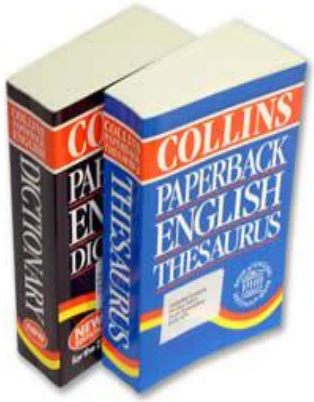
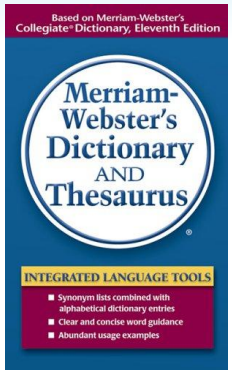
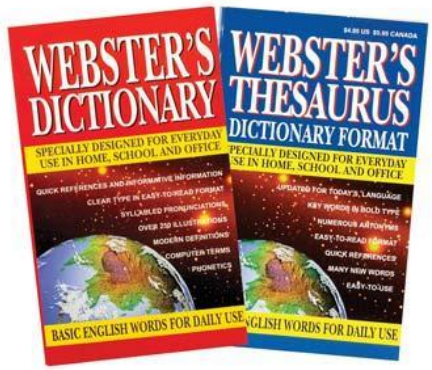
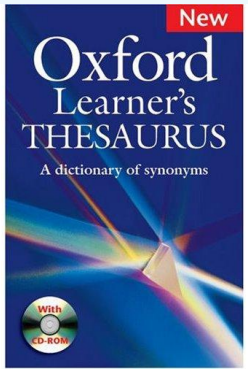
Идеографическая лексикография.

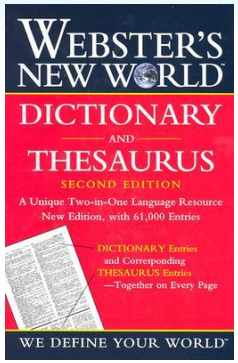
ТЕЗАУРУС











LOGO

ЛЕКСИКОГРАФИЧЕСКОЕ ОПИСАНИЕ

ТЕЗАУРУСНОЕ ОПИСАНИЕ

МЕТОДЫ ФОРМАЛИЗАЦИИ ЗНАНИЙ,

методы представления знаний субъекта о предметной области с использованием знаковых систем

Лексикографическое (словарное) описание – лексические единицы описаны несвязно, непоследовательно и прерывно в виде системы слов и их определений.

Тезаурусное описание – знание фрагментируется и структурируется так, что оно оказывается разделенным на отдельные группы понятий, связанные между собой определенными отношениями [Филиппович, Прохоров, 2002].

LOGO

ТЕЗАУРУС

от греч. τησαυροσ
«хранилище, сокровищница»

LOGO

ТЕЗАУРУС

1) словарь, в котором максимально полно представлены все слова языка с исчерпывающим перечнем примеров их употребления в текстах

2) идеографический словарь, в котором показаны семантические отношения (родовидовые, синонимические и др.) между лексическими единицами

LOGO

Ассоциативные тезаурусы **– лексемы или термины распределяются по тематическим группам на основе ассоциативных связей между называемыми понятиями.**

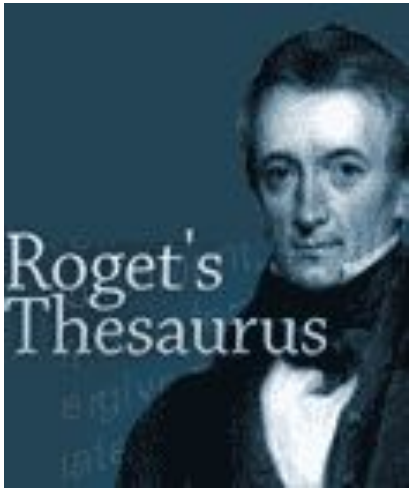
LOGO

Тематический принцип организации лексики используется в

1. учебных идеографических словарях
2. разговорниках для путешественников
3. иллюстрированных словарях
4. приложениях к алфавитным словарям

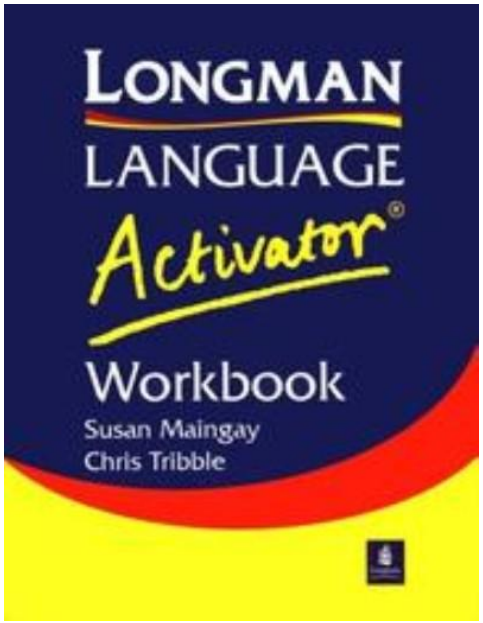
Идеографические (иерархические) тезаурусы **– отражены иерархические связи между понятиями .**

Словарь Роже **(1852, *The Thesaurus of English Words and Phrases*) – первая научно обоснованная попытка создания макета логически упорядоченной лексики языка.**

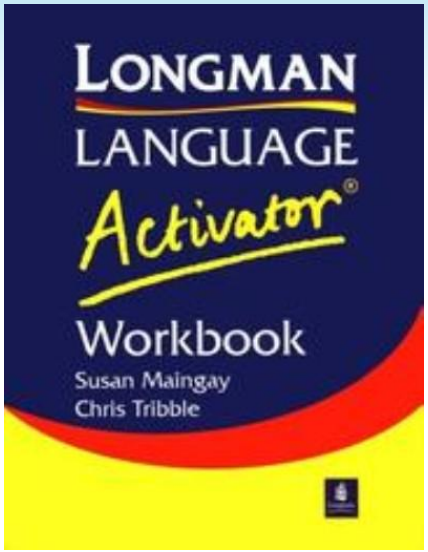


LOGO

Дефинитивный тезаурус **– сочетание идеографического и толкового словарей; включает определения значений группируемых слов (Например, *Longman Language Activator*)**



- «Двойная» семантизация лексической единицы в дефинитивном тезаурусе (понятийная система + толкование)
- вводит читателя в необходимую систему концептов
- разъясняет сущность, сходства и различия понятий



LOGO

Терминологические тезаурусы – включают систематизированную терминологию.

Информационно-поисковые тезаурусы (предназначенные для использования в специфических информационно-поисковых системах) состоят из контролируемого, но изменяемого словаря терминов, между которыми указаны смысловые связи [Караулов, 1981].

LOGO

Информационно-поисковый тезаурус **включает:**

- **дескрипторы (основные термины)**
- **недискрипторы (вспомогательные термины)**
- **алфавитный указатель**
- **систематический указатель**

LOGO

Типы понятийных отношений в терминологическом тезаурусе:

ИЕРАРХИЧЕСКИЕ:

- родовидовая связь
- связь «часть-целое»
- связь «класс-объект из класса»

Родовидовая связь **устанавливается между двумя дескрипторами, если объем понятия нижестоящего дескриптора входит в объем понятия вышестоящего дескриптора.**

Родовое понятие, гипероним

ВЕЩЕСТВО

Видовое понятие, гипоним

ЗАГРЯЗНЯЮЩЕЕ ВЕЩЕСТВО

Родовидовая связь = гиперо-гипонимическая

LOGO

Пометы, указывающие на родовидовую связь:

- BT (Broader Term)
- NT (Narrower Term)
- TT (Top Term)

LOGO

Пометы, указывающие на родовидовую связь:

radiation damage

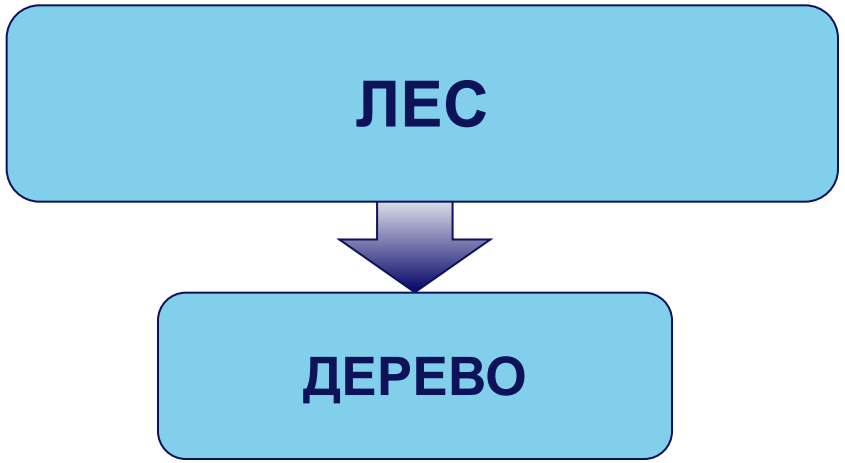
S:5193, T:1381, T:4549, T:6304, T:7259

DEF	Somatic and genetic damage to living organisms caused by exposure to ionizing radiation. (Source: GUNN)
THEME	BIOLOGY; HUMAN HEALTH; NATURAL AREAS, LANDSCAPE, ECOSYSTEMS; RADIATIONS
GROUP	HEALTH, NUTRITION
TT	disease
BT	disease
NT	radiation sickness
RT	nuclear accident radiation effect

Связь «часть-целое» **устанавливается между двумя дескрипторами в том случае, если нижестоящий дескриптор обозначает компонент объекта, обозначаемого вышестоящим дескриптором.**

Целое, холоним

Часть, мероним



Связь «часть-целое» = меронимия, или партонимическая связь

LOGO

Пометы, указывающие на связь «часть-целое»:

- **ВЦ (Выше-целое)**
- **НЧ (Ниже-часть),**
- **BTP (Broader Term Partitive)**
- **NTP (Narrower Term Partitive)**

LOGO

Отношение «класс-элемент класса»

связывает дескриптор-наименование класса объектов или категории и термин, характеризующий элемент класса, или частный случай понятия.

Класс

Элемент класса

LOGO

Пометы, указывающие на связь «класс-элемент класса»:

- **BTI (Broader Term Instance)**
- **NTI (Narrower Term Instance)**

LOGO

Типы понятийных отношений в терминологическом тезаурусе:

ИЕРАРХИЧЕСКИЕ:

- родовидовая связь
- связь «часть-целое»
- связь «класс-объекта из класса»

ЭКВИВАЛЕНТНОСТЬ:

- абсолютная синонимия
- стилистическая синонимия
- относительная синонимия

LOGO

Отношения эквивалентности:

- **абсолютная синонимия:**

  ИПТ – информационно-поисковый тезаурус; нагрев – нагревание

- **стилистическая синонимия**

  пневмония – воспаление легких

- **относительная синонимия**

  диета – питание

LOGO

Отношения эквивалентности:

conservation

	S:2158, T:3907, T:4703
DEF	No definition needed. (Source: NDN)
UF	*preservation*
THEME	FOOD, DRINKING WATER; INDUSTRY
GROUP	INDUSTRY, CRAFTS; TECHNOLOGY; EQUIPMENTS
TT	industrial process
BT	industrial process

LOGO

Ассоциативные отношения **– являются объединением других отношений, не входящих в иерархические отношения или в отношения синонимии.**

- **причина – следствие**
- **процесс – субъект процесса/инструмент**
- **действие – признак**
- **действие – объект действия**
- **объект – свойство объекта**
- **объект – источник/происхождение объекта**
- **объект – единица измерения/измерительный прибор**
- **сырье/материал – продукт**

СТРУКТУРА ТЕЗАУРУСА

LOGO

1. классифи-кационная схема понятий
2. указатели понятий-дескрипто-ров
3. идеогра-фическая часть
4. Пермутаци-онный указатель понятий-дескрипто-ров

LOGO

Классификационная схема понятий **делит всю совокупность понятий на крупные понятийные зоны или области (фасеты) и системно их упорядочивает.**

LOGO

Алфавитный указатель дескрипторов –

последовательное перечисление всех понятий-дескрипторов, характеризующих каждую понятийную зону, причем внутри зон понятия расположены по какому-либо содержательному принципу.

LOGO

Идеографическая часть –

под каждым понятием сгруппирована вся относящаяся к нему лексика; т. е. дескрипторная статья представляет собой семантическое поле, в качестве названия которого и используется понятие, или дескриптор.

Пермутационный указатель –

все значащие слова, входящие в состав терминов тезауруса, расположены в данном указателе по алфавиту, а после каждого значащего слова приводятся в алфавитном порядке все термины тезауруса, которые содержат это значащее слово в качестве терминоэлемента, причем каждый термин встречается в пермутационном указателе столько раз, сколько значащих слов он содержит .

NASA Thesaurus, 2000

(*The National Aeronautics and Space Administration – Национальное агентство по аэронавтике и исследованию космического пространства*)

NASA Thesaurus, 2000

Классификационная схема:

AERONAUTICS

01 Aeronautics (General)
02 Aerodynamics
03 Air Transportation And Safety
04 Aircraft Communications And Navigation
05 Aircraft Design, Testing And Performance
06 Avionics And Aircraft Instrumentation
07 Aircraft Propulsion And Power
08 Aircraft Stability And Control
09 Research And Support Facilities (Air)

NASA Thesaurus, 2000

Классификационная схема:

ASTRONAUTICS

12 Astronautics (General)
13 Astrodynamics
14 Ground Support Systems and Facilities (Space)
15 Launch Vehicles and Launch Operations
16 Space Transportation and Safety
17 Space Communications, Spacecraft Communications, Command and Tracking
18 Spacecraft Design, Testing and Performance
19 Spacecraft Instrumentation and Astrionics
20 Spacecraft Propulsion and Power

NASA Thesaurus, 2000

Классификационная схема:

CHEMISTRY AND MATERIALS .....

23 Chemistry and Materials (General) .....
24 Composite Materials .....
25 Inorganic, Organic, and Physical Chemistry...
26 Metals and Metallic Materials.....
27 Nonmetallic Materials .....
28 Propellants and Fuels .....
29 Space Processing.....

NASA Thesaurus, 2000

Идеографическая часть:

rocket engines

DEF Reaction engines that contain within themselves, or carry along with themselves, all the substances necessary for their operation or for the consumption or combustion of their fuel, not requiring of any outside substance and hence capable of operation in outer space. Used for interplanetary propulsion.

UF *interplanetary propulsion*

GS engines

. **rocket engines**

. . booster rocket engines

. . . AJ-10 engine

. . . Algol engine

. . . apogee boost motors

. . . H-1 engine

. . . LR-87-AJ-5 engine

NASA Thesaurus, 2000

Идеографическая часть:

Sputnik satellites
GS artificial satellites
. Soviet satellites
. . **Sputnik satellites**
. . . Sputnik 1 satellite
. . . Sputnik 2 satellite
. . . Sputnik 3 satellite
. . . Sputnik 4 satellite
. . . Sputnik 5 satellite
Soviet spacecraft
. **Sputnik satellites**
. . Sputnik 1 satellite
. . Sputnik 2 satellite
. . Sputnik 3 satellite
. . Sputnik 4 satellite

NASA Thesaurus, 2000

Пермутационный указатель:

	space exploration
	space flight
extended duration	**space** flight
	use long duration space flight
	space flight feeding
long duration	**space** flight
manned	**space** flight
manned	**space** flight network
planetary	**space** flight
	use interplanetary flight
return to Earth	**space** flight
	space flight stress
	Space Flight Tracking and Data Network
	space flight training
Dyna-Soar	**space** glider
	use X-20 aircraft

LOGO

Тезаурус ЮНЕСКО

ЮНЕСКО — Организация Объединённых Наций по вопросам образования, науки и культуры.

UNESCO Thesaurus

LOGO

1977 г. *– англоязычная версия*

2003 г. *– англ., франц., исп.*

2011г. *– 7000 терминов на английском и русском языках и 8600 терминов на французском и испанском языках*

UNESCO Thesaurus

LOGO

http://databases.unesco.org/thesaurus/

LOGO

Тезаурус, встроенный в учебные словари

MACMILLAN DICTIONARY

http://www.macmillandictionary.com

LOGO

Идеографический словарь русского языка О. С. Баранова, 1995.

- **4166 словарных статей**
- **638 подразделов**
- **144 раздела**
- **17 отделов**

LOGO

Толково-понятийный словарь русского языка: 600 семантических групп. М., 2008.

- **Около 16 500 слов и устойчивых выражений**
- **600 семантических групп**

LOGO

Толково-понятийный словарь русского языка: 600 семантических групп. М., 2008.

- **идеографический словарь активного типа;**
- **включает самую употребительную лексику;**
- **отражает не логически сконструированную картину мира, а языковую картину мира;**
- **каждая группа слов и выражений делится на подгруппы;**
- **расположение слов в подгруппе определяется отношением центр-периферия**

LOGO

Алфавитный указатель слов и выражений, входящих в семантические группы

антракт ⇒ ПЕРЕРЫВ$_2$

антропоморфный ⇒ ЧЕЛОВЕК/ЛЮДИ$_3$

апартаменты ⇒ ДОМ$_3$

апатичный ⇒ ЭМОЦИЯ$_9$

апатия ⇒ РАВНОДУШНЫЙ$_3$

апельсиновый ⇒ ЦВЕТ$_{10}$

аплодировать ⇒ ОДОБРЯТЬ$_7$

аплодисменты ⇒ ОДОБРЯТЬ$_7$

LOGO

Содержание

БЕЖАТЬ 15
1. двигаться очень быстро 15
2. двигаться очень быстро в разные стороны . 15
3. тренироваться, участвовать в спортивных соревнованиях по бегу 15
4. стараться бежать, бегать быстрее кого-либо 15

БЕЗОПАСНЫЙ 15
1. такой, который не причиняет вреда или ущерба 15
2. состояние, когда человеку не угрожает опасность 16
3. давать человеку, группе людей возможность находиться в безопасном месте 16
4. место, обеспечивающее безопасность . . . 16