

CAB Thesaurus 2018 edition report

Anton Doroszenko, Thesaurus Manager
29th June 2018

Introduction

This edition is a major update compared with the previous 2017 edition. Not only has the content grown greatly but also there were improvements made to thesaurus structure and organization to increase its utility. Particular attention was paid to the needs of Plantwise, Compendia, VetMed, Global Health and the CAB Abstracts database. Attention was also given to significantly increasing translation of English content.

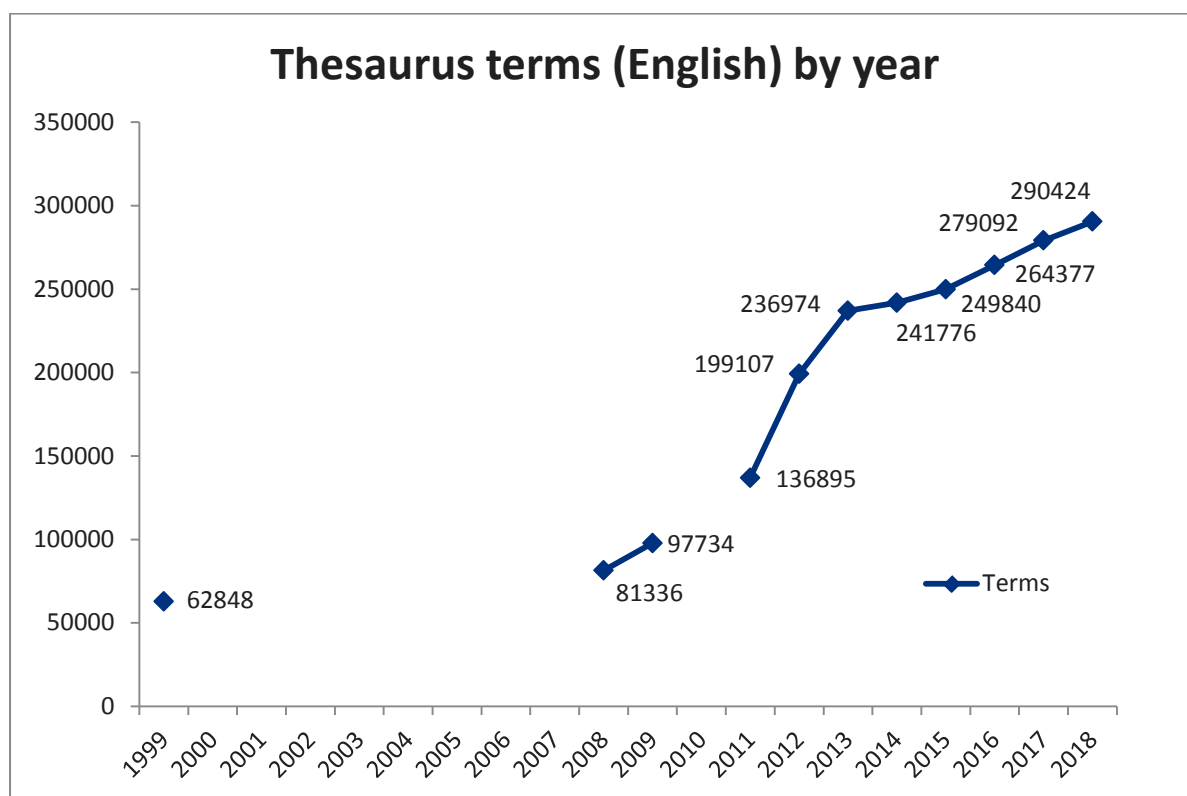
New Content and Features – some highlights

1. A total of 35,940 terms were updated in some way to produce the 2018 edition. Of these, 11,518 were new terms.
2. Revision of dicot plants continued using the current APG IV system of classification. A top level revision of all dicot orders plus revisions of 52 families down to species level in the Ranunculales, Proteales, Buxales, Trochodendrales, Gunnerales, Dilleniales, Saxifragales, Zygophyllales, Vitales, Fabales, Rosales, Fagales, and Cucurbitales were completed.
3. Two new associative relationships — Natural Enemy (where it has been confirmed as a biocontrol agent) ↔ Host and Disease Vector ↔ Vectored Agent — including 945 vectors of 353 vectored agents and 2535 biocontrol agents on 979 hosts, were added. Taxonomic revisions of the relevant organisms were carried out in parallel.
4. Translations of thesaurus relationships and category codes into French and Portuguese were added. Previously we had Dutch, English, German and Spanish. This means that when browsing the thesaurus web site in these languages all the headings will be in the language selected.
5. Added 5201 translations of non-taxonomic terms from English, particularly into German (2805), French (1063), Spanish (567), Portuguese (447) and Dutch (304), focussing on products and commodities, chemicals, and diseases terminology to ensure the hierarchies linked up well.
6. Added a new category code for natural processes; 923 terms were categorized as such.
7. Reduced the number of terms categorized as Miscellaneous from 5695 in 2017 (2.0% of total terms) to 3760 currently (1.3% of total terms). These will be reduced even further in the next thesaurus edition by redistributing them to other categories.
8. Increased the English terms with history notes from 85.1% to 86.5%.

Some statistics and figures

The total number of terms is approaching 2.8 million (2,775,794 to be exact, an increase of 124,593 since the 2017 edition), which includes 172,972 distinct concepts (preferred terms), 136,827 non-preferred terms, as well as translations from English into ten 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 2017	Number in 2018	% compared with English in 2018
Danish	228868	239687	82.5%
Dutch	250724	261840	90.2%
English - UK	279092	290424	100%
English – USA*	717	719	0.2%
Finnish	230729	241543	83.2%
French	231522	243396	83.8%
German	237481	251054	86.4%
Italian	228502	239314	82.4%
Norwegian	228703	239513	82.5%
Portuguese	251205	262460	90.4%
Spanish	254380	265756	91.5%
Swedish	229278	240088	82.7%

*Only where there is a different spelling from UK English

Growth in terms by technical category in English

Category	Number in 2017	Number in 2018	Increase in terms	% Increase since 2017
Scientific names	227115	237925	10810	4.8%
Common names	14747	14926	179	1.2%
Homographs	730	740	10	1.4%
Abbreviations	1145	1169	24	2.1%
Orthographic variants	1088	1433	345	31.7%
Registered names	65	74	9	13.8%

Change in terms by subject category in English

Category	Number in 2017	Number in 2018	Change in terms
Animal Breeds	1381	1450	69
Anatomical and Morphological Structures	964	983	19
Activities	202	206	4
Biogeographic Regions	17	17	0
Chemicals and Chemical Groups	8060	7821	-239
Climate Related	86	87	1
Commodities and Products	2300	3004	704
Disciplines, Occupations and Industries	714	817	103
Diseases, Disorders, and Symptoms	6920	7032	112
Geographic Entities	1522	1550	28
Habitats	66	73	7
Infrastructure	275	286	11
Institutions and Organisations	285	293	8
Miscellaneous	5695	3760	-1935
Natural Processes (new in 2018)	0	923	923
Organism Groups	1872	1877	5
Organism Names	243682	254743	11061
Publications	151	163	12
People Groups	500	501	1
Properties	1672	1800	128
Soil Types	377	377	0
Topographic Features	367	369	2
Techniques, Methodologies and Equipment	3520	3876	356
Time Periods	88	89	1
Vegetation Types	83	84	1

Ranks

29 ranks are included for taxonomic names as well as Unranked, which is used for informal taxonomic names such as clades. The table below lists some of the main ones.

Rank	Number of taxonomic names in the 2017 edition	Number of taxonomic names in the 2018 edition
Phylum	112	113
Class	182	183
Order	718	734
Family	3845	3976
Genus	29363	30155
Species	183450	193146

Thesaurus Subject Coverage – a brief outline

The main areas covered by the CAB Thesaurus are agriculture and human health. In agriculture the coverage includes forestry, horticulture, animal production, aquaculture, animal health and 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 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. Other subject areas covered include food science, leisure, recreation and tourism.

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
- Natural Enemy ↔ Host
- Disease Vector ↔ Vectored Agent