

## CAB Thesaurus 2017 edition report

Anton Doroszenko, Thesaurus Manager  
5<sup>th</sup> July 2017

### Introduction

This edition is a major update compared with the previous 2016 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.

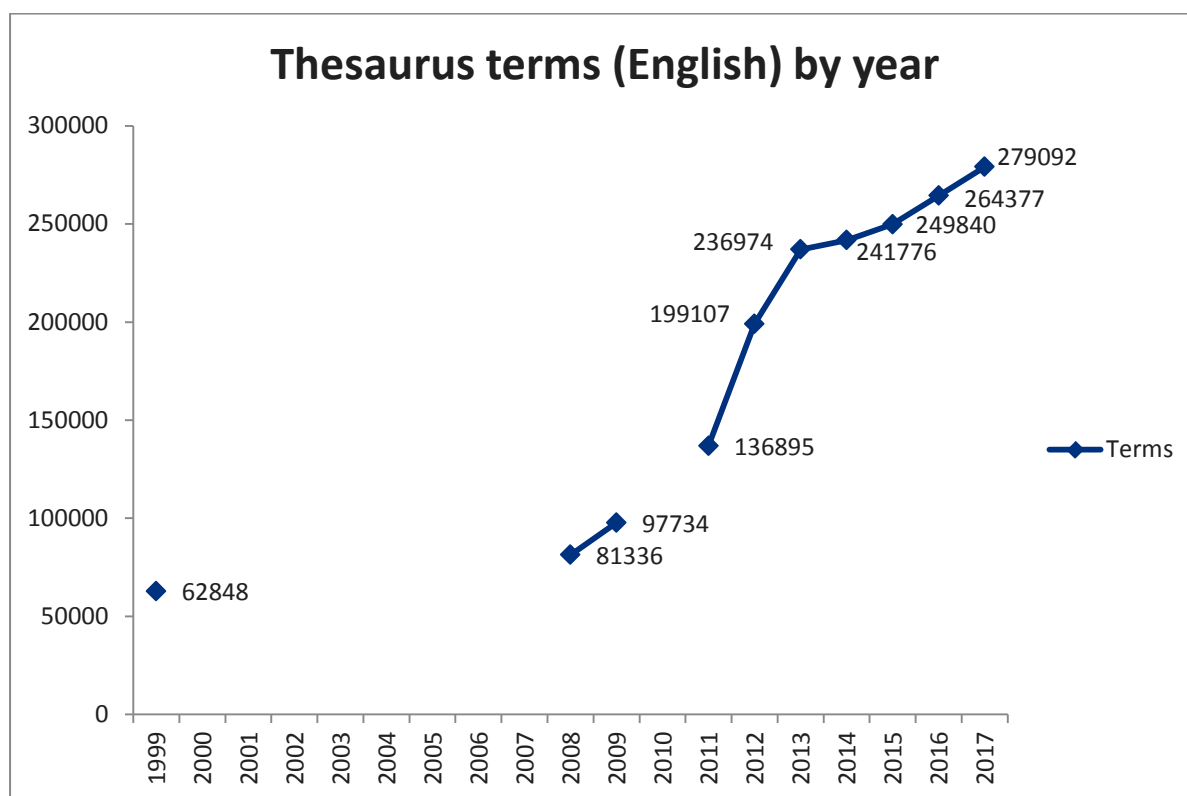
### New Content and Features – some highlights

1. Completed a revision of all biological control agents, using the Natural Enemies and BIOCAT databases – two former CABI projects – as sources. In the process 3361 extra organism names were added to the thesaurus.
2. Completed a second round of updates to the grass family, which is hugely important to agriculture and ecology, and in the process added many hundreds of species. This exercise completed the revision of the entire monocots using the APG III classification system. A start was made to revising the dicot plants using the more current APG IV system of classification. This project will take at least two years to complete given the challenge of revising such a large group of plants.
3. Added another 800 taxonomic names of organisms present in the Invasive Species Compendium not previously included in the thesaurus. This was in addition to the more than 2000 verified names added in 2016.
4. Added over 600 medical terms to improve indexing of Global Health. The particular focus was on neoplasms, medicinal drugs, and metabolic markers related to disease.
5. Added translations of thesaurus relationships and category codes into Dutch and Spanish. Previously we had English and German only. This means that when browsing the thesaurus web site all the headings will be in the language selected.
6. Added many translations of English non-taxonomic terms, particularly into Finnish (1247 added), French (964 added) and German (1559 added), focussing on products and commodities, chemicals, and diseases terminology to ensure the hierarchies linked up well.
7. Reduced the number of terms categorized as Miscellaneous from 7087 in 2016 (2.7% of total terms) to 5695 currently (2.0% 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 83.5% to 85.1%.

### Some statistics and figures

The total number of terms is approaching 2.7 million (2,651,201 to be exact), which includes 166,395 distinct concepts (preferred terms), 132,109 non-preferred terms, as well as translations from English into ten European languages. This is much more than any other life sciences thesaurus in existence. It is more than double the size of the USDA's National Agricultural Library Thesaurus, eight times the size of AGROVOC, and ten times the size of GACS.

## Growth in number of terms in English



## Terms by language

Language	Number in 2016	Number in 2017	% compared with English in 2017
Danish	215204	228868	82.0%
Dutch	237001	250724	89.8%
English - UK	264377	279092	100%
English – USA*	645	717	0.3%
Finnish	215816	230729	82.7%
French	216899	231522	83.0%
German	222213	237481	85.1%
Italian	214838	228502	81.9%
Norwegian	215039	228703	81.9%
Portuguese	237468	251205	90.0%
Spanish	240639	254380	91.1%
Swedish	215614	229278	82.2%

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

## Growth in terms by technical category in English

Category	Number in 2016	Number in 2017	Increase in terms	% Increase since 2016
Scientific names	213455	227115	13660	6.4%
Common names	14491	14747	256	1.8%
Homographs	704	730	26	3.7%
Abbreviations	1101	1145	44	4.0%
Orthographic variants	922	1088	166	18.0%
Registered names	57	65	8	14.0%

*Change in terms by subject category in English*

<b>Category</b>	<b>Number in 2016</b>	<b>Number in 2017</b>	<b>Change in terms</b>
Animal Breeds	1376	1381	5
Anatomical and Morphological Structures	955	964	9
Activities	187	202	15
Biogeographic Regions	12	17	5
Chemicals and Chemical Groups	7628	8060	432
Climate Related	84	86	2
Commodities and Products	2157	2300	143
Disciplines, Occupations and Industries	457	714	257
Diseases, Disorders, and Symptoms	6501	6920	419
Geographic Entities	1560	1522	-38
Habitats	49	66	17
Infrastructure	246	275	29
Institutions and Organisations	260	285	25
Miscellaneous	7087	5695	-1392
Organism Groups	1866	1872	6
Organism Names	229798	243682	13884
Publications	123	151	28
People Groups	481	500	19
Properties	1455	1672	217
Soil Types	376	377	1
Topographic Features	364	367	3
Techniques, Methodologies and Equipment	2939	3520	581
Time Periods	87	88	1
Vegetation Types	78	83	5

*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.

<b>Rank</b>	<b>Number of taxonomic names in the 2016 edition</b>	<b>Number of taxonomic names in the 2017 edition</b>
Phylum	112	112
Class	182	182
Order	713	718
Family	3794	3845
Genus	28623	29363
Species	170722	183450

## 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, and Enzyme Commission Number.

Non-standard associative relationships: Crop Plant ↔ Harvested Product and Disease Agent ↔ Disease Name.