CAB Thesaurus 2020 edition report

Anton Doroszenko, Thesaurus Manager
10th August 2020

Introduction

The 2020 edition of CABT was released on 5th August. It is a significant update compared with the previous edition published in July 2019. There are large increases in content, along with improvements to thesaurus structure and organization. The focus has been on the needs of Plantwise, Compendia, VetMed Resource, and the Global Health and the CAB Abstracts databases. Attention was also given to significantly increasing translation of English content.

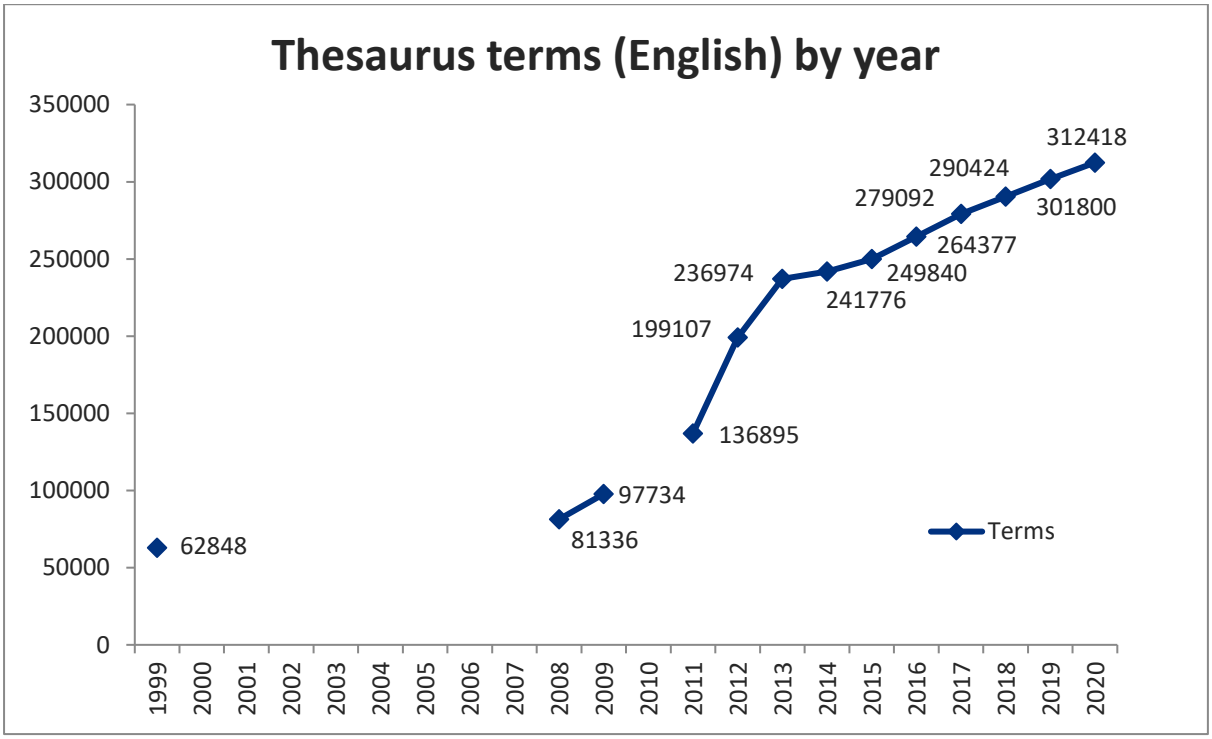
New Content and Features – some highlights

1. The number of terms/labels now exceeds 3 million (3,014,039 to be exact). A major milestone.
2. A total of 25,225 terms were updated in some way to produce the 2020 edition. Of these, 10,713 were new terms in English.
3. Revision of dicot plant families continued, using the current APG IV system of classification. Revisions were completed of all families included in the Rosales, notably the rose family which is hugely important for crops and ornamentals, as well as Fagales (7 families), Crossosomatales (7 families), Huerteales (4 families), and numerous smaller families in other orders.
4. A review of the lesser known crops and culinary fruits, particularly in Central and South America, West Africa and Asia, resulted in a substantial number of new additions. These were also linked to major groupings and products.
5. Many new invasive species were checked/added/updated, especially those impacting the Caribbean, Baltic and Mediterranean.
6. More ontological-style relationships were added. The thesaurus now includes 1366 vectors of 728 vectored agents (medical, veterinary and horticultural), 3363 biocontrol agents of 1380 hosts, and 3567 diseases caused by 2223 disease agents. Taxonomic revisions of the relevant organisms were carried out in parallel.
7. A new associative relationship was added to identify hybrids and their parents.
8. We added 7508 translations of non-taxonomic terms from English, particularly into German (2002), French (897), Spanish (437), Portuguese (336), Finnish (334) and Dutch (217). Russian was added as a new language this year, with 799 translations.
9. The number of terms categorized as Miscellaneous was reduced from 2991 in 2019 (1.0% of total terms in English) to 2631 currently (0.8%). These will be reduced even further in the next thesaurus edition by redistributing them to other categories.
10. English terms with history notes increased from 87.7% to 88.8%.

Some statistics and figures

The total number of terms/labels now stands at 3,014,039, an increase of 109,274 since the 2019 edition, which includes 182,060 distinct concepts (preferred terms), 149,164 non-preferred terms, as well as translations from English into eleven European languages. This is many times more than any other life sciences thesaurus in existence.

Growth in number of terms in English



Terms by language

Language	Number in 2019	Number in 2020	% compared with English in 2020
Danish	250805	260470	83.4%
Dutch	273101	282442	90.4%
English - UK	301800	312418	100%
English – USA*	744	751	0.2%
Finnish	252829	262286	84.0%
French	255231	265204	84.9%
German	265743	276830	88.6%
Italian	250861	260735	83.5%
Norwegian	250639	260317	83.3%
Portuguese	274038	283496	90.7%
Russian	0	799	0.3%
Spanish	277632	287190	91.9%
Swedish	251342	261101	83.6%

*Only where there is a different spelling from UK English

Growth in terms by technical category in English

Category	Number in 2019	Number in 2020	Increase in terms	% Increase since 2019
Scientific names	248855	257978	9123	3.7%
Common names	15145	16482	1337	8.8%
Homographs	748	752	4	0.5%
Abbreviations	1176	1205	29	2.5%
Orthographic variants	1677	1865	188	11.2%
Registered names	75	77	2	2.7%

Change in terms by subject category in English

Category	Number in 2019	Number in 2020	Change in terms
Animal Breeds	1453	1453	0
Anatomical and Morphological Structures	992	995	3
Activities	212	213	1
Biogeographic Regions	17	17	0
Chemicals and Chemical Groups	7855	7891	36
Climate Related	87	88	1
Commodities and Products	3120	3420	300
Disciplines, Occupations and Industries	866	903	37
Diseases, Disorders, and Symptoms	7176	7274	98
Geographic Entities	1544	1544	0
Habitats	76	76	0
Infrastructure	297	301	4
Institutions and Organisations	303	311	8
Miscellaneous	2991	2631	-360
Natural Processes	1066	1101	35
Organism Groups	1880	1887	7
Organism Names	265896	276151	10255
Publications	172	172	0
People Groups	502	502	0
Properties	1904	2016	112
Soil Types	377	377	0
Topographic Features	370	370	0
Techniques, Methodologies and Equipment	4217	4362	145
Time Periods	91	91	0
Vegetation Types	84	85	1

Taxonomic Ranks

29 ranks are included for scientific names of organisms. ‘Unranked’ is used for informal taxonomic names, such as clades. The table below lists some of the main ranks.

Rank	**Number of taxonomic names in the 2019 edition**	**Number of taxonomic names in the 2020 edition**
Phylum	116	119
Class	196	203
Order	785	808
Family	4232	4308
Genus	31075	31447
Species	202214	210573

Thesaurus Subject Coverage – a brief outline

The main areas covered by the CAB Thesaurus are agriculture and human health. Other subject areas covered include food science, leisure, recreation and tourism.

In agriculture, the coverage includes crop production, forestry, horticulture, animal production, animal health, and aquaculture, and as well as their economic, social and environmental aspects. There is comprehensive coverage of the biology, ecology and biotechnology of plants, animals and micro-organisms of economic importance, including invasive species, pathogens, pests and parasites.

In human health and medicine, the emphasis is on human nutrition, community and public health, tropical diseases, communicable diseases, and medical mycology, entomology and parasitology.

Thesaurus classification scheme

1. GENERAL
    - common terms - general processes, properties and characteristics
    - named regions and countries
    - research and methodology
    - mathematics, statistics and computer science
    - communication and information

2. PHYSICAL SCIENCES
    - physics
    - chemistry

3. EARTH SCIENCES
    - geology
    - geomorphology
    - soil science
    - hydrology
    - meteorology and climatology

4. LIFE SCIENCES
    - biology
        - microbiology
        - botany
        - zoology
        - cytology
        - embryology
        - molecular biology
        - biochemistry
        - physiology
        - genetics
        - biological structure and form
        - taxonomy
    - ecology
    - behaviour and psychology
    - organisms

5. APPLIED SCIENCE AND TECHNOLOGY
    - health and pathology
        - diseases
        - pathogenesis and epidemiology
        - health and health protection
        - pharmacology and toxicology
    - applied human and animal nutrition
    - applied genetics and breeding
    - agriculture, forestry and fishery
    - environment and natural resource management
    - technology and engineering
        - materials and equipment
        - materials handling and processing
        - transport
        - energy and power
    - food science

6. SOCIAL SCIENCES AND HUMANITIES
    - social sciences
        - education
        - sociology
        - social welfare
        - policy, law and legislation
        - economics
    - culture and humanities
        - leisure, recreation and tourism

Other Features of the CAB Thesaurus

Notes fields:

- Organism Name Author
- Term Definition
- History Note
- General Notes
- Scope Notes
- Source of Term
- CAS Registry Number
- Enzyme Commission Number

Non-standard associative relationships

- Crop Plant ↔ Harvested Product
- Disease Agent ↔ Disease Name
- Biocontrol Agent ↔ Host
- Disease Vector ↔ Vectored Agent
- Hybrid ↔ Hybrid Parent