

# The Information Retrieval Thesaurus\*

Stella G. Dextre Clarke

Luke House, West Hendred, Wantage, OX12 8RR, UK,  
<stella@lukehouse.org>

Thesauri have been a recurrent theme in Stella Dextre Clarke's career spanning forty-five years, mostly spent on information system development and database publishing. Graduating originally in chemistry, she earned a master's degree in information science before serving in a variety of information centres, where duties typically involved vocabulary design and maintenance, then bibliographic database management and subsequently consultancy. Her work on standards and on vocabulary development was recognized in 2006 when she won the Tony Kent Strix Award for outstanding achievement in information retrieval. Now retired, she is an honorary member of ISKO and the Vice-chair of ISKO UK.



Dextre Clarke, Stella G. 2019. "The Information Retrieval Thesaurus." *Knowledge Organization* 46(6): 439-459. 99 references. DOI:10.5771/0943-7444-2019-6-439.

**Abstract:** In the post-war period before computers were readily available, urgent demand for scientific and industrial development stimulated research and development (R&D) that led to the birth of the information retrieval thesaurus. This article traces the early history, speciation and progressive improvement of the thesaurus to reach the state now conveyed by guidelines in international and national standards. Despite doubts about the effectiveness of the thesaurus throughout this period, and notwithstanding the dominance of Google and other search engines in the information retrieval (IR) scene today, the thesaurus still plays a complementary part in the organization of knowledge and information resources. Success today depends on interoperability, and is opening up opportunities in linked data applications. At the same time, the IR demand from workers in the knowledge society drives interest in hybrid forms of knowledge organization system (KOS) that may pool the genes of thesauri with those of ontologies and classification schemes.

Received: 2 April 2019; Accepted: 10 April 2019

Keywords: thesaurus, thesauri, ISO 25964, terms, information retrieval

\* Derived from the article titled "Thesaurus (For Information Retrieval)" in the ISKO Encyclopedia of Knowledge Organization, Version 1.2 published 2019-03-07. Article category: KOS Kinds.

## 1.0 Introduction and clarification of scope

This article is about thesauri intended for use in information retrieval (IR—see note 1), rather than literary thesauri, which are generally designed for the different purpose of helping and inspiring the choice of words and phrases in normal discourse. *Roget's Thesaurus*, that very well-known literary thesaurus first published in 1852, long pre-dates the first IR thesaurus and probably inspired the invention of the latter. For this reason, there is some reference to literary thesauri in the history section of this article. In other sections, however, the term "thesaurus" invariably refers to the information retrieval thesaurus.

## 2.0 What is a thesaurus?

### 2.1 Purpose

The prime function of a thesaurus is to support information retrieval by guiding the choice of terms for indexing and searching. According to ISO 25964-1 (International Organization for Standardization 2011, Clause 4.1):

The traditional aim of a thesaurus is to guide the indexer and the searcher to choose the same term for the same concept ... a thesaurus should first list all the concepts that might be useful for retrieval purposes in a given domain. The concepts are represented by terms, and for each concept, one of the possible representations is selected as the preferred term ... Secondly, a thesaurus should present the preferred terms in such a way that people will easily identify the one(s) they need. This is achieved by establishing relationships between terms—and/or between concepts—and using the relationships to present the terms in a structured display.

Foskett (1980) lists seven purposes for a thesaurus, of which six could be considered subdivisions or sub-aspects of the main purpose cited above (As for his seventh purpose, a means of standardizing the use of terms in a given subject field, Foskett acknowledges that this is desirable rather than realistic). While the ISO 25964 description dates from 2011, it follows principles established long before. For example, Lancaster (1972, 25) explains:

Schultz (1967) has distinguished the functions of the information retrieval thesaurus from a thesaurus of the Roget type as follows. Roget's purpose was to give an author a choice of alternative words to express one concept; to display a set of words of similar meanings to allow an author to choose one that best suits his need. The information retrieval thesaurus tends to be more prescriptive. The thesaurus compiler chooses one term from among several possible, and directs the user to employ this one by means of references from synonyms and other alternative forms.

The use of preferred terms rather than language-independent codes or character strings is a key feature distinguishing the thesaurus from the classification schemes that were commonly used for IR before the advent of the thesaurus. Retrieval may seem simpler, to the layman, if it can be expressed in words rather than codes. But there is an ambiguity challenge to overcome—in the language of normal discourse one concept can be expressed in many different ways, and conversely one term can have many different meanings. To achieve the aim of always choosing the same term for the same concept, an artificial indexing language has to be established, in which synonyms are controlled, homographs are disentangled, and each preferred term is allowed only one meaning (although some may have very broad meanings). The thesaurus conveys that artificial language.

This *modus operandi* for the thesaurus became established in the 1960s. The computer was then in its infancy: small, primitive, and almost entirely unavailable to the communities of researchers and practitioners needing to retrieve information. Without computers to help, trained human intermediaries were needed at two critical stages of the best IR systems: to index and/or classify the source documents and (in the second stage) to perform searches of the same.

Several classic texts of this period, such as Gilchrist (1971), Lancaster (1972), Aitchison and Gilchrist (1972), and Soergel (1974), make it clear that the thesaurus is just one component in the whole IR system comprising a set of tools and procedures, all of which have to be designed in harmony. In those days, the IR system typically operated in isolation. While modern technology enables many more possibilities, needing even greater attention to compatibility among system components, today's standards still respect and support the original design principles.

## 2.2 Content and structure

The components of a thesaurus are most succinctly laid out in the UML (Unified Modelling Language) model shown in ISO 25964-1 and reproduced as Figure 1 below

(The model may also be seen on the official website at [www.niso.org/schemas/iso25964/](http://www.niso.org/schemas/iso25964/), and downloaded from [http://www.niso.org/schemas/iso25964/Model\\_2011-06-02.jpg](http://www.niso.org/schemas/iso25964/Model_2011-06-02.jpg). Key features are explained in Will (2012)).

Thus, the essential core of a thesaurus is a collection of concepts represented by terms and interlinked by relationships, of which the three main types are equivalence (between terms), hierarchical (between concepts) and associative (also between concepts). By long established convention, the tags USE and UF (Use For) precede preferred and non-preferred terms respectively, and the equally characteristic tags BT, NT and RT indicate broader, narrower and associatively related concepts respectively. Figure 2 illustrates how these simple elements are traditionally displayed. A great many in-house thesauri are built in this minimal way, without calling upon the many optional extras provided for in the data model. The alphabetical display in Figure 2 may optionally be supplemented by other types of presentation, as discussed in Section 4.3 below on “systematization.”

The thesaurus can also be visualized as a complex web of interlinked concepts in which each concept is labelled by one or more terms in one or more languages. It has these main features:

- The semantic scope of a concept is indicated partly by the totality of terms labelling it, partly by the hierarchical relationships linking it to broader and/or narrower concepts, and where this is not enough, by a scope note and sometimes term definitions.
- Admissible hierarchical relationships are of three types: generic, instancial or partitive (subject to some restrictions on the eligible types of partitive link). It is optionally possible to distinguish these types, using the tags BTG/NTG, BTI/NTI, BTP/NTP respectively.
- Admissible associative relationships apply to non-hierarchical situations wherever two concepts are so associated that an indexer or a searcher should consider using one of them as well as, or instead of, the other.
- Concepts may be presented and ordered in arrays with node labels, following the principles of facet analysis
- Concepts may also be grouped in loose structures to suit particular domains or applications
- Concepts not explicitly included in the thesaurus may be represented by combinations of preferred terms (in situations known as “compound equivalence”—for example: “coal mining USE coal + mining”)
- It is also possible to add metadata to terms, to concepts, to relationships and to the thesaurus as a whole, such as dates of introduction or change, version history, house-keeping data, copyright information, etc.



pesticides	UF: fumigants	plant products	NT: cereals
	BT: agrochemicals		fruits
	NT: fungicides		spices
	herbicides		vegetables
	insecticides	RT: plants	
	RT: pests		
pests		plants	
	NT: pest insects	RT: plant products	
	plant pests		
	RT: pesticides	<i>porkers</i>	
		USE: pigs	
pigs		poultry	
	UF: hogs	BT: livestock	
	<i>porkers</i>	NT: chickens	
	<i>sows</i>	ducks	
	BT: livestock	geese	
		turkeys	
plant pests		RT: eggs	
	BT: pests		
		sheep	
		BT: livestock	
		RT: wool	

Figure 2. Brief extract from a minimal monolingual thesaurus.

It should be stressed that many of these features are optional, enabling a variety of sophisticated uses, and should not deter straightforward use of the basics in simple applications. Detailed advice on all of them may be found in national and international standards—principally ISO 25964 and ANSI/NISO Z39.19 (National Information Standards Organization [2005] 2010)—and are further explicated in Aitchison et al (2000), Broughton (2006a) and Will 2012.

Despite availability of the guidance cited above, few current or past thesauri comply with the standards in every detail. Difficulties and divergences commonly occur in the following aspects:

- Rigorous conformity with guidelines for hierarchical relationships
- Rigorous facet analysis
- In a multilingual thesaurus, when and how to establish equivalence across languages
- When and how to admit complex concepts designated by compound terms
- Adoption of the data model

Surmounting such difficulties demands considerable expertise and time, adding to the expense of thesaurus construction and to doubts about cost-effectiveness, as noted in Sections 4.4 and 4.6 below.

### 2.3 Definitions

An authoritative definition of “thesaurus” may be found in the international standard ISO 25964-1 (Clause 2.62):

controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.

Although phrased differently, a broadly compatible definition is that in the American standard ANSI/NISO Z39.19-2005 (R2010) (National Information Standards Organization [2005] 2010, Clause 4.1):

A controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators should be employed reciprocally.

The above definitions derive their authority from the process of drafting and approving a standard, which requires agreement by a committee of experts and extensive consultation among the user community. But copious alternative definitions exist in a variety of texts, illustrating the

extent of confusion that surrounds the thesaurus. Many are intended to counteract loose use of the term “thesaurus,” which is commonly applied to any sort of knowledge organization system (KOS), such as a subject headings list, or to a set of synonym rings. Conversely, some vocabularies that could properly be described as thesauri may instead be called an ontology or a taxonomy.

## 2.4 Why the confusion?

Some of the current confusion may be explained by developments that emulate thesaural conventions in other types of KOS. In the 1990s, for example, a “thesaurification” project explored adaptation of some schedules of the UDC (Universal Decimal Classification) (Riesthuis and Bliedung 1991). This did not interfere with the primary function of the UDC as a classification scheme. Around the same time, the *Library of Congress Subject Headings (LCSH)* began to adopt thesaurus tags such as BT, NT and RT in its display (Thesaural use of these tags is illustrated in Figure 2). Today (3 August 2016), a Wikipedia entry for the *LCSH* claims that “The Library of Congress Subject Headings (LCSH) comprise a thesaurus ... of subject headings, maintained by the United States Library of Congress, for use in bibliographic records.” The use of thesaural conventions and BT/NT tags, however, does not make the *LCSH* a thesaurus, as pointed out long ago by Rolland-Thomas (1993) and illustrated more recently by Spero (2008 and 2012). The *LCSH* is fundamentally a subject headings list (defined as a “structured vocabulary comprising terms available for subject indexing, plus rules for combining them into pre-coordinated strings of terms where necessary” – ISO 25964, clause 2.57) rather than a thesaurus. Differences between the way subject headings and thesauri are used are discussed in De Keyser (2012).

Another explanation is that all the national and international standards for thesauri take the form of guidelines rather than mandatory instructions. Adopters, therefore, have a great deal of liberty to cherry-pick only the recommendations that suit the circumstances of their own thesaurus and ignore the rest.

A third part of the explanation is that very often the person charged with sorting out an organization’s information assets has little or no training in knowledge organization. If the decision is to develop an in-house indexing language or a filing structure, it may be built in whatever way comes easiest, and randomly named a “classification scheme” or a “thesaurus” or an “ontology” to suit the fashion of the day. The misnomer “thesaurus” has spread easily this way, leading to much confusion.

Even for the trained information professional, distinguishing between the different types of KOS can be hard. Over the years, many attempts at clarification have been

made (e.g. Fast et al, 2002; Garshol 2004; Hodge 2000; Kless et al 2012). Useful definitions of several KOS types may be found in ANSI/NISO Z39.19 and ISO 25964; ISO 25964-2 (International Organization for Standardization 2013) also brings out the similarities and the differences to provide for in the context of interoperability. Zeng (2008) casts further light by analysing and comparing features of many different types of KOS.

A different sort of confusion surrounds the basic roles of terms versus concepts. From the early days of thesaurus R&D, the basic aim was to index the semantic content of documents rather than the terminological content. Concepts useful for indexing were collected in a thesaurus, where they were organized and their inter-relationships were established. When a hierarchical relationship was established, the reciprocal links between the broader and narrower concepts might usefully have been designated BC (broader concept) and NC (narrower concept). In practice, however, they were named BT (broader term) and NT (narrower term), and this practice was adopted widely. The 1974 edition of ISO 2788 (International Organization for Standardization 1974) attempted to clarify by explaining “the hierarchical relation is represented by the references BROADER TERM (BT), representing the relation of a concept being superordinated, and NARROWER TERM (NT), indicating the reciprocal relation” (International Organization for Standardization (1974) clause 3.4.3). But it was too late—the misnomers have stuck, to this day.

Over the decades, this confusion has led to much misunderstanding among thesaurus users. Also, the software developed for thesaurus management has often adopted a data model in which the hierarchical and associative relationships are established between terms rather than between concepts, and this has impeded thesaurus interoperability. See discussion in Dextre Clarke and Zeng (2011).

## 3.0 How a thesaurus is used

### 3.1 For post-coordinate indexing and searching

The original thesaurus purpose and mode of use as declared in the standards is confirmed in many texts, such as Wellisch (1995, 475) “thesauri are primarily intended for indexing as well as for searching and retrieval from post-coordinated systems, in which an indexer may assign several descriptors to documents, while users may combine those descriptors to form search statements.” For more background on post-coordinated systems and their underpinning with controlled vocabularies and Boolean logic, see Sharp (1967), Lancaster (1972) or Dextre Clarke (2008). Today’s continuing demand for quality bibliographic databases supporting Boolean retrieval is upheld by Hjørland (2015).

As originally conceived, the act of consulting the thesaurus either for indexing or for searching can be time-consuming. When the user has worked out key concepts of the document to be indexed or the query to be investigated, he or she needs to find an entry point among the terms and/or groups available, and follow the network of relationships to establish the closest possible match in the thesaurus. Skill as well as patience and subject knowledge are needed, since thesauri vary greatly in quality and in format (see Section 5 below). Nowadays it is hard to find indexers with the requisite training, while trained end-users are very rare indeed. Therefore, modern systems tend to automate both indexing and searching, using an electronic version of the thesaurus.

Thesaurus-based indexing functions may be needed in situations such as library cataloguing, compilation of bibliographic databases and tagging/indexing of image collections. Generally, a software package designed for that application is used, with indexing support capabilities that vary from (at the simple end) speeding up the task of thesaurus navigation to (at the sophisticated end) delivering totally automatic indexing. In between are systems that validate and/or switch the indexer's terms, that select candidate terms for the human indexer to accept or reject and "suggester" systems for social tagging.

Forty years ago Caplan (1978) reported a number of failures in trials of thesaurus-based automatic indexing. More recently Lancaster (1998) provided a more promising account of the techniques available, but concluded (294) "even the most sophisticated of current automatic indexing procedures compare unfavourably with skilled human indexing." Eight years later, Tudhope et al (2006) were still calling for more research. Ten years later, research into metadata enrichment with thesaurus terms was outlined in Tudhope and Binding (2016). Kempf and Neubert (2016) describe several modes of implementation, including one that exploits inter-KOS mappings. Clause 16 of ISO 25964-1 advises on the thesaurus features needed to enable such functions.

Meanwhile, as described in Section 7 below, a new breed of KOS is emerging in the enterprise search sector, loosely named "taxonomy," and stimulating a demand for automatic categorization tools. Since some taxonomies share some features with thesauri (See ISO 25964-2, Clause 19), the associated R&D effort is already yielding progress that can be applied to thesaurus-based indexing. Unfortunately, a great many in-house applications go unreported in the research literature, including research by the vendors of software for automatic categorization. In the experience of this author, the support for thesaurus-based indexing in off-the-shelf library management packages is rarely as effective or user-friendly as it could be. Likewise, the quality of automatic or semi-automated indexing tools varies

greatly and much care is needed to obtain reliable outputs. Further discussion of automatic indexing is outside the scope of this article, particularly since most cases do not use a thesaurus.

Turning now to search applications, here too the electronic medium speeds up thesaurus navigation. Furthermore, with suitable software it enables broadening or narrowing a search at will. Consider, for example, a search for "packaging AND fruit." Relevant results would include any items dealing with the packaging of any of potentially hundreds of different types of fruit. The technique known as "search explosion" exploits the hierarchical relationships in a thesaurus to expand the search statement automatically and cover all those hundreds of fruit types. It is similarly possible to extend a search via associative relationships, and this is usually termed "search expansion." These and other search functions are reviewed in Shiri et al (2002), and further discussed in Shiri (2012). The case study of the *STW Thesaurus for Economics* by Kempf and Neubert (2016) illustrates similar techniques, and other ways in which a thesaurus can be used to enhance retrieval, even when the user is unaware of its support.

Evidently indexing and searching have moved on from the early days, when a thesaurus and its IR system could operate usefully in isolation and even without a computer. Thesaurus use in today's IR applications relies on electronic manipulation, involving transfer of data from one subsystem to another. Success depends on interoperability, i.e., the ability of systems and/or components to exchange information and to use the information that has been exchanged. There are now at least two main contexts for thesaurus interoperability:

- "Vertical integration" of the thesaurus with software for indexing or searching or occasionally some other IR function, as already described;
- "Horizontal engagement" of the thesaurus with another KOS (perhaps another thesaurus, or a subject headings list, or a classification scheme), typically requiring conversion of indexing and search expressions between the languages of the different KOSs.

The vertical context sees a thesaurus transformed from a static map of concepts, terms and relationships to a functioning system. The horizontal context crosses a different boundary, to be described next.

### 3.2 Networked uses, especially in the semantic web

A single search across multiple databases would be relatively straightforward if all used the same natural language, the same machine protocols and the same indexing language. To overcome the disparities found in real life, two

approaches to interoperability are especially relevant for KOSs, namely inter-vocabulary mappings and linked data.

A mapping is defined as a “relationship between a concept in one vocabulary and one or more concepts in another” (ISO 25964-2, clause 3.41). For example, an equivalence mapping between the concepts labelled “instant coffee” in one thesaurus and “soluble coffee” in another, would establish that they are viewed as identical for semantic purposes. Existence of mappings like this makes it easy to “translate” search queries for use in the corresponding IR systems, and/or to augment the metadata of resources indexed with either thesaurus. When sets of mappings are available between many KOSs, it opens the prospect of extending searches widely and multilingually.

The value of such mappings is demonstrated in the “Metathesaurus” of the Unified Medical Language System (UMLS) <[www.nlm.nih.gov/research/umls/](http://www.nlm.nih.gov/research/umls/)>, a semantic tool serving research in biomedicine, health care and related fields. It contains concepts from more than 100 KOSs as well as relationships from within the KOSs and many mappings between their respective concepts. Andrade and Lopes Gines de Lara (2016) assess its usefulness in retrieval from relevant databases. The influence of this construct has led some authors to speak of a “metathesaurus” wherever existing thesauri are integrated, linked or mapped together (Shiri 2012)—and a variety of ways is possible.

Not all mappings are as simple as equivalence. Dextre Clarke (2011a) enumerates a variety of mapping types in-

vestigated in research projects such as Renardus, MACS (Multilingual ACcess to Subjects), CrissCross and Ko-MoHe. ISO 25964-2 (International Organization for Standardization 2013) provides for hierarchical and associative mappings as well as equivalence. Hierarchical mappings are directional—either broader or narrower. Equivalence mappings subdivide into simple or compound; compound equivalence has two subtypes (intersecting or cumulative) while simple equivalence can be qualified as exact or inexact. Figure 3 shows the range of mapping types, with an example of each. Mapping statements should be expressed using the conventional tags (EQ, BM, NM etc) and the symbols shown.

Even more subtlety is possible in applications that need to distinguish between subtypes of hierarchical mapping. See Figure 4.

While thesaurus mapping projects have a much longer history (see, for example, Horsnell (1975) or Hood and Eberman (1990) or Hoppe (1996) reporting on UMLS work that began in 1986), the growth of the Internet and the WWW has made them more widely applicable. Thus, Zeng and Chan (2004) drew attention to opportunities emerging in the internet context and Vizine-Goetz et al (2004) described a labour-saving methodology. Mayr and Petras (2008) illustrated the possibilities. Several other mapping projects were reported in Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization (Boteram et al 2011).

Equivalence	
Simple:	Laptop computers EQ Notebook computers
Exact equivalence:	Aubergines =EQ Egg-plants
Inexact equivalence:	Horticulture ~EQ Gardening
Compound:	
Intersecting compound equivalence:	Women executives EQ Women + Executives
Cumulative compound equivalence:	Inland waterways EQ Rivers   Canals
Hierarchical	
Broader:	Streets BM Roads
Narrower:	Roads NM Streets
Associative	e-Learning RM Distance education

Figure 3. Mapping types in ISO 25964-2, with examples of mapping statements.

Hierarchical Subtype	Mapping statement example	Reciprocal example
Generic	rats BMG rodents	rodents NMG rats
Instantial	Paris BMI capital cities	capital cities NMI Paris
Whole-part	fingers BMP hands	hands NMP fingers

Figure 4. Mapping statements that distinguish between subtypes of hierarchical mapping.

Doerr (2000) analysed perceived semantic problems of thesaurus mapping. Confusingly for us today, his use of the term “mapping” differs from the ISO 25964 definition, applying to relationships within one vocabulary rather than between different ones. Thus, he deplored the weakness of thesaurus semantics for hierarchical relationships when compared with class subsumption in an ontology. Subsequent release of SKOS (see note 2) seems to have overcome or at least eased such problems (Tudhope and Binding 2016). According to Isaac and Baker (2015, 2)

The lack of a way to express less formal semantics hindered many early projects that tried to apply Semantic Web technology in the cultural sector by massaging existing knowledge organization systems into formal ontologies. Given the scope of the artifacts considered, this effort required considerable ontological debugging that was ultimately of dubious value. Indeed, most information retrieval scenarios using KOS for searching or browsing collections do not require more than the information that one concept is broader than another.

Establishment of the world wide web (WWW) has brought new opportunities and challenges for IR in general and for KOS use in particular. On the one hand, vast resources have come within our reach; on the other hand, individual resources may be expressed in a multiplicity of languages like the Tower of Babel. As pointed out in the context of the *STW Thesaurus for Economics*, “The Web changes everything” (Kempf and Neubert 2016, 162).

A particular breakthrough for KOS linkage was approval and release of the W3C recommendation *SKOS Reference* (Miles and Bechhofer, 2009) with specific guidance on how to publish mappings between KOSs. SKOS publication followed on from a research report by Miles (2006, 1) aiming “to develop a formal theory of retrieval using controlled vocabularies that have a simple and intuitive structure [such as thesauri, classification schemes, subject heading systems, taxonomies and other types of structured vocabulary], to provide the necessary theoretical foundations for the development of Semantic Web languages and design patterns for distributed retrieval applications.” Since 2009, a number of extensions have been added to SKOS to support interoperability in particular contexts; work on some mapping tools for thesauri is described in note 2.

Turning to the other main interoperability opportunity, the principles of linked data are set out in Tim Berners-Lee’s 2006 paper at [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html). As he explains (1), “The Semantic Web ... is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. Like the web of hypertext,

the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links [sic] between arbitrary things described by RDF.” For a KOS (such as a classification scheme or a thesaurus) the essential starting point is to publish the whole scheme on the web using resource description framework (RDF) syntax and giving each concept or class a uniform resource identifier (URI). Once that is in place, anyone anywhere can set up a direct link to any concept or class. For example, if a web page or a bibliographic record in a database on the web has been indexed with the thesaurus concept “renewable energy,” the person interested in that concept can move directly from the thesaurus to those and other relevant pages. This opens up the prospect for any thesaurus published on the web to act as a connecting hub for an immense literature in the subject field concerned, without any need to assemble the disparate documents in one collection or database. See note 2 for tools for hand-in-hand application of ISO 25964 and SKOS.

A vision of Wikipedia as the connecting linked data hub for hundreds of thesauri and other KOSs is outlined in a speculative paper from Garcia-Marco (2016). Kempf and Neubert (2016) show how the use of linked open data (LOD) is already paying off for the *STW Thesaurus for Economics*. Baca and Gill (2015) describe the challenges and long development path at the Getty Research Institute leading up to publication on the web of three KOSs that are very influential and widely used in the cultural heritage sector: *The Art & Architecture Thesaurus*, the *Union List of Artist Names* and the *Getty Thesaurus of Geographic Names*. They hope, thereby, to enable potential universal access to information in different formats and languages, about the works of art and countless other exhibits in museums, libraries and galleries around the world. The 2017 release by the European Commission of a new European Interoperability Framework (EIF) <[ec.europa.eu/isa2/eif](http://ec.europa.eu/isa2/eif)> further emphasizes the opportunities for public administrations to put linked data to work across the countries of Europe.

### 3.3 Other uses

Although not the primary purpose, thesauri may also be used for precoordinate indexing (Wellisch 1995, ISO 999:1996 (International Organization for Standardization, 1996)). When this is done, users of the precoordinate index (typically found at the back of a book) are not expected to consult a thesaurus (since cross-references to synonyms etc. may be embedded within the index). Conversely, a thesaurus may be used not for indexing but only for searching. This removes the need for compliance with the standards. See Section 5.7 below.

Educationalists sometimes argue that a thesaurus is valuable in its own right, for domain analysis, as a conceptual and terminological guide to a domain, and for development of the mind. Lykke Nielsen (2001, 778) states that “the thesaurus is a tool that helps individual users to get an understanding of the collective knowledge domain.” Broughton (<[www.iskouk.org/sites/default/files/ISKOUKGreatDebate-Broughton\\_0.pptx](http://www.iskouk.org/sites/default/files/ISKOUKGreatDebate-Broughton_0.pptx)>, slide 6) argues “the thesaurus teaches us to take a critical and analytical approach to the domain. It makes us think about the nature of concepts, the form of their labels [and about] their relationships,” and (slide 9) “there’s something fundamental about this approach to modelling information domains that should not be lightly abandoned.” More generally Soergel (2014) has argued that the construction of any sort of knowledge organization schema, particularly with entity-relationship modelling, facet analysis and a graphical presentation of concepts, is a useful learning discipline. Still more uses are emerging as the internet pervades the office and everyday living. To satisfy these new uses, however, the standard thesaurus model may need to evolve.

#### 4.0 History of thesaurus development and use

##### 4.1 Origins

To Peter Mark Roget, working in the middle of the nineteenth century, we owe the insight that it would be valuable to supply (Roget 1952, 559 emphasis original) “a collection of the words [the English language] contains and of the idiomatic combinations peculiar to it, arranged, not in alphabetical order, as they are in a dictionary, but according to the *ideas* which they express.” His aim, rather than information retrieval, was to help “find the word, or words, by which [an] idea may be most fitly and aptly expressed.”

An expanding volume of scientific and other scholarly literature in the first half of the twentieth century brought challenges for classification, the orthodox retrieval technology of the times. It led to developments such as faceted classification, post-coordinate indexing, and experiment with various sorts of cards, all of which were to prove helpful when the idea of an IR thesaurus was conceived.

According to Roberts (1984), the first suggestion of using a thesaurus in the context of IR came from Calvin Mooers in 1947. At around the same time C. L. Bernier and E. J. Crane made a similar, independent suggestion, but (Roberts, 272) “expressed the view that a general thesaurus was not an appropriate form for retrieval purposes.” Much experiment followed over the next decade, but none of the various thesaurus approaches described by Joyce and Needham (1958)—e.g., “term lattices”—seems to have prospered, nor come close to the style of thesaurus that was to emerge in 1959. It was after this ges-

tation period that (Roberts 1984, 281) “the first full-scale, operational in-house retrieval thesaurus [was produced] to solve pressing practical problems at E. I. Du Pont Nemours and Co., Inc., Wilmington, U.S.A.” Krooks and Lancaster (1993) credit Eugene Wall with developing the principles that determined the shape of this pioneering compilation.

##### 4.2 Period of ascendancy

More research and development (R&D) followed the 1959 birth of the IR thesaurus, one of the driving forces being the post-war preoccupation of the US military with a need for effective information management. Progress came with publication of the *Thesaurus of ASTLA Descriptors* (Armed Services Technical Information Agency 1960) and of the *Chemical Engineering Thesaurus* (American Institute of Chemical Engineers 1961), followed in 1967 by the landmark *Thesaurus of Engineering and Scientific Terms* (Office of Naval Research 1967), commonly known as *TEST*. A fuller description of these works can be found in Krooks and Lancaster (1993) and in Aitchison and Dextre Clarke (2004).

Widespread use of thesauri continued throughout the 1960s, 1970s and 1980s in IR systems that mostly relied on cards of various types, sizes and materials, including some that were sorted by machines (Dextre Clarke 2008; Sharp 1967). These were post-coordinate systems, which require each document to be indexed by selecting relevant terms from a controlled vocabulary such as a thesaurus. A thesaurus was used too by many of the bibliographic databases that were hosted online by services such as Lockheed’s Dialog system, followed later by CD-ROM distribution. Notable pioneers of construction methodology included Jean Viet, Jean Aitchison and Donald Leatherdale, who each produced a number of influential thesauri. Dextre Clarke (2008) provides a vivid account of how the tools and technology of those times were used.

Further impetus came from development of national and international standards for thesaurus construction, indicating the extent of interest from the information-using community. The most influential, listed in chronological order of their first editions, included:

Deutsches Institut für Normung. DIN 1463 Guidelines for the establishment and development of monolingual thesauri [translated title] 1972 (Now withdrawn, with ISO 25964-1 recommended in its place).

International Organization for Standardization. ISO 2788-1974 Documentation - Guidelines for the establishment and development of monolingual thesauri. 1st ed. International Organization for Standardization: Geneva, 1974 (Now superseded by ISO 25964-1).

American National Standards Institute. ANSI Z39.19-1974 American National Standard Guidelines for thesaurus structure, construction and use. American National Standards Institute: New York, 1974 (Now superseded by ANSI/NISO Z39.19-2005).

International Organization for Standardization. ISO 5964-1985. Documentation - Guidelines for the establishment and development of multilingual thesauri. International Organization for Standardization: Geneva, 1985 (Now superseded by ISO 25964-1).

These and other KOS standards are discussed in Dextre Clarke (2011b), although this article pre-dates publication of the two parts of ISO 25964, in 2011 and 2013 respectively (International Organization for Standards 2011 and 2013).

### 4.3 Systematization

As noted by Dextre Clarke (2001, 86) “standardisation has not brought uniformity.” Having the status of guidelines rather than mandatory requirements, all the standards left plenty of scope for continuing experiment. While nearly all published thesauri include an alphabetical list of terms (which may be as simple as the extract in Figure 2, or can show additional attributes and relationships) very often the alphabetical list is complemented by other types of display.

The great weakness of any alphabetical list is the need to know a term before one can find the corresponding concept(s). Thus, an alphabetical list does not respect Rogge’s vision that it would be useful to arrange terms systematically according to the ideas or concepts they represent. His literary insight applies equally in the context of information retrieval. A search concerning “wood” for example, could equally be expressed using the term “timber,” and an alphabetical list would place these terms far apart even though the underlying concept may well be the same.

The classic *TEST* thesaurus addressed this weakness by providing three indexes: permuted, hierarchical and by subject category. A derived style, slightly more elaborate, was followed in several thesauri designed by Jean Viet, including the influential *Macrothesaurus* from the Organization for Economic Cooperation and Development (OECD). A different approach was adopted by Aitchison, Gomersall and Ireland in their ground-breaking 1969 vocabulary *Thesaurifacet*, comprising a faceted classification fully integrated with a thesaurus. This approach relies on concept-based analysis from the very start, enabling elaboration of the faceted classification and subsequent derivation of a thesaurus. Aitchison and Dextre Clarke (2004) describe how Aitchison progressively refined and enhanced this technique over the decades to follow, designing a long line of thesauri such as the *UNESCO Thesaurus* (Aitchison 1977), the *BSI ROOT Thesaurus* (British Standards Institution 1981) and the *Inter-*

*national thesaurus of refugee terms* (Aitchison 1996). Biswas and Smith (1989) review a number of other efforts to combine a classification scheme with a thesaurus, especially the “Classaurus” and its variants developed in India by Bhattacharyya, Devadason and others. Broughton (2006a) also advocates facet analysis as the soundest basis for thesaurus construction and claims that “the generation of a thesaurus from its equivalent faceted classification is almost as automatic a process as thesaurus construction can ever hope to be” (Broughton, 2006b, 60).

Rather than a full-blooded classification, the systematic listing of preferred terms in *MeSH (Medical Subject Headings)* was a set of extensive hierarchical “tree structures” with an elaborate expressive notation that served both as a vocabulary look-up device and as a search key in the databases of MEDLARS (Medical Literature Analysis and Retrieval System) and later Medline. The first (1982) edition of the multilingual thesaurus *AGROVOC* (Leatherdale 1982), taking a different approach, avoided the need for a separate hierarchical section by embedding the complete upper and lower hierarchical context of each concept within the alphabetical display. See Figure 5.

#### HORSES

*uf* *equus caballus*  
**BT1** equidae  
**BT2** perissodactyla  
**BT3** mammals  
**BT4** vertebrates  
**BT1** livestock  
**BT2** domestic animals  
**BT3** animals  
NT1 draught horses  
NT1 foals  
NT1 mares  
NT1 racehorses  
NT1 saddle horses  
NT2 ponies  
NT1 stallions  
NT2 geldings  
**rt** meat animals

Figure 5. Entry for a preferred term in the English version of AGROVOC’s 1982 edition.

Throughout the 1960s, 1970s and 1980s, much of the experiment was constrained by the need to provide users with printed copies of the thesaurus, and update them regularly. Even after bibliographic databases such as Medline and AGRIS became available online through host services such as Dialog, or on CD-ROM discs, the corresponding thesauri were still widely distributed in hard copy. As late as 1990 the first edition of the influential *Art & Architecture*

*Thesaurus* was published conventionally, and even followed in 1994 by a second edition (Getty Art History Information Program 1994) in five weighty volumes, each over 500 pages. But after that, only the electronic version has been maintained. From the 1990s onwards, most new thesauri have been published in electronic media only.

If the focus is on an electronic version, not only are the costs and hassle of printed distribution eliminated, but also there is greater freedom to change the presentation frequently in response to feedback, and develop features that support indexing and searching of any associated databases. For example, the *STW Thesaurus for Economics* is nowadays published on the web (see <http://zbw.eu/stw/version/latest/about.en.html>), enabling immediate searching of the EconBiz database and at the same time supporting linked open data applications. Similarly, *AGROVOC* has in the twenty-first century undergone huge redevelopment, exploiting SKOS to enable linked data applications and incorporating some new relationship types in its “agrontology” (Caracciolo and Keizer 2014). Figure 6 shows how an entry in *AGROVOC* looks in 2017, with simultaneous views of hierarchy and all the language equivalents, etc., on one screen, plus easy hyperlinks to all related concepts.

#### 4.4 Maturity, senescence or rejuvenation?

A tailing-off in the popularity of thesauri has occurred from approximately the end of the 1980s, probably due to in-

creasing availability of desktop computers, as well as the rise of the internet (Dextre Clarke 2008). The new technologies have enabled alternative retrieval methods that for most applications appear less expensive than post-coordinate indexing plus thesaurus development and maintenance. From that time onwards, while a good thesaurus works no less effectively than before, its role has been relegated to relatively fewer search applications (Dextre Clarke 2016), such as retrieval from image collections (MacFarlane 2016), cultural heritage collections and bibliographic databases. In these situations, it still brings benefits, especially when implemented in linked data mode (Tudhope and Binding 2016). Shiri (2012) paints an optimistic picture of the opportunities.

Latest versions of the standards Z39-19, ISO 25964-1 and ISO 25964-2 are dated 2005 (reaffirmed 2010), 2011 (confirmed 2017) and 2013 respectively. While in them the basic principles of thesaurus design show little change from previous versions, it is clear the context in which a thesaurus operates has changed markedly. Interoperability is now the key to success—and is reflected in the content of the standards. See note 3 for some clarification of the differences between these standards.

Some authors believe the way relationships are treated in a thesaurus could usefully evolve. Alexiev et al (2014) suggest that interoperability, inferencing capabilities and the reliability of search explosion would all be improved by more rigorous discrimination between the three types of hierarchical relationship allowed by the standards. Hjørland (2016)

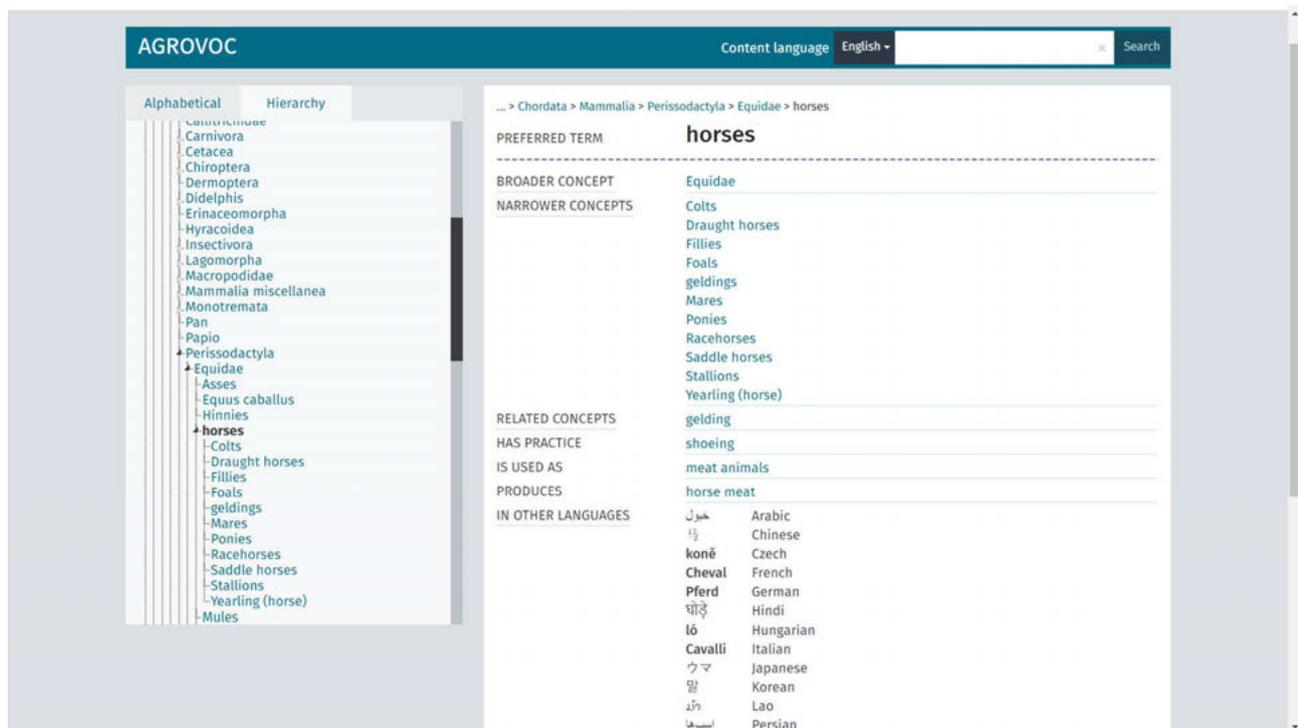


Figure 6. Online display of the entry for “horses” in AGROVOC, January 2017.

asks why thesauri “bundle” different kinds of semantic relations into one relationship type—the associative—and suggests (151) that “thesauri would probably be improved” if they adopted some attributes from ontologies, in particular the avoidance of “standardized limitations on the kind of semantic relations used.” He points out further that the most useful types of relationship to specify may vary from one domain to another. Vernau ([www.iskouk.org/sites/default/files/190215Debate\\_1-JudiVernau.mp3](http://www.iskouk.org/sites/default/files/190215Debate_1-JudiVernau.mp3)) too has called for changes in the approach to relationships.

Enthusiasts for change may like to note that the current standards are already permissive of developments, e.g., the inclusion of new customized relationships, that do not transgress the existing rules. All international standards are reviewed on a five-year cycle, enabling proponents to make the case for revision as soon as such developments have proved their worth. The passage of time will tell whether the thesaurus continues as before in its relatively few niche applications, or whether it blossoms into new networked opportunities, perhaps revitalised by an infusion of ideas from ontologies and other types of KOS. Section 7 below summarizes the challenges and opportunities for continuing exploitation and evolution.

## 5.0 Types and styles of thesaurus

### 5.1 Overview

The Basel Register of Thesauri, Ontologies & Classifications (BARTOC) at [www.BARTOC.org](http://www.BARTOC.org) and the Taxonomy Warehouse <[taxonomywarehouse.com](http://taxonomywarehouse.com)> list hundreds of thesauri among other types of KOS. Despite their astonishing variability in aspects such as subject scope, size, specificity, function, format, layout, language, quality of construction, etc., it is hard to divide them into distinct species or types. Much of the variation seems stylistic rather than fundamental, with one style borrowing features from another and a proliferation of hybrids. This section will, therefore, start by describing the “bare minimum” that can be expected in any IR thesaurus, and continue with some discussion of frequently observed differences in style, before discussing some categories of thesaurus that might or might not be considered distinct types.

### 5.2 The bare minimum

These features are indispensable in a traditional functioning thesaurus:

- For every concept deemed worth indexing/searching, inclusion of as many as possible of the terms that might represent it, with one of these selected as “preferred”;
- Any hierarchical or close associative relationships between the concepts should be shown;
- Some kind of display or index must enable users to look up the terms and concepts.

Taken together these three requirements typically lead to a list of all terms and relationships, with entries alphabetically arranged, in the style of the extract in Figure 2. Alongside these traditional requirements, it is worth noting a trend towards applications in which the thesaurus is implemented behind the scenes; this reduces or obviates the need for any kind of display or, indeed, for designating the preferred term for a concept.

While the vocabulary illustrated in Figure 2 complies with the standards, a more ambitious thesaurus would also incorporate scope notes, history notes, faceted arrays introduced by node labels, concept groups and other optional features. The data model in Figure 1 points to very many opportunities for enhancing a thesaurus in ways that are standards-compliant and supportive of interoperability in networked applications. As to format, the alphabetic list is often supplemented by other displays to help users find the right term, such as a classified display, a set of hierarchical trees, a permuted index or even a graphical display. Thesauri that were developed to serve a particular database sometimes show extras, such as the number of postings for each term. In lieu of explicit display, some of the extra features may be hidden, invoked only as functions of a retrieval system.

### 5.3 Different styles for different communities

In this section, we discuss presentational differences, which may not be fundamental to thesaurus operation but can still influence user acceptance and hence retrieval effectiveness. In certain domains, a particularly influential thesaurus has influenced the development of subsequent vocabularies. For example, *MeSH* (the *Medical Subject Headings*, list of the National Library Medicine, first issued in 1960) early on developed a set of “tree structures” having a distinctive style of notation directly functional in database retrieval (and still visible in today’s MeSH online, see <https://meshb.nlm.nih.gov/#/treeSearch>). Both the trees and the notation were emulated in later thesauri for medical applications, such as *EMTREE*, the thesaurus of Elsevier’s *Excerpta Medica* database. Similarly, the *Art & Architecture Thesaurus* first published in 1990, with a very distinctive style of facet-driven hierarchical display incorporating “guide terms,” has inspired much thesaurus development work in the heritage sector worldwide.

Other historical influences have been thesaurus maintenance software and preferences of the original designers. Thus, thesauri managed with the CAIRS or the TIKIT

package characteristically presented “stop terms” and “go terms,” and some user communities still look for these in every new thesaurus. Many thesauri funded in the twentieth century by the Commission of the European Communities used the ASTUTE software, which generated alphabetical displays with entries in the style of Figure 5. The style and conceptual approach of pioneers Jean Viet and Jean Aitchison (see 4.3 above) inspired many followers to apply the principles of classification to thesaurus development.

A good many other variations on format and layout of printed thesauri, including some graphical representations, are described and/or illustrated in Foskett (1980). Shiri (2012) provides an update, including screen layouts for electronic thesauri, to be discussed next.

#### 5.4 Electronic thesauri

Arguably, an electronic format is just another stylistic variation, not affecting the fundamentals. Electronic thesauri have been around from the early days of online bibliographic databases such as AGRIS, CAB Abstracts, INSPEC, Engineering Index, ERIC, etc, that chose to provide their search vocabularies as a printed thesaurus and as an electronic version of the same, integrated to greater or lesser degree with the search functions of the database. In such cases, the underlying content and structure of both online and printed versions are the same.

That said, the electronic medium offers enhanced opportunities for thesaurus design, maintenance, presentation and implementation, enabling interactive retrieval functions for the users as described in Section 3 above. To exploit linked data and other semantic web applications, the electronic thesaurus should be published on the web in the format of the W3C standard *SKOS Simple Knowledge Organization System Reference* (see <https://www.w3.org/TR/skos-reference/>). ISO 25964-2 (International Organization for Standardization 2013) gives further advice on semantic interoperability between thesauri and other KOSs. Shiri (2012) discusses several examples and offers guidelines for the design of thesaurus-enhanced search interfaces.

#### 5.5 Multilingual vs monolingual thesauri

All the stylistic variations described so far can apply to monolingual or to multilingual thesauri. The inclusion of more than one language is not just another variable—it makes a big difference to design, maintenance and use. Compare the illustration in Figure 2 with that in Figure 7, for a bilingual thesaurus (English/Spanish). The display illustrated is for use by speakers of English; an alternative, language-inverted display for speakers of Spanish would show all the terms and relationships for that language.

Multilingual thesauri can be subdivided into two types—symmetrical or not. In a symmetrical thesaurus, every concept has a preferred term in each of the languages, and the scope and relational structure is identical in each. In a non-symmetrical thesaurus, not every concept need be represented in all the languages, and the hierarchical structure may vary from one language to another to accommodate cultural differences. See more discussion and examples in Working Group on Guidelines for Multilingual Thesauri of IFLA Classification and Indexing Section (2009) and Hudon (2001).

#### 5.6 Macro- and micro-thesauri

An original aim of the OECD’s *Macrothesaurus* published in 1972 was to “create a documentary language for processing information in the broad field of economic and social development, while striving for compatibility with sectoral thesauri serving agriculture, industry, labour, education, population, science, technology, culture communication, health and the environment” (Viet 1972, v). Both the name and the aim were popular, and so years later the term “macrothesaurus” with a small “m” was borrowed as a generic name for any broad-level thesaurus that either contains or is aligned with a number of “microthesauri” having greater specificity in a more limited field.

The normal situation according to Aitchison et al (2000, 177) is for built-in compatibility, with the specialized “microthesaurus” being “mapped onto, and entirely integrated within, the hierarchical structure of some broader thesaurus, the macrothesaurus.” They acknowledge, however, that sometimes the macrothesaurus is a separate entity, managed independently of any corresponding microthesauri.

In practice, it is not easy to maintain alignment between specialized thesauri used by different communities, unless management is centralized. Among the successful examples today is *EUROVOC*, the Multilingual Thesaurus of the European Union, which is structured into twenty-one “domains,” each of which is subdivided into a number of “microthesauri.” Concepts belong to more than one microthesaurus if appropriate. Each microthesaurus contains a hierarchically structured list of concepts, terms and relationships, and can be downloaded separately (see and browse at <http://eurovoc.europa.eu/drupal/>).

#### 5.7 The search thesaurus

The search thesaurus is one designed for use, not in indexing, but only at the search stage. (see more discussion in Aitchison et al (2000) and Lykke Nielsen (2004)). At first glance, this would not seem to make it very different. And indeed, sometimes a normal standards-compliant thesau-

<p>pesticides</p> <p>es: plaguicidas</p> <p>UF: fumigants</p> <p>BT: agrochemicals</p> <p>NT: fungicides</p> <p>herbicidas</p> <p>insecticidas</p> <p>RT: pests</p> <p>pests</p> <p>es: plagas</p> <p>NT: pest insects</p> <p>plant pests</p> <p>RT: pesticides</p> <p>pigs</p> <p>es: cerdos</p> <p>UF: hogs</p> <p>porkers</p> <p>sows</p> <p>BT: livestock</p> <p>plant pests</p> <p>es: plagas de plantas</p> <p>BT: pests</p>	<p>plant products</p> <p>es: productos de origen vegetal</p> <p>NT: cereals</p> <p>fruits</p> <p>spices</p> <p>vegetables</p> <p>RT: plants</p> <p>plants</p> <p>es: plantas</p> <p>RT: plant products</p> <p>porkers</p> <p>USE: pigs</p> <p>poultry</p> <p>es: aves de corral</p> <p>BT: livestock</p> <p>NT: chickens</p> <p>ducks</p> <p>geese</p> <p>turkeys</p> <p>RT: eggs</p> <p>sheep</p> <p>es: ovinos</p> <p>BT: livestock</p> <p>RT: wool</p>
--	---

Figure 7. Extract from one Alphabetical display of a Spanish/English thesaurus.

**Pioneers.** Pioneer(s). Early settler(s). Pilgrim(s). Frontiers(man,men). Backwoods(man,men). Early colonist(s). Homesteader(s). Early immigrant(s). Consider also: discoverer(s), explorer(s), pathfinder(s), scout(s), trailblazer(s), leader(s). See also: Explorers; Pioneering; Scientists.

Figure 8. Sample entry from Knapp’s search thesaurus (Knapp 1993).

rus is applied only at the search stage, and then described as a “search thesaurus.”

A deeper study, however, reminds us of the way a standard thesaurus is designed to work (ISO 25964-1, Clause 4.1): “The concepts are represented by terms, and for each concept, one of the possible representations is selected as the preferred term.” In the case of the thesaurus shown in Figure 2, for example, an indexer would assign the term “pigs” to every item in the collection that deals with pigs or sows or hogs or porkers. The searcher would use only the term “pigs” to retrieve all these items. But if the same tool was being used as a search thesaurus, indexing would not have taken place. The searcher would have to look for “pigs OR sows OR hogs OR porkers.”

Thus, the notion of a “preferred term” is inapplicable to a search thesaurus designed as such. Standards such as ISO 25964 become irrelevant, allowing even greater freedom of content, style and structure. Lopez-Huertas (1997) proposes one example, structured very differently from the standard thesaurus. Another fully worked example is Knapp’s “The contemporary thesaurus of social science terms and synonyms” which attempts to remind readers of many alternative ways of expressing the same idea, using a layout quite different from the standard (see Figure 8). In practice, not many such works have been published.

The converse of the search thesaurus is the “indexing thesaurus,” to be used for indexing and not for search. While applications are sometimes found in which indexing

is enhanced automatically with the help of a thesaurus, a standard thesaurus is usually applied, rather than one designed for indexing alone.

## 6.0 Performance and evaluation

While the criteria presented by Mader and Haslhofer (2015) apply to a range of KOSs and not specifically thesauri, they do help evaluate interoperability in the context of SKOS use, for any controlled vocabulary. Much earlier, Owens and Cochrane (2004) described four approaches—structural, formative, observational and comparative—to thesaurus evaluation. None of these directly measures the effectiveness with which a thesaurus succeeds in the purpose for which it was intended—retrieving information. Such a measure is difficult if not impossible to devise, partly because the thesaurus is only one of several components in the retrieval system, and partly because there are so many variables in the context of use. Lengthy experiments in the 1960s and 1970s studied the effects on precision and recall as different features of indexing languages were tested, but ultimately failed to provide conclusive support for the use of any controlled vocabulary (Keen 1973; Soergel 1994; Svenonius 1986; Dextre Clarke 2001). Despite efforts over many years, we still do not have definitive proof that development and use of a thesaurus is a worthwhile investment. Dextre Clarke (2016) provides an account of the continuing debate.

The long debate was highlighted at an event run by the UK Chapter of ISKO in February 2015 (see proceedings at <http://www.iskouk.org/content/great-debate>), with a subsequent special issue of *Knowledge Organization* (v. 43, no.3 2016) devoted wholly to questioning the future of thesauri. Despite reservations expressed by Hjørland (2016), that depiction of the future makes it clear the context of KOS use is changing, and the thesaurus evolving to occupy new roles and opportunities. Although quantitative proof of efficacy may be lacking, there is plenty of qualitative evidence of thesauri prospering and supporting users in some key areas of a changing environment. Modes of evaluation may have to adapt to reflect the new context.

## 7.0 The future of thesauri

The thesaurus as conceived by the current national and international standards is still based on the assumption “that human intellect is usually involved in the selection of indexing terms and in the selection of search terms. If both the indexer and the searcher are guided to choose the same term for the same concept, then relevant documents will be retrieved. This is the main principle underlying thesaurus design” (ISO 25964-1, Introduction vi). Nowadays, opportunities to apply the thesaurus may shrink because the

trained indexer and searcher are increasingly scarce. End-users are largely unaware of thesauri (this is confirmed, for example, by Greenberg (2004)); trained indexers and searchers are usually deemed unaffordable. Areas where the thesaurus seems most likely to survive and flourish include:

- Applications where no other IR technology is effective (e.g. indexing of still images)
- Applications with an income to pay the costs of indexing and thesaurus maintenance (e.g. profitable bibliographic databases)
- Applications with new benefits to spread the costs over more outcomes (e.g., via linked data and/or mapping services)
- Enhanced implementation behind the scenes, so that users get the benefits without the discomforts of look-up (e.g., with automatic-aided indexing);
- Evolved or hybrid KOSs, with new characteristics that are now in demand (e.g., by incorporation of domain-specific relationships).

Examples of developments like these may be found in the special issue of *Knowledge Organization* (2016) mentioned above and in Shiri (2012).

Simultaneously as true thesauri still thrive in the types of application just listed, a parallel future may lie in their gradual transformation under the banner “taxonomy.” This term, long applied to the practice and science of classification and especially the Linnaean classification of biological organisms, has been widely applied since the 1990s to a variety of KOSs found in electronic media. Applications include corporate intranets, online retail sales outlets, digital libraries, public sector advice websites, as well as displacement of the thesaurus in some of its traditional occupations. White (2016) points to the value of KO tools and techniques in some of these contexts.

There’s still little uniformity among the “taxonomies” being developed for such applications, which may be simple heading lists, or may be complex hybrids that combine features from thesauri, traditional classification schemes, faceted schemes, ontologies and other types of KOS. In comparison with the widespread adoption of web search engines, their value is barely recognized. But there certainly is a very large need and opportunity for the principles of knowledge organization to be applied towards helping millions of workers in the knowledge society to find information resources of all kinds. The names we shall find for the emerging hybrid vocabularies are hard to predict, but we can safely say that the thesaurus will pass some of its genes into new tools for searching the cyberworld to come.

## 8.0 Further reading

Full details of the publications in this list may be found in the references, listed at the end of this article.

a) Useful registers of thesauri (among other types of KOS) may be found in the Basel Register of Thesauri, Ontologies & Classifications at [www.BARTOC.org](http://www.BARTOC.org) and the Taxonomy Warehouse at [taxonomywarehouse.com](http://taxonomywarehouse.com). The latter site also lists relevant events, blogs, publishers and links to some associated products such as software.

b) Two guides to thesaurus construction are recommended:

- Aitchison, Jean, Alan Gilchrist and David Bawden (2000) *Thesaurus construction and use: a practical manual*.
- Broughton, Vanda (2006) *Essential thesaurus construction*.

The first of these also carries an extensive bibliography. Both guides draw heavily on the then current national and international standards for thesauri: ISO 2788, ISO 5964, BS8723 and ANSI/NISO Z39.19 (of which the first three have since been withdrawn, superseded by ISO 25964).

c) Specialist software is needed for thesaurus construction and maintenance. While ISO 25964-1 and ANSI/NISO Z39.19 both advise on the functionality required, see also the following article and its list of references:

- Will, Leonard (2010) “*Thesaurus Management Software*.”

d) A special issue of *Cataloging & Classification Quarterly* (Roe and Thomas 2004) was devoted to “The thesaurus: review, renaissance and revision,” in which all the articles have useful reference lists. Contents include:

- Aitchison and Dextre Clarke (2004) “The thesaurus: a historical viewpoint, with a look to the future.”
- Greenberg (2004) “User comprehension and searching with information retrieval thesauri.”
- Johnson (2004) “Distributed thesaurus web services.”
- Landry (2004) “Multilingual subject access: the linking approach of MACS.”
- Lykke Nielsen (2004) “Thesaurus construction: key issues and selected readings.”
- Owens and Cochrane (2004) “Thesaurus evaluation.”
- Riesland (2004) “Tools of the Trade: Vocabulary Management Software.”
- Shearer (2004) “A practical exercise in building a Thesaurus.”
- Thomas (2004) “Teach yourself thesaurus: Exercises, readings, resources.”

- Will (2004) “Thesaurus consultancy.”

e) A special issue of *Knowledge Organization* (v. 43, no. 3 2016) was devoted to a continuation of the ISKO-UK debate “This house believes that the traditional thesaurus has no place in modern information retrieval.” All the articles have useful reference lists. Contents include:

- Dextre Clarke (2016) “Origins and trajectory of the long thesaurus debate.”
- Dextre Clarke and Vernau (2016a) “Guest editorial: the thesaurus debate continues.”
- Dextre Clarke and Vernau (2016b) “Questions and answers on current developments inspired by the thesaurus tradition.”
- Garcia-Marco (2016) “Enhancing the visibility and relevance of thesauri in the Web: searching for a hub in the linked data environment.”
- Hjørland (2016) “Does the traditional thesaurus have a place in modern information retrieval?”
- Kempf and Neubert (2016) “The role of thesauri in an open Web: a case study of the STW Thesaurus for Economics.”
- MacFarlane (2016) “Knowledge Organisation and its role in multimedia information retrieval.”
- Tudhope and Binding (2016) “Still quite popular after all those years - the continued relevance of the information retrieval thesaurus.”
- White (2016) “The value of taxonomies, thesauri and metadata in enterprise search.”

f) The interest group NKOS (Networked Knowledge Organization Systems/Services/Structures) runs projects and activities enabling all sorts of KOS as networked interactive information services, especially through the internet. On its website at [nkos.slis.kent.edu](http://nkos.slis.kent.edu) are links to many relevant publications and the proceedings of past events. The presentations from the NKOS workshops in USA and Europe are especially helpful in pointing to the future of thesauri and other KOSs. For an overview, see Tudhope and Lykke Nielsen 2006.

## Notes

1. “Information retrieval” is used here broadly to mean “the activity of obtaining information resources relevant to an information need from one or more collections of information resources” (definition adapted from Wikipedia). It is not limited to use in systems held on computer.
2. As explained on the W3C (World Wide Web Consortium) website, “SKOS [Simple Knowledge Organization Systems] is an area of work developing specifica-

tions and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.” A key product of this programme is the *SKOS Simple Knowledge Organization System Reference* (Miles and Bechhofer 2009), a common data model for sharing and linking knowledge organization systems via the web. Development work on this specification took place around the same time as the development of BS 8723 and ISO 25964, with regular communication between the corresponding teams, so that a high degree of compatibility was achieved. The main difference between them can be summarized as follows: While ISO 25964-1 serves as a standard for construction of thesauri, SKOS is a standard for publishing thesauri and other types of KOS on the web. Likewise, ISO 25964-2 recommends the sort of mappings that can be established between one KOS and another; SKOS presents a way of expressing these when published to the web.

The data models of these two standards are not identical, because ISO 25964 must provide for the needs of all sorts of thesauri (whether for web use or for other applications) while SKOS [29] must provide for all sorts of KOS (including classification schemes and many others that do not comply with ISO 25964). Good alignment between the two made possible a set of correspondences <ISO25964-SKOSXL-MADS-2013-12-11.pdf> between components of the data models, developed by the same teams of authors. Where the basic SKOS data model lacked a construct corresponding to a feature of the ISO 25964 model, the SKOS-XL <[www.w3.org/TR/skos-reference/skos-xl.html](http://www.w3.org/TR/skos-reference/skos-xl.html)> model was used, supplemented by additional proposals where necessary. Care was taken to avoid incompatibility with another project to align SKOS with MADS [www.loc.gov/standards/mads/](http://www.loc.gov/standards/mads/). Based on the documented correspondence table, an RDF schema that provides a machine-readable version for these mappings as well as for the elements from the ISO 25964 model was developed and made available at <http://purl.org/iso25964/skos-thes>. At the time of writing (July 2017), this latter site is unavailable but work is in hand to restore it. More background on all the above developments is provided at <[www.niso.org/schemas/iso25964/#skos](http://www.niso.org/schemas/iso25964/#skos)>.

It should be noted that exploitation of these interoperability standards and opportunities demands skill and attention to detail. Some practical examples and cautionary tales are provided in De Smedt (2012) and Lindenthal (2012).

3. The American standard ANSI/NISO Z39.19 and the International standard ISO 25964 are broadly aligned. Here are some key points of similarity or difference:

- The latest version of Z39.19 is a single document, issued in 2005 and reaffirmed in 2010; whereas ISO 25964 is a two-part standard, of which the first part (ISO 25964-1) was issued in 2011 and confirmed in 2017 while the second (ISO 25964-2) was issued in 2013.
- The scope of Z39.19 is broadly comparable with that of ISO 25964-1, but Z39.19 covers several types of monolingual controlled vocabulary—lists of controlled terms, synonym rings, taxonomies and thesauri—while ISO 25964-1 deals only with thesauri, both monolingual and multilingual.
- ISO 25964-1 provides a data model but Z39.19 does not
- The whole of ISO 25964-2 (ninety-nine pages) deals with interoperability between thesauri and other types of KOS, including classification schemes, taxonomies, subject heading schemes, ontologies, terminologies, name authority lists and synonym rings. The treatment of interoperability in Z39.19 is contained in one clause of eight pages (plus a five-page appendix), in which multilingual thesauri are treated as a special case of interoperability.
- Z39.19 may be downloaded free of charge from the NISO website, whereas each part of ISO 25964 currently (2017) costs 198 Swiss francs, from the ISO store.

In view of the relatively wider scope of ISO 25964 in respect of thesauri, including in-depth treatment of interoperability and provision of a data model, it has been referenced more often than Z39.19 in this article.

## References

- Aitchison, Jean. 1977. *UNESCO Thesaurus*. Paris: UNESCO.
- Aitchison, Jean. 1996. *International Thesaurus of Refugee Terms*. 2nd ed. New York: United Nations High Commissioner for Refugees.
- Aitchison, Jean and Stella Dextre Clarke. 2004. “The Thesaurus: A Historical Viewpoint, with a Look to the Future.” *Cataloging & Classification Quarterly* 37, no. 3/4: 5-21.
- Aitchison, Jean and Alan Gilchrist. 1972. *Thesaurus Construction: A Practical Manual*. London: Aslib.
- Aitchison, Jean, Alan Gilchrist and David Bawden. 2000. *Thesaurus Construction and Use: A Practical Manual*. 4th ed. London: Aslib.

- Aitchison, Jean, Alan Gomersall and Ralph Ireland. 1969. *Thesaurifacet: A Thesaurus and Faceted Classification for Engineering and Related Subjects*. Whetstone, Leicester: English Electric.
- Alexiev, Vladimir, Antoine Isaac and Jutta Lindenthal. 2014. "On Compositionality of ISO 25964 Hierarchical Relationships (BTG, BTP, BTI)." Paper presented at the 13th European Networked Knowledge Organization Systems (NKOS) Workshop held during the DL2014 Conference. <https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2014/content/NKOS2014-abstract-alexiev-isaac-lindenthal.pdf>
- American Institute of Chemical Engineers. 1961. *Chemical Engineering Thesaurus: A Wordbook for Use with the Concept Coordination System of Information Storage and Retrieval*. New York: American Institute of Chemical Engineers.
- American National Standards Institute. 1974. *American National Standard Guidelines for Thesaurus Structure, Construction and Use*. ANSI Z39.19-1974. New York: American National Standards Institute.
- Andrade, Juliatti de and Marilda Lopes Ginez de Lara. 2016. "Interoperability and Mapping between Knowledge Organization Systems: Metathesaurus – Unified Medical Language System of the National Library of Medicine." *Knowledge Organization* 43: 107-12.
- Armed Services Technical Information Agency. 1960. *Thesaurus of ASTLA Descriptors*. Arlington, VA: Armed Services Technical Information Agency.
- Baca, Murtha and Melissa Gill. 2015. "Encoding Multilingual Knowledge Systems in the Digital Age: The Getty Vocabularies." *Knowledge Organization* 42: 232-43.
- Berners-Lee, Tim. 2006. "Linked Data." [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
- Biswas, Subal C. and Fred Smith. 1989. "Classed Thesauri in Indexing and Retrieval: A Literature Review and Critical Evaluation of Online Alphabetic Classaurus." *Library and Information Science Research* 11, no. 2: 109-41.
- Boteram, Felix, Winfried Goedert and Jessica Hubrich, eds. 2011. *Concepts in Context: Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization*. Würzburg: Ergon.
- BSI (British Standards Institution). 1981. *BSI ROOT Thesaurus*. Milton Keynes, England: British Standards Institution.
- Broughton, Vanda. 2006a. *Essential Thesaurus Construction*. London: Facet.
- Broughton, Vanda. 2006b. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." *Aslib Proceedings* 58: 49-72.
- Caplan, Priscilla Louise. 1978. "Thesaurus-Based Automatic Indexing: A Study of Indexing Failure." PhD diss., University of North Carolina.
- Caracciolo, Caterina and Johannes Keizer. 2014. "What KOS can do, with the Proper Tools Available. About AGROVOC, Edited in Vocbench and Used in the AGRIS Web Application." Paper presented at Knowledge Organization - Making a Difference: ISKO UK Biennial Conference. <http://www.iskouk.org/sites/default/files/CaraccioloSlidesISKO-UK2015.pdf>
- De Keyser, Pierre. 2012. "Introduction to Subject Headings and Thesauri." In *Indexing: From Thesauri to the Semantic Web*, ed. Pierre De Keyser. Oxford: Chandos, 1-37.
- De Smedt, Johan. 2012. "Exchanging ISO 25964-1 Thesauri Data using RDF, SKOS and SKOS-XL." Paper presented at the 11th European Networked Knowledge Organization Systems (NKOS) Workshop, Sept 2012. [https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2012/presentations/ISO25964-mapping-to-SKOS-XL\\_TPDL-2012-09-27v6.pdf](https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2012/presentations/ISO25964-mapping-to-SKOS-XL_TPDL-2012-09-27v6.pdf)
- Dextre Clarke, Stella G. 2001. "Organising Access to Information by Subject." In *Handbook of Information Management*. 8th ed., ed. Alison Scammell. London: Aslib, 72-110.
- 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.
- Dextre Clarke, Stella G. 2011a. "In Pursuit of Interoperability: Can we Standardize Mapping Types?" In *Concepts in Context: Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization*, ed. Felix Boteram, Winfried Goedert and Jessica Hubrich. Würzburg: Ergon, 91-109.
- Dextre Clarke, Stella G. 2011b. "Knowledge Organization System Standards." In *Encyclopedia of Library and Information Science*. 3rd ed., ed. Marcia J. Bates and Mary Niles Maack. Boca Raton, FL: CRC Press, 3176-83.
- Dextre Clarke, Stella G. 2016. "Origins and Trajectory of the Long Thesaurus Debate." *Knowledge Organization* 43: 138-44.
- Dextre Clarke, Stella G. and Marcia Lei Zeng. 2011. "Standard Spotlight: From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling." *Information Standards Quarterly* 25, no. 1: 20-6.
- Dextre Clarke, Stella G. and Vernau, Judi. 2016a. "Guest Editorial: The Thesaurus Debate Continues." *Knowledge Organization* 43: 135-7.
- Dextre Clarke, Stella G. and Vernau, Judi. 2016a. "Questions and Answers on Current Developments Inspired by the Thesaurus Tradition." *Knowledge Organization* 43:203-9.
- Doerr, Martin. 2000. "Semantic Problems of Thesaurus Mapping." *Journal of Digital Information* 1, no. 8. <http://jodi.ecc.soton.ac.uk/Articles/v01/i08/Doerr/>

- Elsevier Science. 1991- *EMTREE: The Life Science Thesaurus*. Amsterdam: Elsevier Science.
- Fast, Karl. 2003. "Controlled Vocabularies: A Glosso-Thesaurus." *boxesandarrows* (blog), October 25. <http://boxesandarrows.com/controlled-vocabularies-a-glosso-thesaurus/>
- Foskett, Douglas J. 1980. "Thesaurus." *Encyclopedia of Library and Information Science*, ed. Allen Kent, Harold Lancour and J. E. Daily. New York: Marcel Dekker, 30: 416-62.
- Garcia-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: 193-202.
- Garshol, Lars Marius. 2004. "Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it All." *Journal of Information Science* 30: 378-91.
- Getty Art History Information Program. 1994. *Art & Architecture Thesaurus*. 2nd ed. Oxford: Oxford University Press.
- Gilchrist, Alan. 1971. *The Thesaurus in Retrieval*. London: Aslib.
- Greenberg, Jane. 2004. "User Comprehension and Searching with Information Retrieval Thesauri." *Cataloging & Classification Quarterly* 37, no. 3/4: 103-20.
- Hjørland, Birger. 2015. "Classical Databases and Knowledge Organization: a Case for Boolean Retrieval and Human Decision-Making during Searches." *Journal of the Association for Information Science and Technology* 66, no. 8: 1559-75.
- Hjørland, Birger. 2016. "Does the Traditional Thesaurus Have a Place in Modern Information Retrieval?" *Knowledge Organization* 43: 145-59.
- Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington DC: Council on Library and Information Resources. <http://www.clir.org/pubs/reports/pub91/contents.html>
- Hood, Martha W. and Christine Ebermann. 1990. "Reconciling the CAB Thesaurus and AGROVOC." *LAALD Quarterly Bulletin* 35, no. 3: 181-5.
- Hoppe, Stephan. 1996. "The UMLS – a Model for Knowledge Integration in a Subject Field." In *Compatibility and Integration of Order Systems: Research Seminar Proceedings of the TIP/ISKO Meeting*, ed. Ingetraut Dahlberg and Krystyna Siwek. Warsaw: Wydawnictwo SBP, 97-110.
- Horsnell, Verina. 1975. "The Intermediate Lexicon: An Aid to International Co-Operation." *Aslib Proceedings* 27: 57-66.
- Hudon, Michele. 2001. "Relationships in Multilingual Thesauri." In *Relationships in the Organization of Knowledge*, ed. Carol A Bean and Rebecca Green. Dordrecht: Kluwer, 67-80.
- IFLA Working Group on Guidelines for Multilingual Thesauri. 2009. "Guidelines for Multilingual Thesauri." <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>
- ISO (International Organization for Standardization). 1974. *Guidelines for the Establishment and Development of Monolingual Thesauri*. ISO 2788-1974. Geneva: International Organization for Standardization.
- ISO (International Organization for Standardization). 1985. *Guidelines for the Establishment and Development of Multilingual Thesauri*. ISO 5964-1985. Geneva: International Organization for Standardization.
- ISO (International Organization for Standardization). 1996. *Guidelines for the Content, Organization and Presentation of Indexes*. ISO 999:1996. Geneva: International Organization for Standardization.
- ISO (International Organization for Standardization). 2011. *Thesauri for Information Retrieval. Information and Documentation*. Part 1 of *Thesauri and Interoperability with Other Vocabularies*. ISO 25964-1. Geneva: International Organization for Standardization.
- ISO (International Organization for Standardization). 2013. *Interoperability with Other Vocabularies*. Part 2 of *Thesauri and Interoperability with Other Vocabularies*. ISO 25964-2. Geneva: International Organization for Standardization.
- Isaac, Antoine and Thomas Baker. 2015. "Linked Data Practice at Different Levels of Semantic Precision: The Perspective of Libraries, Archives and Museums." *ASIS&T Bulletin*. [http://www.asis.org/Bulletin/Apr-15/AprMay15\\_Isaac\\_Baker.html](http://www.asis.org/Bulletin/Apr-15/AprMay15_Isaac_Baker.html)
- Joyce, T. and R. M. Needham. 1958. "The Thesaurus Approach to Information Retrieval." *American Documentation* 9: 192-97.
- Keen, E. Michael. 1973. "The Aberystwyth Index Languages Test." *Journal of Documentation* 29: 1-35.
- Kempf, Andreas Oscar and Joachim Neubert. 2016. "The Role of Thesauri in an Open Web: a Case Study of the STW Thesaurus for Economics." *Knowledge Organization* 43: 160-73.
- Kless, Daniel, Simon Milton and Edmund Kazmierczak. 2012. "Relationships and Relata in Ontologies and Thesauri: Differences and Similarities." *Applied Ontology* 7, no. 4: 401-28.
- Knapp, Sara D. 1993. *The Contemporary Thesaurus of Social Science Terms and Synonyms. A Guide for Natural Language Computer Searching*. Phoenix: Oryx.
- Krooks D. A. and Lancaster F. W. 1993. "The Evolution of Guidelines for Thesaurus Construction." *Libri* 43, no. 4: 326-42.
- Lancaster, F. W. 1972. *Vocabulary Control for Information Retrieval*. Washington, D.C.: Information Resources Press.
- Lancaster, F. W. 1998. *Indexing and Abstracting in Theory and Practice*. London: LA Publishing.

- Leatherdale, Donald. 1982. *AGROVOC*. Rome: Food and Agriculture Organization of the United Nations and Commission of the European Communities.
- Lindenthal, Jutta. 2012. "Ambiguities in Representing Thesauri Using Extended SKOS - Examples from Real Life". Paper presented at the 11th European Networked Knowledge Organization Systems (NKOS) Workshop, Sept 2012. [https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2012/presentations/TPDL2012\\_NKOS\\_Ambiguities\\_R1.pdf](https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2012/presentations/TPDL2012_NKOS_Ambiguities_R1.pdf)
- Lopez-Huertas, M. J. 1997. "Thesaurus Structure Design: A Conceptual Approach for Improved Interaction." *Journal of Documentation* 53: 139-77.
- Lykke Nielsen, Marianne. 2001. "A Framework for Work Task Based Thesaurus Design." *Journal of Documentation* 57: 774-97.
- Lykke Nielsen, Marianne. 2004. "Thesaurus Construction: Key Issues and Selected Readings." *Cataloging & Classification Quarterly* 37, no. 3/4: 57-74.
- MacFarlane, Andrew. 2016. "Knowledge Organisation and its Role in Multimedia Information Retrieval." *Knowledge Organization* 43: 180-3.
- Mader, Christian and Bernard Haslhofer. 2015. "Quality Criteria for Controlled Web Vocabularies." Paper presented at International Conference on Theory and Practice of Digital Libraries 2011, NKOS Workshop. [http://eprints.cs.univie.ac.at/2923/1/tpdl\\_workshop.pdf](http://eprints.cs.univie.ac.at/2923/1/tpdl_workshop.pdf)
- Mayr, Philipp and Vivien Petras. 2008. "Building a Terminology Network for Search: the KoMoHe Project." In *Metadata for Semantic and Social Applications: Proceedings of the International Conference on Dublin Core and Metadata Applications*, ed. Jane Greenberg and Wolfgang Klas. Singapore: Dublin Core Metadata Initiative, 177-82.
- Miles, Alistair. 2006. "Retrieval and the Semantic Web: Incorporating a Theory of Retrieval Using Structured Vocabularies." PhD diss., Oxford Brookes University. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.7212&rep=rep1&type=pdf>
- Miles, Alistair and Bechhofer, Sean, eds. 2009. "SKOS Simple Knowledge Organization System Reference: W3C Recommendation." <http://www.w3.org/TR/skos-reference>
- NISO (National Information Standards Organization). (2005) 2010. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. ANSI/NISO Z39.19-2005 (R2010). Bethesda, MD: National Information Standards Organization.
- National Library of Medicine. 1960-. *Medical Subject Headings (MeSH)*. Bethesda, MD: National Library of Medicine.
- Office of Naval Research. 1967. *Thesaurus of Engineering and Scientific Terms (TEST)*. New York: Engineers Joint Council and US Department of Defense.
- Owens, Lesley Ann and Pauline Atherton Cochrane. 2004. "Thesaurus Evaluation." *Cataloging & Classification Quarterly* 37, no. 3/4: 87-102.
- Riesthuis, Gerhard J. A. and Steffi Bliedung. 1991. "The-saurification of the UDC." In *Tools for Knowledge Organization and the Human Interface. Proceedings of the 1st International ISKO Conference*, ed. Robert Fugmann. Frankfurt/Main: Indeks Verlag, 2, 109-17.
- Roberts, N. 1984. "The Pre-History of the Information Retrieval Thesaurus." *Journal of Documentation* 40: 271-85.
- Roe, Sandra K. and Alan R. Thomas, eds. 2004. "The Thesaurus: Review Renaissance and Revision." Special issue, *Cataloging & Classification Quarterly* 37, nos. 3-4.
- Roget, Peter. 1952. *Everyman's Thesaurus of English Words and Phrases*, ed. D. C. Browning. London: J. M. Dent.
- Rolland-Thomas, P. 1993. "Thesaural Codes: An Appraisal of their Use in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 16, no. 2: 71-91.
- Sharp, J. R. 1967. "Information Retrieval." In *Handbook of special librarianship and information work*. 3rd ed., ed. Wilfred Ashworth. London: Aslib, 141-232.
- Shiri, Ali. 2012. *Powering search. The Role of Thesauri in New Information Environments*. Medford, New Jersey: Information Today.
- Shiri, Ali Asghar, Crawford Revie and Gobinda Chowdhury. 2002. "Thesaurus-Enhanced Search Interfaces." *Journal of Information Science* 28: 111-22.
- Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles, CA: Melville.
- Soergel, Dagobert. 1994. "Indexing and Retrieval Performance: The Logical Evidence." *Journal of the American Society for Information Science* 45: 589-99.
- Soergel, Dagobert. 2014. "Knowledge Organization for Learning." In *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects: Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014, Kraków, Poland*, ed. Wieslaw Babik. Advances in Knowledge Organization 14. Würzburg: Ergon, 22-32.
- Spero, Simon E. 208. "LCSH is to Thesaurus as Doorbell is to Mammal: Visualizing Structural Problems in the Library of Congress Subject Headings." Paper presented at the DCMI International Conference on Dublin Core and Metadata Applications in Berlin, 22-26 September 2008. <http://dcpapers.dublincore.org/pubs/article/view/937>
- Spero, Simon E. 2012. "What, if Anything, is a Subdivision?" In *Facets of Knowledge Organization: Proceedings of the ISKO UK Second Biennial Conference*, ed. Alan Gilchrist and Judi Vernau. Bingley, UK: Emerald, 69-83.
- Svenonius, Elaine. 1986. "Unanswered Questions in the Design of Controlled Vocabularies." *Journal of the American Society for Information Science* 37: 331-40.
- Tudhope, Douglas and Ceri Binding. 2016. "Still Quite Popular after All Those Years - the Continued Relevance of

- the Information Retrieval Thesaurus.” *Knowledge Organization* 43: 174-79.
- Tudhope, Douglas, Traugott Koch and Rachel Heery. 2006. “Terminology Services and Technology: JISC State of the Art Review.” <http://www.ukoln.ac.uk/terminology/JISC-review2006.html>
- Tudhope, Douglas and Marianne Lykke Nielsen. 2006. “Introduction to Knowledge Systems and Services. *New Review of Hypermedia and Multimedia* 12, no. 1: 3-9.
- Viet, Jean. 1972. *Macrothesaurus for Information Processing in the Field of Economic and Social Development*. Paris: OECD Development Centre.
- Vizine-Goetz, Diane, Carol Hickey, Andrew Houghton and Roger Thompson. 2004. “Vocabulary Mapping for Terminology Services.” *Journal of Digital Information* 4, no. 4. <https://journals.tdl.org/jodi/index.php/jodi/article/view/114/113>
- Wellisch, Hans H. 1995. *Indexing from A to Z*. 2nd ed. New York: H W Wilson.
- White, Martin. 2016. “The Value of Taxonomies, Thesauri and Metadata in Enterprise Search.” *Knowledge Organization* 43: 184-92.
- Will, Leonard. 2010. “Thesaurus Management Software.” In *Encyclopedia of Library and Information Sciences*. 3rd ed., ed. Marcia J. Bates and Mary Niles Maack. Boca Raton, FL: CRC Press, 5238-46.
- Will, Leonard. 2012. “The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus.” *Bulletin of the American Society for Information Science and Technology* 38, no. 4. [https://www.asis.org/Bulletin/Apr-12/AprMay12\\_Will.pdf](https://www.asis.org/Bulletin/Apr-12/AprMay12_Will.pdf)
- Zeng, Marcia Lei. 2008. “Knowledge Organization Systems (KOS).” *Knowledge Organization* 35: 160-82.
- Zeng, Marcia Lei and Lois Mai Chan. 2004. “Trends and Issues in Establishing Interoperability among Knowledge Organization Systems.” *Journal of the American Society for Information Science and Technology* 55: 377-95.