Exploring semantic connections through a thesaurus in the earth observation domain[1]

Erika Pasceri
University of Calabria, Italy
Claudia Lanza
University of Calabria, Italy
Anna Perri
University of Calabria, Italy

Abstract

This paper presents the creation of a thesaurus for the Earth Observation (EO) systems domain which also includes terms representing the potential clinical effects caused by the pollutants. The goal is to provide a terminological means of support for professionals and public audiences in understanding specialized terms and their semantic connections with other representatives of this area of study (Bernier-Colborne 2012). The paper presents, in the first instance, the steps carried out in compiling the source corpus from which the specialized information about the contaminant substances has been retrieved; subsequently, the work will address the extraction of the specific terminology with the aid of pre-trained semantic software as well as the identification of a sub-methodological procedure to improve term configuration, i.e., semantic relation configuration by isolating the domain-specific connective verbal constructions (Auger 2008) and exploiting the textual contextualization to broaden the conceptual system and enhance the semantic accuracy.

1.0 Introduction

To achieve interoperability and avoid semantic ambiguity, the healthcare community has largely identified the need to develop standard codes for commonly used concepts and terms in healthcare delivery. Code-based vocabularies, terminologies, and classifications have been thoroughly developed: the standardization task in this area involves the development (or the constant update) of standard code-sets for generally used concepts, terms,

1 Authors have equally contributed to this work, however Erika Pasceri particularly focused on "Introduction", "The Method", "Corpus compilation" and "Text Processing"; Claudia Lanza focused on "Thesaurus construction", "Semantic interoperability" and "Results"; Anna Perri focused on "Discussion" and "Conclusion".

disease names, procedures, entity names, laboratory tests, observations, devices, clinical findings, pharmaceutical products, organisms, etc. The most relevant application of this kind of information is to assist in the meaningful interpretation of healthcare information exchanged across healthcare systems (Sinha et al. 2012). The more information we share, the more knowledge we achieve and that is particularly true in using knowledge organization systems that embed cross information useful for those systems created to support decision making processes.

The thesaurus is one of the most common Knowledge Organization Systems (KOSs) and it has been designed to improve information retrieval and to reduce the ambiguity of natural language in describing items for searching goals. The main purpose of thesauri is to provide a form of organization with respect to the subject into logical, semantic divisions as well as to index document collections in order to retrieve them. Moreover, basic controlled vocabularies have been developed to reduce ambiguity also by defining terms with scopes notes, while more complex vocabularies provide a set of synonyms for each concept (Tudhope and Binding 2007). The final goal of such a system is to guide final users – through a semantic structure – to someone else's terminology that in specialized domains is required to be extremely specific.

To date, the health data ecosystem comprises a wide array of complex heterogeneous data sources. A wide range of clinical, health care, social and other clinically relevant information are stored in these data sources and in this sense the challenge is to find the right way to access these data in a conscious way. The semantic net on which thesauri are implemented allows to connect concepts to each other in a meaningful way, especially when they refer to heterogeneous data. In this paper we present a source that semantically connects the impact of pollutants on human health.

Pollutants are substances, solids, liquids or gases, mainly produced in high concentrations from human activities, that exert adverse effects on the environment, by polluting the water, the air and the soil (Manisalidis et al. 2020). The negative impact of pollution on human health has been extensively demonstrated. The toxic effects exerted by pollutants are strictly related to their physical and chemical properties. Aerosol compounds have a greater toxicity than gaseous compounds because their tiny size increases their penetration capacity (Colbeck 2009). Many pollutants act as major factors in human diseases, such as particulate matter, nitrogen oxide, carbon monoxide, sulfur dioxide, Volatile Organic Compounds, dioxins and Polycyclic Aromatic Hydrocarbons. The short and long exposure to these pollutants can promote respiratory and cardiovascular diseases, reproductive and central nervous system dysfunctions, cutaneous diseases and cancer, significantly increasing the mortality rate (Manisalidis et al. 2020). Persistent organic pollutants (POPs) include many industrial chemicals with long half-lives, from a few weeks to a dozen years or more, depending on physiological conditions as well as on the environment. The lipophilic properties and the stability of POPs, allow them to persist in soil, air, water and in animal tissue, where they can accumulate and biomagnify.[2] Exposure of human and animals to POPs can lead to a variety of adverse effects, including carcinogenicity and teratogenicity, genotoxicity, reproductive- and endocrine-disrupting effects, immunotoxicity and neurotoxicity (Kailun et al. 2021). Furthermore, heavy metals are included within the group of highly emitted contaminants and their adverse effect on living organisms has been widely studied in recent decades (Houessionon et al. 2021). Mercury is one of the most toxic heavy metals, which can migrate around the globe and magnify through the food chain, ultimately harming human health (Yang et al. 2020).

2.0 The method

In this section the methodological tasks followed towards the construction of a thesaurus for the EO systems domain of study is described. This knowledge organization tool has been aimed at formalizing the terminological arrangement of a specialized lexicon and at providing a semantic resource able to support the indexing operations as well as the retrieval of documentations on this specific reference area of study. The thesaurus creation process is a horizontal activity meant to improve the semantic interoperability and information management procedures for the Earth observation systems domain, and it represents one task included within the two projects, Igosp, part of the ERA-PLANET main program, and E-SHAPE. Igosp focuses on the identification and classification of pollutant hot-spots, while E-shape on global earth sustainability models.

2.1 Corpus compilation

In order to obtain a list of representative technical terms to be used as a starting point for the creation of a semantic network proper to thesauri, the first phase of the activity covered the retrieval of the scientific literature about the domain of study (Trigari 1993). Therefore, the initial procedure started by compiling the source corpus, the collection of documents from which to start to process the textual contents and extract the representative

2 Stockholm Convention. The POPs, available at: <http://www. pops. int/TheConvention/ ThePOPs/ tabid/ 673/ Default.aspx>.

terms meant to be inserted in the thesaurus and be connected with each other by means of the semantic relationships stated by the ISO norms (ISO 25964:2011 and ISO 25964:2013). The documents which have been gathered are included in Pubmed as data sources. This web-based portal uses the The Medical Subject Headings (MeSH) thesaurus to index scientific papers and automatically maps terms to find the ones included within the MeSH controlled vocabulary. In particular, documents have been selected starting from 2000 to 2019 and they refer to original scientific articles (including in vitro and in vivo studies), case reports, systematic reviews and meta-analysis of epidemiological studies.

The approach followed to perform an exhaustive data mining research (Ecker 2010; Bramer 2018) has been based on the establishment of a clear and focused question, the identification of the key concepts addressing the different elements of these latter, and of the elements that should be used to obtain the best results and ordered them by their specificity and importance to determine the best search approach.

In particular, terminologists and domain-experts agreed on the following wildcards criteria (Mishra et al. 2009) to achieve a reliable source corpus for the technical EO domain: (i) identification of the appropriate terms and synonyms as well as combined terms with connector units (AND, OR or NOT); (ii) detection of experts' profiles in the mercury toxicity field to search for their most cited articles; (iii) selection of terms or combined terms to execute the investigation, such as, *environment contamination*, *occupational human exposure to mercury*, *mercury and human health*, *methylmercury OR MeHg AND human health*, *mercury toxicity*, *mercury OR MeHg AND cardiovascular diseases*, *mercury OR MeHg AND neurodegenerative diseases*, *mercury OR MeHg AND endocrine diseases*, *mercury OR MeHg AND respiratory diseases*, *mercury and fetal toxicity*, *pregnancy AND mercury OR MeHg*, *molecular mechanism of mercury toxicity*, *global mercury OR Hg cycle*, *mercury OR MeHg and Kidney diseases*, *mercury AND tubular renal toxicity*, *mercury AND endothelial damage*, *methylmercury OR MeHg AND brain*, *methylmercury OR MeHg AND placenta*, *anthropogenic mercury sources*, *mercury OR MeHg bioaccumulation*, *mercury AND urine*, *mercury tissue accumulation*, *mercury exposure AND cancer*, *gold mining AND mercury*, *dental amalgam mercury*, *atmospheric mercury concentrations*, *atmospheric mercury emissions*, *Minamata disease*, *Minamata convention*.

A total of 250 studies have been reviewed and they represent the source corpus used to run the next steps related to the semantic analysis and creation of a list of candidate terms useful to build a specialized knowledge-domain thesaurus.

2.2 Text processing

Once having finalized the population of the domain-oriented documentation, the following task has concerned the text processing operations to achieve a precise controlled terminological list to analyze. The source corpus has been processed with the aid of a semi-automatic term extractor tool, TextToKnowledge (T2K) (Dell'Orletta et al. 2014), with which the most representative terms have been identified and filtered out by means of the document inverse statistical measures. The level of granularity the EO candidate terms bear with respect to this field of knowledge has undergone a validation phase with the experts of the domain who set a specificity treebank for the term base. Indeed, as Trigari (1993) in one or more stages underlines, terms included in thesauri should be selected according to their compliance with the controlled language rather than the natural one, meaning that terms need to have morphological features that correspond to standard conventions widely used in the documentary procedures, and that they have to be unambiguous dealing with the area of study. The author also points out that term inclusion in thesauri should start by taking into account their *economic value* within the thesaural structure based on exhaustivity and specificity criteria, synonymy preferences, the use of notes which could demark their semantic scope in those cases where the hierarchical and associative structures are not explicitly functional.

The semantic resource developed is the result of the terminological analysis starting from the controlled lists of candidate lexical units obtained with the execution of T2K linguistic analysis. In particular, this software is based on the exploitation of Natural Language Processing (NLP) functionalities, statistical text analysis and machine learning methods, and reflects the domain-oriented information. The output provided by this semantic software is a list of the terms sorted according to their frequency level within the documents and it represents the first step to select the candidate lexical units meant to become part of the higher structured system of a thesaurus (Hudon 2009). Indeed, the terminologist's activity in identifying the salient information in the domain-oriented texts is facilitated starting from a controlled list of terms representing the key domain concepts accompanied by their frequency scores retrieved according to their occurrences in the source texts. Other elements to take into account when selecting the candidate terms according to Houdon are linked to the intrinsic scientific value of the terms, their structural compatibility, their contextual dependency degree and coherence. With the latter criterion the author underlines how the selection of the terms should follow the same logic, meaning that if the terminologists in accordance with the domain-experts start choosing a vulgar form for the representation of the concepts instead of the scientific one, this methodology should be pursued along the entire thesaural construction. In the specific use case presented in this paper, the preferred configuration has been driven towards the systematization of the terminology in the thesaurus framework from a scientific perspective.

2.3 Thesaurus construction

The thesaurus has been chosen with respect to other KOSs (Souza 2012) because of its semantic fixity in defining the associations among domain-representative terms, i.e., managing the unstructured technical information on contaminant substances in an entangled and controlled network of semantic relationships of hierarchy, synonymy and association that can support the understanding of the specialized type of information.

> "[...] thesauri consist of a selection of concepts supplemented with information about their semantic relations (such as generic relations or ('associative relations"). Some words in thesauri are "preferred terms" (descriptors), whereas others are "lead-in terms." The descriptors represent concepts. The difference between "a word" and "a concept" is that different words may have the same meaning and similar words may have different meanings, whereas one concept expresses one meaning." (Hjørland 2007, 367).

Indeed, thesauri represent semantic tools that systematize the information proper to technical fields of knowledge in an interconnection of semantic relationships. The method applied for the construction of the thesaurus has followed the principles contained in the ISO 25964:2011 and ISO 25964:2013 standards where a detailed overview of the thesaurus' construction process starting from the definition of its semantic relationships system is provided. The three main types of relationships refer to hierarchy, equivalence and association connective structures (Broughton 2008). The hierarchical relationship is based on the "degrees or levels of superordination and subordination, where the superordinate concept represents a class or whole, and subordinate concepts refer to its members or parts" (ISO 25964:2011, 58) and it supports the identification of suitable levels of specificity when looking for a more general or more specific term representing a domain-oriented concept. The ISO standard tags for this kind of interconnection are Broader Term (BT) and Narrower Term (NT). An example in the developed EO thesaurus can be the following: BT *contamination*, NT *mercury contamination* NT2 *atmospheric mercury contamination.*

The equivalence relationship stands for the synonymy link to be detected within the knowledge domain, they are tagged as Use (USE) and Used For

(UF) where the first relates to the preferred entry term while the second to the synonyms, e.g. USE *mercury*, UF *Hg.*

Still according to the ISO standard, the associative connection, marked as Related Term (RT) refers to terms that are semantically or conceptually linked to each other but not hierarchically, e.g., *waste incineration* RT *incinerators.*

Following the terminological extraction process the analysis has covered the selection of the main representative terms to be considered as the best candidate entries to include in the thesaurus and to interconnect according to the standards' guidelines. This phase involves a strict form of collaboration with the domain experts responsible for providing a high-level support in the identification of the salient and updated information, in the sense of usability within the specific groups of experts in the domain, to take into account (Shultz 1968).

From a semantic point of view, the specific goals addressed within this activity rely on a formalization of the terminology proper to the EO field of study. The treatment of the specialized languages commonly refers to multiple operations dealing with the semantic disambiguation, with the definition of terminological specificity level and univocity in the way concepts are represented, thus providing a fixed and unambiguous model to be shared with a community of users.

The thesaurus is exploited as a search engine tool to retrieve the documentation needed on the environmental domain focusing on the impacts of mercury and atmospheric pollutants on human health (Aronson 1994). This is how the resource is structured: 979 terms have been included with different levels of depth according to the types of relations set out.

2.4 Semantic interoperability

The perspective of this paper addresses the improvement of semantic interoperability, usually granted by the exploitation of the markup languages or by consulting external vocabularies dealing with the same thematic thread.

For this specific field of study there are several knowledge organization systems available and officially shared by the experts of this domain. For instance, as comparative standards to which to refer to execute the mapping operations, aimed at evaluating the semantic coverage reached in the proposed terminological asset, the existing knowledge organization resources General Multilingual Environmental Thesaurus (GEMET),[3] AGROVOC

3 "GEMET," last accessed September 28, 2021, https://www.eionet.europa.eu/gemet/en/themes/.

Multilingual Thesaurus,[4] Earth and Inspire have been taken into account. Table I shows the best coverage result obtained with GEMET thesaurus (*gemet*) compared to the terminological extraction hereby proposed (*igosp*), and this represents an encouraging result since this latter contains technical information normalized as access points to understand the domain of earth, as the thesaurus developed for this research study yearns to be. The score obtained by mapping the term extraction list with Agrovoc (*agrovoc*) is also encouraging considering that it covers a multidisciplinary range of agricultural sectors: the list from which the EO system thesaurus started to create its internal structure reached a quite low level of the recall measurement, and this should prove the specificity encapsulated with respect to the Agrovoc wide breadth framework.

Corpus term consistency = 93852		
Standards	**n. terms**	**Recall**
Earth	13968	15,8%
Inspire	558	9,8%
Agrovoc	45500	6%
Gemet	5526	20%

Table 1. Coverage with standards

The compliance with the informative texture included in the standards implies a reliable adaptation of the thesaurus to the knowledge domain to be shared, but there are other methods that have been implemented in order to guarantee a deeper level of interoperability. A first factor that impacted the enhancement of the interconnection degree covered the significant collaboration with the experts of this field of study from an environmental perspective as well as medical-scientific one. This interdisciplinary framework facilitated the selection of appropriate terms and semantic structures that could be disseminated throughout several sectors of interest increasing the intra-inter communicability among users with specific skills.

Moreover, the thesaurus has been forged with the objective of creating a resource able to help the indexing of documents, and to be integrated, in this sense, on the GOS⁴M platform (http://www.gos4m.org). In order to implement the thesaurus on this web service, it has been converted into SKOS-RDF language since the flexibility provided by this grammar allows systems

4 FAO, "AGROVOC," last accessed September 28, 2021, https://agrovoc.uniroma2.it/agrovoc/agrovoc/en/?clang=fr.

Terms 979

Terms for each depth level

Deep level	# of terms
Deep level 1	673
Deep level 2	190
Deep level 3	28
Deep level 4	8
Deep level 5	1

Relations between terms 952
Non-preferred terms 96

Figure 1. Thesaurus elements

to communicate with each other and facilitate the access to the information (Van Assem et al. 2004).

3.0 Results

In this section some representative examples of the outputs obtained from the previous tasks are presented.

a. *Terms and semantic relationships*

The thesaurus for EO systems domain is currently constituted by 979 terms and it contains 952 structures of semantic relationships configured through the synergic work of the scientific experts and terminologists, a summary is provided by Figure 1 above.

The next three figures (Figure 2, Figure 3, Figure 4) depict the branching of the semantic connections starting from a preferred term at the top of the lines (*POPs*, *aqueous mercury*, *mercury*).

In each tab for selected terms the visualization can be dynamic by expanding the entry terms with their more specific relations, thus elaborating the hierarchical relationship functionality to retrieve knowledge-domain information (McMath 1989), or having a more consistent and flatter alphabetical list where it is possible to see all the relations with the other terms.

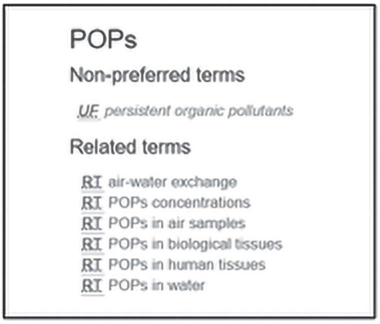

Figure 2. POPs term-set

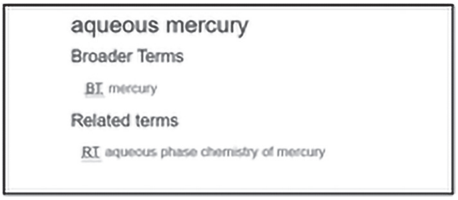

Figure 3. Aqueous mercury term-set

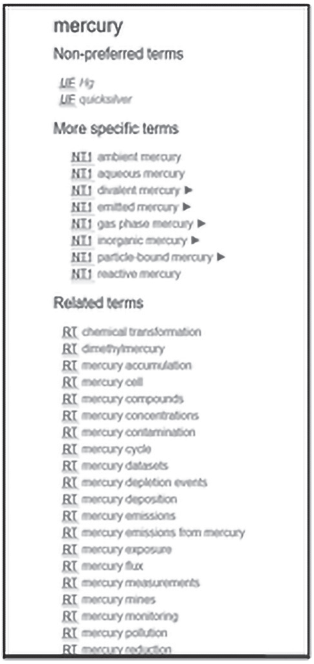

Figure 4. Mercury term-set

4.0 Discussion

According to the final goal of the project, for which the tool has been created,[5] the thesaurus is meant to be integrated on the GOS[4]M web-based platform having as its main goal that of helping users in the indexing and retrieving operations over a large set of documentation about the EO domain. Indeed, by exploiting its inner structure through RDF language, the system can run queries extracting texts where the informative needs are satisfied in a more precise way. The advantages in creating and using a resource such as a thesaurus are intrinsically connected to the way it can be employed to retrieve domain-oriented information by relying on its officially accepted, and consequently reliable, structure of semantic connections. In fact, the searches run over a web-platform with the aid of inferences extracted from the hierarchical, associative and synonymous batteries can help users to directly access a precise group of information (Aitchison 2003). The thesaurus has been built – as mentioned above – with terms coming from scientific literature selected from PubMed that contains studies specifically referring to the impact of mercury on human health conditions, giving in this way an added value to the semantic tool.

Thanks to the support of interdisciplinary teamwork, the integration of information referred to the domain-oriented terms has been achieved. Through the help guaranteed by the experts a list of significant verb clusters has been identified in order to investigate within the source corpus the contexts by which these connections would have offered a detailed mosaic of new data to be imported in the thesaural structure. This activity helped to obtain a richer consistency in the terminology asset by relying on a contextual structure belonging to the texts from which the candidate terms of the thesaurus have been retrieved. In this way, the compliance with the documentation has been maintained and, at the same time, enhanced by exploiting the supplementary information in the same texts through targeted verbal constructions that would have specifically led to precise term associations. In some cases, these patterns have created a new configuration of terms, meaning that terms that were not present in the thesaurus (but they did exist within the output list of the terminological extraction) have been included because of their relationships with the preferred entries, now part of the verbal constructions under examination, e.g., FGDs *UF flue gas desulfurization*, *hazardous waste RT nonhazardous waste*, *developing brain RT methylmercury toxicity*.

5 The semantic source created is part of the activities of the e-Shape project, Grant Agreement number 820852.

Finally, the main goal of this procedure is to elaborate an enhanced model for the associative relationship proper to thesauri configurations, usually ambiguous and not strictly precise in the connections it forges among the domain-oriented terms.

In addition to this we intend to elaborate an enhanced model for the associative relationship proper to thesauri configurations, usually ambiguous and not strictly precise in the connections it forges among the domain-oriented terms. Some examples of new candidate term connections obtained by using verbal pattern banks to improve the specificity of the associative relation are given in Table 2:

Original term (from corpus)	**Semantic relationship created**	**Term connection**
inorganic mercury	RT (is transformed to)	methylmercury
deposited divalent species	RT (transformed to)	gaseous elemental mercury
methylmercury	RT (cause adverse effects in)	developing brain
methylmercury	RT (absorbed through)	gastrointestinal tract
toxic metal	RT (Contaminated)	seafood
methylmercury	RT (excreted in)	breast milk
air-sea exchange	RT (has a large impact on)	GEM concentrations
biological samples	RT (used to assess)	mercury exposure

Table 2. Terms connection examples

5.0 Conclusion

Specialized competences are constituted by specialized terminologies (Cabré 1996), and this is a key aspect to bear in mind when dealing with highly scientific domains of study. Indeed, these specific sectors are characterized by a likewise specific lexicon that implies a series of normalization operations to allow this information to be diffused and create a knowledge representation to be reproduced. In this sense, the paper proposed the creation of a terminological resource, a thesaurus, to be used as a baseline composed by technical terms selected alongside the support and agreement of domain-experts that can represent (i) a structured organization of the EO system knowledge and (ii) a reliable starting point for the information retrieval web-based tasks.

References

Aitchison, Jean, Bawden David, and Alan Gilchrist. 2001. *Thesaurus construction and use: a practical manual*. London: Routledge, 2001.

Aronson, Alan R., Thomas C. Rindflesch, and Allen C. Browne. 1994. "Exploiting a Large Thesaurus for Information Retrieval." *RIAO '94: Intelligent Multimedia Information Retrieval Systems and Management* 1: 197-216.

Auger, Alain, and Barriere Caroline. 2008. "Pattern-based approaches to semantic relation extraction: A state-of-the-art." *Terminology* 14: 1–19.

Bernier-Colborne, Gabriel. 2012. "Defining a gold standard for the evaluation of term extractors". In *Proceedings of the Eighth International Conference on Language Resources and Evaluation,* (LREC 2012): 15–18. https://doi.org/10.1075/term.14.1.

Bramer, Wichor M., Gerdien B. de Jonge, Melissa L. Rethlefsen, and Frans Mast. 2018. "A systematic approach to searching: an efficient and complete method to develop literature searches." *Journal of the Medical Library Association* 106, no. 4. https://doi: 10.5195/jmla.2018.283.

Broughton, Vanda. 2008. *Costruire tesauri*. Editrice bibliografica.

Cabré, Maria Teresa. 1996. "Terminology today". In *Terminology, LSP and Translation: Studies in language engineering in honour of Juan C. Sager*, edited by Harold Somers, 15-34. Benjamins Translation Library. https://doi.org/10.1075/btl.18.

Colbeck, Ian, and Mihalis Lazaridis. 2010. "Aerosols and environmental pollution". *Naturwissenschaften* 97:117-31, https://doi.org/10.1007/s00114-009-0594-x.

Dell'Orletta, Felice, Giulia Venturi, Andrea Cimino, and Simonetta Montemagni. 2014. "T2k^ 2: a system for automatically extracting and organizing knowledge from texts." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (ELRA 2014): 2062-70.

Ecker, Erika, and Andrea C. Skelly. 2010. "Conducting a winning literature search." *Evidence-Based Spine-Care Journal* 1, no. 1. https://doi: 10.1055/s-0028-1100887.

FAO. "AGROVOC." Last accessed September 28, 2021. https://agrovoc.uniroma2.it/agrovoc/agrovoc/en/?clang=fr.

"GEMET." Last accessed September 28, 2021. https://www.eionet.europa.eu/gemet/en/about/.

Hjørland, Birger. 2007. "Semantics and Knowledge Organization." *Annual Review of Information Science and Technology* 41, no. 1: 367-405. https://doi.org/10.1002/aris.2007.1440410115.

Houessionon, M. G. Karel, Edgard-Marius D. Ouendo, Catherine Bouland, Sylvia ATakyi, Nonvignon Marius Kedote, Benjamin Fayomi, Julius N Fobil, and Niladri Basu. 2021. "Environmental Heavy Metal Contamination from Electronic Waste (E-Waste) Recycling Activities Worldwide: A Systematic Review from 2005 to 2017". *International Journal Environment Research and Public Health* 18, no. 7: 3517. https://10.1002/aris.2007.1440410115.

Hudon, Michèle. 2009. *Guide pratique pour l'élaboration d'un thésaurus documentaire*. Montréal: Les éditions Asted.

ISO 25964:2011, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval.

ISO 25964:2013, Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies.

Kailun, Sun, Yan Song, Falin He, Mingyang Jin, Jingchun Tang, and Rutao Liu. 2021. "A review of human and animals exposure to polycyclic aromatic hydrocarbons: Health risk and adverse effects, photo-induced toxicity and regulating effect of microplastics". *Science of the Total Environment* 773: 145403. http://doi:10.1016/j.scitotenv.2021.145403.

Manisalidis, Ioannis, Elisavet Stavropoulou, Agathangelos Stavropoulous, and Eugenia Bezirtzoglou. 2020. "Environmental and Health Impacts of Air Pollution: A Review". *Frontiers in Public Health* 8, no. 14. http://doi:10.3389/fpubh.2020.00014.

McMath, Charles F., Robert S. Tamaru, and Roy Rada. 1989. "A graphical thesaurus-based information retrieval system." *International Journal of Man-Machine Studies* 31, no. 2: 121-47.

Mishra, Shruti, Sandep Kumar Satapathy, and Debahuti Mishra. 2009. "Improved search technique using wildcards or truncation." *International Conference on Intelligent Agent & Multi-Agent Systems*, 1-4. https://doi:10.1109/IAMA.2009.5228080.

Shultz, Claire K., Richard H. Orr, and Peter Henderson. 1968. *Evaluation of Indexing by Group Consensus.* Washington: Bureau of Research Office of Education, U.S. Department of Health, Education and Welfare.

Sinha, Pradeep K., Sunder Gaur, Bendale Prashant, Mantri Manisha, and Dande Atreya. 2012. *Electronic health record: standards, coding systems, frameworks, and infrastructures*. New York: Wiley-IEEE Press.

Souza, Renato R., Douglas Tudhope, and Mauricio Barcellos Almeida. 2012. "Towards a taxonomy of KOS: Dimensions for classifying knowledge organization systems". *Knowledge Organization* 39, no. 3: 179-92. http://doi:10.5771/0943-7444-2012-3-179.

Tudhope Douglas, and Ceri Binding. 2009. "Faceted Thesauri." *Anxiomathes* 18, no. 2: 18:211. http://doi:10.1007/s10516-008-9031-6.

Van Assem, Mark, Maarten R. Menken, and Guus Schreiber. 2004. "A method for converting thesauri to RDF/OWL." *The Semantic Web – ISWC 2004: Third International Semantic Web Conference,Hiroshima, Japan, November 7-11* 3298, (Springer 2004): 17-31. http://doi:10.1007/978-3-540-30475-3_3.

Yang, Lixin, Yuanyuan Zhang, Feifei Wang, Zidie Luo, Shaojuan Guo, and Uwe Strähle. 2019. "Toxicity of mercury: Molecular evidence." *Chemosphere* 245: 125586. https://doi.org/10.1016/j.chemosphere.2019.125586.