



Thesauri Supported by the Elsevier Fingerprint Engine™



For Pure, SciVal Experts, Reviewer Finder, Overlap
Check and Portfolio Analysis

Date: August 2014

Empowering Knowledge

Elsevier Fingerprint Engine Thesaurus Support

Introduction

The Elsevier Fingerprint Engine provides indexing for all documents in all scientific domains. The Elsevier Fingerprint Engine identifies concepts in free text documents such as publication abstracts, funding opportunities and awarded grant announcements automatically. These concepts are well-defined terms extracted from domain-related thesauri and vocabularies.

How it works

The Elsevier Fingerprint Engine mines the free text using a variety of Natural Language Processing (NLP) techniques, identifies the concepts that define the text, and produces a semantic index, or *fingerprint*, of weighted terms that represents the text.

The products using the Elsevier Fingerprint Engine can present these fingerprints as document, author, department or institution profiles; moreover, these fingerprints can be used for search and classification as well.

For more information, please visit <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/elsevier-fingerprint-engine>

Thesaurus Overview

This document lists all thesauri currently supported by both the Elsevier Fingerprint Engine (and the products based on this text analytics software). In applications, the following scientific domains are usually distinguished; supported thesaurus mentioned next to it.

Domain	Thesaurus
Agriculture & Biology	NAL
Arts & Humanities	Humanities Vocabulary
Business & Economics	Economics Vocabulary <i>or</i> STW Thesaurus (Expanded)
Chemical Substances & Compounds	Compendex (Ei) Thesaurus <i>and</i> MeSH and Supplementals Thesaurus
Earth & Environmental Sciences	Geobase Thesaurus
Engineering & Materials Science	Compendex (Ei) Thesaurus
Mathematics	Cambridge Math Thesaurus (Expanded)
Medicine & Life Sciences	MeSH Thesaurus
Physics & Astronomy	NASA Thesaurus
Social Sciences	Gesis Thesaurus (Expanded)

In the following pages, essential data are provided for each thesaurus. The focus for each thesaurus is indicated in broad terms, designating the scientific domain to which this thesaurus is applied. The designations for scientific domains may vary slightly between applications, but rarely very significantly.

The content of documents in the Business & Economics subject area are expressed by either the STW thesaurus *or* the Economics vocabulary – this is the client’s choice. The contents of Chemistry domain documents are expressed by both the Compendex thesaurus (for engineering terminology) as well as the MeSH (and Supplementals) thesaurus (for the compounds). A particular slice of the MeSH thesaurus is selected to annotate compounds in a normalized way, viz. the section containing drugs and compounds – which is the largest [by far] subset of concepts in MeSH.

The descriptive terminology used to describe thesauri is kept as general as possible. Nevertheless, in some circles one term may be more easily understood than another; for instance, we use ‘concepts’ or ‘descriptors’ or ‘preferred terms’ interchangeably. Thesaurus maintainers and users may be used to other terminology, or even use ‘descriptors’ or ‘concepts’ in distinct meanings (for MeSH, particularly, this distinction has meaning). The default, intended meaning of ‘the number of preferred terms’, of course, is ‘the number of unique conceptual indices’, i.e. concepts that we distinguish in indexing content.

Compendex

Name (short, full)	Compendex (COMPUTerized ENgineering inDEX)
Actual version	Compendex 2013
Focus	Engineering & Materials Science Chemistry
Short description	Compendex is the computerized version of the Engineering Index (Ei for short) originated in the 19th century. More than 10,000 terms model all subject areas of engineering. Compendex has hierarchical and associative relationships between its preferred terms as well as synonym relations between preferred and non-preferred terms constituting a comprehensive, fine-grained domain-specific thesaurus. Recently, it was enhanced with a variety of linguistic filters for use in the Elsevier fingerprint engine.
Language(s)	English
Top Level Categories	Chemical Engineering Civil Engineering Electrical Engineering General Engineering Mechanical Engineering Mining Engineering
# of Top Level Categories	6
# of Hierarchy levels	11
# of descriptors (= preferred terms = concepts)	10,903
total # of terms (Synonyms)	19,678
Publisher / Owner	Elsevier (Scopus and EV Content Management)
Revision Cycle	annually

Geobase

Name (short, full)	Geobase
Actual version	Geobase 2014
Focus	Earth & Environmental Sciences
Short description	<p>Geobase is a domain-specific thesaurus, i.e. a controlled vocabulary whose terms are structured by hierarchical and equivalence relations. Hierarchical relations connect terms to broader and to narrower terms, equivalence relations connect them to synonyms. It contains about 11,000 terms from the areas of human and physical geography, geology, geomechanics, ecology, international development, and oceanography.</p> <p>A variety of linguistic filters increases the precision of indexing with Geobase.</p>
Language(s)	English
Top Level Categories	<ul style="list-style-type: none"> • agriculture, technology and engineering • biological phenomena • communication and information • earth and environment • environmental phenomena • equipment, man-made structures, and transport vehicles • management, planning, policy and legislation • materials, products and resources • mathematical and general phenomena • organisms and biological structures • physical, chemical and earth phenomena • research and methodology • science and scientific disciplines • society and economy • space and time • type of article or study
# of Top Level Categories	16
# of Hierarchy levels	10
total # of terms (Synonyms)	9,517
# of Non-preferred terms (synonyms)	11,146

Publisher / Owner	Elsevier
Revision Cycle	annually

GESIS + Worldbank/Law thesaurus

Name (short, full)	Thesaurus for the Social Sciences (GESIS Thesaurus) enhanced with the Law branch of the WorldBank (WB) thesaurus
Actual version	0.9.2 (RDF/Skos version) 03/2013, WB thesaurus: 10 / 2011
Focus	Social Sciences + Law
Short description	<p>GESIS is a crucial instrument for the content-oriented search by keywords in SOFIS (Social Science Research Information System) and SOLIS (Social Science Literature Information System) developed by the InformationsZentrum (IZ) Sozialwissenschaften, Bonn. Topics in all of the social science disciplines are included. [fr Gesis homepage]</p> <p>The WorldBank thesaurus is a large collection of concepts (~ 500,000 preferred terms) covering all subject areas relevant in politics. Law and Justice is one out of 30 main categories. It contains descriptors from more than 30 aspects of the legal domain including Corporate Law, International Law, Administrative and Regulatory Law, Labor and Employment Law, Tax law etc.</p>
Language(s)	German, English, Russian (Gesis), English (WorldBank Law)
Top Level Categories	<p>After re-organizing the 171 original Gesis top level categories into 9 we added a Law category for the WorldBank Law terms:</p> <ul style="list-style-type: none"> • Fundamentals of the Social Sciences • Society • Social Change • Fundamentals and Manifestations of Social Behavior • Individual, Personality • Social Role • Interdisciplinary Application Areas of Social Sciences • Special Social Sciences • Law • Special Lists of Terms
# of Top Categories	10
# of Hierarchy levels	10

# of descriptors per language (Gesis11122016)		Gesis	WBLaw	Together
	en	9728	10497	20225
	de	9930		
	fr	9919		
total # of terms (Gesis11122016)		Gesis	WBLaw	Together
	en	11058	10497	21555
	de	12382		
	fr	12105		
# of orphans	0			
Publisher / Owner	GESIS – Leibniz Institute for the Social Sciences / WorldBank			
Revision Cycle	2-3 years / irregular			
url	http://www.gesis.org/en/services/tools-standards/social-science-thesaurus/#c10193 (Gesis), http://multites.net/mtsql/wb/site/ (WorldBank thesaurus)			



Math thesaurus

Name (short, full)	Math thesaurus, Connecting Mathematics Thesaurus
Actual version	2004 / 2013
Focus	Mathematics
Short description	<p>The Connecting Mathematics Thesaurus has grown from a proposal made by the NRICH project at the University of Cambridge to develop an online dictionary of mathematics to support its other online publications. The thesaurus contains more than 4,500 concepts with descriptors in English and other European languages, in a hierarchical structure.</p> <p>[http://mathforum.org/library/view/61686.html, modified]</p> <p>To enhance coverage, Elsevier has merged the Cambridge Mathematics thesaurus with a generated vocabulary containing 20000 terms.</p>
Language(s)	English, Danish, Finnish, Hungarian, Lithuanian, Polish, Slovak, Spanish (some)
Top Level Categories	<ul style="list-style-type: none"> • Attribute • Method • Statistical thing • Mathematics • Algebraic object • Number • Limit • Book • Game • Imaginary objects • Logical thing • Dimension • Place • Number pattern • Physical thing • Unit • Person • Equipment • Geometric object • Notation • Process • Financial term
# of Top Level Categories	22
# of Hierarchy levels	11
# of Concepts (English)	3,993 (Math thes.) / 22,425 (incl. Math vocabulary)
total # of Terms (English)	3,818 (Math thes.) / 36,767 (incl. Math vocabulary)
Publisher / Owner	University of Cambridge and Partners

Revision Cycle	None, discontinued
url	http://thesaurus.maths.org/mmkb/view.html?resource=downloads&msglang=en

MeSH

Name (short, full)	MeSH Medical Subject Headings MeSH and MeSH Supplemental terms.
Actual version	MeSH 2014
Focus	MeSH: Life Sciences. <ul style="list-style-type: none"> • Experts: Medicine & Life Sciences • JRT: Life and Health Sciences MeSH Supplementals: Chemistry <ul style="list-style-type: none"> • Experts: Chemical Substances & Compounds JRT: Chemistry
Short description	MeSH is the National Library of Medicine's controlled vocabulary for biology, medicine and neighboring fields, used, among other things, for indexing articles for the MEDLINE® database. More than 170,000 synonyms are assigned to more than 26,000 descriptors, arranged in a hierarchical structure with 15 top categories. In addition to the descriptors, MeSH contains some 140,000 Supplementary Concepts which are mapped to the closest fitting MeSH descriptors. Many of these records describe chemical substances. We defined a variety of linguistic filters to optimize indexing with MeSH.
Language(s)	English (translated into numerous other languages by different publishers)
Top Level Categories	<ul style="list-style-type: none"> • Anatomy • Organisms • Diseases • Chemicals and Drugs • Analytical, Diagnostic & Therapeutic Techniques & Equipment • Psychiatry and Psychology

	<ul style="list-style-type: none"> • Phenomena and Processes • Disciplines and Occupations • Anthropology, Education, Sociology and Social Phenomena • Technology, Industry, Agriculture • Humanities • Information Science • Named Groups • Health Care • Publication Characteristics • Geographicals
# of Top Level Categories	16
# of Hierarchy levels	11
# of Concepts (English)	<p>C graph (with MeSH Supplementary Concepts) : 245775</p> <p>B graph (without MeSH Supplementary Concepts): 26771</p> <p>H graph (Chemical Compounds): 166207</p>
# of Terms	<p>C graph (with MeSH Supplementary Concepts) : 739923</p> <p>B graph (without MeSH Supplementary Concepts): 113642 (without MeSH Permuted terms)</p> <p>H graph (Chemical Compounds, including 76903 regularly patterning codes (CAS, EC, etc.)): 511136</p>
# of orphans	0
Publisher / Owner	United States National Library of Medicine (NLM)
Revision Cycle	annually
url	http://www.nlm.nih.gov/mesh/filelist.html

NAL Thesaurus

Name (short, full)	NALT National Agricultural Library Agricultural Thesaurus
Actual version	NALT 2014
Focus	Agriculture + Biology
Short description	<p>NAL Agricultural Thesaurus is the National Agricultural Library's controlled vocabulary. The thesaurus is intended for indexing materials and for aiding retrieval in agricultural information systems. Currently, the thesaurus is, inter alia, the indexing vocabulary for the National Agricultural Library's bibliographic database of citations to agricultural resources, AGRICOLA. The subject scope of agriculture includes terminology in the supporting biological, physical and social sciences. Biological nomenclature comprises a majority of the terms in the thesaurus.</p> <p>[http://agclass.nal.usda.gov/background.shtml]</p> <p>It was enhanced with a variety of linguistic filters for use in the Elsevier Fingerprint Engine.</p>
Language(s)	English, Spanish
Top Level Categories	<ul style="list-style-type: none"> • Animal Science and Animal Products • Biological Sciences • Breeding and Genetic Improvement • Economics, Business and Industry • Farms and Farming Systems • Food and Human Nutrition • Forest Science and Forest Products • Geographical Locations • Government, Law and Regulations • Health and Pathology • Insects and Entomology • Natural Resources, Earth and Environmental Sciences • Physical and Chemical Sciences • Plant Science and Plant Products • Research, Technology and Engineering • Rural and Agricultural Sociology • Taxonomic Classification of Organisms
# Top Level Categories	17
# of Hierarchy levels	14

# of concepts	53,268 en
total # of terms (Synonyms)	98,564 en
Publisher / Owner	U.S. National Agricultural Library and the Inter-American Institute for Cooperation on Agriculture (IICA)
Revision Cycle	annually [http://agclass.nal.usda.gov/background.shtml]
url	http://agclass.nal.usda.gov/download.shtml



Arts and Humanities Vocabulary

Name	Arts and Humanities Vocabulary
Version	2013
Focus	Arts & Humanities
Short description	Domain-relevant contents of a document are expressed by the Elsevier Fingerprint Engine as a so-called semantic fingerprint, a set of concepts ranked by importance. As an interim solution for subject domains for which a thesaurus is not yet available, a vocabulary is constructed using keywords mined from Scopus records for a particular domain. By means of a number of preprocessing steps, similar keywords are merged into concepts. With statistical approaches, the relevant keywords are identified to form the vocabulary.
Language(s)	English
Top Level Categories	<p>The Arts and Humanities vocabulary covers these subject areas:</p> <ul style="list-style-type: none"> • History • Language and Linguistics • Archeology • Classics • Conservation • Philosophy of Science • Literature and Literary Theory • Museology • Music • Philosophy • Religious Studies • Visual Arts and Performing Arts
Hierarchy levels	2
Preferred terms	25848
Total terms	34447
Publisher / Owner	The Arts and Humanities vocabulary is owned and maintained by Elsevier and is updated regularly.

Business and Economics

Name	Business and Economics Vocabulary
Version	2013
Focus	Business & Economics
Short description	Domain-relevant contents of a document are expressed by the Elsevier Fingerprint Engine as a so-called semantic fingerprint, a set of concepts ranked by importance. As an interim solution for subject domains for which a thesaurus is not yet available, a vocabulary is constructed using keywords mined from Scopus records for a particular domain. By means of a number of preprocessing steps, similar keywords are merged into concepts. With statistical approaches, the relevant keywords are identified to form the vocabulary.
Language(s)	English
Top Level Categories	The Business and Economics vocabulary covers these subject areas: <ul style="list-style-type: none"> • Accounting • Business and International Management • Marketing • Organizational Behavior and Human Resource Management • Strategy and Management • Industrial Relations • Economics and Econometrics • Finance
Hierarchy levels	2
Preferred terms	20890
Total terms	30851
Publisher / Owner	The Business and Economics vocabulary is owned and maintained by Elsevier and is updated regularly.

NASA thesaurus

Name (short, full)	National Aeronautics and Space Administration (NASA) <i>Thesaurus</i>
Actual version	2013
Focus	Aerospace Engineering + supporting areas [cf. http://www.sti.nasa.gov/thesfrm1.htm]
Short description	The NASA Thesaurus contains the authorized subject terms by which the documents in the NASA Aeronautics and Space Database are indexed and retrieved. The scope of this controlled vocabulary includes not only aerospace engineering, but all supporting areas of engineering and physics, the natural space sciences (astronomy, astrophysics, and planetary science), Earth science, and to some extent, the biological sciences. The Thesaurus contains over 18,460 terms, complemented with 4,480 synonyms, organized within a hierarchical structure.
Language(s)	English
Top Level Categories	At present, there are 1,623 Top Level Categories in the NASA thesaurus which can't be listed here. The terms of the NASA thesaurus are gathered into these subject divisions: [http://www.sti.nasa.gov/sscg/subcat.html] <ul style="list-style-type: none"> • Aeronautics • Astronautics • Chemistry and Materials • Engineering • Geosciences • Life Sciences • Mathematical and Computer Sciences • Physics • Social and Information Sciences • Space Sciences • General
# of Top Level Categories	1,623
# of Hierarchy levels	8
# of preferred terms per language	18470

Total # of terms per language	23493
Publisher / Owner	NASA Scientific and Technical Information (STI) program
Revision Cycle	monthly
url	http://www.sti.nasa.gov/products.html#data

