Drug Information Portal in Europe: Information Retrieval with Multiple Health Terminologies

Saoussen SAKJI [a,b,1], Catherine LETORD [a], Suzanne PEREIRA [a,b,c], Badisse DAHAMNA [a], Michel JOUBERT [b], Stéfan J. DARMONI [a]

[a] *CISMeF, GCSIS, TIBS, LITIS EA 4108, University of Rouen, France*
[b] *LERTIM EA 3283, University of Marseille, France*
[c] *Vidal Company, Issy les Moulineaux, France*

Abstract. The objective of this work is the creation of a bilingual (French/English) drug information portal (DIP), in a multi-terminological context. The development of the DIP was based on the CISMeF portal, which catalogues and indexes the most important and quality-controlled sources of institutional health information in French. DIP has created specific functionalities related to drug and used specific drugs terminologies and classifications: the ATC classification, the CAS numbers, the French codes CIS, and CIP, as well as trade names and the International Nonproprietary Names of the drugs. DIP is the result of collaboration between the CISMeF team and the VIDAL private Company, specialized in drug information. DIP is conceived to facilitate the user information retrieval using several health terminologies. In the framework of the drug field, using multi-terminological context, indexing drugs with the appropriate codes or/and terms revealed to be very important to have the appropriate information storage and retrieval.

Keywords. drug information services, Europe, information storage and retrieval, information, Internet, terminology as subject

1. Introduction

To have precise information about drugs, an end-user should limit the number increasingly significant of drugs confusion. Since several years, international on-line drug databases (e.g., Vidal database) have been established to give to health professionals drugs information more relevant than it is [1]. To go further, we develop a drug portal to respond as much as possible to drug-related questions. A portal generally includes a search engine, and is a Web site that catalogues the main resources available in a given field. A health portal has several characteristics: it must respect precise quality standards; in the health domain, the reference is the Health on the Net code of conduct [2]. This code of conduct was selected in France in 2007 by the High health official to certify the e-health sites of this country.

---

[1] Corresponding Author: Court Leschevin, Door 21, 3rd floor, 1 rue de Germont 76031 Rouen Cedex, France; E-mail: Saoussen.Sakji@chu-rouen.fr.

Within the framework of the European project PSIP (Patient Intelligent Safety through Procedures in medication)[2] [3], we aim at developing a bilingual (French/English) Drug Information Portal (DIP) allowing to improve information retrieval about drugs. DIP was established to help health science information specialists, pharmacists and lay people to improve recall and comprehensiveness when searching for drug information. DIP is the result of collaboration between the CISMeF team and the private company VIDAL, specialist in drug information.

2. Material and Methods

CISMeF (URL: http://www.chu-rouen.fr/cismef) is the French acronym of Catalogue et Index des Sites Medicaux Francophones. It is a health portal which was conceived to catalogue and index the most important and quality-controlled sources of institutional health information in French (N ≈ 64.000), with the aim of providing the most relevant resources to user according to his context [4].

CISMeF uses two standard tools to organize information: the MeSH thesaurus and Dublin Core meta data [5]. The heterogeneity of Internet health resources led the CISMeF team to enhance the MeSH thesaurus, with the introduction of two new concepts: resource types (RT) and metaterms. These two new concepts were added to the MeSH descriptors (terms which allow the resource indexing) and the MeSH qualifiers (terms which specify the descriptor sense, and underline one of its particular aspect). CISMeF resource types are an extension of the publication types of MEDLINE and are used to categorize the nature or genre of the content of the resource. A metaterm is generally a medical specialty or a biological science, which has semantic links with one or more MeSH descriptors, qualifiers and RTs.

To adapt to drug information, CISMeF improved its terminology server: users can access all chemical substances (including drug substances), pharmacological actions, and new resources types related to the drugs such as: information on the drugs, drug evaluation, guidelines for drug use, monograph pharmacy, … A definition of each one of these resource types was provided either by the CISMeF team or the Vidal experts.

Within the MeSH thesaurus, the chemical substance names may correspond either to hierarchical MeSH descriptors or to MeSH supplementary concept records (SC) or synonyms of these terms. The MeSH SCs are terms of reference, not hierarchical, making it possible to describe chemical substances. Without being regarded as descriptors of the MeSH thesaurus, the SCs are semantically related to the latter: on the one hand, for each SC, MeSH recommends a projection towards descriptors; in addition it mentions the descriptor(s) corresponding with the pharmacological(s) action(s) of the described substance. For example, for the SC "cetuximab", MeSH recommends the mapping towards the descriptor "monoclonal antibodies" and specifies that the corresponding pharmacological action is "antineoplastic agents".

With regard to drugs, the most important is to retain the concept of "substance" and not just the concept of MeSH descriptor or MeSH SC. This is why, for the needs of the DIP, we created the concept "Substance" which makes it possible to gather the chemical substances: the MeSH descriptors or the MeSH SCs.

The development of a terminology adapted to the drugs is also made by the integration of trade (or brand or commercial) names, the International Nonproprietary Names (INNs) and of the various national and international codes related to the drugs and chemical substances such as the French CIP (Presentation Identifying Code) [6], the French CIS (Specialty Identifying Code) [7], the French UCD code (Common Dispensation Unit) [8], and the international ATC (Anatomical, Therapeutic and Chemical) code [9] and CAS (Chemical Abstract Service) code [10].

The ATC classification, controlled by the Collaborating Centre for Drug Statistics Methodology of the World Health Organization (WHO), is used to classify the drugs. The drugs are divided into various groups according to the organ or the system on which they act and/or their therapeutic and chemical characteristics. The ATC code has the general form LCCLLCC where (L represents a letter and C a number). In this system, the drugs are classified in five groups at different levels: The 1st level: anatomical group (1 alphabetical character). The 2nd level: principal therapeutic group (2 numerical characters). The 3rd level: therapeutic/pharmacological sub-group (1 alphabetical character). The 4th level: chemical/therapeutic/pharmacological sub-group (1 alphabetical character). The 5th level: sub-group for chemical substance: the individual active ingredient or the association of active ingredients (2 numerical characters). Each level of this classification corresponds to an ATC code and an ATC label. The label of the 5th level corresponds to the INN of the substance, when it exists. This code is allocated according to its principal indication. However, the latter can vary from one country to another, which explains why there can be several ATC codes for the same drug according to the country concerned. For example, thanks to an internal study carried out by the Vidal Company, approximately 10% of the drugs do not have the same ATC code between France and Denmark (one of the partners of the project PSIP). It was thus necessary for us to adapt to the French and Danish context to overcome with the problem of the varying ATC.

The CAS registry number of polymeric chemicals, a biological sequence or an alloy corresponds to its unique recording number in the Chemical Abstract Service database (CAS), a division of American Chemical Society (ACS). The CAS assigns these numbers, identifiable by an algorithm. Within the MeSH thesaurus, the CAS codes are connected to the MeSH descriptors and the MeSH SCs, which correspond to substances and thus were integrated into the CISMeF terminology server.

In France, the CIS includes eight digits and is allocated to the pharmaceutical specialties being or having made the subject of a drug approval [11]. It is managed in France by the AFSSAPS (French acronym for French Agency of sanitary security of health products) and in European Union by the EMEA (European Medicines Evaluation Agency). The CIP (code with 7 digits and recently with 13 digits) identifies the presentation of a pharmaceutical specialty. It is also managed by the AFSSAPS. A drug can be identified by several CIS numbers, which refer to a different dosage and/or a different dosage form for a specific drug. For the same CIS code, we can have several CIP codes according to the various existing presentations (size and/or conditioning). The UCD code characterizes the smallest unit used for the drugs dispensation in the care establishments. The UCD code is formed by seven digits established by the Club Inter Pharmaceutical, and published in the Official Journal.

3. Results

The construction of the Drug Information Portal (http://doccismef.chu-rouen.fr/servlets/DIP) is drawn on the CISMeF portal, of which it is a component with specific functionalities.DIP was developed in four steps:

Step 1: Creation of a metaterm "drug"

The CISMeF team has manually attached to the new metaterm "drugs" all the MeSH descriptors in relation with the drug such as: "pharmacological actions", "drug marketing authorization", "drug contamination". Then, we selected the qualifiers which are used to index the documents concerning drugs namely: "drug action and chemical substances", "pharmacokinetics", "drug therapy", … Lastly, we bound to this metaterm all the documents which had been described with the resource type "information on the drug" and implicitly all the hierarchically lower resource types. The regrouping of these terms on the level of the metaterm "drug" makes it possible to join together all the documents related to the drug.

Step 2: Creation of the portal site

DIP contains primarily search tools inspired largely by the Doc'CISMeF search engine, but with some specificities focused on drugs. The DIP search tools contain a bilingual (French and English) simple and advanced search. These two search modes are specific of a country, because of the varying ATC codes. Simple search can be done by the generic or brand name or the INN, or by any code related to drugs. Advanced search allows a specific search by a combination of all the elements describing the drugs: generic name/Brand name, INN, pharmacological action, codes related to drugs. Thus, a search can be conducted in different manners thanks to these various codes. These various accesses are devoted to various users (pharmacists, pharmacologists, toxicologists, …). For this reason it was necessary to adapt the search forms according to the needs of the user.

Step 3: Creation of the contextual links

The third step allows the implementation of the contextual links towards English speaking data and information bases in particular Drug Information Portal of the US National Library of Medicine (NLM) [12], and search tools of the National Center for Biotechnology Information (NCBI, NLM, NIH) in health sciences, which includes in particular Pub Med/MEDLINE and the NIH chemical databases.

Step 4: Development of the "Google selection DIP"

Lastly, we created a specific Google drug search, named "Google DIP", making it possible to carry out Google searches limited to a selection of health quality-controlled web sites relating to drugs and previously selected by the CISMeF team. We provided to Google a list of quality sites editors, limited to the field of the drug. These editor sites are the most important French institutional sites (like the High Health Authority or the National Drug Safety Agency). As the Google crawler covers at least all the static pages of a site, the corpus of Google DIP should include all the resources of DIP, but also the other pages which were not selected manually by the CISMeF team. For the Danish version of the DIP, a specific list will have to be defined by the Danish partners of the project PSIP. The query results are presented as descriptive notes giving directly access to information. Within each resource, a field named "substance" was defined making it possible to gather and locate the respective substances, the MeSH descriptors linked to it as well as their contextual pharmacological actions.

4. Conclusion and Discussion

In this paper, we have described a specific drug information portal, with specific emphasis on its multi-terminological information retrieval. To our knowledge, this approach is quite innovative. In 2008, the National Library of Medicine (NLM) set up the "NLM Drug Information Portal". This portal gives access to information concerning more than 12,000 drugs. Search can also be made with the generic name or the trade name. To our knowledge, research by CAS code or ATC code is not yet possible in this US Drug Information Portal. The development of an advanced search form in the DIP makes it also possible to better refine the search, crossing, for example, an INN with a pharmacological action. The results underscore the need for a basic understanding of the substances characteristics; effective use of chemical dictionary files; awareness of indexing by codes; and a well-planned search strategy that includes flexibility to make changes as necessary in order to complete a successful search.

The collaboration with the Vidal Company may lead in the near future to a multi-lingual Drug Information Portal: it needs the addition of European brand names and the translation of the various terminologies (MeSH, ATC) in European languages. The Health French Ministry is planning to design a Governmental Drug Information Portal in 2009, where the content has to be limited to French government sources. The current PSIP DIP is sufficiently scalable to answer to this reduced content compared to the current content described in this paper.

References

[1] Darmoni, S.J., Dufour, F., Massari, P., Arnoudts, S., Dieu, B., Alizon, B., Hantute, N., Baldenweck, M. Consultation of the Vidal electronic dictionary in the Rouen University Hospital: Analysis of the first year of utilisation (1992–1993). In Barahona, P., Veloso, M., Bryant, J. (Eds.) *Proceedings of the Twelfth International Congress of the European Federation for Medical Informatics MIE 1994*, IOS Press, Amsterdam, 384–388.

[2] Boyer, C., Gaudinat, A., Baujard, V., Geissbühler, A. (2007) Health on the Net Foundation: Assessing the quality of health web pages all over the world. *Studies in Health Technology and Informatics* 129(Pt 2):1017–1021.

[3] PSIP Web site. https://www.psip-project.eu/.

[4] Darmoni, S.J., Amsallem, E., Haugh, M., Lukacs, B., Leroux, V., Thirion, B., Weber, J., Boissel, J.P. (2003) Level of evidence as a future gold standard for the content quality of health resources on the internet. *Methods of Information in Medicine* 42(3):220–225.

[5] Dekkers, M., Weibel, S. (2003) State of the Dublin Core Metadata Initiative. *D-Lib Magazine* 9(4), http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april03/weibel/04weibel.html.

[6] Standardisation de la codification et du marquage des produits de santé URL: http://www.cipclub.org/institutionnel/fr/real/pdf/CIP%20Recommandation%20Codification%20V1-2.pdf.

[7] Health Administration systems, http://www.has-sante.fr/portail/jcms/c_576417/referentiel-de-certification-par-essai-de-type-des-logiciels-daide-a-la-prescription-en-medecine-ambulatoire.

[8] L'assurance maladie en ligne, http://www.ameli.fr/professionnels-de-sante/directeurs-d-etablissements-de-sante/codage/medicaments/medicaments-a-code-ucd.php.

[9] Skrbo, A., Begović, B., Skrbo, S. (2004) Classification of drugs using the ATC system (Anatomic, Therapeutic, Chemical Classification) and the latest changes. *Medicinski Arhiv* 58(1 Suppl 2):138–141. (in Bosnian).

[10] Dittmar, P.G., Stobaugh, R.E., Watson, C.E. (1976) The Chemical Abstracts Service Chemical Registry System. I. General Design. *Journal of Chemical Information and Computer Sciences* 16(2):111.

[11] Autorisation de Mise sur le Marché de médicaments (AMM), http://admi.net/jo/20070316/SANM0720920V.html.

[12] US NLM Drug Information Portal. http://druginfo.nlm.nih.gov/drugportal/drugportal.jsp.