

## **CAB Thesaurus Q2 2021 edition report**

Anton Doroszenko, Thesaurus Manager  
6<sup>th</sup> July 2021

### Introduction

The Q2 2021 edition of CABT was released on 5<sup>th</sup> July. It is a significant update compared with the previous quarterly edition. As always, the focus has been on the needs of [Plantwise](#), [Compendia](#), [VetMed Resource](#), and the [Global Health](#) and the [CAB Abstracts](#) databases.

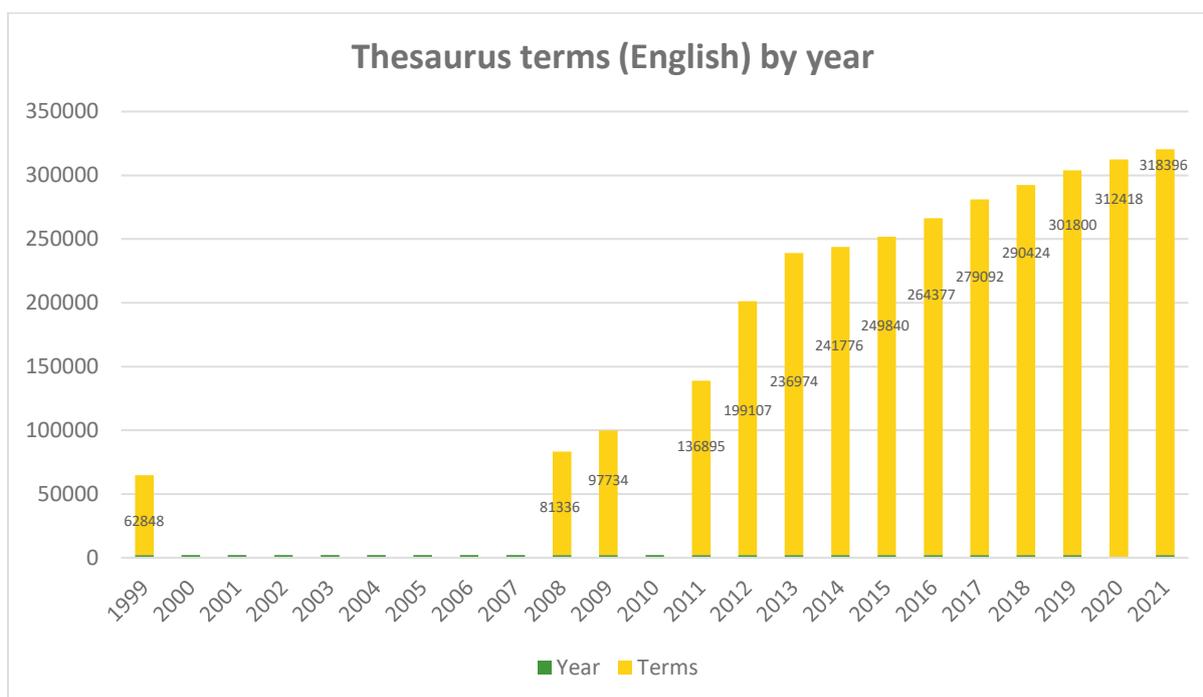
### New Content and Features – some highlights

1. The number of terms/labels is approaching 3.1 million (3,087,873 to be exact).
2. In the last year a total of 15,971 terms were updated in some way to produce the current edition. Of these, 6741 terms, mostly organism names, but also including about 200 terms relevant to Global Health, were newly added in English.
3. Revision of dicot plant families continued, using the current [APG IV system](#) of classification. In 2021, Cunoniaceae, Oxalidaceae and Elaeocarpaceae were revised, to complete revisions of all families included in Oxalidales. Additionally, Combretaceae, Begoniaceae, Vochysiaceae, twelve families included in Brassicales, as well as various smaller families in other orders, were revised.
4. All newly approved pesticides (latest update in March 2021) were added.
5. More ontological-style relationships were added. The thesaurus now includes 1372 vectors of 740 vectored agents (medical, veterinary and horticultural), 3363 biocontrol agents of 1380 hosts, and 3590 diseases caused by 2257 disease agents. Taxonomic revisions of the relevant organisms were carried out in parallel.
6. In the last year, 5241 translations of non-taxonomic terms from English were added, particularly into German (1449), Dutch (348), French (324), Spanish (286), Portuguese (248), and Finnish (213). For Russian, a new language since 2020 and not yet with much content, 1463 translations were added.
7. The number of terms categorized as Miscellaneous was reduced from 2631 in August 2020 to 2259 currently (0.7% of total terms in English).
8. Since August 2020, English terms with history notes increased from 88.8% to 90.4%.
9. American English spelling variants were removed as separate entries in the thesaurus and the relationship converted to a translation of British English. This has reduced the number of terms included in some Subject Categories (see below).

### Some statistics and figures

The total number of terms/labels now stands at 3,087,873, an increase of 73,834 since the August 2020 edition, which includes 184,651 distinct concepts (preferred terms), 152,949 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 2020	Number in 2021	% of UK English in 2021
Danish	260470	266924	83.8%
Dutch	282442	289051	90.8%
English - UK	312418	318396	100%
English – USA*	751	778	0.2%
Finnish	262286	268762	84.4%
French	265204	271791	85.4%
German	276830	284502	89.4%
Italian	260735	267325	84.0%
Norwegian	260317	266729	83.8%
Portuguese	283496	290008	91.1%
Russian	799	2262	0.7%
Spanish	287190	293739	92.3%
Swedish	261101	267606	84.0%

\*Only where there is a different spelling from UK English

## Growth in terms by technical category in English

Category	Number in 2020	Number in 2021	Increase in terms	% Increase since 2020
Scientific names	257978	264242	6264	2.4%
Common names	16482	16520	38	0.2%
Homographs	752	754	2	0.3%
Abbreviations	1205	1215	10	0.8%
Orthographic variants	1865	1955	90	4.8%
Registered names	77	77	0	0%

*Change in terms by subject category in English*

<b>Category</b>	<b>Number in 2020</b>	<b>Number in 2021</b>	<b>Change in terms</b>
Animal Breeds	1453	1453	0
Anatomical and Morphological Structures	995	991	-4
Activities	213	216	3
Biogeographic Regions	17	19	2
Chemicals and Chemical Groups	7891	7989	98
Climate Related	88	86	-2
Commodities and Products	3420	3332	-88
Disciplines, Occupations and Industries	903	940	37
Diseases, Disorders, and Symptoms	7274	7158	-116
Geographic Entities	1544	1548	4
Habitats	76	80	4
Infrastructure	301	292	-9
Institutions and Organisations	311	305	-6
Miscellaneous	2631	2259	-372
Natural Processes	1101	1138	37
Organism Groups	1887	1883	-4
Organism Names	276151	282469	6318
Publications	172	168	-4
People Groups	502	493	-9
Properties	2016	2095	79
Soil Types	377	372	-5
Topographic Features	370	370	0
Techniques, Methodologies and Equipment	4362	4377	15
Time Periods	91	90	-1
Vegetation Types	85	85	0

*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.

<b>Rank</b>	<b>Number of taxonomic names in the 2020 edition</b>	<b>Number of taxonomic names in the 2021 edition</b>
Phylum	119	119
Class	203	204
Order	808	811
Family	4308	4318
Genus	31447	31649
Species	210573	216547

## 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
- 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