Thesaurus Newsletter

Quarterly update from the CAB Thesaurus management team

Issue H2 – 2019

Upcoming meetings

QURATOR 2020: 20-21 January 2020, Berlin, Germany. Theme: digital curation technologies, using AI to support knowledge workers, media managers and business strategists to enrich and structure digital content with machine-readable data. https://qurator.ai/conference-qurator-2020/

5th LODLAM Summit: 3-4 February 2020, The Getty Center, Los Angeles, California, USA. Theme: Linked Open Data in Libraries, Archives and Museums. https://lodlam.net/

14th IEEE International Conference: 3-5 February 2020, San Diego, California, USA. Theme: Semantic computing. https://www.ieee-icsc.org/

OCTA 2019: 6-8 February 2020, Tunis, Tunisia. Theme: Organization of Knowledge and Advanced Technologies. https://multiconference-octa.loria.fr/

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. ISKO 2020

Progress on CABT 2020

Rapid progress has been made on the next edition of CAB Thesaurus, due for public release in July 2020. Since the last edition was published in July 2019:

- 11,644 terms were updated
- 4657 new terms in English were added
- The total number of terms (labels), in all languages combined, is now 2,955,967
- 4352 new translations were added from English, including
    - 1636 into German
    - 576 into French
    - 302 into Spanish
    - 247 into Portuguese
    - 210 into Finnish
    - 133 into Dutch

Culinary fruits update

An effort has been made to include a complete global list of culinary fruits, i.e. fruits which are either grown or harvested for human consumption. More than 100 new local and regional species/names have been added, cross-linked to their preferred botanical names, synonyms and other uses. Until recently, the main focus had been the more widespread species of commercial and nutritional importance. New species include the pome fruit *Flacourtia inermis*, the drupes *Eugenia involucrata* and *Nephelium hypoleucum*, and the berries *Billardiera scandens* (common apple-berry) and *Grewia retusifolia* (emu berry). Species already in the thesaurus have been checked and updated.

Forage database

Grasslands and Forage Abstracts has been published by CABI since 1971 and contains over 255,000 research summaries in its database. Grasslands are important in the agriculture of many regions worldwide and CAB Thesaurus covers the subject well. However, we realised that the tropics could do with improvement. Therefore, we reviewed all the Fact Sheets in the Tropical Forages database and filled in any gaps in species coverage.

New version of MultiTes

Multisystems Inc. released two updates to the Unicode compliant version of MultiTes Pro in November 2019. The latest is v.2019.11.14u. A new feature is hierarchy expansion for narrower terms on the term record display. There were also various code optimizations, visual enhancements, and bug fixes.

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

How we revise and prioritise new content

CAB Thesaurus (CABT) is used worldwide by many different organizations for their own internal use, but our primary customer is CABI itself, to make sure that our products fulfil the needs of our customers, particularly researchers and librarians.

CABT is huge. It is nearing 3 million term labels. Clearly, we can't review everything in any given year. So how do we prioritise what to update?

Partly we look at what hasn't been revised in some time. An example of this is revising the plant families. The classification of plants has changed greatly in the light of advances in genetic analysis. Originally CABT used the Cronquist System of taxonomic classification, which was developed between the 1960s and 1980s. It was widely adopted by plant taxonomists, though other systems co-existed, such as those by Engler, Bentham & Hooker, and Takhtajan. These systems were based primarily on morphological and chemical characteristics. Currently CABT is converting to using the APG System, which is being developed by the Angiosperm Phylogeny Group. The APG System attempts to create a classification based on monophyletic groups, i.e. groups consisting of all the descendants of a common ancestor. We revised the monocots based on APG III (published 2009) and other families using APG IV (published 2016). Revising all the plants is a huge job and won't be completed probably until 2022.

Other taxonomic projects in previous years were revising the fungi, bacteria, nematodes, and viruses. These were given priority because of their importance in animal, plant and human disease. CABI's products are particularly strong in these areas.

www.cabi.org