

Title: Development, dissemination, and applications of a new terminological resource the Q-Code taxonomy for professional aspects of General Practice / Family Medicine.

Authors

Marc Jamoulle¹, Melissa Resnick², Julien Grosjean³, Ashwin Ittoo⁴, Elena Cardillo⁵, Robert Vander Stichele⁶, Stefan Darmoni⁷, Marc Vanmeerbeek⁸

1. MD. Family physician. Department of General Practice. University of Liège, Belgium

2. MS. MLS. Medical librarian, PhD Candidate. Health Sciences Center, University of Texas, Houston, USA

3. PhD. Professor of Computer Science, Department of Information and Medical Informatics (D2IM), University of Rouen, France

4. PhD. Professor of Information Science. HEC School of Management. University of Liège, Belgium

5. PhD. Researcher of Information Science and Medical Informatics. National Research Council, Italy Institute of Informatics and Telematics, National Research Council, Cosenza, Italy

6. MD. PhD. Family physician. Professor of Pharmacology. Heymans Institute, University of Ghent, Belgium

7. MD. PhD. Professor of Medical Informatics. Department of Information and Medical Informatics (D2IM), University of Rouen, France

8. MD. PhD. Family physician. Professor of family medicine. Department of General Practice. University of Liège, Belgium

Type of Article: Background Paper

Keywords: General practice – Primary health care - Terminology as Topic – Semantic web - Qualitative Research.

Key messages. (Word count: 49)

- ✓ A taxonomy for professional aspects of GP/FM is proposed as an extension to the ICPC classification.
- ✓ 182 Q-Codes are available as an on line semantic web resource in 8 languages.
- ✓ Face validity of this new taxonomy is illustrated by the rapidly growing list of current applications.

Word count: abstract; 239, key message; 49, text; 2622

References: 31 Table ; 2 Figure; 2

Abstract of Background Paper:

Background: While documentation of clinical aspects of General Practice/Family Medicine (GP/FM) is assured by the International Classification of Primary Care (ICPC), there is no taxonomy for the professional aspects (context and management) of GP/FM.

Aim: To present the development, dissemination, applications, and resulting face validity of the Q-Codes taxonomy specifically designed to describe contextual features of GP/FM, proposed as an extension to the ICPC

Development: The Q-Codes taxonomy was developed from Lamberts' seminal idea for indexing contextual content (1987) by a multi-disciplinary team of knowledge engineers, linguists and general practitioners, through a qualitative and iterative analysis of 1702 abstracts from six GP/FM conferences using Atlas.ti software. A total of 182 concepts, called Q-codes, representing professional aspects of GP/FM, were identified and organised in a taxonomy.

Dissemination: The taxonomy is published as an on-line terminological resource, using semantic web techniques and Web ontology language (OWL) (www.hetop.eu/Q). Each Q-code is identified with an Unique Resource Identifier (URI), and provided with preferred terms, formal definitions in eight languages (pt, es, en, fr, nl, ko, vi, tr) and search filters for Medline and web searches.

Applications. .

This taxonomy has already been used to support queries in bibliographic databases (e.g. Medline), to facilitate indexing of grey literature in GP/FM as congress abstracts, master theses, websites and as an educational tool in vocational teaching,

Conclusions: The rapidly growing list of practical applications provides face-validity for the usefulness of this freely available new terminological resource.

INTRODUCTION

Several attempts have been made to define the distinct concepts of General Practice and Family Medicine (GP/FM), as in statements of professional organisations [1, 2, 3, 4, 5] and medical textbooks [6, 7, 8, 9, 10, 11]. These definitions have been reviewed previously [12], in an attempt to delineate the scope of the field.

For the documentation of clinical data, the International Classification of Primary Care [13] (ICPC-2) is used, often in conjunction with the International Classification of Diseases (ICD). ICPC-2 has been developed to identify reasons for encounters, such as symptoms, acts performed or requested and diagnosis in daily practice. However, it was not constructed to reflect the professional context of GP/FM, such as organizational, managerial, ethical, environmental, educational, and investigational factors.

The absence of adequate normalized keywords for identifying specific professional contextual features of GP/FM, is an impediment to proper indexing of grey literature in GP/FG.

More than 50% of the scientific outputs of General Practitioners at congresses are never published [14, 15]. Since McWynney famous initiative FAML I [16, 17], there is no knowledge management system dedicated to GP/FM grey literature (research which has not been published in academic journals, nor indexed by the major bibliographic databases) [18]. Consequently study reports, congress abstracts, or collections of dissertations are often not globally available to researchers and practitioners, due to this lack of indexation [19]. For example, the WONCA Europe website maintains a database of more than 30,000 abstracts of conferences in GP/FM. However, these contents are not indexed, which hinders their utility to foster research and support the daily operations of GP/FM.

A taxonomy of professional contextual features of GP/FM has been developed to address this issue using present day technology for information storage and retrieval [20], such as machine learning, semantic web technologies and natural language processing (NLP).

This new taxonomy was named Q-Codes, in honour of the late Henk Lamberts, a Dutch Professor of Family Medicine (University of Amsterdam), who proposed to use the letter Q (not yet used in ICPC) for the development of a rudimentary coding system for his departmental research library.

Aim

The aim of this background paper is: i) to shortly describe the development of the Q-Codes taxonomy, designed to cover the professional contextual aspects of GP/FM; ii) to present its dissemination tools available for accessing the taxonomy; and iii) to describe a number of recent applications, as an exploration of the face-validity of this new terminological resource for primary care.

Development of the taxonomy

A qualitative and iterative analysis of 1,702 abstracts in English or French from 6 GP/FM congresses held between 2007-2014 (see Table 1) has been performed from a practicing GPs perspective, in a bottom up approach, to identify relevant concepts focusing on the professional features of GP/FM (e.g. teaching, ethics, or environmental hazard issues). The choice of conferences was pragmatic. Abstracts needed to be available for analysis by the first author (MJ) in due time before the conference, so that the results could be presented and discussed with attendees [21]. The study of Kruschinski et al. [22] on the content of 614 abstracts of the European General Practice Research Database used a similar methodology to delineate the aspects of GP/FM, and was reused in the development process of the taxonomy.

Year	Data Origin	Language(s)	Source	#
2007	Online. Access to reviewer pages before the congress	English	WONCA Europe conference	998
2013	Print, received during the congress	Portuguese	Portuguese 18th national conference of family medicine	128
2013	Excel file, abstracts received after the congress	French	Congres de la Confédération des Généralistes Enseignants (CNGE)	205
2014	Online. Access to reviewer pages before the congress	French	Congres de la Confédération des Généralistes Enseignants (CNGE)	289
2014	Print, published in Primary care, 2014	English, French, German	SwissFamilyDocs 2014	45
2014	Excel file	French	Belgian Congress 2014	37
2010	Print, published Family Practice 27(4) 459 :67	English	EGPRN Study	614

Table 1 Sources and languages of the abstracts analyzed

Starting point for the analysis of the congress abstracts were the original Q-Codes created by H. Lamberts. To identify what is really at stake and respect the bottom-up approach, only themes addressed by GPs during the conferences have been classified. The data was analysed in a grounded theory approach with Atlas.ti software (<http://atlasti.com/>). Each abstract is read and code attributed along the classification ICPC-2 for clinical aspects and with a new Q-Code for emerging professional contextual aspects. By grouping these identified concepts and their Q-Codes into subcategories and categories, gradually the taxonomy (a hierarchical ontological structure) takes shape. Cimino's standard set of desiderata for terminological resources was applied [23]. For a full description of this development, the reader is referred to a prior publication [24].

Each concept was illustrated by a carefully chosen sample of pertinent articles, to illustrate and deepen the understanding of the each individual Q-Code. Each Q-Code was linked to other terminological systems reference terminologies, such as BabelNet.org, DBpedia, and to carefully chosen controlled vocabulary MeSH and to a selected bibliography in open access. Twenty-four GPs and terminology experts from ten countries were involved in the translation of the terms and definitions into 8 languages, thereby contributing to the further validation of the concept description of each Q code.

In the end, a taxonomy of 182 Q-Codes was created to represent professional contextual features of GP/FM. There are 8 domains in the taxonomy: doctor's issues, patient's issues, patient categories, structure, knowledge management, and research and development. The two remaining domains are Planetary health issues and Medical Ethics. Each domain is further divided in categories and subcategories (see the electronic annex for a full list of the 182 codes).

Dissemination of the taxonomy

The taxonomy with 182 Q codes was published as a semantic web resource on the multilingual, multi-terminology portal HETOP after free inscription (www.hetop.eu) [25] and in Unique Resource Identifier (URI) format (see Table 2),

<ul style="list-style-type: none">• To reach the hierarchy ICPC-2 http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBO&tab=1 ICPC-2 Process http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBOPROC&tab=1 Q-Codes http://www.hetop.eu/hetop/Q?la=en&rr=CGP_CO_Q&tab=1• To reach each rubrics ICPC RFE and diagnosis: http://www.hetop.org/hetop/?la=en&rr=CIP_D_A01 ICPC Process http://www.hetop.org/hetop/?la=en&rr=CIP_P_30 Q-Codes http://www.hetop.eu/hetop/Q?la=en&rr=CGP_QC_QC1

Table 2 Free access to ICPC-2 and Q-Codes in URI format. To change the language; change the ISO 639 for the language ; Ex: =pt for Portuguese (en,fr,es,pt,tr,vi,ko,nl allowed, more expected). To change the rubric; change the code at the end.

Applications of the taxonomy

Numerous practical are currently in use or under development (see <http://3cgp.docpatient.net> for a constantly updated overview). We will limit ourselves here to describe examples of recent applications for three domains: facilitating bibliographic searching; facilitating indexing of congress abstracts and dissertations, and e-learning applications.

Facilitating bibliographic searching.

Within the HeTOP website, a search and retrieval engine was build, which allows for multilingual bibliographic queries on PubMed. These queries facilitate the search for relevant literature on issues in primary care.

We will illustrate this with an example. Suppose a medical student or general practitioner is interested in studies, pertaining to the concept of “work-life balance”. This concept is a Q-Code, but not a keyword in Medical Subject Headings (MeSH), which makes it difficult to conduct a search in Medline. By clicking on the Q-Code on www.hetop.eu/Q, relevant MeSH key words will be suggested, with detailed information on their scope, and relationships in the MeSH hierarchy. In addition, a relevant search profile for this concept will be automatically provided, turning 182 Q-Codes into search filters for LiSSa (<http://www.lissa.fr>), a French publications database and for Medline (PubMed (see fig 1).

Trainees in GP/FM have used this search engine to explore concepts at stake during a consultation, and to have a prepared bibliography search for each important themes used in daily practice. An example of this approach is given in http://docpatient.net/3CGP/QC/clinical_exercise.htm

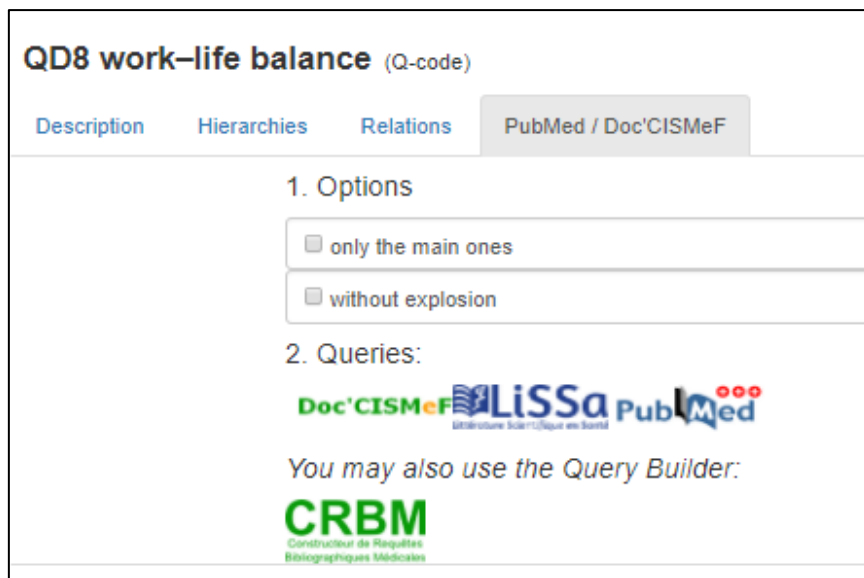


Figure 1 The multilingual Inforoute info-button of HeTOP allows access to an automatic query on CISMeF and LiSSa (French knowledge bases) and on PubMed.

Indexation of dissertations and congress abstracts

Q-Codes taxonomy and ICPC-2 are used to index the master theses in GP/FM in French speaking Belgium (see www.mgtfe.be (in French) and at the Department of General Practice of the University of Coimbra, Portugal). The results of long working hours devoted to the realization of an exciting final master thesis will remain unread, if not properly indexed. Students are offered ICPC, the Q-Code taxonomy, an indexing tool from the HeTOP portal, and a user guide to assist them in the task to index their work[26].

This approach is now reused by researchers from Houston University, Texas, US, for indexing of question answer pairs in telehealth Brazil [27] and by the Brazilian Society of Family and Community Medicine (SBMFC) to support the indexing of the congress abstracts of the next SBMFC congress (<http://www.cbmfc2017.com.br/trabalhos/>).

Indexing of congress abstracts allows congress participants to identify more quickly particular contributions of interest during the conference. Researchers not attending the conference can also more easily pin point interesting contributions and their authors. In addition, indexing all abstracts allows for a careful study of the scope and main points of the conference. As an example, we present shortly the results of such an analysis of the 212 abstracts of the congress of the French Society of teachers in GP/FM (CNGE), held in Lille in 2014. ICPC-2 was used to index clinical issues, and the Q-Codes taxonomy for professional contextual features. The distribution of themes of the QT domain (training and teaching) gives an insight in the abstracts presented (Fig 2) . In the abstracts presented at this congress, 45 address the general issue of Quality assurance, 14 are about EBM and 38 about guidelines, etc.

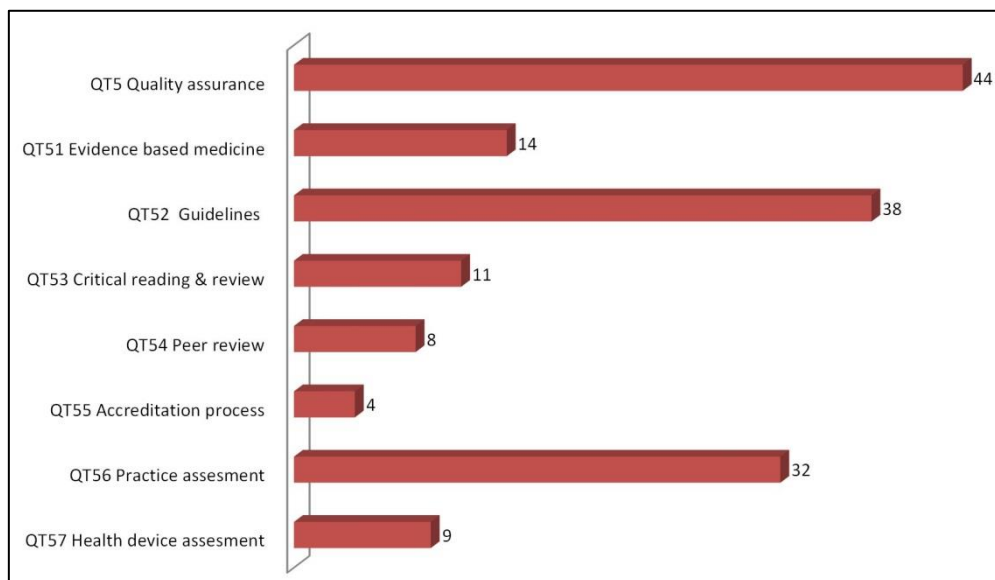


Figure 2 Themes addressing the issue of Quality in the Domain QT Knowledge Translation encountered in the 212 abstracts of the CNGE congress in 2014 (Q-Codes version 2.3)

In France, a Multi-Terminology Concepts Extractor (ECMT) is created for the full automation of the indexing process for conference abstracts, dissertations, and abstracts from the French LiSSa base [28]. A project to combine the Q-Codes Taxonomy with this tool for application in the domain of GP/FM is initiated.

E-Learning applications

In Vietnam, the Q-Codes Taxonomy has been used in an e-learning program for GP/FM students (<http://3cgp.docpatient.net/action-research/q-codes-in-an-e-learning-program-in->

vietnam). In this project, Natural Language Processing techniques are used to guide the students to the right Q-codes for their documentation needs.

DISCUSSION

Main findings

The Q-Codes Taxonomy for professional aspects of GP/FM has been developed as a new terminological resource for primary care. It is proposed as an extension to the clinically oriented International Classification for Primary Care (ICPC). This taxonomy is available on the web on www.hetop.eu/Q (free inscription) and in Unique Resource Identifier (URI) format (see Table 1), as a professional terminological resource in 8 languages, and fit for semantic web applications.

This taxonomy and its tools of dissemination are available at the point of care, to practicing physicians, researchers, trainers and trainees in GP/FM, as an illustration of the extent and complexity of this profession.

A growing list of recent applications of this taxonomy illustrates its rapid uptake, and provides face-validity to its usefulness for scientific communication in the field of GP/FM, by facilitating indexing of dissertations, master theses and congress abstracts, and facilitating bibliographic searches in grey literature repositories of Primary Care and in general bibliographic systems such as Medline.

Strengths and limitations.

The Q-Codes Taxonomy has been developed from a strong empirical basis in a bottom up approach, and applying terminological expertise and sound qualitative methods. To provide free access to the taxonomy, a database was published on the web, created with a standardized terminological structure and advanced semantic web techniques. Great effort was invested in linking and mapping the newly identified Q-codes to existing concepts in existing terminologies, both linguistic (Babelnet) and conceptual (Pubmed).

Further ontological research is needed to determine whether the two main rules of taxonomical thinking have been respected: completeness (all identified) and exclusivity (one place for each concept) [29]. A taxonomy is never complete and must evolve with the discipline. In this development process, we stopped at 182 Q-Codes, but future congress will certainly reveal new concepts.

The value of a taxonomy is determined by the quality and speed of the maintenance and updating process. Although Q-Codes have not been officially endorsed by the World Organization of Family Doctors International Classification Committee (WICC), the maintenance will be handled by a working group of WICC (www.ph3c.org/Q) who are also the custodians of ICPC-2.

Interpretation of study results in relation to existing literature

To the best of our knowledge, this is the first systematic, internationally accepted description of professional aspects in GP/FM. An historic predecessor was FAMILI, a short-lived extension of MeSH, available only in print, operational since 1980, and governed by Jan Mc Whydney (a founding father of GP/FM) [16, 17]. This effort was abandoned in 1992 [30]. This approach to the doctor's professional context should not be confused with the contextual approach of the patient's universe, as developed by Schrans [31], who studied the elements of the patient's life context (ideas, concerns, and expectations) that influence his or her state of health, the health problems offered to the doctor, with substantial impact on the process of care.

One could argue that the development of a profession-specific taxonomy for professional aspects of GP/FM is unnecessary, as the aims for such a taxonomy could be met with existing resources, such Medical Subject Headings (MeSH). However, the experience gained during the development of the taxonomy has shown that many profession-specific concepts are not properly represented in MeSH. Furthermore, the learning curve for the correct use of a general and vast thesaurus is steeper than for the use of a simple profession-specific taxonomy.

The feasibility to translate the Q-Codes labelling and definitions in 8 languages illustrates that the taxonomy is robust to cultural diversity. The rapid acceptance of this new terminological tool and its integration in a number of recent applications is also an indication of face-validity. However, more in depth research is needed to underpin the validity of this taxonomy for its intended uses. Studying the impact of the use of the taxonomy in search and indexing efficiency of medical students, general practitioners, librarians, and automated systems alike will be the next steps.

The use of automatic ontology learning techniques, relying on term extraction and semantic relation extraction, should be explored to identify concepts and relations from literature and incorporate them in the taxonomy, and facilitate the maintenance process.

Supplemental online material.

Terminology in eight languages <http://3cgp.docpatient.net/terminology/>

Q-Codes in eight languages <http://3cgp.docpatient.net/tabular/>

Acknowledgement

To all contributors who names are listed on <http://3cgp.docpatient.net/contributors/>

Conflict of interest

None

Ethical issues

None

Sources

1. AAFP. Definition of Family Medicine 1984 [cited 2017]. Available from: <http://www.aafp.org/about/policies/all/family-medicine-definition.html>
2. Olesen F, Dickinson J, Hjortdahl P. General practice--time for a new definition. *BMJ*. 2000 Feb 05;320(7231):354-7. PubMed PMID: 10657333; PubMed Central PMCID: PMC1127143. eng.
3. Allen J, Gay B, Crebolder H, et al. The European Definition of General Practice / Family Medicine. WONCA Europe; 2011.
4. AFPI, editor The concept of Family Medicine. National Conference of Family Medicine and Primary Care (FPMC 2015); 2015.
5. Copolillo F, Urquiza R, Jure H. Avances de la Medicina Familiar y Comunitaria en Iberoamerica: V Cumbre de WONCA iberoamerica CIMF y la Carta de Quito. *Archivos de Medicina Familiar y General*. 2014;11(1):33-38.
6. Vicente VC, Muñoz SC. Tratado de medicina de familia y comunitaria. Médica Panamericana; 2012.
7. Gusso G, Lopes JMC. Tratado de Medicina de Família e Comunidade: 2 Volumes: Princípios, Formação e Prática. Artmed Editora; 2012.
8. Kochen MM. Allgemeinmedizin und Familienmedizin. Georg Thieme Verlag; 2017.
9. Druais P-L, Le Goaziou M-F. Médecine générale. Elsevier Masson; 2009.
10. Švab I, Katić M. International textbook of family medicine: The application EURACT teaching agenda. *Acta medica academica*. 2014;43(1).
11. RCPSC. CanMEDS: Better standards, better physicians, better care [cited 2017]. Available from: <http://www.royalcollege.ca/rcsite/canmeds/canmeds-framework-e>

12. Jamouille M, Resnick M, Vander Stichele R, et al. Analysis of definitions of general practice, family medicine, and primary health care: a terminological analysis. *BJGP Open*. 2017. doi: 10.3399/bjgpopen17X101049.
13. WICC. International Classification of Primary Care. 2nd, revised ed.: Oxford Medical publications; 2005.
14. Hummers-Pradier E. Which Abstracts Do Get Published? - Output Of German GP Research 1999-2003. *WONCA Europe*. Paris2007.
15. Van Royen P, Sandholzer H, Griffiths F, et al. Are presentations of abstracts at EGPRN meetings followed by publication? *Eur J Gen Pract*. 2010 Jun;16(2):100-5. doi: 10.3109/13814788.2010.482582. PubMed PMID: 20504264; eng.
16. WONCA. FAMILI. Family Medicine Literature Index. London, Ont.1980.
17. Challis EB. Famli. *Can Fam Physician*. 1980 Nov;26:1470. PubMed PMID: 21293717; PubMed Central PMCID: PMCPMC2383786. eng.
18. Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Research synthesis methods*. 2014 Sep;5(3):221-34. doi: 10.1002/jrsm.1106. PubMed PMID: 26052848; eng.
19. Lawrence A, Houghton J, Thomas J, et al. Where is the evidence? Realising the value of grey literature for public policy and practice: *Australian Policy Online*; 2012 [cited 2017]. Available from: <https://researchbank.rmit.edu.au/view/rmit:39652>
20. Cardillo E, Roumier J, Jamouille M, et al. Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology: A use case for Hearth Failure. *Terminology and Artificial Intelligence TIA 2013*. 2013.
21. Jamouille M, Dekeuster D. Workshop: Towards an e-archive for Wonca documents. *WONCA Europe conference*; Paris2007.
22. Kruschinski C, Lange M, Lionis C, et al. Themes and methods of research presented at European General Practice Research Network conferences. *Fam Pract*. 2010 Aug;27(4):459-67. doi: 10.1093/fampra/cmq023. PubMed PMID: 20427325; eng.
23. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med*. 1998 Nov;37(4-5):394-403. PubMed PMID: 9865037; PubMed Central PMCID: PMCPMC3415631. eng.
24. Jamouille M, Grosjean J, Resnick M, et al. A Terminology in General Practice/Family Medicine to Represent Non-Clinical Aspects for Various Usages: The Q-Codes. *Stud Health Technol Inform*. 2017;235:471-475. PubMed PMID: 28423837; eng.

25. Grosjean J, Griffon N, Dahamma B, editors. Multiterminology cross-lingual model to create the European Health Terminology / Ontology Portal. 9th International Conference on Terminology and Artificial Intelligence; 2011 8-10 nov; Paris.
26. Jamouille M. Towards Knowledge management in General Practice & Family Medicine; Guide to the indexing of master thesis. 2017.
27. Resnick MP, Santana F, de Araujo Novaes M, et al. Representing second opinion requests from primary care within the Brazilian tele-health program: international classification of primary care, second edition. *Stud Health Technol Inform.* 2013;192:1190. PubMed PMID: 23920964; eng.
28. Cabot C, Soualmia LF, Grosjean J, et al. Evaluation of the Terminology Coverage in the French Corpus LiSSa. *Stud Health Technol Inform.* 2017;235:126-130. PubMed PMID: 28423768; eng.
29. Ittoo A, Bouma G. Term extraction from sparse, ungrammatical domain-specific documents. *Expert Systems with Applications.* 2013;40(7):2530-2540.
30. Dunikowski L. Family Medicine Literature Index. FAMILI. The college of Family Physicians, Canada; 1992.
31. Schrans D, Avonts D, Christiaens T, et al. The search for person-related information in general practice: a qualitative study. *Fam Pract.* 2016 Feb;33(1):95-9. doi: 10.1093/fampra/cmz099. PubMed PMID: 26787770; eng.