy. User's knowledge of the problem domain deeply influences the way the initial problem is turn into needs, along with the way the query will be formulated, and the results evaluated. User's knowledge of the technology supporting the search impacts the formulation and the execution of the query in the IR system.

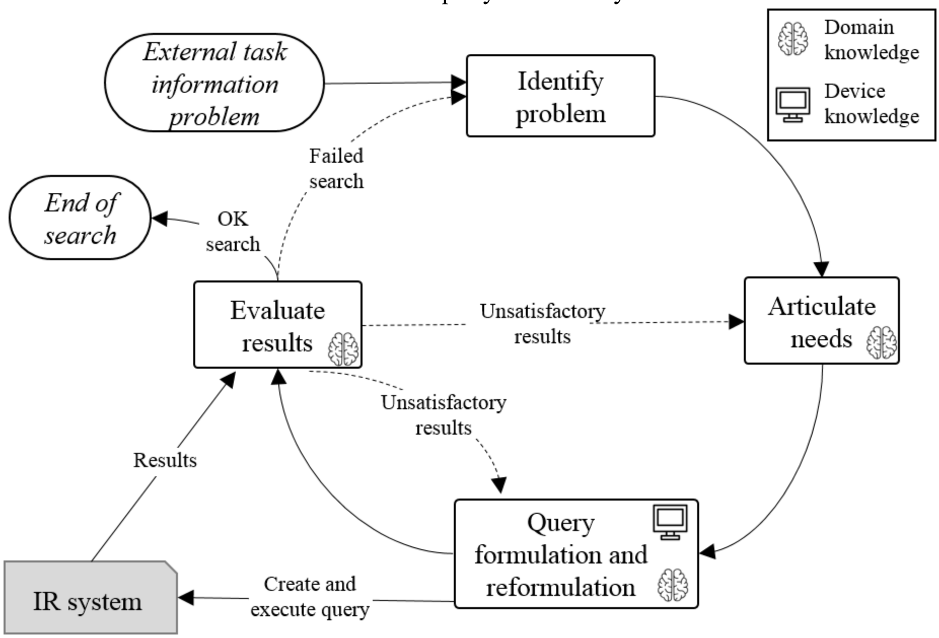


**Figure 1.** Model of information searching activities and knowledge sources adapted from Sutcliffe and Ennis [4]. Icons represent the type of knowledge that impact the steps of the process.

A crucial element in the IR strategy is the choice of the keywords forming the query. It has been shown that different users with the same background may use different keywords to formulate a query in response to a unique question [5]. This heterogeneity in the keywords chosen may questions the relevance and the expressiveness of the results. Yet, despite the importance of this step, as far as we know, no study explains how users search and choose keywords from a terminology to formulate a query. This study aims to describe the keywords' search strategy into a thesaurus in the field of pharmacovigilance.

## 2. Background: the PEGASE project

Pharmacovigilance is "the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem" [6]. In France, pharmacovigilance specialists (PVS) of regional pharmacovigilance centers are in charge of analyzing and reporting the adverse drug reactions (ADR) noticed by healthcare professionals and patients to healthcare authorities. For this purpose, they enter the reports into a national online database and codify the ADR using one or several terms that closely capture the original verbatim. They also search in the database to answer either clinicians' questions about side effects of a medication, or authorities' questions about unexpected drug safety problems (signal detection task). For this, they search existing reports in the database with the keywords used to code the ADR in conjunction with the name of the drug incriminated. In this context, a standardized medical terminology thesaurus, MedDRA® (Medical Dictionary for Drug Regulatory Activities), is useful to unify the codification of the ADR in the reports.

Yet, depending on the PVS' background (e.g. physician or pharmacist) and experience, and on the initial description of the ADR (symptoms of the patient, level of precision, wording), different MedDRA terms may be used for coding this ADR. Then, when PVS need to identify all reports related to a given type of ADR, they first need to identify all MedDRA terms that may have been used to code it. Those terms may be scattered all over the MedDRA hierarchy which makes difficult to find them all and achieve a good expressiveness of the results. The PEGASE national project aims to develop usable tools that help PVS to find all relevant MedDRA terms and to ensure searches achieve a good expressiveness. The present paper reports on the first part of this project: analyzing and modelling PVS' strategies to search and choose MedDRA terms that will form a query.

## 3. Method

In 4 French regional pharmacovigilance centers partners of the project, 2 ergonomics experts (LD & RM) independently shadowed PVS during their daily work focusing on their strategies for searching and choosing MedDRA terms while searching reports. PVS were asked to think-aloud while performing their searches. Their interaction with the online report database, their behavior and their verbalizations were transcribed for analysis.

Besides, during complementary individual interviews, 4 PVS (one per center) were asked to explain how they would have chosen MedDRA terms in 3 realistic pharmacovigilance scenarios designed by a pharmacist expert in pharmacovigilance (CB): 2 searches to answer clinicians' questions (one "easy", the MedDRA term is mentioned in the question, and one "complex", the case includes multiple symptoms and no MedDRA term is mentioned), and one search in the context of signal detection. Participants were asked to justify each step of their decision-making process. Their decisions and explanations were recorded and transcribed for analysis.

Finally, all data were analyzed to identify the invariant strategies implemented for searching and choosing MedDRA terms along with the variations and their causes. The strategies were modelled through the analytical description method (MAD [7]).

## 4. Results

A total of 22 PVS were observed (36h10). The complementary interviews lasted 4h. The strategies for searching MedDRA terms for detecting a signal and for answering a clinical question differ from each other on the need for results' exhaustivity. In the context of the signal detection, PVS select the MedDRA terms so that they are sure the query identifies all relevant reports. When answering a clinician's question, PVS may be content with a sample of relevant reports as soon as they enable them to answer the question asked and give insights on how to fix or manage the ADR. Apart from this difference, the search strategy is roughly the same.

When searching reports of common ADR, PVS know the MedDRA terms and do not have to search them. On the contrary, for unusual or complex searches, PVS do not know the relevant MedDRA terms and must proceed iteratively to find the right one(s). They proceed in 3 main phases (Figure 2):

- *Preparation of a list of terms and of query evaluation criteria.* Based on the verbatim of the ADR notification, on their experience and on their expertise, PVS think to a term or several terms that may represent the ADR. They also mentally estimate the number of reports they expect to find in the database (e.g. dozen or thousands) based on the characteristics of the drug incriminated (date of marketing authorization, class of drug) and of the ADR (known frequency of the effect, amount of documentation about the pair drug-effect).
- *Exploration of the MedDRA hierarchy and choice of a term.* PVS enter successively each MedDRA term they have in mind into the online report database and explore the MedDRA hierarchy around this term by checking its filiation and relatedness[2]. From the term initially entered, they move up and down the hierarchy and, at each level, they explore the terms above and beneath it until they find a MedDRA term that includes all terms that seem relevant to describe the ADR, but not too many irrelevant ones. In the context of signal detection, PVS do not consider the number of irrelevant terms: they prefer to get irrelevant terms and therefore irrelevant reports than missing a relevant one.
- *Evaluation of the results against the evaluation criteria.* The term chosen is combined with the name of the drug incriminated into a query in the online report database. Several (combinations of) MedDRA terms may be tested till the number of reports identified with the query meets PVS' expectation or till all possible terms have been exhausted. That way, choosing a MedDRA term is closely intertwined with the query building.

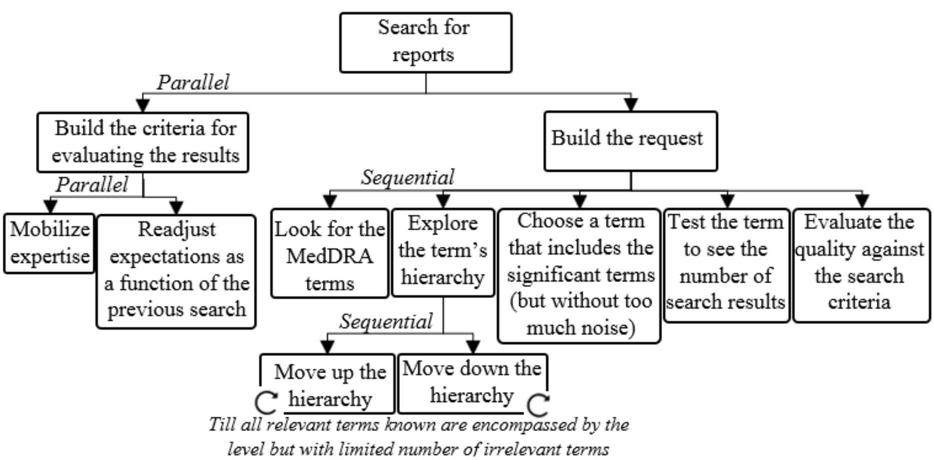


**Figure 2.** Simplified PVS' MedDRA terms search strategy for answering a clinician's question modelled with the analytical description method.

[2] The section of the French online report database in which the MedDRA term is entered enables to explore the levels of the MedDRA hierarchy.

## 5. Discussion

As far as we know, this study is the first to explore how PVS choose MedDRA terms to perform a query in an online report database. To a larger extent, as far as we know, it is the first attempt to model how users of an IR system search and choose standardized terms from a thesaurus to perform a query. Results show that, in the contexts of PVS' answer to clinicians' questions and of signal detection, the search and the choice of keywords within a thesaurus shares similarity with the IR activity as it is usually modelled (e.g. [4]): it is an iterative process based on an informed trial and error approach. Besides, this process is closely interwoven with the query building process and cannot be pulled apart from it. The keyword search and choice strategy could be considered as a subprocess of the IR process and described by Sutcliff and Ennis [4]. For now, the model elaborated from the results of this study applies only to the field of French pharmacovigilance. Similar analyses in other contexts will help generalize it to other pharmacovigilance contexts and finally to any kind of search strategies into a thesaurus.

In the frame of the PEGASE project, this model was used to develop design specifications for new interfaces supporting the identification of MedDRA terms so that reports' searches achieve a good level of expressiveness. For instance, to help the PVS choose the best MedDRA term(s) amongst several, the search interface should present each MedDRA term at its place within the hierarchy with its filiation(s) and relatedness to enable PVS to compare them. Moreover, after having entered the drug name into the online database, as soon as PVS click on a MedDRA term, the interface should display the number of reports that are indexed with this term. It will inform them directly about the relevance of the term in comparison to the defined evaluation criteria. A total of 22 recommendations have been given to designers. Three different interfaces are under development and will undergo user testing.

## Acknowledgment

The PEGASE project has been funded by the French National Agency for research (ANR-16-CE23-0011). The authors would like to thank all PVS who were shadowed and interviewed for the kindness and their patience.

## References

[1] J. Dinet, A. Chevalier, A. Tricot, Information search activity: An overview, *European Review of Applied Psychology* **62** (2012), 49-62.

[2] G. Marchionini, *Information seeking in electronic environments*, Cambridge University Press, Cambridge, 1995.

[3] P. Ingwersen, Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory, *Journal of Documentation* **52** (1996), 3-50.

[4] A. Sutcliffe, M. Ennis, Towards a cognitive theory of information retrieval, *Interacting with computers*, **10** (1998), 321-51.

[5] M. Bates, The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* **13** (1989), 407-24.

[6] WHO, *Pharmarmacovigilance*, Available from: http://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/.

[7] S. Sebillotte, D.L. Scapin, From users' task knowledge to high-level interface specification, *Journal of Human-Computer Interaction* (2009), 1-9.