Computational Terminology in eHealth

Federica Vezzani[1][0000−0003−2240−6127] and Giorgio Maria Di Nunzio[2][0000−0001−9709−6392]

[1] Dept. of Linguistic and Literary Studies – University of Padua
[2] Dept. of Information Engineering – University of Padua
federica.vezzani@phd.unipd.it, giorgiomaria.dinunzio@unipd.it

Abstract. In this paper, we present a methodology for the development of a new eHealth resource in the context of Computational Terminology. This resource, named TriMED, is a digital library of terminological records designed to satisfy the information needs of different categories of users within the healthcare field: patients, language professionals and physicians. TriMED offers a wide range of information for the purpose of simplification of medical language in terms of understandability and readability. Finally, we present two applications of our resource in order to conduct different types of studies in particular in Information Retrieval and Literature Analysis.

1 Introduction

Computational Terminology (CT) is a recent field of study gathering the interest of researchers who have experienced the need to improve communication, or to access domain-related information [3]. CT is closely related to the organization and management of linguistic data for different purposes and, for this reason, several types of terminological resources have been implemented in order to satisfy these needs, such as specialized dictionaries, terminological databases and glossaries. Moreover, scientific needs in fast growing domains (such as biomedicine, chemistry and ecology) and the overwhelming amount of textual data published daily demand that terminology is acquired and managed systematically and automatically.

In the medical field, the need to manage data is increasingly evident, in particular in the context of health practice through the support of mobile and portable tools and in the physician-patient interactions. According to GSMA,[3] there are four major customers in the eHealth market: health-care providers, payers (both public and private health-care insurers), governments, and Health-care consumers, each of whom have different priorities and needs. Focusing on health-care consumers, the main priority is to manage one's own health in order to have access to reliable and comprehensible medical information. For this reason, there are freely available resources [35, 18], as for example *Everyday Health*[4]

[3] https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/05/Role-and-Value-of-MNOs-in-eHealth1.pdf
[4] https://www.everydayhealth.com/

which offers personalized health advice, tools, and communities, or *Medscape*,[5] offering medical information for specialists, physicians, and industry professionals. These tools are based on the need for effective communication between various actors and the transmission of information in a clear and understandable way. Therefore, terminology is an important issue to be considered because medical language is often characterized by a complex lexicon difficult to understand. Many studies [6, 7, 30, 14] point out problems related to medical terminology such as semantic ambiguity, incorrect use of suffixes, archaism maintaining, redundancy in the formation of compounds, and etymological inconsistencies. As a result, patients and in general non-experts in medicine are often exposed to medical terms that can be semantically complex and hardly understandable [38].

Public libraries have a long history of providing community outreach programs and services to their diverse user population, including aiding access to reliable consumer health information and electronic health resources and offering health-information literacy programs [39]. Acquiring this knowledge, from public libraries as well as the availability of free and publicly available online resources, and organizing it into a digital library would ease the problem of health literacy defined as 'the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions' [16]. Such digital library would also facilitate medical practitioners in the discovery of novel treatments or diseases. However, building an ontology for a medical digital library is not a trivial task since it requires a significant amount of time and manual effort.

In this paper, we describe the motivations and the methodology behind the development of TriMED, an eHealth resource based on terminological and linguistic features [41]. This terminological database responds to the information needs of different categories of users within the health-care field: patients, language professionals and physicians. This tool is composed of multilingual terminological records offering a wide range of information depending on the user identification in three languages: English, Italian and French. This kind of terminological resource and, in particular, the model of terminological record designed for this tool, can be exploited in order to conduct different types of studies in a variety of research areas: not only medical terminology analysis and simplification, but also identification and retrieval of relevant medical documents and linguistic analysis of medical terms in literature.

The reminder of this paper is organized as follows: in Section 2, we present the state-of-the-art from two points of view in the context of eHealth: CT and Digital Library (DL). In Section 3, we describe the requirements of the terminological eHealth resource focusing on the definition of a medical terminological record and its structure. Then, in Section 4, we show the applications of these terminological records in two different domains: Information Retrieval and Literature analysis. Finally in Section 5, we give our conclusions and some hints on future works.

[5] https://www.medscape.com/

2 Related Works

In this section, we present the state-of-the-art relating to the eHealth field from two different perspectives: Computational Terminology and Digital Libraries.

2.1 Computational Terminology

The medical field gathers people of different social statuses, such as students, pharmacists, biologists, nurses and mainly doctors and patients [34]. Despite their different levels of expertise, these people need to interact and understand each other; but, unfortunately, the communication is not always easy and effective [31, 15, 24]. The recent contributions of CT aim to i) the simplification of medical texts in terms of readability and understandability and to ii) the implementation of resources and applications in order to facilitate patient-doctor dialogue in situations of medical diagnosis. Text simplification is closely linked to the readability studies [12], the purpose of which is to address the ease with which a document can be read and understood by people. In medical texts, one source of difficulty may be due to the specific and specialized notions that are used. Indeed, the medical field conveys very specific and often opaque notions (e.g., *myocardial infarction, cholecystectomy, abdominal strangulated hernia, galactose urine*), that are difficult to understand by lay people. Numerous studies address these issues from different perspectives and following different computational approaches for the analysis of medical terminology such as:

1. the automatic identification of specialized and non-specialized terms;
2. the identification of equivalents lay terms;
3. the identification of the level of technicality;
4. the realization of terminological resources.

In [42], the authors propose an automatic method in order to distinguish between specialized and non-specialized occurrences of verbs in medical corpora. This method uses a contrastive automatic analysis of such verbs based on the semantic annotation of the verbs nominal co-occurrents. The results show that some verbs regularly co-occur with specialized terms in a given context or corpus, while the same verbs mostly occur with general language words in a different corpus. These kinds of observations fit in the context of readability: verbs which co-occur frequently with specialized terms can be considered as sources of reading difficulties for non-experts in the medical field. Other approaches [2] try to assess the level of technicality of a term through the combination of learning-to-rank techniques with statistical and linguistic features, that is the specialization degree to each of the entries given in a list of synonym terms. In [22], the authors propose a method for the classification of term technicality considering term variation in the medical field. Authors focused on the different degree of specialization of a term and term variation in the characterization of clinical sublanguages by analyzing the language used in a corpus of Belgian Electronic Health Records (EHRs). Each sections of these records vary systematically with

regard to their lexical, terminological and semantic composition, as well as their potential for term variation, so that they distinguished between vernacular (e.g. 'buik', 'belly') and specialized (e.g. 'abdomen') terms.

The study presented in [20] highlights the need for reliable terminological resources in this field. This study presents how to build specific lexicon in which the words are rated according to whether they are *understandable* or *non-understandable*. French medical words have been rated by human annotators on a scale with three positions: *I can understand, I am not sure, I cannot understand*. In this way, the *I cannot understand* category may be associated with specialized words, while the *I can understand* category may be associated with non-specialized words.

At present, one of the most reliable terminological resource providing patient-oriented information is SNOMED CT,[6] that is an international standard for the interoperability in Digital Health; it can be used to represent clinically relevant information consistently and to support the development of comprehensive high-quality clinical content in health records. Another open access and collaborative resource is the Consumer Health Vocabulary (CHV) [7] developed by the Department of Biomedical Informatics at the University of Utah. The purpose of this project is to translate technical terms into popular language by linking informal and common health words and expressions to technical terms used by health professionals. Such kind of resources may help both in the detection of parts of medical documents containing complex and non-understandable terms, and in the patient-physician communication by highlighting such terms that should be explained to patients in order to make the communication more successful and easy.

2.2 Digital Libraries

The existence of large digital libraries containing medical documents has given the opportunity to develop text mining techniques for the automatic extraction of knowledge from unstructured text. In [27], the authors identified typical cue-phrases and structural indicators that introduce definitions of medical terms through an analysis of a set of consumer-oriented medical articles. Building an ontology of medical concepts requires times and effort; the authors of [28] propose a method to reduce these costs by starting with a small (or seed) ontology and enrich it with concepts and semantic relations acquired from medical abstracts. Concept Mapper [45] combines an ontology together with an automatic generated thesaurus based on document term co-occurrences in order to provide users with suggestions for query expansion.

Patients usually experience difficulty when they use health information retrieval systems due to the vocabulary gap between their request and the corresponding controlled vocabulary terms used to index the health information retrieval system [44]. It is therefore important to evaluate the difficulty of a

[6] `https://www.snomed.org/`
[7] `http://consumerhealthvocab.chpc.utah.edu/CHVwiki/`

term in order to provide recommended terms for an effective query formulation. In [29], the authors propose a metric to label text difficulty levels and focus on a lexical measure about term familiarity. The work presented in [25] analyzes user generated terms and compare them to those generated by professionals. This study also show that most health information users have low levels of health literacy which means that users are not able to access health information effectively and understand the information.

There are examples of successful systems that support users in the search of medical digital libraries. PERSIVAL (PErsonalized Retrieval and Summarization over Images, Video and Language) is a medical digital library which provides access to literature that is clinically relevant to the patient under their care at the point of patient care [32]. This information includes the medical history, laboratory results, procedures performed and diagnoses, which can be used to pinpoint articles that can provide the physician with the latest results relevant to the patient under care. MedSearch is a retrieval system for medical literature based on a Semantic Similarity Retrieval Model (SSRM) which suggests discovering semantically similar terms in documents and queries using term ontologies [23]. The SINAMED and ISIS projects focused on information access on patient clinical records and related scientific documentation [4]. These projects integrate automatic text summarization and categorization algorithms to improve access to bilingual information in the biomedical domain. In this way, the combination of a medical information system with the electronic clinical record would help doctors to take decisions, to decrease the mistakes and the clinical variability and to increase the patient's safety.

3 TriMED: A Terminological eHealth Resource

In this context, we are developing a new eHealth resource in order to tackle the problem of the complexity of medical terminology by considering different level of communications. The multilingual resource is named TriMED [41]: it is a database collecting terminological records compiled over a set of technical terms manually extracted by experts in linguistics and terminology from a corpus of documents. These documents concerning the oncology field and, in particular, breast cancer treatments. are selected from specialized online magazine reviews based with the highest impact factor value, such as "Breast Cancer research and treatment"[8] for English language, from national associations websites such as "AIMaC - Associazione Italiana Malati di Cancro"[9] for Italian, and the "Association Francophone pour les Soins Oncologiques de Support (AFSOS)"[10] for French.

Terminological records are the core of our methodology and provide different kinds of information depending on the user identification.

[8] http://www.springer.com/medicine/oncology/journal/10549
[9] https://www.aimac.it
[10] http://www.afsos.org

3.1 Users

We have identified three categories of people that are mostly affected by the complexity of medical language and they can benefit from the use of this resource: patients, language professionals (translators and interpreters) and physicians.

Patients: Patients, or more in general lay people, find a considerable difficulty in understanding information, both oral and written, about their own health [13, 21, 26, 1]. As a consequence, they need to understand medical technical terms using their correspondent in the popular language or using an appropriately calibrated language for the communication to be effective. For this reason, they are the first category of users who can benefit from the use of TriMED because this eHealth resource provides the equivalent of the technical term in the popular language, that is the term most frequently used (as for example *fever* for *pyrexia*) and provides an informative definition respecting a non-specialized level of register.

Translators and Interpreters: There are several language related factors that need to be taken into consideration in the correct transmission of health information [18]. The need to use interpreters in the health-care domain becomes increasingly evident especially referring to the immigration phenomena. TriMED is designed to support 'language professionals', such as translators and interpreters who work in emergency situations. TriMED offers terminological records providing the translation into three languages (English, French and Italian) of the technical term and all the linguistically relevant information for the process of decoding and transcoding it in its oral and written forms. In this way, professionals can immediately consult a reliable translation of the term in an easy and quick way. The objective is indeed to try to reduce the time of terminological research by offering a regularly updated and reliable digital resource.

Physicians: The international scale release of medical knowledge implies that most of the scientific texts are produced in English [33]. In terms of spreading new health care protocols and scientific discoveries, language could be a barrier to service transactions among medical specialists speaking different languages because perfect knowledge and mastery of the foreign language is not an expected outcome. In order to overcome these language barriers and to satisfy the peer-to-peer communication, TriMED offers the possibility to consult the translation of the technical term respecting the specialized linguistic register.

The definition of the three categories of users allowed us to design the structure of TriMED as a user-friendly resource. A user can select one of the three category and then access the related information need. Currently, the eHealth resource contains about 328 terminological records that, at present time, are under a review process performed by terminologists and translators. After the review process, records will be openly available; a demo of the application with some examples of terminological records in English an French is available online. [11]

[11] https://gmdn.shinyapps.io/TriMED/

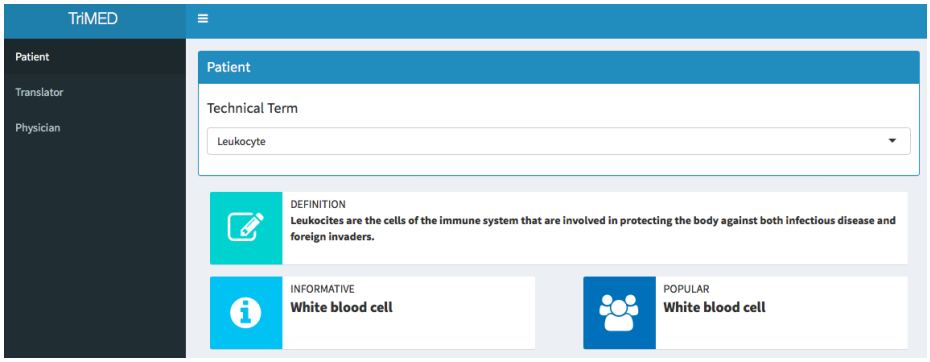

Fig. 1: Patient visualization for the technical term *Leukocyte* and its equivalent in popular language *White Blood Cell*.

3.2 Terminological Record

Our methodology is based on the formulation of a new model of terminological record which is designed to satisfy the information needs of the above mentioned categories of users both from inter and intra-linguistic viewpoints. Terminological records are commonly used in terminology and linguistics as a tool for the collection of linguistic data referring to a specific concept [19]. We designed the new records in order to extend the SNOMED CT records and Consumer Health Vocabulary (CHV) records. In particular, we link our terminological records to the SNOMED CT concepts by means of the SNOMED concept identifiers; then, on the basis of the CHV we propose not only the technical term and its equivalent in popular language, but also intra and inter-linguistic medical information in French and Italian which are not currently supported by those terminological resources. In Fig. 1 we present the terminological record visualization for a patient, while in Fig. 2 the linguistic analysis for technical terms in source and target languages are shown for language professionals. All the information provided are necessary for the technical-scientific translator in order to decode (interpret) and then transcode (transfer) the meaning of a technical term and help the professional choosing the correct translating candidate. TriMED terminological record is structured around four axes of analysis of the technical term:

1. Formal features;
2. Semantics;
3. Corpus;
4. References.

Regarding the formal and lexical framework of the term, we provide information such as: gender, spelling, pronunciation in the International Phonetic Alphabet (IPA) and other information about the etymology, such as derivation

and composition of the term. In addition, we propose the spelling variant and the related acronyms which are currently used in medical language. Finally, based on the WordNet resource,[12] the record contains all the nouns, verbs, adjectives, and adverbs deriving from the analyzed term and which fall into the same semantic sphere.

The second section focuses on the semantic features of the term. First, we propose a definition extracted from reliable resources such as Merriam-Webster Medical Dictionary,[13] MediLexicon[14] especially for acronyms and abbreviations, TERMIUM Plus,[15] or "Enciclopedia Salute" from the Italian Ministry of Health.[16] In addition, we provide the semic analysis of the term [36] that is a methodology used in compositional semantics in order to decompose the meaning of technical terms (lexematic or morphological unity) into minimal units of meaning: the semes. Moreover, in order to evaluate the semantic behavior of a term, we collect the phraseology of the term by considering cases of collocations [17] and colligations [37]. Finally, we provide the synonymic variants of the term: in this way, we categorize terms and their semantic relations.

In the corpus section, we provide all specialized contexts where technical terms have been extracted and then we proceed through the identification of the domain and the register of communication of the term (popular, slang, familiar, current or standard and specialized). The term and its definition, therefore, take on meaning when they are connected to a specific domain: in our analysis, we identify the domain and subdomains of the text (such as surgery, pathology, pharmacology, etc). Finally, since all of this information has been extracted from different sources, we provide references to each source.

4 Applications

The TriMED application was realized in R and the user interface implemented with the Shiny R package [5]. The 328 records were loaded by experts in linguistics and terminology and the current structure of each record follows the tidy data approach [43] that uses dataframes as the main unit of data storage and analysis.

In this Section, we present two different applications of the TriMED terminological records; our aim is to show that a structured collection of terminological data can be useful and effective in order to conduct researches related to the medical field for different domain of interest: Information Retrieval and Literature Analysis.

[12] https://wordnet.princeton.edu/
[13] https://www.merriam-webster.com/
[14] https://www.medilexicon.com/
[15] https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html
[16] http://www.salute.gov.it/portale/home.html

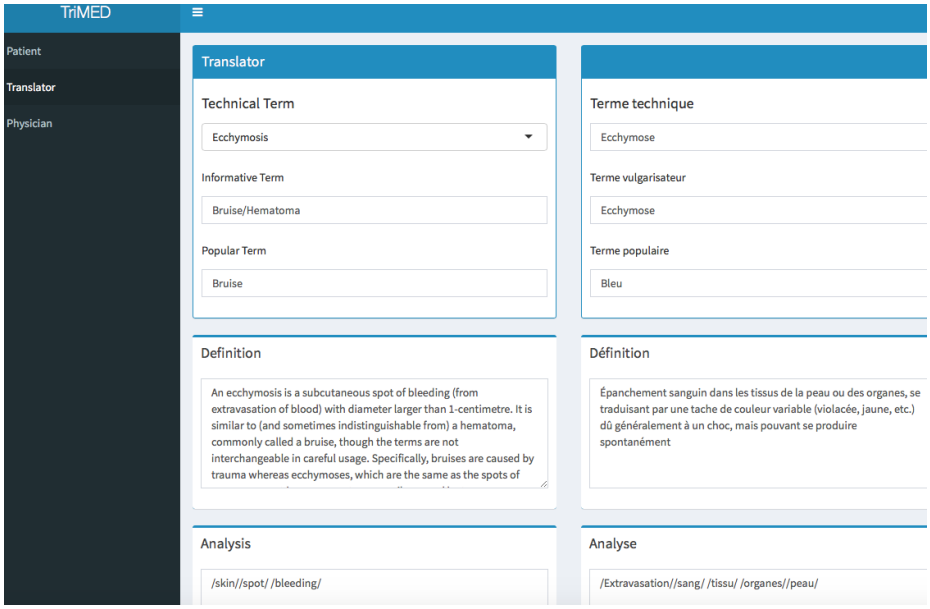

Fig. 2: Translator visualization for the technical term in the source language (EN) *Ecchymosis* and its equivalent in the target language (FR) *Ecchymose*.

4.1 Information Retrieval

Our first experiment in the application of detailed terminological records in order to improve the description of the medical lexicon was the participation to the Cross-Language Evaluation Forum eHealth Task on "Technologically Assisted Reviews in Empirical Medicine" [9, 11, 10]. The main goal of the task was to find the most relevant medical publications by means of a manual query rewriting approach. Our query reformulation methodology was based on a purely terminological approach, proposing variants based on a detailed linguistic analysis of the technical terms presented in the initial query. This starting point allowed us to see the usefulness of a digital terminological and multilingual resource for the retrieval of medical information [8].

In these experiments, we asked two experts in linguistics to rewrite the initial query provided with the experimental collection through a terminological methodology based on the following steps:

1. Identification of technical terms;
2. Manual extraction of technical terms;
3. Linguistic and semantic analysis;
4. Formulation of terminological records;
5. Query rewriting.

Variant	Query
information need	First rank symptoms for schizophrenia
expert keywords	diagnosis, diagnostic, first rank symptoms, symptom, schizophrenia, FRS, international pilot study, IPSS, schneider, schneiderian, schizophrenics, non-schneiderian
expert readable	Diagnostic accuracy of one or multiple FRS for diagnosing schizophrenia as a psychotic disorder

Table 1: Query reformulation: keyword and readable variants

After the extraction of technical terms, the two linguists started to formulate TriMED terminological records in order to write two variants of the original query. The first variant was written with the aim of creating a list of keywords resulting from the semic analysis, that is the study of meaning in linguistic units, of the technical terms extracted from the initial query. The second variant was written with the aim of reformulating the information need into a humanly readable sentence using alternative terms such as synonyms, orthographic variants, related forms and/or acronyms. The two experts worked independently from each other by following this structured linguistic methodology and focusing on different terminological aspects. We name these two experiments with "keywords" and "readable": in Table 1, an example of query reformulation is shown for the initial query about *First rank symptoms for schizophrenia*.

This approach in combination with Technology-Assisted Review system allowed us to achieve a perfect recall on almost all the topics provided for the CLEF task [8]. Therefore, the linguistic approach based on terminological record has contributed to an effective and efficient reformulation for the retrieval of the most relevant documents for the research.

4.2 Literature Analysis

In this section, we present the second application of the new model of terminological record for linguistic analysis in the literature domain [40]. This is an innovative approach combing quantitative and qualitative analyses in order to study medical terminology in literary texts. In particular, we focused on the works of Conan Doyle and our case study was the entire collection of adventures of Sherlock Holmes, starting from 'A Study in Scarlet' (1887) to 'The Casebook of Sherlock Holmes' (1927), freely available on the Project Gutenberg website.[17] We initially proceeded with the semi-automatic extraction of 98 English technical terms, as well as their collocations by means of the tidytext R package for text analysis.[18] After the identification of medical terms, we proceeded with the formulation of TriMED terminological records by focusing on different linguistic aspects of such literary texts such as:

1. The level of technicality of a term, dividing popular terms from technical ones such as 'St. Vitus's dance' for *Chorea minor*, 'Heart Disease' for *Cardiopathy* or 'Nosebleed' for *Epistaxis*. In this way, we could analyze changes

[17] https://www.gutenberg.org
[18] https://cran.r-project.org/web/packages/tidytext/

in the linguistic register by focusing on the diastratic variation resulting from the specialized-popular dualism in order to bridge the gaps between various registers;
2. The semantic behavior of a term in a diachronic sense, evaluating how a specialized term can change its meaning over time and by comparing the use of the term in the past and its current use, such as 'Consumption' used to indicate the process of general decay of the organism in place of the current *cachexia* or *wasting syndrome*. Likewise, we can evaluate the disuse of a term as 'brain fever' used until the first half of the nineteenth century to indicate the association between an irregular set of neurological symptoms;
3. The syntactic behavior of technical terms in the literary corpus through the analysis of collocations, as a sequence of terms that frequently co-occur, as for example *asphyxia by confinement* vs *asphyxia in a confined space*.

This innovative study constitutes a first attempt aiming to show that TriMED and its terminological records are valid digital supports not only for specialized documents, but also for literary corpora. Moreover, this kind of study led us to evaluate different features of medical terminology, in particular, by considering the diachronic variation of the technical term consisting in a formal or semantic change over time.

5 Conclusions and Future Works

In this paper, we described the development of an eHealth resource, named TriMED, that is a digital library of multilingual terminological records designed to satisfy the information needs of different categories of users within the health-care field. This resource provides information at multiple levels of linguistic register with the main purpose of simplification of medical language in terms of understandability and readability. Indeed, the new model of terminological record we propose allows to provide different information in order to i) satisfy the peer-to-peer communication between medical experts, ii) facilitate the comprehension of medical information by patients, and iii) provide a regularly updated resource for scientific translators. The data as well as the source code of the application will be available under OSI-approved licenses. [19]

Moreover, we presented two different applications in order to show that a structured collection of terminological data can be effective in multiple fields of study related to the medical domain for different purposes, such as Information Retrieval and Literature Analysis. TriMED terminological record proved to be a useful support tool for the identification of the most relevant medical publications according to a specific topic and for the linguistic analysis of medical terminology in Conan Doyle literary works.

As future work, in order to ease the manual work of the terminologists and translators, we are studying a method to create automatically a draft version of a record by reusing pieces of information already available in the online resources.

[19] `https://opensource.org/licenses`

References

1. Ali, T., Schramm, D., Sokolova, M., Inkpen, D.: Can i hear you? sentiment analysis on medical forums. In IJCNLP, pages 667-673. (2013)
2. Bouamor, D., Llanos, L.C., Ligozat, A., Rosset, S., Zweigenbaum, P.: Transfer-based learning-to-rank assessment of medical term technicality. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016. (2016)
3. Bourigault, D., Jacquemin, C., L'Homme, M.: Recent Advances in Computational Terminology. Natural language processing, J. Benjamins Publishing Company (2001), https://books.google.it/books?id=0eX8HtA7mQIC
4. de Buenaga Rodríguez, M., López, M.J.M., Gachet, D., Vázquez, J.M.: The SINAMED and ISIS projects: Applying text mining techniques to improve access to a medical digital library. In: Gonzalo, J., Thanos, C., Verdejo, M.F., Carrasco, R.C. (eds.) Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Lecture Notes in Computer Science, vol. 4172, pp. 548–551. Springer (2006). https://doi.org/10.1007/11863878_65, https://doi.org/10.1007/11863878_65
5. Chang, W.: Shiny: Web Application Framework for R (2015), http://CRAN.R-project.org/package=shiny, r package version 0.11
6. Cimino, J.J.: Terminology tools: state of the art and practical lessons. Methods of information in medicine **40**(04), 298–306 (2001)
7. Detmar, S.B., Muller, M.J., Schornagel, J.H., Wever, L.D., Aaronson, N.K.: Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. Jama **288**(23), 3027–3034 (2002)
8. Di Nunzio, G.M.: A study of an automatic stopping strategy for technologically assisted medical reviews. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) Advances in Information Retrieval - 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings. Lecture Notes in Computer Science, vol. 10772, pp. 672–677. Springer (2018). https://doi.org/10.1007/978-3-319-76941-7_61, https://doi.org/10.1007/978-3-319-76941-7_61
9. Di Nunzio, G.M., Beghini, F., Vezzani, F., Henrot, G.: An interactive two-dimensional approach to query aspects rewriting in systematic reviews. IMS unipd at CLEF ehealth task 2. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017. (2017), http://ceur-ws.org/Vol-1866/paper_119.pdf
10. Di Nunzio, G.M., Ciuffreda, G., Vezzani, F.: Interactive sampling for systematic reviews. IMS unipd at CLEF 2018 ehealth task 2. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018. (2018)
11. Di Nunzio, G.M., Vezzani, F.: Using R markdown for replicable experiments in evidence based medicine. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings. pp. 28–39 (2018)
12. DuBay, W.H.: The principles of readability. Online Submission (2004)
13. Elhadad, N., Sutaria, K.: Mining a lexicon of technical terms and lay equivalents. In Proc BioNLP Workshop, pages 49-56. ACL. (2007)
14. Epstein, R.M., Franks, P., Fiscella, K., Shields, C.G., Meldrum, S.C., Kravitz, R.L., Duberstein, P.R.: Measuring patient-centered communication in patient–physician

consultations: theoretical and practical issues. Social science & medicine **61**(7), 1516–1528 (2005)
15. Eysenbach, G.: Poverty, human development, and the role of ehealth. Journal of medical Internet research **9**(4) (2007)
16. Ferguson, L.A., Pawlak, R.: Health literacy: The road to improved health outcomes. The Journal for Nurse Practitioners **7**(2), 123 – 129 (2011). https://doi.org/https://doi.org/10.1016/j.nurpra.2010.11.020, http://www.sciencedirect.com/science/article/pii/S1555415510005519
17. Firth, J.R.: A synopsis of linguistic theory, 1930-1955. Studies in linguistic analysis (1957)
18. Gibbons, M.C.: eHealth Solutions for Healthcare Disparities. Springer Publishing Company, Incorporated, 1st edn. (2007)
19. Gouadec, D.: Terminologie: constitution des données. AFNOR gestion, AFNOR (1990), https://books.google.it/books?id=cRrnAAAAMAAJ
20. Grabar, N., Hamon, T.: A large rated lexicon with french medical words. In: LREC (Language Resources and Evaluation Conference) 2016 (2016)
21. Grabar, N., Van Zyk, I., De La Harpe, R., Hamon, T.: The comprehension of medical words. In: Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5. pp. 334–342. BIOSTEC 2014, SCITEPRESS - Science and Technology Publications, Lda, Portugal (2014). https://doi.org/10.5220/0004803803340342, https://doi.org/10.5220/0004803803340342
22. Grön, L., Bertels, A.: Clinical sublanguages: Vocabulary structure and its impact on term weighting. Terminology **24**(1) (2018)
23. Hliaoutakis, A., Varelas, G., Petrakis, E.G.M., Milios, E.E.: Medsearch: A retrieval system for medical information based on semantic similarity. In: Gonzalo, J., Thanos, C., Verdejo, M.F., Carrasco, R.C. (eds.) Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Lecture Notes in Computer Science, vol. 4172, pp. 512–515. Springer (2006). https://doi.org/10.1007/11863878_56, https://doi.org/10.1007/11863878_56
24. Jimison, H., Gorman, P., Woods, S., Nygren, P., Walker, M., Norris, S., Hersh, W.: Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) **175**, 1–1422 (2008)
25. Joo, S., Choi, Y.: Content analysis of social tags generated by health consumers. In: II, P.L.B., Allard, S., Mercer, H., Beck, M., Cunningham, S.J., Goh, D.H., Henry, G. (eds.) Proceedings of the 15th ACM/IEEE-CE Joint Conference on Digital Libraries, Knoxville, TN, USA, June 21-25, 2015. pp. 249–250. ACM (2015). https://doi.org/10.1145/2756406.2756959, http://doi.acm.org/10.1145/2756406.2756959
26. Keselman, A., Tse, T., Crowell, J., Browne, A., Ngo, L., Zeng, Q.: Assessing consumer health vocabulary familiarity: an exploratory study. J Med Internet Res, 9(1): e5. (2007)
27. Klavans, J., Muresan, S.: Evaluation of DEFINDER: a system to mine definitions from consumer-oriented medical text. In: ACM/IEEE Joint Conference on Digital Libraries, JCDL 2001, Roanoke, Virginia, USA, June 24-28, 2001, Proceedings. pp. 201–202. ACM (2001). https://doi.org/10.1145/379437.379488, http://doi.acm.org/10.1145/379437.379488
28. Lee, C., Na, J., Khoo, C.S.G.: Towards ontology enrichment with treatment relations extracted from medical abstracts. In: Sugimoto, S., Hunter, J., Rauber, A.,

Morishima, A. (eds.) Digital Libraries: Achievements, Challenges and Opportunities, 9th International Conference on Asian Digital Libraries, ICADL 2006, Kyoto, Japan, November 27-30, 2006, Proceedings. Lecture Notes in Computer Science, vol. 4312, pp. 419–428. Springer (2006). https://doi.org/10.1007/11931584_45, https://doi.org/10.1007/11931584_45
29. Leroy, G., Endicott, J.E.: Term familiarity to indicate perceived and actual difficulty of text in medical digital libraries. In: Xing, C., Crestani, F., Rauber, A. (eds.) Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation - 13th International Conference on Asia-Pacific Digital Libraries, ICADL 2011, Beijing, China, October 24-27, 2011. Proceedings. Lecture Notes in Computer Science, vol. 7008, pp. 307–310. Springer (2011). https://doi.org/10.1007/978-3-642-24826-9_38, https://doi.org/10.1007/978-3-642-24826-9_38
30. M., R.: La terminologie médicale et ses problèmes. Tribuna, Vol. IV, n. 12 (2003)
31. McCray, A.: Promoting health literacy. J of Am Med Infor Ass, 12:152-163. (2005)
32. McKeown, K.R., Elhadad, N., Hatzivassiloglou, V.: Leveraging a common representation for personalized search and summarization in a medical digital library. In: ACM/IEEE 2003 Joint Conference on Digital Libraries (JCDL 2003), 27-31 May 2003, Houston, Texas, USA, Proceedings. p. 159. IEEE Computer Society (2003). https://doi.org/10.1109/JCDL.2003.1204856, https://doi.org/10.1109/JCDL.2003.1204856
33. Neveol, A., Grosjean, J., Darmoni, S., Zweigenbaum, P.: Language resources for french in the biomedical domain. In: Proceedings of the LREC 2014. ELRA, Reykjavik, Iceland (May 2014)
34. Parker, R.M., Williams, M.V., Weiss, B.D., Baker, D.W., Davis, T.C., Doak, C.C., Doak, L.G., Hein, K., Meade, C.D., Nurss, J., et al.: Health literacy-report of the council on scientific affairs. Jama-Journal of the American Medical Association **281**(6), 552–557 (1999)
35. Ranck, J.: Disruptive Cooperation in Digital Health. Springer International Publishing (2016), https://books.google.it/books?id=mGfVjwEACAAJ
36. Rastier, F.: Sémantique interprétative. Formes sémiotiques, Presses universitaires de France (1987), https://books.google.it/books?id=BnuyAAAAIAAJ
37. Sinclair, J.: Reading concordances: An introduction. Pearson Longman (2003)
38. Street Jr, R.L.: Information-giving in medical consultations: the influence of patients' communicative styles and personal characteristics. Social science & medicine **32**(5), 541–548 (1991)
39. Tu-Keefner, F.: The value of public libraries during a major flooding: - how digital resources can enhance health and disaster preparedness in local communities. In: Morishima, A., Rauber, A., Liew, C.L. (eds.) Digital Libraries: Knowledge, Information, and Data in an Open Access Society - 18th International Conference on Asia-Pacific Digital Libraries, ICADL 2016, Tsukuba, Japan, December 7-9, 2016, Proceedings. Lecture Notes in Computer Science, vol. 10075, pp. 10–15. Springer (2016). https://doi.org/10.1007/978-3-319-49304-6_2, https://doi.org/10.1007/978-3-319-49304-6_2
40. Vezzani, F., Di Nunzio, G.M., Henrot, G.: (not so) elementary, my dear watson! In: Coronato, R., Gesuato, S. (eds.) Filling Gaps, Building Bridges: Qualitative and Quantitative Approaches to the Study of Literature. Padova (June 2018)
41. Vezzani, F., Di Nunzio, G.M., Henrot, G.: Trimed: A multilingual terminological database. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation LREC 2018, Miyazaky, Japan, May 7-12, 2018. (2018 In press)

42. Wandji Tchami, O., Grabar, N.: Towards automatic distinction between specialized and non-specialized occurrences of verbs in medical corpora. In: Proceedings of the 4th International Workshop on Computational Terminology (Computerm). pp. 114–124. Association for Computational Linguistics and Dublin City University, Dublin, Ireland (August 2014), http://www.aclweb.org/anthology/W14-4814
43. Wickham, H.: Tidy data. The Journal of Statistical Software **59** (2014), http://www.jstatsoft.org/v59/i10/
44. You, S., DesArmo, J., Mu, X., Lee, S., Neal, J.C.: Visualized related topics (VRT) system for health information retrieval. In: IEEE/ACM Joint Conference on Digital Libraries, JCDL 2014, London, United Kingdom, September 8-12, 2014. pp. 429–430. IEEE Computer Society (2014). https://doi.org/10.1109/JCDL.2014.6970209, https://doi.org/10.1109/JCDL.2014.6970209
45. Zhu, B., Leroy, G., Chen, H., Chen, Y.: Medtextus: an intelligent web-based medical meta-search system. In: ACM/IEEE Joint Conference on Digital Libraries, JCDL 2002, Portland, Oregon, USA, June 14-18, 2002, Proceedings. p. 386. ACM (2002). https://doi.org/10.1145/544220.544333, http://doi.acm.org/10.1145/544220.544333