An Iconic Approach to the Browsing of Medical Terminologies

Jean-Baptiste Lamy[a], Van Bui Thuy[a], Agnès Lillo-Le Louët[b], Cédric Bousquet[a,c]

[a] LIMICS, Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, INSERM UMRS 1142, Sorbonne Université, Paris, France
[b] Centre Régional de Pharmacovigilance HEGP, AP-HP, Paris, France.
[c] SSPIM, Unit of Public Health and Medical Information, CHU University Hospital of Saint Etienne, Saint-Étienne, France

Abstract

Medical terminologies are the basis of interoperability in medicine. They allow connecting the various systems and data and facilitate searches in databases. An example is the MedDRA terminology, used in particular for coding drug adverse events. However, these terminologies are often complex and involve a huge number of terms. Consequently, it is difficult to browse them or find the desired terms. Traditional approaches consist of lexical search, with the problems of synonymy and polysemy, or tree-based navigation, but the user often gets "lost" in the tree.

Here, we propose a new approach for browsing medical terminologies: the use of pictograms and icons, for formulating the query in complement to a textual search box, and for displaying the search results. We applied this approach to the MedDRA terminology. We present both the methods and search algorithms and the resulting browsing interface, as well as the qualitative opinions of two pharmacovigilance experts.

Keywords:

Terminology as Topic, Pharmacovigilance, Nonverbal Communication.

Introduction

Medical terminologies are the basis of interoperability in medicine [1]. They allow connecting the various systems and data, and facilitate searches in databases [2]. An example is the MedDRA terminology, used in particular for coding drug adverse events. Traditional search approaches are either based on lexical search or on the navigation in a hierarchy [3]. However, lexical searches suffer from synonymy, polysemy and metonymy, e.g. a search with "auricular" return terms related to both heart and ear. Navigation in a tree is usually complex and the user often gets "lost" in the tree [4]. Brown [5] described the various difficulties encountered when performing searches in MedDRA-coded pharmacovigilance database.

In the literature, several authors suggest the use of pre-coordination for defining terms in terminologies, i.e. the decomposition of disorders into anatomic location and morphology, e.g. "nephritis," defined as an inflammation of the kidney, and post-coordination for selecting terms [6]. Bakhshi-Raiez et al. [7] proposed to take advantage of the compositional aspect of the SNOMED CT terminology to facilitate the searches in the terminology. Lee et al. [8] proposed the use of post-coordinated expressions to retrieve ICD codes. Third-generation, knowledge-based terminologies, aim at formalizing semantic relations and decompositions between terms, which can help with searches [9]. The use of semantic web standards such as RDF has also been proposed to facilitate data retrieval from MedDRA [10].

In previous works [11], we proposed an iconic compositional language for medical concepts. In this paper, we propose an iconic approach to the browsing of medical terminologies. Our approach uses icons at two levels: (1) for formulating the query, in combination with a traditional textual search box, and (2) for displaying the search results, by grouping similar terms using icons. This work is part of the Pegase project, which aims at helping pharmacovigilance experts with the use of the MedDRA terminology.

Methods

MedDRA is a medical classification often used for coding drug adverse effects and events. It has 5 levels, from the more general to the more specific: system organ class (SOC), high-level group terms (HLGT), high-level terms (HLT), preferred terms (PT) and low-level terms (LLT). The classification is multi-axial: a term can have more than one parent. We worked on the French translation of MedDRA 19.

VCM [11] is an iconic language for representing the patient main clinical conditions, including symptoms, diseases, physiological states (*e.g.* age class or pregnancy), risks and history of diseases, drug and non-drug treatments, lab tests and follow-up procedures. VCM includes a set of graphical primitives (colors and pictograms) and a grammar to combine these elements for creating icons.

For representing clinical signs and disorders, a VCM icon is made of a color, a basic shape, a central pictogram and a set of modifier pictograms. The color indicates the temporal aspect of the icon: red for current states of the patient, orange for risk of future states, and brown for past states. The basic shape is a circle for physiological states or a square for pathological states (diseases or symptoms). The central pictogram indicates the anatomico-functional location (e.g. endocrine system) or the patient characteristic (*e.g.* pregnancy) involved; and special pictograms are available for a few specific disorders associated with a specific anatomico-functional location (*e.g.*

```
function lexico_icono_hierarchical_search(keywords, pictograms, selected_levels):
    if only keywords (i.e. pictograms = ∅):
        terms = { t such that MedDRA(t) and match(t.label, keyword) and t.levels ∈ selected_levels }
        icons = { i such that Icon(i) and ∃ t ∈ terms with has_icon(t, i) }
    else if only pictograms (i.e. keywords = ∅):
        icons = { i such that Icon(i) and ∀ p ∈ pictograms we have has_part(i, p) }
        terms = { t such that MedDRA(t) and t.levels ∈ selected_levels and ∃ i ∈ icons with has_icon(t, i) }
        icons = { i such that i ∈ icons and ∃ t ∈ terms with has_icon(t, I)}
    else (both keywords and pictograms):
        terms = { t such that MedDRA(t) and match(t.label, keyword) and t.levels ∈ selected_levels }
        icons = { i such that Icon(i) and ∀ p ∈ pictograms we have has_part(i, p) }
        terms = { t such that t ∈ terms and ∃ i ∈ icons with has_icon(t, i) }
        icons = { i such that i ∈ icons and ∃ t ∈ terms with has_icon(t, I)}
    return (icons, terms)
```

Figure 2– Algorithm performing the lexico-icono-hierarchical search in pseudo-code. The search takes three input parameters: a set of lexical keywords, a set of pictograms, and a set of selected levels in the hierarchy (e.g. {PT, LLT}). It returns two values: the set of icons and the set of MedDRA terms to display. We used a first-order logic syntax in the algorithm, i.e. classes are unary predicates and properties are two-argument predicates, e.g. "MedDRA(t)" means that it is a MedDRA term and "has_icon(t, i)" means that it has icon i. Finally, match() is a function that performs keyword-based string comparisons.

diabetes for the endocrine system). Modifier pictograms can be added to specify (a) a general pathological process (*e.g.* tumor or infection), and (b) a "transversal" anatomical structure that can be present in many anatomico-functional locations (*e.g.* blood vessels and nerves are present in most organs).

The VCM ontology has been designed for formalizing the semantics of VCM icons [12]. This ontology includes 3 parts: (1) graphical concepts corresponding to VCM graphical primitives (*i.e.* the various shapes, colors and pictograms), (2) medical concepts (*i.e.* the main anatomical structures, biological functions, pathological processes, *e.g.* liver, hepatic function and inflammation, but *not* the various disorders, such as hepatitis), and (3) relations between the graphical and the medical concepts (*e.g.* the "liver" central pictogram is associated with both the liver anatomic structure and the hepatic biological function). In a previous work [13], we started to design a mapping between MedDRA and VCM.

Model

We designed an OWL ontology including: (1) all MedDRA terms with their codes, labels and parent-child relations, (2) MedDRA to VCM mapping and (3) decomposition of VCM icons in pictograms. Figure 1 shows the general model of the resulting ontology.
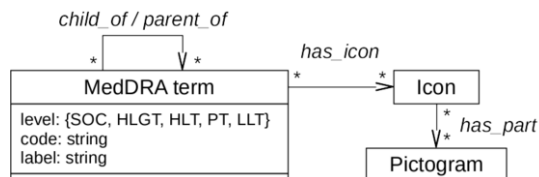

Figure 1– General model of the ontology, as a UML class diagram.

We used PyMedTermino [14] to extract the MedDRA terminology and to deal with VCM icons. The resulting OWL ontology includes 531,095 RDF triples (46 Mb) and belongs to the $\mathcal{ALIF}(D)$ family of description logics. The ontology was stored in an optimized quadstore based on a SQLite3 database, using the Owlready ontology-oriented programming module

[15, 16] for Python. Owlready was used because it allows an object-oriented access to ontologies with performances similar to SQL databases. It also supports a full-text search.

Search Approaches

The proposed interface combines three approaches for searching terms; they are detailed in the following paragraphs.

The **full-text search** approach consists of searching for one or more keywords in the labels of the MedDRA terms. Each keyword has to match the beginning of one word of the term, *e.g.* keyword "Abs" match term "Renal abscess" but not "Bile acid malabsorption". We used Owlready2 to perform the search, which relies on the FTS5 implementation of Sqlite3. In the interface, keywords are entered into a search bar.

The **iconic approach** consists of selecting one or more pictograms and searching for all terms that are associated with icons that include all the selected pictograms. To facilitate the selection of pictograms, we organized them on a schematic anatomical sketch, called "Mister VCM". It shows 37 pictograms, grouped in 6 regions: the head, the hat (for social medicine), the thoughts (for psychology and addictions), the trunk, the arm (only one limb is detailed) and the etiologies. The entire VCM language has more than 150 pictograms, however, many of them are specialized versions of a more general pictogram, and "Mister VCM" shows only the general pictograms. The user can click on a pictogram to select it, and the gray pictogram is replaced by a red icon. By clicking the same pictogram again, the user can deselect it. When several pictograms are selected, we consider their intersection, *i.e.* the terms have to match all the selected pictograms.

The **hierarchical approach** considers the hierarchical structure of medical terminologies. In the present work, we limited this approach to the ability to filter MedDRA terms according to their depth level in the terminology, *e.g.* search only for PT, or for both PT and LLT, ... The interface includes 5 checkboxes, one per MedDRA level; they allow the user to select the desired levels.

The output of the search is a set of MedDRA terms, but also a set of VCM icons. The set of icons are be used to display the search results in a synthetic and visual way. Figure 2 gives the algorithm that combines the three search approaches. If only keywords are present in the query, we search for the terms of

Figure 3– Screenshot of the results of an iconic search for PT and LLT with the "eye" pictogram.

the desired levels whose label matches the keywords; then we search for their icons. If only pictograms are present, we search for the icons having these pictograms; then we search for the terms of the desired levels that have at least one of these icons; finally, we filter the set of icons to those associated with at least one term in the resulting set of terms (due to the filtering on hierarchical levels, some icons can be associated with no terms). When both keywords and icons are present, we search for terms (taking into account keywords and levels) and for icons (taking into account pictograms) independently, and then we restrict both resulting sets to the elements associated with at least one element in the other set. In addition, the proposed algorithm is designed so that each step involves at the most two classes out of the three in the model (*i.e.* MedDRA terms, Icons, and Pictogram). This improves performances of the search.

Displaying the Search Results

Search results can be numerous; therefore, we use icons to organize them. Icons are sorted by the number of associated terms. For each icon, we display a panel with the icon on the left and the list of the corresponding MedDRA terms on the right. LLT are displayed in gray and other levels in black. When more than 6 terms are present, we display only the first 5 ones (in alphabetical order) and the total number of terms; the entire list of terms can be obtained by clicking on the panel. Whenever a given term is associated with more than one icon in the resulting set of icons, the term appears in several panels (one per icon), with a gray label indicating the number of icons (*e.g.* "2 icons"). By mouse-hovering this label, the user can see a popup bubble showing all the icons.

Moreover, when there are more than 10 icons, we reduce the list of icons by removing all icons that are more specific than another. In fact, the VCM ontology defines is-a relation between icons. For example, the icon "renal infection" is more specific than "renal disorder" and "infection". The terms associated with the removed icons are re-associated with their parent icon (if several parents are present, the terms are associated with all of them and will be displayed in several panels, as previously). When the panel of the parent icon is clicked, the terms initially associated with the parent icon are displayed, as well as other panels for the child icons.

Preliminary Evaluation

As a preliminary evaluation, we showed the proposed interface to two pharmacovigilance experts working daily with the MedDRA terminology. We gathered their opinions and their comments using a qualitative form including the following 7 general questions: Is the interface useful for finding MedDRA terms when coding a new case? when gathering codes for performing an exhaustive search in a pharmacovigilance database? Do you search mainly at the PT level or the LLT level? Is the iconic search interesting? alone? in combination with the lexical search? Is a training necessary for using VCM icons?

Results

Iconic Interface for Browsing MedDRA

Figure 3 shows a screenshot of the proposed interface. On the left, the user can enter keywords, click pictograms on "Mister VCM" and select the desired MedDRA levels. These three

operations can be done in the order of his choice. The proposed system has a good performance: even when searching over the largest part of MedDRA (the tumor pictogram), the response time does not exceed 0.6 seconds.

Figure 3 shows the results of an iconic search with the "eye" pictogram. 3,650 terms were found and grouped according to 25 icons (more specific, children, icons exist but are not shown). When clicking on an icon panel, the entire list of associated terms is shown, as well as more specific icons. Figure 4 shows an example of such a panel, obtained after clicking on the "allergic inflammation of eye" icon on Figure 3 (it is the icon marked with a green dot). The panel shows 2 more specific child icons: "allergic inflammation of the optic nerve" and "allergic edema of eye".

Figure 4– Screenshot of the panel shown when clicking the "allergic inflammation of eye" icon in Figure 3 (the one marked with a green dot).

Figure 5 shows an example of a search mixing lexical and iconic queries. We wanted to search for abscess in the urinary tract. "Abscess" is a clearly defined and specific keyword, so we typed it into the lexical search box. On the contrary, "urinary tract" is more general: many synonyms exist, and it also includes several subparts (*e.g.* ureter). Thus, we clicked on the "kidney" pictogram that represents the urinary system. The results show 2 icons with 17 terms.

When a term is clicked, it is displayed on a page with its label, its MedDRA code, its icons and its position in the MedDRA hierarchy, including ancestors and children terms. Ancestors and children can be clicked. As a consequence, the proposed interface also allows a classical tree-based navigation in the terminology.

Expert Opinions

The purely iconic search was not found very interesting by the two experts. On the contrary, the combination of the lexical and iconic search was described by them as very interesting, for example when dealing with synonymy (*e.g.* "auricular" keyword + heart pictogram to avoid auricular terms referring to ear) or for increasing the sensibility of the search, since most pictograms are rather general. This second point makes our interface particularly appealing to them when gathering codes for performing an exhaustive search. Moreover, the ability to select the depth level allows starting at a high level and then going down to lower levels if there is no result on the higher level.
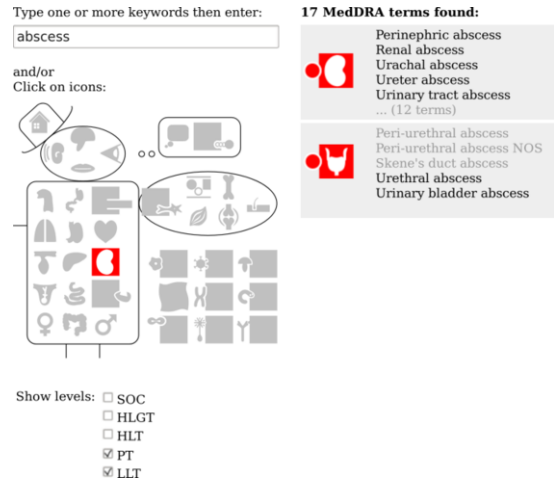

Figure 5– Screenshot of a lexico-iconic search: we searched for PT and LLT in MedDRA with the keyword "abscess" and pictogram "urinary".

On the contrary, when coding a new case, experts have to "stick" to the diagnostic mentioned in the case, and thus a lexical search through LLT is more appropriate. The iconic interface is thus less interesting for them. However, one expert qualified VCM icons as an "Esperanto of medical language". She considered that the iconic interface would be very useful for clinical research associates in the industry, because they are often non-specialists (*i.e.* not a physician or pharmacist) and could benefit from the guidance of icons: they may help to clearly identify the anatomic location or the morphology of a complex pathology, *e.g.* VCM icons explicitly represent "glomerulonephritis" as an inflammation of a part of the kidney while it may not be obvious to a non-specialist. In addition, many of them have to use MedDRA in English, even if it is not their native language.

Finally, both experts considered that training is needed before using VCM icons.

Discussion

In this paper, we proposed an original approach for browsing and searching medical terminologies, using pictograms and icons in addition to lexical keywords and terms. We applied this new approach to MedDRA. The resulting tool was found interesting by pharmacovigilance experts, in particular for gathering terms before performing an exhaustive search in a pharmacovigilance database, and for non-specialist users.

"Mister VCM" was initially designed to gather and organize a small set (5-50) of VCM icons. Here, we used it in a very different way, for allowing the selection of one or more pictograms in order to formulate a query in a visual way. This is the first time we use VCM for a querying purpose.

The icon enables a semantic search in the terminology. For example, the heart pictogram allows searching for all cardiac-related terms, without having to deal with the problems of synonymy or polysemy, while a lexical search would have to consider the many existing synonyms: "heart", "cardiac",

"card-", but also parts of the heart, such as "ventricular" or "coronary", and polysemy, *e.g.* "cardia" is not related to the heart but to the stomach, despite its short lexical distance which "cardiac" would suggest.

Some medical terminologies have an ontological structure that could be used for performing a semantic search. For instance, SNOMED CT relates all disorders to their finding site. These relations could be used to find all disorders related to the heart (including the heart subparts). However, the huge number of concepts in SNOMED CT makes this approach difficult and tedious to use. Moreover, ontological definitions are usually stricter than a human expert would expect, especially when performing broad search: *e.g.* in SNOMED CT, pericardium is not included in the heart because, strictly speaking, it is not a part of the heart. On the contrary, VCM pictograms are far less numerous, and they have a broader definition: *e.g.* the heart pictogram also covers the heart-related structure such as pericardium.

In the literature, Massari et al. [17] proposed meta-terms based on medical specialties to enrich terminologies and facilitate searches. Meta-terms can be combined with other types of searches. This approach is similar to the one we proposed here, but neither iconic nor visual.

Conclusions

Icons are a new, promising approach for browsing and searching medical terminologies. The perspectives of this work are twofold. First, the proposed interface will be properly evaluated in pharmacovigilance settings, using quantitative methods. Second, it could also be adapted to other medical terminologies such as ICD10 or SNOMED CT and used for helping physicians associate coded terms to their patients. Applications in medical education are also possible.

Acknowledgements

References

[1] Cimino J J. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998;37(4-5):394–403.

[2] Degoulet P, Sauquet D, Jaulent M C, Zapletal E, Lavril M. Rationale and design considerations for a semantic mediator in health information systems. Methods Inf Med. 1998;37(4-5):518–26.

[3] Bousquet C, Trombert B, Souvignet J, Sadou E, Rodrigues JM. Evaluation of the CCAM Hierarchy and Semi Structured Code for Retrieving Relevant Procedures in a Hospital Case Mix Database. In: AMIA Symposium. vol. 2010; 2010. p. 61–5.

[4] Sundvall E, Nyström M, Petersson H, Åhlfeldt H. Interactive visualization and navigation of complex terminology systems, exemplified by SNOMED CT. Stud Health Technol Inform. 2006;124:851–856.

[5] Brown EG. Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug Saf. 2003;26(3):145–58.

[6] Cornet R, Nyström M, Karlsson D. User-directed coordination in SNOMED CT. Stud Health Technol Inform. 2013;192:72–6.

[7] Bakhshi-Raiez F, de Keizer NF, Cornet R, Dorrepaal M, Dongelmans D, Jaspers MWM. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. IJMI. 2012;81:351–362.

[8] Lee D, Cornet R, Lau F, de Keizer N. A survey of SNOMED CT implementations. J Biomed Inform. 2013;46(1):87–96.

[9] Souvignet J, Declerck G, Asfari H, Jaulent MC, Bousquet C. OntoADR a semantic resource describing adverse drug reactions to support searching, coding, and information retrieval. J Biomed Inform. 2016;63:100–107.

[10] Alecu I, Bousquet C, Degoulet P, Jaulent MC. PharmARTS: terminology web services for drug safety data coding and retrieval. Stud Health Technol Inform. 2007;129(Pt 1):699–704.

[11] Lamy JB, Duclos C, Bar-Hen A, Ouvrard P, Venot A. An iconic language for the graphical representation of medical concepts. BMC Medical Informatics and Decision Making. 2008;8:16.

[12] Lamy JB, Soualmia LF. Formalization of the semantics of iconic languages: An ontology-based method and four semantic-powered applications. Knowledge-Based System. 2017;135:159–179.

[13] Lamy JB, Tsopra R. Combining semantic and lexical methods for mapping MedDRA to VCM icons. In: Stud Health Technol Inform. vol. 247; 2018. p. 905–909.

[14] Lamy JB, Venot A, Duclos C. PyMedTermino: an open-source generic API for advanced terminology services. Stud Health Technol Inform. 2015;210:924–928.

[15] Lamy JB. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif Intell Med. 2017;80:11–28.

[16] Lamy JB. Ontology-Oriented Programming for Biomedical Informatics. Stud Health Technol Inform. 2016;221:64–68.

[17] Massari P, Pereira S, Thirion B, Derville A, Darmoni S J. Use of super-concepts to customize electronic medical records data display. Stud Health Technol Inform. 2008;136:845–50.

Address for Correspondence
Jean-Baptiste Lamy <jibalamy@free.fr>
Bureau 149, LIMICS, UFR SMBH, Université Paris 13
74 rue Marcel Cachin
93017 Bobigny, France