

CAB Thesaurus 2019 edition report

Anton Doroszenko, Thesaurus Manager
2nd July 2019

Introduction

This edition is a significant update compared with the previous edition published in July 2018. 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, Global Health and the CAB Abstracts database. Attention was also given to significantly increasing translation of English content. The major innovation this year was conversion from an ANSI- to a Unicode-compliant software platform. This opens the way to include non-Roman scripts, such as Chinese.

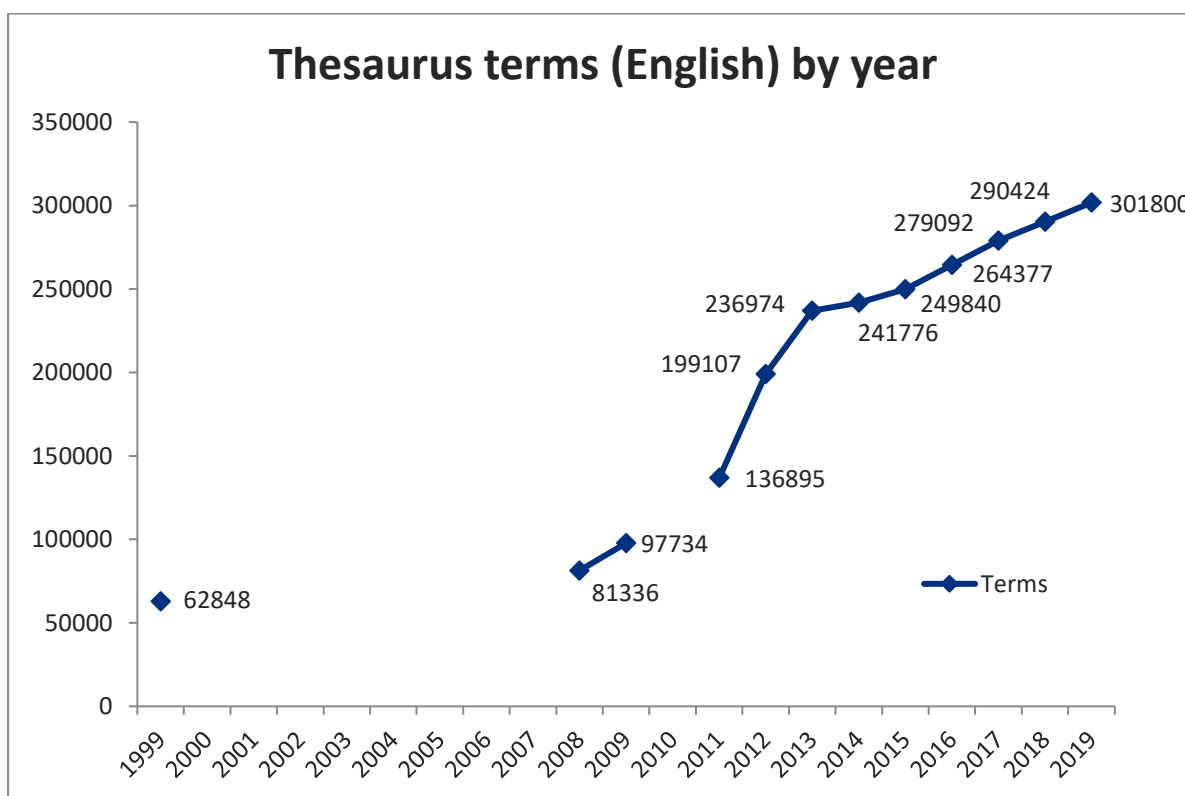
New Content and Features – some highlights

1. A total of 33,383 terms were updated in some way to produce the 2019 edition. Of these, 12,859 were new terms in English.
2. Revision of dicot plant families continued using the current APG IV system of classification. Revisions were completed of Moraceae, Urticaceae, Betulaceae as well as numerous small families, and a start was made on Rosaceae, which is hugely important for crops and ornamentals.
3. In excess of 2000 invasive species, and other species threatened by them, were added to support Invasive Species Compendium. About half of these were aquatic.
4. More ontological-style relationships were added. The thesaurus now includes 1344 vectors of 693 vectored agents, 3314 biocontrol agents of 1332 hosts, and 3541 diseases caused by 2186 disease agents. Taxonomic revisions of the relevant organisms were carried out in parallel.
5. We added 8297 translations of non-taxonomic terms from English, particularly into German (3787), Spanish (943), French (905), Portuguese (646), Finnish (356) and Dutch (332).
6. The number of terms categorized as Miscellaneous was reduced from 3760 in 2018 (1.3% of total terms in English) to 2991 currently (1.0%). These will be reduced even further in the next thesaurus edition by redistributing them to other categories.
7. English terms with history notes increased from 86.5% to 87.7%.
8. Usage data analytics were added to the thesaurus web site (www.cabi.org/cabthesaurus/) in December 2018. To date, the site has had 15,438 unique visitors, making 30,897 visits, and viewing 228,976 pages.

Some statistics and figures

The total number of terms/labels exceeds 2.9 million (2,904,765 to be exact, an increase of 128,971 since the 2018 edition), which includes 178,235 distinct concepts (preferred terms), 142,912 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 2018	Number in 2019	% compared with English in 2019
Danish	239687	250805	83.1%
Dutch	261840	273101	90.5%
English - UK	290424	301800	100%
English – USA*	719	744	0.2%
Finnish	241543	252829	83.8%
French	243396	255231	84.6%
German	251054	265743	88.1%
Italian	239314	250861	83.1%
Norwegian	239513	250639	83.0%
Portuguese	262460	274038	90.8%
Spanish	265756	277632	92.0%
Swedish	240088	251342	83.3%

*Only where there is a different spelling from UK English

Growth in terms by technical category in English

Category	Number in 2018	Number in 2019	Increase in terms	% Increase since 2018
Scientific names	237925	248855	10,930	4.6%
Common names	14926	15145	219	1.5%
Homographs	740	748	8	1.1%
Abbreviations	1169	1176	7	0.6%
Orthographic variants	1433	1677	244	17.0%
Registered names	74	75	1	1.4%

Change in terms by subject category in English

Category	Number in 2018	Number in 2019	Change in terms
Animal Breeds	1450	1453	3
Anatomical and Morphological Structures	983	992	9
Activities	206	212	6
Biogeographic Regions	17	17	0
Chemicals and Chemical Groups	7821	7855	34
Climate Related	87	87	0
Commodities and Products	3004	3120	116
Disciplines, Occupations and Industries	817	866	49
Diseases, Disorders, and Symptoms	7032	7176	144
Geographic Entities	1550	1544	-6
Habitats	73	76	3
Infrastructure	286	297	11
Institutions and Organisations	293	303	10
Miscellaneous	3760	2991	-769
Natural Processes	923	1066	143
Organism Groups	1877	1880	3
Organism Names	254743	265896	11153
Publications	163	172	9
People Groups	501	502	1
Properties	1800	1904	104
Soil Types	377	377	0
Topographic Features	369	370	1
Techniques, Methodologies and Equipment	3876	4217	341
Time Periods	89	91	2
Vegetation Types	84	84	0

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 2018 edition	Number of taxonomic names in the 2019 edition
Phylum	113	116
Class	183	196
Order	734	785
Family	3976	4232
Genus	30155	31075
Species	193146	202214

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
- Biocontrol Agent ↔ Host
- Disease Vector ↔ Vectored Agent