

Contents

Special Issue:

The Great Debate:

“This House Believes that the Traditional Thesaurus has no Place in Modern Information Retrieval.”

Background to the Debate

Stella G. Dextre Clarke and Judi Vernau.

Guest Editorial:

The Thesaurus Debate Continues 135

Stella G. Dextre Clarke.

Origins and Trajectory of the Long Thesaurus

Debate 138

Article *for* the Proposition

Birger Hjørland.

Does the Traditional Thesaurus Have a Place

in Modern Information Retrieval? 145

Articles *against* the Proposition

Andreas Oskar Kempf and Joachim Neubert.

The Role of Thesauri in an Open Web:

A Case Study of the STW Thesaurus for

Economics 160

Douglas Tudhope and Ceri Binding.

Still Quite Popular After all Those Years—

The Continued Relevance of the Information

Retrieval Thesaurus 174

Andrew MacFarlane.

Knowledge Organisation and its Role

in Multimedia Information Retrieval 180

Martin White.

The Value of Taxonomies, Thesauri

and Metadata in Enterprise Search 184

Francisco-Javier García-Marco.

Enhancing the Visibility and Relevance

of Thesauri in the Web: Searching for a Hub

in the Linked Data Environment 193

Points of view

Stella G. Dextre Clarke and Judi Vernau.

Questions and Answers on Current Developments

Inspired by the Thesaurus Tradition 203

Letters to the Guest Editors 210

Contents pages

Dextre Clarke, Stella G. 2016. “Origins and Trajectory of the Long Thesaurus Debate.” *Knowledge Organization* 43 no. 3: 138-144. 33 references.

Abstract: The information retrieval thesaurus emerged in the 1950s, settled down to a more-or-less standard format in the 1970s and has continued to evolve marginally since then. Throughout its whole lifetime, doubts have been expressed about its efficacy with emphasis latterly on cost-effectiveness. Prolonged testing of different styles of index language in the 1970s failed to settle the doubts. The arena occupied by the debate has moved from small isolated databases in the post-war era to diverse situations nowadays with the whole Internet at one extreme and small in-house collections at the other. Sophisticated statistical techniques now dominate the retrieval landscape on the Internet but leave opportunities for the thesaurus and other knowledge organization techniques in niches such as image libraries and corporate intranets. The promise of an ontology-driven semantic web with linked data resources opens another opportunity. Thus much scope remains for research to establish the usefulness of the thesaurus in these places and to inspire its continuing evolution.

Hjørland, Birger. 2016. “Does the Traditional Thesaurus Have a Place in Modern Information Retrieval?” *Knowledge Organization* 43 no. 3: 145-159. 82 references.

Abstract: The introduction (1.0) of this article considers the status of the thesaurus within LIS and asks about the future prospect for thesauri. The main following points are: (2.0) Any knowledge organization system (KOS) is today threatened by Google-like systems, and it is therefore important to consider if there still is a need for knowledge organization (KO) in the traditional sense. (3.0) A thesaurus is a somewhat reduced form of KOS compared to, for example, an ontology, and its “bundling” and restricted number of semantic relations has never been justified theoretically or empirically. Which semantic relations are most fruitful for a given task is thus an open question, and different domains may need different kinds of KOS including different sets of relations between terms. (4.0) A KOS is a controlled vocabulary (CV) and should not be considered a “perfect language” (Eco 1995) that is simply able to remove the ambiguity of natural language; rather much ambiguity in language represents a battle between many “voices” (Bakhtin 1981) or “paradigms” (Kuhn 1962). In this perspective, a specific KOS, e.g. a specific thesaurus, is just one “voice” among many voices, and that voice has to demonstrate its authority and utility. It is concluded (5.0) that the traditional thesaurus does not have a place in modern information retrieval, but that more flexible semantic tools based on proper studies of domains will always be important.

Kempf, Andreas Oskar and Joachim Neubert. 2016. “The Role of Thesauri in an Open Web: A Case Study of the STW Thesaurus for Economics.” *Knowledge Organization* 43 no. 3: 160-173. 21 references.

Abstract: This paper illustrates the changing role of thesauri interlinked with overall changes of modern information infrastructure services, referring to “STW Thesaurus for Economics” as a case study. It starts with an overview of the history and development of the STW and describes the far-reaching changes brought about by its publication on the Web, with regard to subject indexing, retrieval and new uses for Linked Open Data. It argues that only the most recent technological developments help thesauri to exploit their full potential which is why they more than ever have a place in current information retrieval and infrastructure.

Tudhope, Douglas and Ceri Binding. 2016. “Still Quite Popular After All Those Years: The Continued Relevance of the Information Retrieval Thesaurus.” *Knowledge Organization* 43 no. 3: 174-179. 40 references.

Abstract: The recent ISKO-UK conference considered the question of whether the traditional thesaurus has any place in modern information retrieval. This note is intended to continue in the spirit of that good-natured debate, arguing that there is indeed a role today and highlighting some recent work showing the continued relevance of the thesaurus, particularly in the linked data area. Key functionality that a thesaurus makes possible is discussed. A brief outline is provided of prominent work that employs thesauri in three key areas of infrastructure underpinning advanced retrieval functionality today: metadata enrichment, vocabulary mapping and web services.

MacFarlane, Andrew. 2016. Knowledge Organisation and its Role in Multimedia Information Retrieval. *Knowledge Organization* 43 no. 3: 180-183. 28 references.

Abstract: Various kinds of knowledge organisation, such as thesauri, are routinely used to label or tag multimedia content such as images and music and to support information retrieval, i.e. user search for such content. In this paper, we outline why this is the case, in particular focusing on the semantic gap between content and concept based multimedia retrieval. We survey some indexing vocabularies used for multimedia retrieval, and argue that techniques such as thesauri will be needed for the foreseeable future in order to support users in their need for multimedia content. In particular, we argue that artificial intelligence techniques are not mature enough to solve the problem

of indexing multimedia conceptually and will not be able to replace human indexers for the foreseeable future.

White, Martin. 2016. "The Value of Taxonomies, Thesauri and Metadata in Enterprise Search." *Knowledge Organization* 43 no. 3: 184-192. 28 references.

Abstract: Although the technical, mathematical and linguistic principles of search date back to the early 1960s and enterprise search applications have been commercially available since the 1980s; it is only since the launch of Microsoft SharePoint 2010 and the integration of the Apache Lucene and Solr projects in 2010 that there has been a wider adoption of enterprise search applications. Surveys carried out over the last five years indicate that although enterprises accept that search applications are essential in locating information, there has not been any significant investment in search teams to support these applications. Where taxonomies, thesauri and metadata have been used to improve the search user interface and enhance the search experience, the indications are that levels of search satisfaction are significantly higher. The challenges faced by search managers in developing and maintaining these tools include a lack of published research on the use of these tools and difficulty in recruiting search team members with the requisite skills and experience. There would seem to be an important and immediate opportunity to bring together the research, knowledge organization and enterprise search communities to explore how good practice in the use of taxonomies, thesauri and metadata in enterprise search can be established, enhanced and promoted.

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

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