Enhancing the Visibility and Relevance of Thesauri in the Web: Searching for a Hub in the Linked Data Environment†

Francisco-Javier García-Marco

Universidad de Zaragoza, Departamento de Ciencias de la Documentación e Historia de la Ciencia, Facultad de Filosofía y Letras, C/ Pedro Cerbuna 12, Zaragoza, Spain, 50009, <jgarcia@unizar.es>

Francisco Javier García Marco achieved his Ph.D. in philosophy and arts in 1994 and has been Professor of information and library science at the University of Zaragoza since 1996.

García-Marco, Francisco-Javier. 2016. "Enhancing the Visibility and Relevance of Thesauri in the Web: Searching for a Hub in the Linked Data Environment." *Knowledge Organization* 43 no. 3: 193-202. 51 references.

Abstract: Thesauri have triumphed in many domains that require precise and exhaustive information because of their representational power, their capability to integrate the concept-based and alphabetical approaches to organizing information, and their standardization and, more recently, formalization. Nevertheless, there is room to improve their relevance in the digital age by embracing the open linked data initiatives and by taking advantage of their structural and functional proximity to some of the big collaborative knowledge repositories in the Internet, notably the Wikipedia environment. With a focus on its implications for enhanced interoperability, this structural proximity is analysed, and the benefits of such collaboration for the different potential stakeholders are considered. It is proposed that better devices for ensuring semantic browsing are provided when necessary, and that an open hub for thesauri interconnection is developed, perhaps using existing big open Internet semantic facilities, such as Wikipedia.

Received: 27 January 2016; Accepted 30 January 2016

Keywords: Wikipedia, thesauri, knowledge, terms, relations, semantic

† This paper presents results of the project CSO2015-65448-R, "Possibilities and Requirements of Knowledge Organization Systems (KOS) for the Interoperability between the Institutions of Memory and the Cultural Tourism Sector in the Internet," funded by the Ministry of Economy and Competitiveness of the Government of Spain.

1.0 Introduction: looking for a global impact in the Internet environment

The context, history and current digital challenges of thesauri are well known (see, among others, Foskett 1982; Lancaster 1986; Gilchrist 1991, 2003; Dahlberg 1991, 1992, 1995; Williamson 1996; Aitchison and Dextre Clarke 2004; Dextre Clarke 2008) and also their most recent achievements, remarkably the road to the new ISO 25964 standard (International Organization for Standardizaion 2011 and 2013), summarized by Dextre Clarke and Zeng (2012). Thesauri and other vocabulary control tools are already important in information systems and databases where the information, the users, or the costs of misinformation have great economic value, for example, in chemical, pharmaceutical, medical or heritage databases. But what roles could remain for thesauri in this new environment dominated by Google and other new generation search engines, even in academic contexts (Georgas 2013, 2014, 2015; Kemman, Kleppe and Scagliola 2014)?

2.0 The need for thesaurus "hubs" and the big Internet "knowledge graphs"

For a global impact in the current digital context, thesauri and other knowledge organization tools must find a way to support the big search engines that users prefer as portals to the Internet. Only if thesauri can be used to improve searches made through the main search engines, will they become increasingly relevant in the global Internet arena. In fact, big search engines have been incorporating an explicit semantic dimension in their search

models in response to the ever-increasing size of the web and the related growing demand of users for greater precision, mainly by exploiting big knowledge repositories that have been formalized in "knowledge graphs." For example, Google incorporated synonym "rings" in its latent semantic indexing model in February 2004 (Mooz 2015), and in May 2012 began to implement its own "Knowledge Graph" that codifies people, places, things and the relations among them, remarkably presented (Singhal 2012) as retrieving "things, not strings."

The semantic web has seen clear developments where such projects of inserting knowledge organization systems into the heart of the web can be made possible (e.g., Library of Congress 2015). For this task, thesauri have three big strengths: the simplicity of their model, their ability to integrate different perspectives and traditions in the field of knowledge organization and their standardization. Thesauri correctly and fully identify, synthesize and express the basic relations that users (and search engines) need for successful terminological or conceptual navigation. They can be used both to express relations among terms to the users in search interfaces and to feed relevant relations into the "knowledge graphs" that search engines increasingly use to improve relevance and specificity. In addition, thesauri have the status of an international standard that is key to simplifying operations and ensuring interoperability. Because of this, from a commercial point of view, they have an authority in the quest to fulfil the need for structured vocabularies in the semantic web.

ISO 25964-2 (International Organization for Standardization 2013, 16-19) proposes three strategies to interconnect structured vocabularies: the structural unity, direct-linked and hub models. All of them can be useful and even complementary in the Internet environment. In principle, the last two can be considered bottom-up, in the sense that no central authority or agent is needed to implement the support. The opposite obtains in the first model, in which the structure is common to the different vocabularies, and this normally requires very strong coordination.

The direct-linked model can be considered as a horizontal strategy that creates a network of interconnected ontologies. This distributed model allows for the different agents to cooperate in gradually building a more interlinked web, project by project. The mapping does not need to be in both directions, at least to begin with. So, providing there is a license or even a tacit approval, only one of the parts must act to start with the net. Such a totally distributed approach has a very similar philosophy to that of hyperlinking and is especially useful for improving retrieval in all of those well-catalogued databases that are becoming increasingly accessible and usable, thanks to the open data and semantic web movements.

In the hub model, all of the different vocabularies are mapped on to one vocabulary, which functions as a central node. As this vocabulary functions as a spine and central point, there is a clear reference point and a structured network can grow. At the same time, this structure is more efficient and cheaper to maintain, because each vocabulary must be mapped only once. Moreover, there is also a single point of reference for all the systems and bots that want to reuse this information, for example, search engines. Thus, related administrative interfaces, e.g. the semantic web application program interfaces (APIs), must be maintained at only one point.

On the other hand, any successful hub must comply at least with three important prerequisites: size, openness and endurance. First, the core vocabulary must be big enough to make possible the interconnection of vocabularies from many different fields, languages and organizations. Second, their politics and legal terms must be sufficiently open, so that a myriad of legal problems and control issues do not derail the efforts, and so that the system can be enriched and linked in both directions. Third, the organization behind it must be sufficiently solid and have a strong and demonstrated commitment to the core vocabulary, so that its long-term preservation is ensured.

Thanks to the clear and unambiguous formalization that ISO 25964-1 provides (International Organization for Standardization 2011, 104), concepts and their relations can be easily and automatically imported into the knowledge graphs of the search engines. So, ISO 25964 can function as a reliable interoperable standard to feed the knowledge bases of the search engines from the existing knowledge repositories.

3.0 The case for Wikipedia as a potential thesaurus "hub"

Could an extended Wikipedia constitute such a platform? After all, lexical thesauri, dictionaries, encyclopaedias and thesauri for information retrieval are very close siblings in their etymology, function and history. The underlying etymology of "treasury" (from the Greek θησαυρός, thesauros, storehouse, treasure) that all these tools share denotes an effort to select the best of their domain—a language, a terminology, a collection of concepts, citations or texts—and to offer it in an organized way, usually not only alphabetically but systematically, enriched with devices for controlling synonymy and polysemy. They only differ in their most specific aims: encyclopaedias facilitate learning and education, with an emphasis on referencing particular persons, places, events, things and processes (knowledge about the natural and cultural worlds); dictionaries and classical thesauri assist in writing and reading and are focused on compiling and organizing lexical knowledge

(knowledge about words and their use); and thesauri, retrieving from repositories, catalogues and documents.

Wikipedia among the encyclopaedias and WordNet among the dictionaries and thesauri are arguably the strongest exemplars of their respective kinds in the Internet domain. In particular, Wikipedia has become the Internet portal for encyclopaedic reference and, thanks to its open and collaborative philosophy, has become a huge semantic repository. So it is not strange that knowledge experts (Okoli et al. 2014; Mesgari et al. 2015) from different semantic-related and many other disciplines are busy exploiting Wikipedia exactly as a knowledge base. Also, some scientific communities (e.g., Gardner et al. 2011) have been exploring the potentialities of the Wikipedia as a social semantic hub, notably in the field of genome research. These communities are also busily involved in building or using lexical thesauri around Wikipedia to overcome very similar problems to those in many other fields, not only with a generic linguistic approach but also in very specific scientific domains, e.g., using Wikisaurus or Rapisardi (Di Franco and Giardino 2014). In the information science field, part of this research (e.g., Hu et al. 2009) has been related to enhancing document clustering and information retrieval. Furthermore, Wikipedia has been successfully used to harvest selected references and contextual information for enhanced news systems (Lee 2008; Howard and Oliver 2011; Evans 2015) and interconnecting library resources and reference (Connaway and Faniel 2014).

But, notably for our discussion above, Wikipedia is also the main information source of the main linked open data (LOD) services, including Google Knowledge Graph, IBM's Watson, Apple's Siri, and, of course, Freebase, DBpedia, Wikidata and Yago, extensively used in all kinds of knowledge-powered and natural language processing (NLP) applications (Bergmann 2015).

4.0 Potential benefits and problems

So, the benefits of explicitly interlinking thesauri and the Wikipedia environment can be diverse for all the applications involved: the Wikipedia and information retrieval thesaurus communities, the end users, the databases and curators and designers of repositories, as well as the big search engines.

Repositories and databases get enhanced positions within the World Wide Web ecosystem. Connaway and Faniel (2014, 71) report several cases in which connecting digital library resources to relevant Wikipedia entries has offered excellent results in improving their discovery and use. This strategy should be generalized (Connaway and Faniel 2014, 16): "We need to provide better, more natural links to library resources in places like Wikipedia and other network services."

Using thesauri, such a connection would provide a wide vocabulary whereon to hang semantic web-compliant thesauri and other knowledge organization systems (KOSs), functioning as a big bi-directional hub. Moreover, thesaurus editors and contributors might immediately benefit from Wikipedia's definitions and contextual information. Also, they could use Wikidata and DBPedia graphs to automatically assist editorial tasks, such as proposing and assessing intralingual and multilingual preferred terms and equivalences, disambiguating, establishing hierarchical and associative relations and as a complementary reference for the evaluation of the scope and coverage of a thesaurus. Being represented in Wikidata, they could be easily interconnected so that third parties could exploit this open semantic repository. Finally, the hub would serve to take advantage of social knowledge sharing and editing and become a source of data about this tagging and linking activity that would be very useful for evaluation purposes and for enhancing relevance judgements. In fact, Wikipedia mappings are actually being provided in some thesauri, like the GEMET Thesaurus (2012). Such a semantic network of KOS around a large tool such as Wikipedia would be very easy to exploit by the "big animals" of Internet search, ensuring that thesauri and other KOSs do not become isolated in the emerging Web 3.0, and that they find their way forward.

Finally, Wikipedia itself could benefit from a tool to assist editors and bots in assigning categories and providing structure, a task in which Wikipedia's contributors and editors appear to be falling behind, in contrast to the huge growth in the number of entries. Thesauri and other KOSs could even be useful in providing different alternative structures and knowledge maps, depending on the user's domain of reference. Besides this, by delivering such highly specific domain maps, thesauri could be used to improve some of Wikipedia's administrative processes, such as identifying new potential articles ("red links") or assessing how well represented is a specific domain. To sum up, setting a large social semantic medium, like Wikipedia or a similar social platform, as an anchor for thesauri could serve very different stakeholders of the information ecology.

However, though the standards are already available, there are also certain problematical obstacles in the way. They are mainly of a social nature. As Berners-Lee (2007) and later Schindler and Vrandeččić (2011) discussed in detail, the web is not only a technical project but also a social one, and both dimensions must be properly addressed to ensure the advancement of the semantic web project. In particular, there are important legal implications in collaborating in these platforms that must be taken into account. If the concepts, terms and relations of a thesaurus are published there, open licenses apply. Recognition is granted, but any control of the use of this information is

lost. Depending on the needs, it would be necessary to examine the legal status of a thesaurus and find ways to protect it if there is a need, for example, by publishing only subsets or old editions. In any case, this is a problem that requires careful research, source by source.

Cooperation would be very much enhanced if the usual one-by-one approach was complemented by a more structured cooperation of all the interested parties. An international organization such as the International Society for Knowledge Organization (ISKO) could promote and support a working group to bring together thesauri and other KOS editors and users, looking for coordination with the Wikipedia Foundation, WordNet bodies, the relevant committees of World Wide Web Consortium and the knowledge graphs governing structures. This could lead to a more ambitious "World Wide Web Thesaurus Foundation" in the future.

5.0 Mapping Wikipedia and thesauri: some clues

Information retrieval thesauri, lexical dictionaries and encyclopaedias can be used together in the Internet to complement their respective weaker points and to enhance their administrative operations and effectiveness, allowing for innovative uses of these tools (Figure 1). On the one hand, Wikipedia, WordNet and thesauri have very specific weaknesses, mainly as a result of their distinctive missions: thesauri are frequently very technical and more limited in scope; Wikipedia has an incomplete system of categories and does not provide any morphological or lexical representation of its terms; WordNet is oriented towards standard lexical knowledge, and it is not rich enough for representing technical or scientific knowledge in specific disciplines. On the other hand, their strengths are complementary to a great degree: thesauri provide highly systematized hierarchical relations; Wikipedia, a very complete and ever-increasing set of conceptual relations; WordNet, a complete description of the morphological and lexical properties of each lexical unit. So, in addition to a very solid common ground for cooperation, their specific strengths can complement each other's weaknesses. Thus, it is not strange that a network of interconnections is actually being developed between virtual encyclopaedias and dictionaries, as can be seen in Figure 1, and that a great potential exists to bring thesauri into this dynamic.

The interconnection among information retrieval thesauri, lexical dictionaries and encyclopaedias is usually more efficiently achieved through semantic web compliant versions or sister applications, like SKOS-expressed thesauri (with some potential problems, see Pastor-Sánchez 2013) or the Wikidata and DBPedia projects in the case of Wikipedia (for pros and cons, see Lee 2008). Furthermore, there are different bridge projects that interconnect big knowledge repositories like Wikipedia and Wordnet. For example, YAGO maps Wordnet, the virtual terminological thesaurus, to Wikipedia; and WikiTax2Wordnet interlinks Wordnet and Wikipedia's categories (Ponzetto and Navigli 2009). Färber et al. (2015) have recently discussed the pros and cons of different components of Wikipedia's semantic environment and other LOD projects that are relevant to this task.

In the case of Wikipedia and thesauri, the mapping could be done directly and with enhanced precision from several of Wikipedia's structural elements, at least for the core relations (Table 1), though substantial intellectual input is likely to be needed to achieve conformity with thesaurus standards in other aspects, like the form of the terms. More of Wikipedia's elements could be useful for more complicated and indirect mappings, which usually will require parsing and other data mining techniques (Bergman 2011).

Thesauri	Wikipedia
Concept	Page with content (article)
Terms	Any page, including redirect and disambiguation pages
Description	First paragraph or section
Preferred term	Category, title
Equivalent terms	Redirect, disambiguation pages
Broader and narrower terms, groups of concepts	Category, hyponyms, navigation templates pages and links, stand-alone lists and links
Top terms, subject areas	Category:Contents
Related terms	See also, further reading, internal hyperlinks
Inter-linguistic equivalents	Alternate language versions

Table 1. Wikipedia and thesaurus structural elements showing some correspondence.

As Table 1 shows, each content page in Wikipedia corresponds to a concept. All the possible alternative terms that can label the content page must be codified in an empty page that must redirect to the main page (Wikipedia 2015a). Polysemy and homonymy are controlled through disambiguation pages that list all the possible senses of a term, provide a short definition, contextualize it in a domain, and redirect the user to the content page. Each content page is titled with a preferred and unique term (Wikipedia 2015b), which is constructed with rules very similar to those in ISO 25964-1 (International Organization for Standardization 2011, 21-23).

The main problem with mapping of equivalence relations between a thesaurus and Wikipedia is that domain-oriented thesauri are frequently more specific than Wiki-

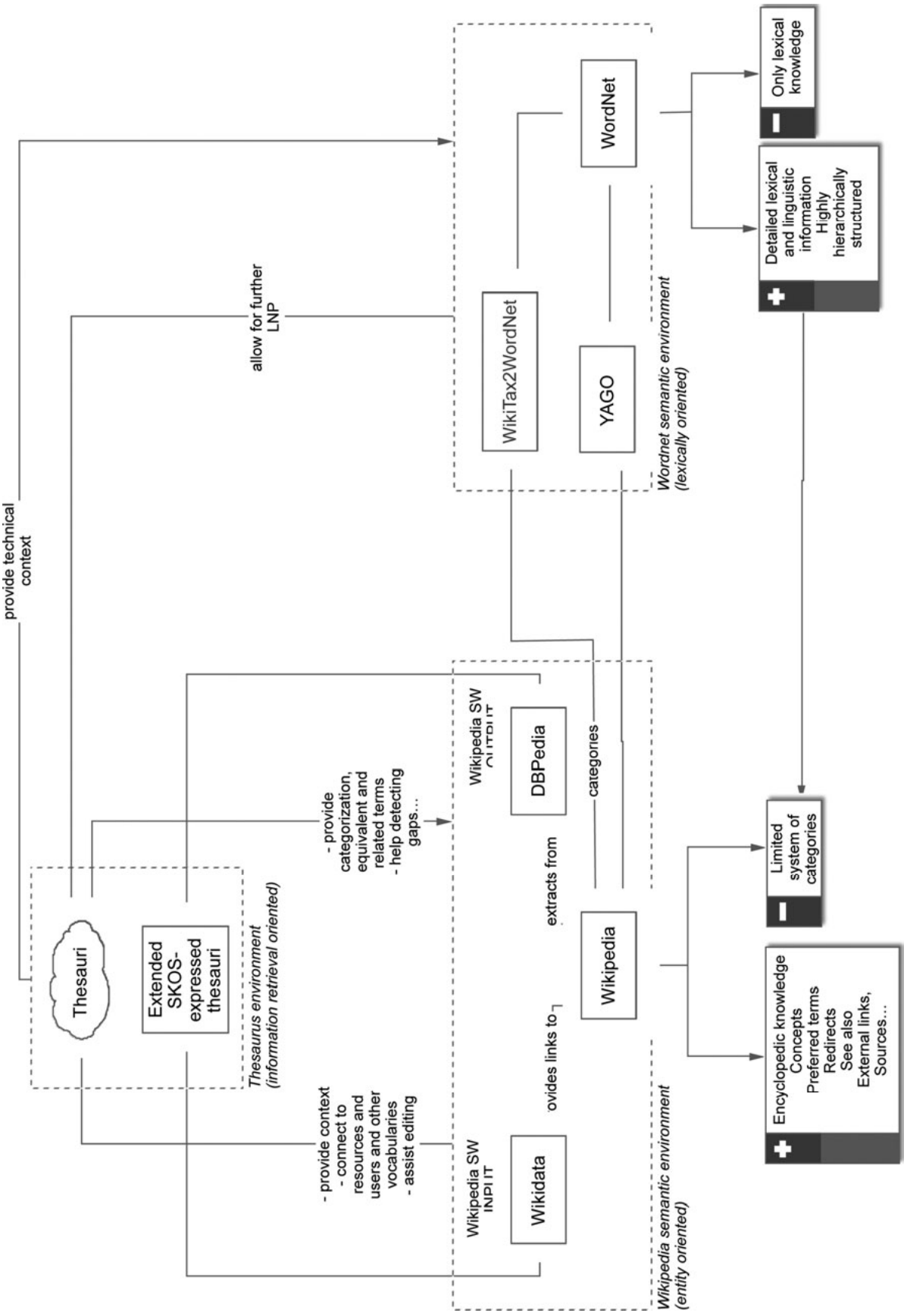

Figure 1. Thesauri, encyclopaedias and dictionaries cooperating in the Internet.

pedia. So, in many cases the terms will not correspond to Wikipedia articles, but to sections or even terms inside their text.

Hierarchical relations (hypernyms, hyponyms, siblings) are also very well codified inside Wikipedia, though with variable success and peculiarities. These relations are established through different devices: categories, lists and navigation templates, each with its advantages and disadvantages (Wikipedia 2015c). They allow for both top-down and bottom-up classificatory approaches. In particular, Wikipedia lists pages (Wikipedia 2015d) are very flexible and provide for the ordering of the concepts in customary or predefined forms.

The strongest classificatory device is the system of "categories" (Wikipedia 2015e, 2015f), added to Wikipedia in a later stage, in 2004. "Contents" were situated in the upper layer of categories, and behind them, categories and subcategories. Categories can be topics (and usually articles themselves) or sets. Though categories are hierarchical, they allow for poly-hierarchical structures of concepts, because the trees can overlap. This has prompted Voss (2006, 5) to state that Wikipedia categories are more like a thesaurus than a classification, because they cannot be considered as a real taxonomy, with mutually exclusive classes. In any case, the system is very flexible, and its characteristics are very similar to those of the hierarchical relations in ISO 25964-1.

Unfortunately, the Wikipedia system of categories presents several practical problems. The ontologies are frequently idiosyncratic, not based on the existence of sufficient definitions or even confusing, and very seldom and poorly used by Wikipedia contributors (Kittur, Chi and Suh 2009; Zazo-Rodríguez, Figuerola and Alonso-Berrocal 2015). As a result, Wikipedia is much more useful as a terminological and encyclopaedic database than as an explicit system of categories or taxonomy. On the other hand, this shortcoming offers an opportunity for thesauri and other structured vocabularies for information retrieval: "anchored" KOSs could also serve as potential alternative systematic access tools for Wikipedia, reciprocally benefiting the Wikipedia project, or be useful to complete and improve the Wikipedia's system of categories.

All hyperlinks that cannot be codified as any of the other relations can be potentially considered as associative relations (RT), but those listed in Table 2 provide controlled terms inside Wikipedia for associative relations. In Table 2, an example is provided to illustrate the equivalences between a thesaurus and Wikipedia by comparing the concept "bioethics" in the *Thesaurus Ethics in the Life Sciences* and Wikipedia.

6.0 An exploratory overview of the interconnection technicalities

Technically, the actual tagging of the mapping could be done by including any semantic web compliant statement in hidden text as invisible comments at the end of the reference section. However (Wikipedia 2015g), "Wikipedia is not a mirror or a repository of links, images, or media files." So, the proper way to do it is by using Wikipedia templates, a device to include complementary information inside a page in a Wikipedia project (Wikipedia 2015h, 2015i). Such a "KOS ref" template should at least point to Wikipedia's article for the KOS (every KOS should have an article in Wikipedia), and to the permanent URI of the corresponding term:

{{KOS ref |label= |wikititle= |kostermuri= }}

A label parameter would allow for the presentation of an acronym of the KOS, if the template contents were to be shown.

A Dublin Core based template to describe KOSs in Wikipedia would also be extremely convenient. Each KOS should have redirect pages from all its alternative name forms, and necessarily from a univocal acronym (e. g., preferred and unique for each KOS) that could also be used as a property value in graphs.

An alternative strategy would be to include the sentences in Wikidata, and then export them to the pertinent Wikipedia articles through infoboxes (Wikipedia 2015j). It would be more efficient, because for semantically formatting the information and therefore making it ready for automatic linking and reuse, Wikidata should store all the sentences that declare the thesaurus terms, and also the relations among them and towards Wikipedia pages. So, ensuring that all the information is properly codified in Wikipedia would be an important goal. This is being done with great success in closely-related fields, e.g., authority control of library subjects and natural and administrative toponyms, for which Wikidata's set of properties is very advanced (Wikidata 2015). Wikidata has the advantage that even the generic properties are very much more defined than in the Wikipedia structural elements. Thus, it allows for the finer mapping of the specific relations (partitive, instantial) that advanced semantic applications usually require (Aitchison and Dextre Clarke 2004, 15-16).

7.0 Conclusions

Through their history, thesauri have occupied important niches in the realms of heritage, pharmaceutical, and legal or scientific documentation, among other fields. In such fields, both the users and indexers are usually highly

	Thesaurus Ethics in the Life Sciences	*Wikipedia*
Concept	http://www2.drze.de/BELIT/thesaurus/show_tree.html?nr=1105&la=en	https://en.wikipedia.org/wiki/Bioethics
Term	bioethics	Bioethics, bioethical
Description	Branch of applied ethics that studies the value implications of practices and developments in the life sciences, medicine, health care and the environment (based on bioethics); cf. more specific terms as "medical ethics," "environmental ethics" or "animal ethics"	Bioethics is the study of the typically controversial ethical issues emerging from new situations and possibilities brought about by advances in biology and medicine. It is also moral discernment as it relates to medical policy and practice. Bioethicists are concerned with the ethical questions that arise in the relationships among life sciences, biotechnology, medicine, politics, law and philosophy. It also includes the study of the more commonplace questions of values ("the ethics of the ordinary") which arise in primary care and other branches of medicine.
Preferred term	bioethics	Bioethics
Equivalent terms	bioethical aspects bioethical issues biomedical ethics	[Redirect:] bioethical biomedical ethics
Broader and narrower terms, groups of concepts	BT: applied ethics NT: animal ethics medical ethics (8 terms under) neuroethics public health ethics	Ethics of science and technology, Social ethics, Philosophy of biology [87 "issues", 15 subcategories of Category:Bioethics, 135 pages (concepts) 3 navigation templates pages and links Stand-alone lists and links]
Top terms, subject areas	Subject area: Subject Area I: Ethics, Philosophy, Theology TT: Ethics	Category:Applied ethics Category:Ethics Category: Philosophy of biology (etc.)
Related terms	agricultural ethics environmental ethics	[See also:] Bioethics (journal) Cytoplasmic transfer Eugenics Feminist Approaches to Bioethics Hastings Center Report (journal) Johns Hopkins Berman Institute of Bioethics Kennedy Institute of Ethics Journal Medical law Neuroethics Preimplantation genetic diagnosis Resources for clinical ethics consultation The Convention on Human Rights and Biomedicine Yeshiva University Medical Ethics Society [plus further reading, other internal hyperlinks]
Inter-linguistic equivalents	Ger: Bioethik Fr: bioéthique	DE: Bioethik FR: Bioéthique [and other 40 languages]

Table 2. Comparison between "Bioethics" in the *Thesaurus Ethics in the Life Sciences* and Wikipedia.

competent. They also tend to require vocabulary control and very flexible, specific and exhaustive searches, because target information is either very expensive to produce, is not replicated in multiple redundant sources, or the costs of dismissing it can be very high. The thesaurus model has proved so successful and flexible that other alphabetical access tools that are used widely in these contexts have become converted to a thesaurus format, like certain subject heading lists. These niches have been preserved in the new Internet environment, and the thesaurus standards have developed toward dealing with an increasingly globalized and networked world by becoming multilingual by default and addressing the interoperability issue.

But to gain increased relevance for the general Internet user and, therefore, to achieve an extended user base, thesauri can also become ancillary to the big search engines, profiting from the opportunities that the development of the semantic web offers. So, *a priori*, cooperating with the big public knowledge graphs of the semantic

web seems another fruitful goal for future research and development efforts in the field of thesauri and other KOSs.

Complementarily, there is a need for a "hub" where thesauri and other KOSs can be connected, searched, exploited and commented on in different ways in the new digital environment. The semantic web allows for completely distributed interconnection, but humans and their deeds need reference points. This task could be pursued by the biggest KOS editors and in conjunction with the big Internet institutions and firms. Wikipedia could be a good candidate, because of its size, openness, social and collaborative nature, careful edition auditing, popularity, free foundation status, orientation towards preservation and knowledge organization ambitions.

To sum up, thesauri are arguably very well positioned to cooperate in fulfilling the need for interoperable category-based systems in the Internet and to thrive in the semantic environment. There are at least three reasons to think so. First, they have solved the problem of integrating concept-based and alphabetical access, leaving sufficient freedom for more taxonomic or terminological approaches to coexist. Because of that, they have been successfully proposed as a multilingual and interoperable hub for other KOSs. Second, they are currently the most standardized KOS model and have a clear and fully programmable formalization, so they are completely prepared to be understood by computer scientists and used in the development of actual applications. Third, their natural proximity to other knowledge repository tools, like virtual encyclopaedias and dictionaries, facilitates ways to interoperate with them. Because of this natural familiarity, thesauri find the knowledge graphs that are built from the big Internet encyclopaedias and dictionaries hospitable, and can also contribute to their sister tools by providing them with alternative structures and relations. So, exploring this synergy could contribute to assuring the relevance of thesauri in the new digital environment.

References

Aitchison, Jean and Stella Dextre Clarke. 2004. "The Thesaurus: A Historical Viewpoint, with a Look to the Future." *Cataloging & Classification Quarterly* 37 nos. 3/4: 5-21.

Bergman, Mike. 2011. "In Search of 'Gold Standards' for the Semantic web." *AI³*. http://www.mkbergman.com/947/in-search-of-gold-standards-for-the-semantic-web/

Bergman, Mike. 2015. "Shaping Wikipedia into a Computable Knowledge Base." *AI³*. http://www.mkbergman.com/1847/shaping-wikipedia-into-a-computable-knowledge-base/

Berners-Lee, Tim. 2007. "The Process of Designing Things in a Very Large Space." Web presentation slides. http://www.w3.org/2007/Talks/0509-www-keynote-tbl/

Berners-Lee, Tim, James Hendler and Ora Lassila. 2001. "The Semantic Web." *Scientific American* May 2001 (0501). http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Connaway, Lynn Silipigni and Ixcel M. Faniel. 2014. *Reordering Ranganathan: Shifting User Behaviors, Shifting Priorities*. Dublin, OH: OCLC Research. http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-reordering-ranganathan-2014.pdf

Dahlberg, Ingetraut. 1991. "Knowledge Organization, Thesauri, and Terminology." *International Classification* 18:133.

Dahlberg, Ingetraut. 1992. "Knowledge organization and Terminology: Philosophical and Linguistic Bases." *International Classification* 19:65-71.

Dahlberg, Ingetraut. 1995. "Current Trends in Knowledge Organization." In *Organización del conocimiento en sistemas de información: actas del i encuentro de ISKO 1*, pp. 7-25. http://www.iskoiberico.org/wp-content/uploads/2014/07/007-026_Dahlberg.pdf.

Dextre Clarke, Stella G. and Marcia Lei Zeng. 2012. "From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modelling." *Information standards quarterly* 24 no. 1:20-26.

Dextre Clarke, Stella G. 2008. "The Last 50 Years of Knowledge Organization: A Journey through my Personal Archives." *Journal of Information Science* 34:427-37.

Evans, Navino. 2015. "Wikidata: the potential of structured data to enable applications such as Histropedia." In *Knowledge Organization – Making a Difference: The Impact of Knowledge Organization on Society, Scholarship and Progress, ISKO UK Biennial Conference, 13th-14th July 2015, London.* http://www.iskouk.org/content/wikidata-potential-structured-data-enable-applications-such-histropedia

Färber, Michael, Basil Ell, Carsten Menne and Achim Rettinger. 2015. "A Comparative Survey of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO." *Semantic Web* 1:1–5.

Foskett, A. C. 1982. *The Subject Approach to Information*, 4th ed. London: Bingley.

Gardner P. P., J. Daub, J. Tate , B. L. Moore, I. H. Osuch, S. Griffiths-Jones, R. D. Finn, E. P. Nawrocki, D. L. Kolbe, S. R. Eddy and A. Bateman. 2011. "Rfam: Wikipedia, clans and the "decimal" release." *Nucleic Acids Research* 39:D141-5. doi:10.1093/nar/gkq1129

GEMET. 2012. General Multilingual Environmental Thesaurus, Version 3.1, 2012-07-20. https://www.eionet.europa.eu/gemet/

Georgas, Helen. 2013. "Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool." *portal: Libraries and the Academy* 13:165-85. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v013/13.2.georgas.html

Georgas, Helen. 2014. "Google vs. the Library (Part II): Student Search Patterns and Behaviors when Using Google and a Federated Search Tool." *portal: Libraries and the Academy* 14:503-32. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v014/14.4.georgas.html

Georgas, Helen. 2015. "Google vs. the Library (Part III): Assessing the Quality of Sources Found by Undergraduates." *portal: Libraries and the Academy* 15:133-61. https://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v015/15.1.georgas.html

Gilchrist, Alan. 1991. "The Standards Jungle." in *Standards for the International Exchange of Bibliographic Information: Papers Presented at a Course Held at the School of Library, Archive and Information Studies, University College London, 3-18 August 1990,* ed. Ia C. Mcllwaine, London: Library Association, 1991, p. 12.

Gilchrist, Alan. 2003. "Thesauri, taxonomies and ontologies - an etymological note." *Journal of Documentation* 59:7-18. doi:10.1108/00220410310457984

Howard, James and Silver Oliver. 2011. "Enhancing the BBC's news and sports coverage with an ontology-driven information architecture." In *Facets of Knowledge Organization: a tribute to Professor Brian Vickery 1918-2009. ISKO UK Biennial Conference, London, 4th-5th July 2011.* http://www.iskouk.org/content/enhancing-bbcs-news-and-sports-coverage-ontology-driven-information-architecture.

Hu, Xiaohua., Xiaodan Zhang, Caimei Lu, E. K. Park and X. Zhou. 2009. "Exploiting Wikipedia as External Knowledge for Document Clustering." In *KDD-09: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, Jun 28-Jul 01, 2009*. New York: ACM, pp. 389-96.

International Organization for Standardization. 2011. ISO 25964-1: *Information and Documentation—Thesauri and Interoperability with other Vocabularies - Part 1: Thesauri for Information Retrieval*. Geneva: International Organization for Standardization.

International Organization for Standardization. 2013. ISO 25964-2: *Information and Documentation—Thesauri and Interoperability with other Vocabularies - Part 2: Interoperability with other Vocabularies*. Geneva: International Organization for Standardization.

Kemman, Max, Martijn Kleppe and Stef Scagliola. 2014. "Just Google It." In *Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. Sheffield: HRI Online Publications, 2014*, ed. Clare Mills, Michael Pidd and Esther Ward.. Sheffield: HRI Online Publications. http://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman

Kittur, Aniket, Ed H. Chi and Bongson Suh. 2009. "What's in Wikipedia? Mapping Topics and Conflict Using Socially Annotated Category Structure." In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*. New York: AMC, pp. 1509-12. http://www-users.cs.umn.edu/~echi/papers/2009-CHI2009/p1509.pdf

Lancaster, F. W. 1986. *Vocabulary Control for Information Retrieval*, 2nd ed. Arlington, Virginia: Information Resources Press.

Lee, Rob. 2008. "Connecting Concepts: Joining up the BBC." *Semantic analysis technology in search of categories, concepts and context, London, 3 November 2008*. Presentation. http://www.iskouk.org/content/connecting-concepts-joining-bbc

Library of Congress. 2015. *LC Linked Data Service: Authorities and Vocabularies*. http://id.loc.gov.

Mesgari, Mostafa, Chitu Okoli, , Mohamad Mehdi, Finn Årup Nielsen and Arto Lanamäki. 2015. "The Sum of All Human Knowledge": A Systematic Review of Scholarly Research on the Content of Wikipedia. *Journal of the Association for Information Science and Technology* 66:219-45. doi:10.1002/asi.23172

Moz. 2015. "Google Algorithm Change History." https://moz.com/google-algorithm-change.

Okoli, Chitu, Mostafa Mesgari, Mohamad Mehdi, Finn Årup Nielsen and Arto Lanamäki. 2014. "Wikipedia in the Eyes of Its Beholders: A Systematic Review of Scholarly Research on Wikipedia Readers and Readership." *Journal of the Association for Information Science and Technology* 65:2381-2403. doi:10.1002/asi.23162

Pastor-Sánchez, Juan Antonio. 2013. "ISO-THES: Ampliando Skos a partir de la norma de tesauros ISO 25964." *Anuario ThinkEPI* 7:189-93.

Ponzetto, Simone Paolo. and Roberto Navigli. 2009. "Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia." *IJCAI* 9:2083-88.

Rapisardi, Elena, Sabina Di Franco and Marco Giardino. 2014. "Web Participatory Framework for Disaster Resilience: Coping with Information Deluge. Engineering Geology for Society and Territory." In *Engineering Geology for Society and Territory*. New York: Springer, pp.169-73. doi10.1007/978-3-319-09303-1_33

Schindler, Mathias and Denny Vrandecčić. 2011. "Introducing New Features to Wikipedia: Case Studies for Web Science." *IEEE Intelligent Systems* 26:56-61. doi. ieeecomputersociety.org/10.1109/MIS.2011.17

Singhal, Amit. 2012. "Introducing the Knowledge Graph: things, not strings." *Google Official Blog*, May 16, 2012.

https://googleblog.blogspot.com.es/2012/05/introducing-knowledge-graph-things-not.html

Voss, Jakob. 2006. "Collaborative Thesaurus Tagging the Wikipedia Way." http://arxiv.org/ftp/cs/papers/0604/0604036.pdf

Williamson, Nancy J. 1996. "Standards and Rules for Subject Access." *Cataloging & Classification Quarterly* 21 nos. 3-4:155-76.

Wikidata. 2015. *Wikidata:List_of_properties*. https://www.wikidata.org/wiki/Wikidata:List_of_properties/all.

Wikipedia. 2015a. *Wikipedia:Redirect*. https://en.wikipedia.org/wiki/Wikipedia:Redirect

Wikipedia. 2015b. *Wikipedia:Manual of Style/Disambiguation pages*. https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Disambiguation_pages

Wikipedia. 2015c. *Wikipedia:Categories, lists, and navigation templates*. https://en.wikipedia.org/wiki/Wikipedia:Categories,_lists,_and_navigation_templates#WP:NAVBOX

Wikipedia. 2015d. *Wikipedia:Stand-alone lists*. https://en.wikipedia.org/wiki/Wikipedia:Stand-alone_lists

Wikipedia. 2015e. *Wikipedia:Categorization*. https://en.wikipedia.org/wiki/ Wikipedia:Categorization

Wikipedia. 2015f. *Wikipedia:FAQ/Categorization*. https://en.wikipedia.org/wiki/ Wikipedia:FAQ/Categorization

Wikipedia. 2015g. *Wikipedia:What Wikipedia is not*. https://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not#Wikipedia_is_not_a_mirror_or_a_repository_of_links.2C_images.2C_or_media_files

Wikipedia. 2015h. *Help:Template*. https://en.wikipedia.org/wiki/Help:Template

Wikipedia. 2015i. *Help:A quick guide to templates*. https://en.wikipedia.org/wiki/Help:A_quick_guide_to_templates#Examples

Wikipedia. 2015j. *Wikipedia:Wikidata*. https://en.wikipedia.org/wiki/Wikipedia:Wikidata

Rodriguez, Ángel F. Zazo , Carlos G. Figuerola and José Luis Alonso Berrocal. 2015. "Edición de contenidos en un entorno colaborativo: el caso de la Wikipedia en español." *Scire* 21 no.2:57-67. http://www.ibersid.eu/ojs/index.php/scire/article/view/4243