

Thesaurus Newsletter

Quarterly update from the CAB Thesaurus management team

Issue Q1 – 2020

Upcoming meetings

Many meetings are being rescheduled due to the COVID-19 epidemic. Please check the conference web sites for the latest information. Thank you.

European DDC Users Group Meeting. 2020: Was 7-8 May 2020, Edinburgh, Scotland. Rescheduled: [check web site](#).

COMPUTERM 2020: 16 May 2020, Marseille, France. Theme: Terminology processing within automatic and human applications. [COMPUTERM web site](#)

European Semantic Web Conference 2020: 31 May-4 June 2020, Heraklion, Crete, Greece. <https://2020.eswc-conferences.org/>

TOTH 2020: 2-5 June 2020, University Savoie Mont-Blanc, Chambéry, France. Theme: Terminology & Ontology – theories and applications. <http://toth.condillac.org/conference>

16th International ISKO Conference: 6-8 July 2020, Aalborg, Denmark. Theme: Knowledge organization at the interface. Web site: [ISKO 2020](#)

SemTech2020: 6-10 July 2020, Indian Statistical Institute, Bangalore. Theme: Semantic techniques and technologies for data and knowledge representation <https://drtc.isibang.ac.in/semtech/>

European Networked Knowledge Organization Systems workshop: 25th August 2020, Lyon, France. Main themes: KOS mappings and user interactions with KOS. <https://nkos-eu.github.io/2020/>

COVID-19 resources

To provide reliable, verified information, CABI's [Global Health](#) – the go-to bibliographic database for the study and practice of national, regional and international public health – is making relevant content available for **free** to support the international effort to fight the outbreak of novel coronavirus.

A [recent CABI blog post](#) reports on the rise in misinformation surrounding the disease, which is hampering efforts to contain and treat the virus. CABI aims to counteract that with the best science.

Our colleagues at [SciDevNet](#) are also putting out many reliable news stories on the coronavirus outbreak.

Of course, CAB Thesaurus has been updated with the latest terminology to aid retrieval from our databases.

Progress on CABT 2020

The next edition of CAB Thesaurus is due for public release in July 2020. In the last 9 months 20,281 terms have been updated, 7721 new terms in English were added; and 6652 new translations were added from English.

Russian added to CABT

Before adopting the Unicode-compliant version of [MultiTes](#) nearly two years ago we weren't able support non-Roman scripts. Now that we can, we recently added Russian as one of our languages. In the first month we made 534 translations into Russian, mostly of geographic and pesticides terms.

Food plants update

In July 2019 CABI published the book [*Food Plants of the World*](#). It is a comprehensive tome (520 pages) covering the identification, culinary uses and nutritional value of food plants. We were somewhat surprised that CABT was distinctly lacking in certain species, local names and classifying the culinary and other uses of many tropical plants. To address that issue, we went through the book from beginning to end, and consulted several similar references, to fill the gaps.

Invasive species update

In our ongoing programme to cover invasive species, we added another 1100 aquatic species and 500 other species in the last three months.

Feedback

We welcome suggestions for improvements to thesaurus content as well as corrections. You may email a team member directly or use the dedicated general [email](#) address.

Obtaining thesaurus files

If your organization is interested in obtaining thesaurus data for your projects they are available in multiple formats, including plain text, CSV, XML and SKOS. If you wish to see beforehand what to expect [sample data](#) are available for download via the thesaurus web site. Please [contact us](#) to discuss your requirements. CABI has offices in a dozen countries.

The thesaurus team

[Anton Doroszenko](#)

[Tony Pittaway](#)

The problem with RTs

Related terms or RTs have been a standard feature of thesauri from the beginning. *Roget's Thesaurus*, first published in 1852, had lists of words linked to similar words, not always exact synonyms, arranged into conceptual divisions and sections.

The problem with RTs is that there is no explicit explanation of how two terms thus linked are related. A human reading a thesaurus has a good chance of understanding what the relationship is, if they have enough background knowledge. But a computer can't, at least not directly.

An example of an RT is a disease and the causal agent of the disease, usually an organism, such as a bacterium or a fungus. We can say that tuberculosis is related to *Mycobacterium tuberculosis* because the latter is the causal agent of the disease. An expert will know how those two concepts are related. They are not placed in the same hierarchy because, in this case, tuberculosis belongs in a disease hierarchy and *Mycobacterium tuberculosis* belongs in a bacterial organism hierarchy.

A computer needs to be 'told' how terms are related. Hence the development of ontologies, which explicitly specify the relationships between concepts. Traditional thesauri haven't really implemented this idea until recently. But CABT is one of them.

Most modern thesaurus management software allows the definition of custom relationships other than the traditional broader, narrower, and related term. We have previously explained how CABT has implemented these, such as disease agent and disease name. The latest addition is HYB and PAR, to indicate that a species has a hybrid and conversely to specify the parents of a hybrid organism.