

**Гендина Н. И., доктор пед. наук, профессор
Кемеровского государственного университета
культуры и искусств**

Виды информационно- поисковых тезаурусов

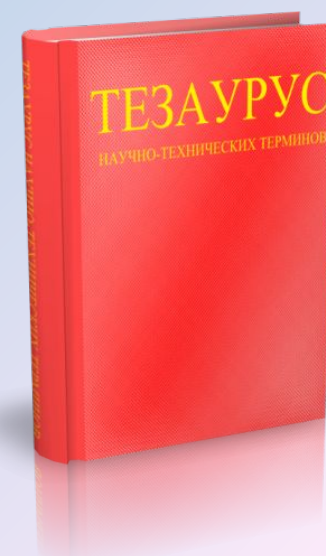
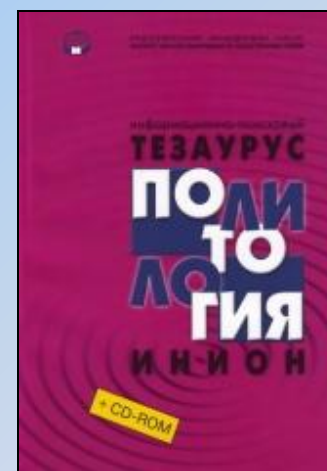
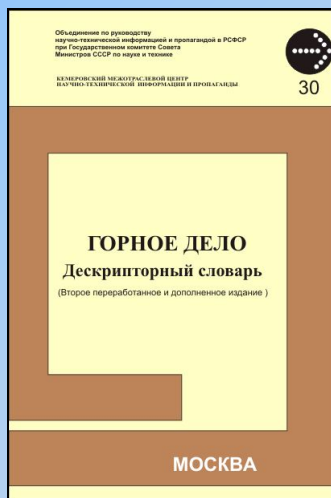
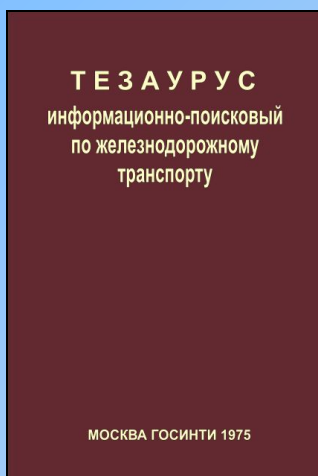
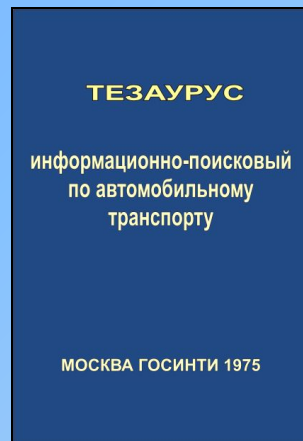
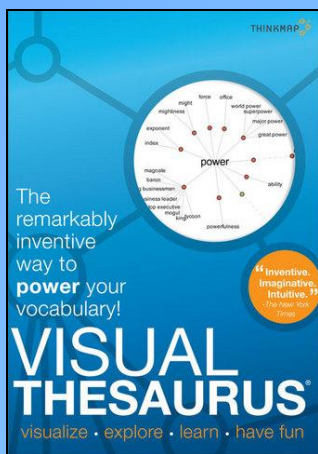
План

1. Представление о множестве отечественных и зарубежных информационно-поисковых тезаурусов
2. Многоаспектная классификация информационно-поисковых тезаурусов
3. Функции информационно-поисковых тезаурусов

Список литературы

1. ГОСТ 7.24-2007 Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению. – Взамен ГОСТ 7.24-90 ; введ. 2008-07-01 / Межгос. совет по стандартизации, метрологии и сертификации. – Москва : Стандартинформ, 2008. – 7 с.
2. ГОСТ 7.25-2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. – Взамен ГОСТ 7.25-80 ; введ. 2002-07-01 / Межгос. совет по стандартизации, метрологии и сертификации. – Минск, 2002. – 16 с.
3. ГОСТ Р 7.0.47-2008. Формат для представления на машиночитаемых носителях словарей информационных языков и терминологических данных. Содержание записи. – Введ. 2009-01-01 / Федерал. агенство по техн. регулированию и метрологии. – Москва : Стандартинформ, 2008. – 17 с.
4. Жмайло, С. В. К разработке современных информационно-поисковых тезаурусов / С. В. Жмайло // НТИ. Сер. 1.– 2004. – №1. – С.23–31.

Представление о множестве отечественных и зарубежных информационно-поисковых тезаурусов (ИПТ)



Многоаспектная классификация ИПТ

№	Основание деления	Виды ИПТ
1	Широта тематического охвата	1.1. Политематические (многоотраслевые) 1.2. Отраслевые 1.3. Узкоотраслевые (проблемные)
2	Назначение	2.1. Базисные (базовые) 2.2. Рабочие
3	Системность построения	3.1. Комплексы 3.2. Отдельные
4	Особенности состава лексики и внутривидового построения	4.1. Включающие дескрипторы и аскрипторы 4.2. Включающие только дескрипторы 4.3. Фасетизированные
5	Количество естественных языков	5.1. Одноязычные 5.2. Многоязычные
6	Форма представления	6.1. В печатном виде 6.2. На электронных носителях
7	Знаковая природа информации	7.1. Текстовые 7.2. Визуальные 7.3. Смешанные

Политематические ИПТ

Тезаурус научно-технических терминов

Автор: Шемакин Ю. И.

Год выпуска: 1972

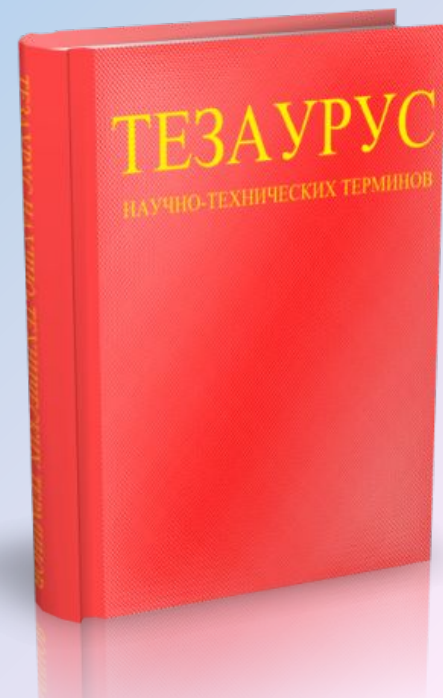
Издательство: Военное издательство Министерства обороны СССР

Включает около 19 тыс. терминов

Тезаурус по общественно-политической тематике для автоматического индексирования в Университетской информационной системе «РОССИЯ»

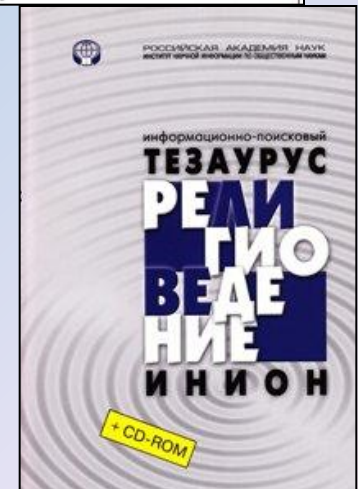
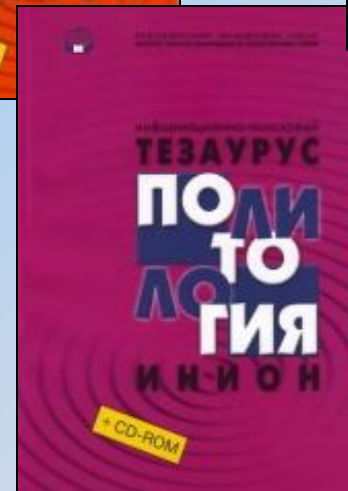
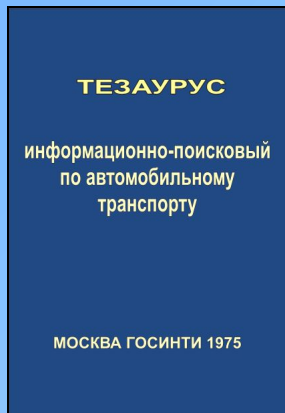
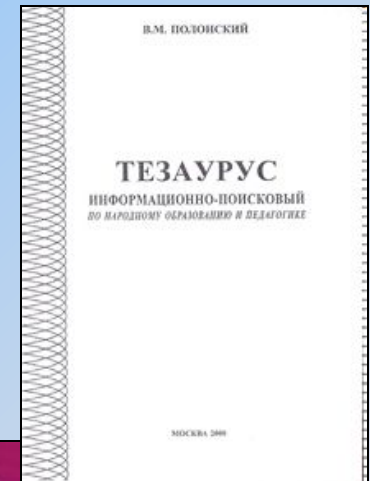
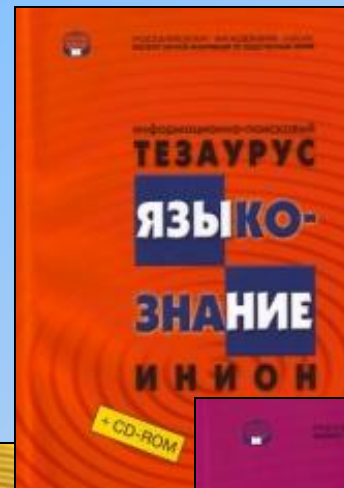
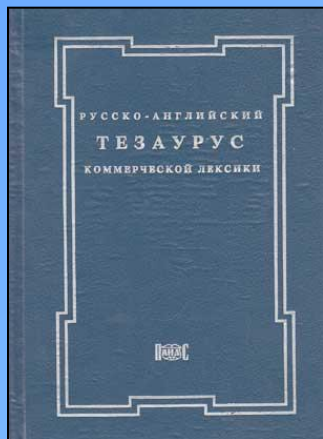
Разработан в 1995 г. Научно-исследовательским вычислительным центром МГУ им. М. В. Ломоносова (<http://www.cir.ru>).

Включает 62 тыс. терминов, более 26 тыс. понятий, отражает около 100 тыс. отношений между понятиями.



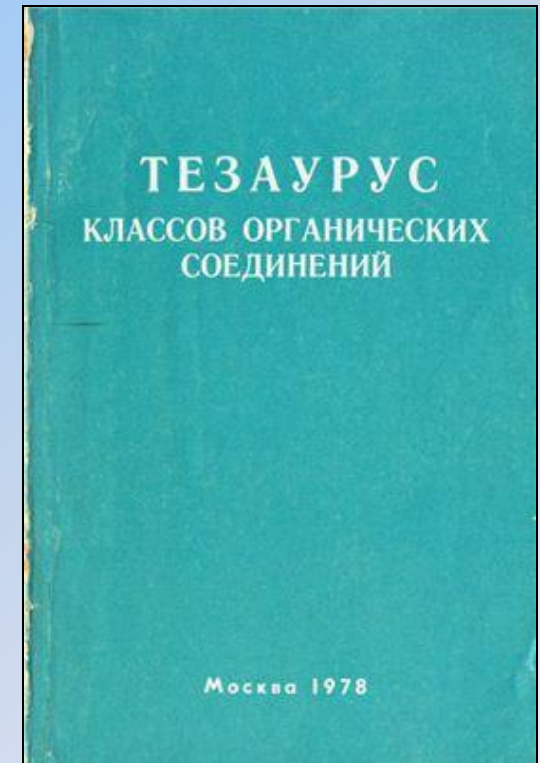
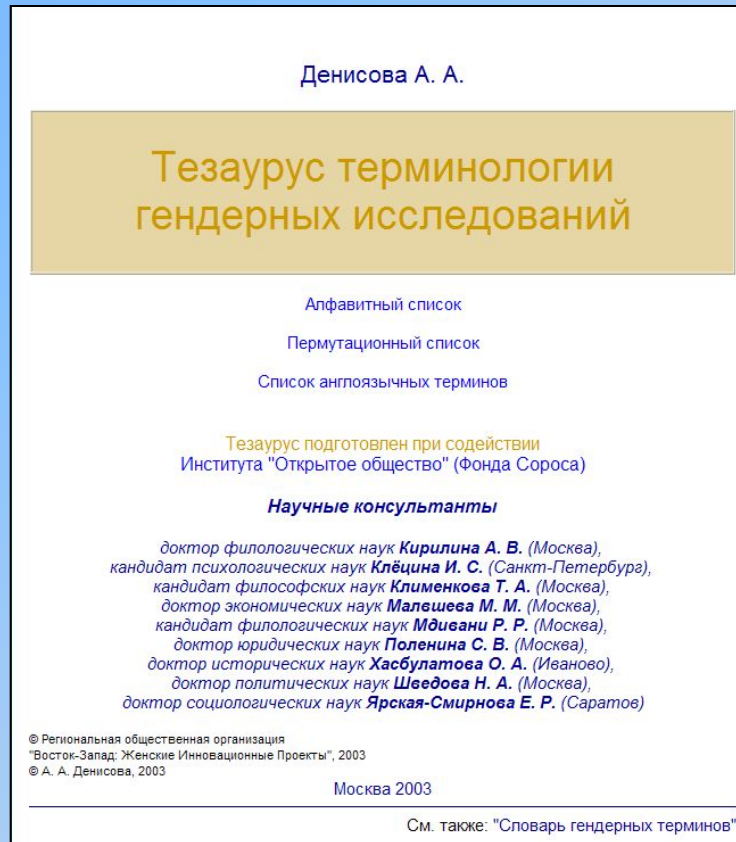
Отраслевые ИПТ

Наиболее многочисленная группа тезаурусов по различным отраслям знания



Узкоотраслевые (проблемные)

Предназначены для поиска информации по отдельному внутриотраслевому направлению



См. также: "Словарь гендерных терминов"

Комплекс тезаурусов Фонда Гетти (США)



Тезаурус по искусству и архитектуре (*The Art and Architecture Thesaurus – AAT*). Содержит около 30 тыс. дескрипторов и более 130 тыс. англоязычных терминов (http://www.getty.edu/research/conducting_research/vocabularies/aa);

Словарь имен авторов (*The Union List of Artist Names – ULAN*). Включает 375 000 имен, в т. ч. варианты имен, псевдонимов и вариантов написания имен на разных языках. Содержит, биографическую и библиографическую информацию о художниках и архитекторах (http://www.getty.edu/research/conducting_research/vocabularies/ulan/);

Тезаурус географических названий (*The Getty Thesaurus of Geographic Names – TGN*). Включает около 1 115 000 названий, типов местностей, координат, описательных заметок, фокусирующихся на местах, важных для изучения искусства и архитектуры (http://www.getty.edu/research/conducting_research/vocabularies/tgn/).

Комплекс тезаурусов ИНИОН

Пример дескрипторной статьи из тезауруса по экономике и демографии

Дескриптор: аграрные кризисы

Вышестоящий: экономические кризисы

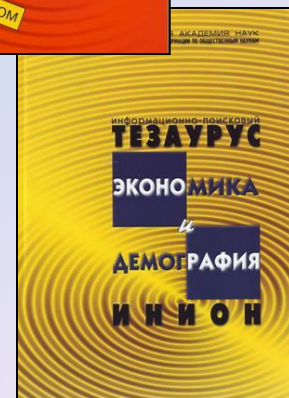
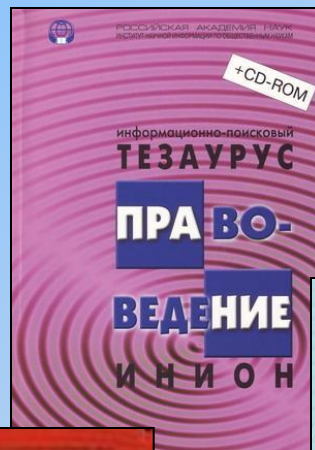
Нижестоящий: продовольственные кризисы

Ассоциативный: промышленные кризисы

Синоним(ы): относительное перепроизводство сельскохозяйственной продукции

Рубрика: A065245; A067107

ББК: 65.012.2; 65.32; 65.2/4-97

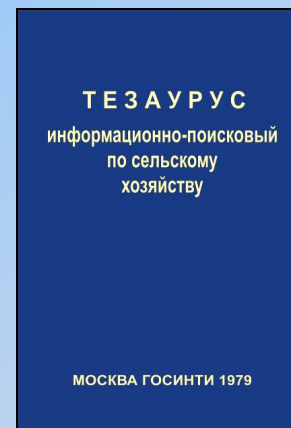
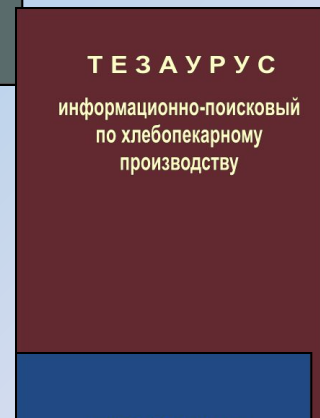
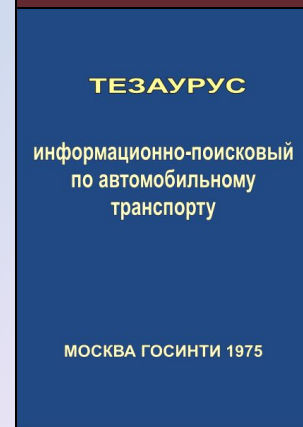
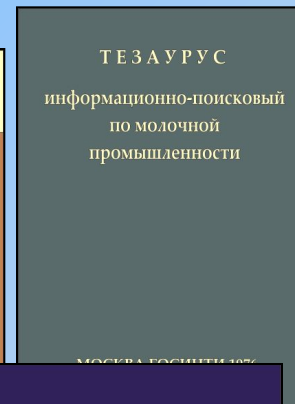
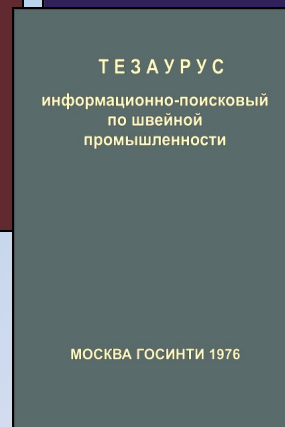
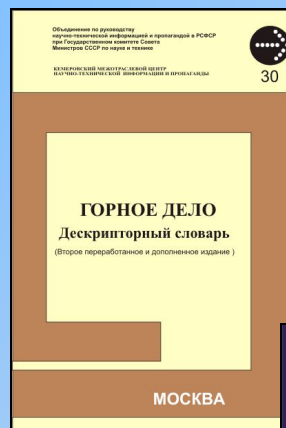
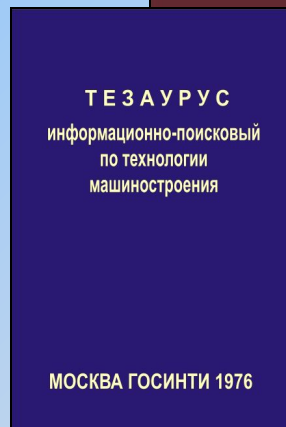
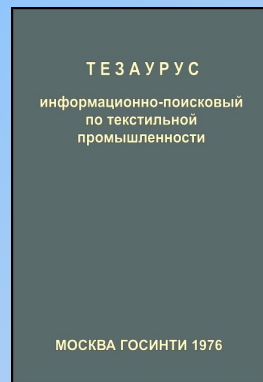


Комплекс тезаурусов Кемеровского ЦНТИ

Пример дескрипторной статьи из ИПТ по электротехнике

Альтернаторы

См. Генераторы электрические
ГЕНЕРАТОРЫ ЭЛЕКТРИЧЕСКИЕ
с Альтернаторы
Электрогенераторы
в Машины электрические
н Автогенераторы
Гидрогенераторы
Дизель-генераторы
а Источники электрической энергии



Различные способы представления лексики в разных комплексах ИПТ

ИПТ по технологии машиностроения

ТЕЛЬФЕРЫ 1202

в Механизмы грузоподъемные
Оборудование подъемно-
транспортное периодического
действия

н Тельферы электрические

а Тали электрические

ИПТ ИНИОН по философии

Дескриптор: Абдукция

Вышестоящий: Логический
метод исследования;
Умозаключение

Ассоциативный: Гипотеза;
Индукция; Логика открытия

Рубрика: A0221; A023121

ББК: 87; 87.4

Различные способы представления лексики в разных ИПТ

Геологический ИПТ многоязычный

Дескриптор: ЖЕЛЕЗО

с ФЕРРО

в МЕТАЛЛ ЧЕРНЫЙ

ЭЛЕМЕНТ

н АВАРУИТ

ГЕМАТИТ

ГЕТИТ

КУБАНИТ

ЛИМОНИТ

МАГНЕТИТ

СИДЕРИТ

а КВАРЦИТ ЖЕЛЕЗИСТЫЙ

ТРОИЛИТ

ТЮРИНГИТ

ШАМОЗИТ

Часть речи: СУЩ

Фасет: 102

Англ. эквивалент: *IRON*

Нем. эквивалент: *EISEN*

Франц. эквивалент: *FER*

Тезаурус по педагогике и образованию (руско-английская версия Тезауруса ЮНЕСКО)

ОБРАЗОВАНИЕ НЕПОЛНОЕ ВЫСШЕЕ
(дескриптор)

LOWER HIGHER EDUCATION (английский
эквивалент)

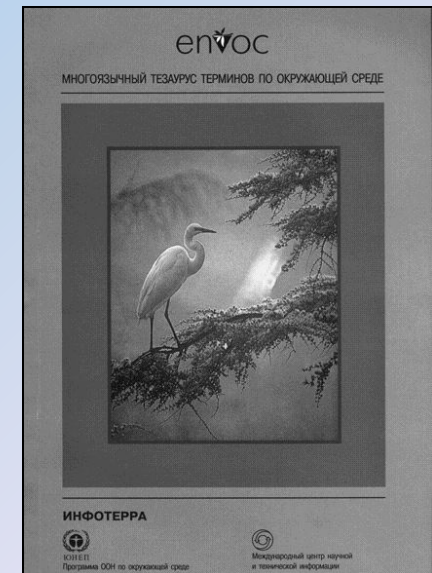
185.546.374 (цифровой код)

В РФ первый уровень высшего профессионального образования, осуществляемый вузом в течение первых двух лет обучения. По завершении этого срока студент имеет возможность продолжить высшее образование или получить диплом о неполном высшем образовании без итоговой аттестации (ограничительная помета, аннотация).

НЕПОЛНОЕ ВЫСШЕЕ ОБРАЗОВАНИЕ
(синоним, входной термин)

Многоязычные тезаурусы

- Тезаурус ООН – тезаурус ЮНБИС (*UNBIS Thesaurus*) – содержит терминологию на 6 официальных языках ООН;
- Тезаурус ЮНЕСКО по образованию (*UNESCO: IBE education thesaurus*), содержит термины на 3 языках;
- Тезаурус Европейского Сообщества – тезаурус *EUROVOC*, содержит лексику на 9 языках;
- Сельскохозяйственный тезаурус *AGROVOC* международной Продовольственной и сельскохозяйственной организации ООН – ФАО (от англ. *Food and Agriculture Organization, FAO*) и Европейского сообщества ЕС, содержит лексику на 3 языках.
- Тезаурус Организации сотрудничества железных дорог – «Информационно-поисковый тезаурус «Транспорт»», отражает терминологию на 3 языках;
- Русский Викисловарь – проект фонда Викимедиа, открыт 1 мая 2004 г. Многоязычный свободно пополняемый словарь и тезаурус. Содержит 285 742 статьи о словах, словообразовательных единицах и словосочетаниях более чем 300 языков мира.



Тезаурус ООН или Тезаурус ЮНБИС (*UNBIS Thesaurus*)


Содержит многоотраслевую терминологию на всех официальных языках ООН: арабском, китайском, английском, французском, русском и испанском, используемую в качестве дескрипторов при анализе и поиске документов ООН.

<http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf>



Тезаурус ООН или Тезаурус ЮНБИС (UNBIS Thesaurus)

БИБЛИОГРАФИЧЕСКО-ИНФОРМАЦИОННАЯ СИСТЕМА ООН

 **ТЕЗАУРУС ЮНБИС** [Поиск](#) [Алфавитный список](#) [Список категорий](#) [? Помощь](#)

БИБЛИОТЕКА ИМ. ДАГА ХАММАРШЕЛЬДА

[Главная страница ООН](#) [Главная страница Библиотеки](#) [ЮНБИСНЕТ](#) [Информация о тезаурусе](#) [Новые термины](#) [Ссылки](#)

ПОИСК

1 Поиск по словам

4 [Поиск](#) [Очистить форму](#)

Слова могут быть усечены при помощи звездочки (*)

2 Выберите тип поиска:

- Простой поиск** (Поиск по терминам и перекрестным ссылкам)
- Полнотекстовой поиск** (Поиск по всем полям, включая смежные термины и ограничительные пометы)

3 Выберите логику поиска:

- Точное соответствие фразе
- Найти все слова (логический оператор AND)**
- Найти любое из слов (логический оператор OR)

[Обращайтесь с вопросами или комментариями](#) [عربي](#) [中文](#) [English](#) [Français](#) [Русский](#) [עברית](#) © Организация Объединенных Наций 2001-2011 годы

Тезаурус ООН или Тезаурус ЮНБИС (UNBIS Thesaurus)

БИБЛИОГРАФИЧЕСКО-ИНФОРМАЦИОННАЯ СИСТЕМА ООН

 **ТЕЗАУРУС ЮНБИС** [Поиск](#) [Алфавитный список](#) [Список категорий](#) [? Помощь](#)

БИБЛИОТЕКА ИМ. ДАГА ХАММАРШЕЛЬДА

[Главная страница ООН](#) [Главная страница Библиотеки](#) [ЮНБИСNET](#) [Информация о тезаурусе](#) [Новые термины](#) [Ссылки](#)

Поиск терминов: "конфликт"

Количество найденных терминов: 7

Показ терминов: 1

Щелкните на термине для получения полной записи

- [Вьетнамский конфликт \(1961-1975 гг.\)](#)
Используется вместо: Положение во Вьетнаме
- [Конфликт](#)
- [Конфликт интересов](#)
Используется вместо: Коллизия интересов
- [Конфликт поколений](#)
- [Культурный конфликт](#)
Используется вместо: Столкновение культур
- [Социальный конфликт](#)
- [Этнический конфликт](#)
Используется вместо: Межэтнический конфликт

[Общайтесь с вопросами
или комментариями](#)



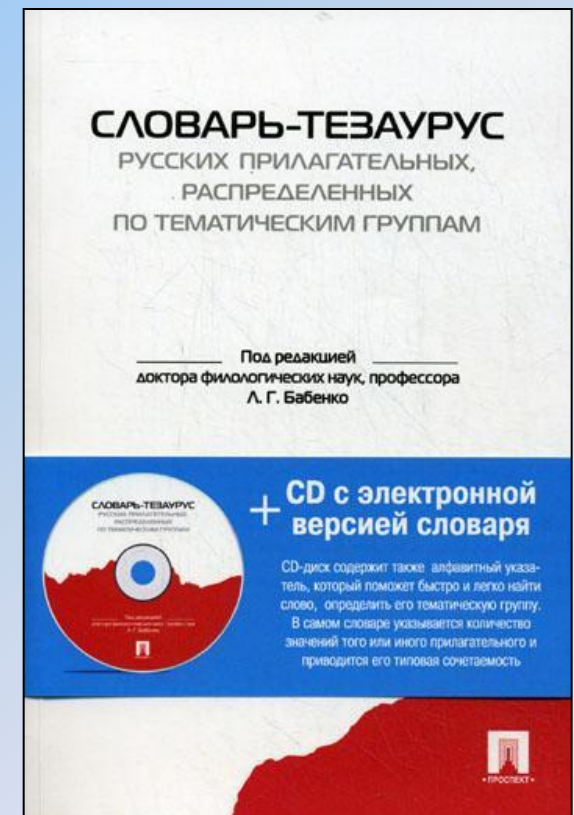
© Организация Объединенных
Наций 2001-2011 годы

Тезаурус ООН или Тезаурус ЮНБИС (UNBIS Thesaurus)

01.03.00 - حفظ السلم والأمن	<p>النزاع العرقي 族裔冲突</p> <p>01.03.00 - 维持和平和安全</p> <p>إستخدام ل: 用于: صراع عرقي 种族冲突</p>	<p>ETHNIC CONFLICT</p> <p>01.03.00 - MAINTENANCE OF PEACE AND SECURITY</p> <p>Used For: <i>INTER-ETHNIC CONFLICT</i></p>
<p>روابط المصطلحات: مصطلحات أعم: النزاع مصطلحات مترابطة: الأقليات التطهير العرقي الصراع الاجتماعي الصراع الثقافي العلاقات العنصرية القومية المجموعات العرقية والعنصرية المفاوضات بين طائفتين النشاط العرقي كره الأجانب منع الصراع الطائفي مصطلحات أعلى: النزاع</p>	<p>▼ 词间关系: 广义词: 冲突 相关词: 两族间谈判 仇外 少数民族 文化冲突 民族主义 民族和种族群体 社会冲突 种族关系 种族清洗 防止团体冲突 族首词: 冲突</p>	<p>▼ Term Relationships: Broader terms: CONFLICT Related terms: CULTURE CONFLICT ETHNIC AND RACIAL GROUPS ETHNICITY GROUP CONFLICT PREVENTION INTERCOMMUNAL NEGOTIATIONS MINORITIES NATIONALISM RACE RELATIONS SOCIAL CONFLICT XENOPHOBIA ETHNIC CLEANSING Top terms: CONFLICT</p>
<p>CONFLIT ETHNIQUE</p> <p>01.03.00 - MAINTIEN DE LA PAIX ET DE LA SECURITE</p>	<p>Этнический конфликт</p> <p>01.03.00 - ПОДДЕРЖАНИЕ МИРА И БЕЗОПАСНОСТИ</p> <p>Используется вместо: <i>Межэтнический конфликт</i></p>	<p>CONFLICTO ETNICO</p> <p>01.03.00 - MANTENIMIENTO DE LA PAZ Y DE LA SEGURIDAD</p>
<p>▼ Relations: Termes génériques: CONFLIT Termes associés: CONFLIT CULTUREL CONFLIT SOCIAL EPURATION ETHNIQUE ETHNICITE GROUPES ETHNIQUES ET RACIAUX MINORITES NATIONALISME NEGOCIATIONS INTERCOMMUNAUTAIRES PREVENTION DES CONFLITS DE GROUPE RELATIONS RACIALES XENOPHOBIE Termes génériques supérieurs: CONFLIT</p>	<p>▼ Связь с другими терминами: Более широкие термины: Конфликт Смежные термины: Ксенофобия Культурный конфликт Межобщинные переговоры Межрасовые отношения Меньшинства Национализм Национальность Предотвращение групповых конфликтов Социальный конфликт Этническая чистка Этнические и расовые группы Головные термины: Конфликт</p>	<p>▼ Relaciones: Términos genéricos: CONFLICTO Términos relacionados: CONFLICTO CULTURAL CONFLICTO SOCIAL DEPURACION ETNICA ETNICIDAD GRUPOS ETNICOS Y RACIALES MINORIAS NACIONALISMO NEGOCIACIONES INTERCOMUNALES PREVENCION DE CONFLICTOS DE GRUPO RELACIONES RACIALES XENOFOBIA Términos genéricos superiores: CONFLICTO</p>

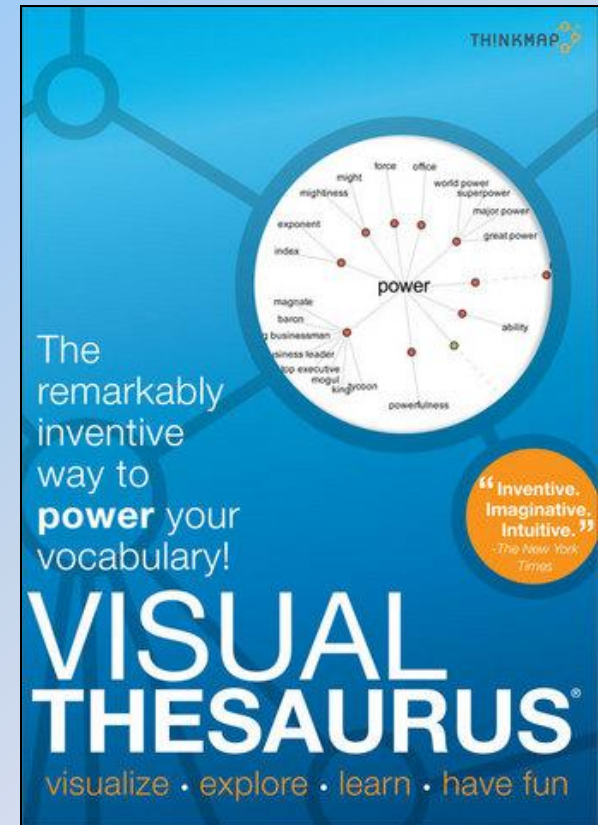
Виды ИПТ по форме представления информации

- ✓ Печатные ИПТ
- ✓ ИПТ на электронных носителях: компакт-дисках, в форме электронных сетевых версий и др.



Виды ИПТ по знаковой природе информации

- ✓ Текстовые ИПТ
- ✓ Визуальные ИПТ
- ✓ Смешанные ИПТ



Визуальный словарь (<http://vslovar.org.ru>)



[О проекте](#) | [Помощь](#)

Найти

в словарях в Интернет

[Энциклопедия](#)

[Компьютеры](#)

[Финансы](#)

[Психология](#)

[Право](#)

[Философия](#)

[Культура](#)

[Медицина](#)

[А](#) [Б](#) [В](#) [Г](#) [Д](#) [Е](#) [Ж](#) [З](#) [И](#) [Й](#) [К](#) [Л](#) [М](#) [Н](#) [О](#) [П](#) [Р](#) [С](#) [Т](#) [У](#) [Ф](#) [Х](#) [Ц](#) [Ч](#) [Ш](#) [Щ](#) [Ъ](#) [Ы](#) [Ь](#) [Э](#) [Ю](#) [Я](#)

На правах рекламы: Недвижимость у Черного моря -квартира Сочи . Сам-плюс: упаковочные пакеты - лучшие цены Профессиональные переводчики в бюро переводов . Обширный словарный запас! Профилактика диабет Ваш дом в поселке Яхромские горки Электростанции, бензогенераторы Распродажа мягкая мебель угловые диваны от фабрики Электростанции, бензогенераторы, дизель-генераторы мебель детская комната от фабрики Большой выбор ремонт кпк королев в Москве бизнес тренинг линкомаулия Болезнь, мор и бремя, иго автомойка требования к печати Sex интернет-казино, виртуальное казино, азартные игры, рулетка бассейны спа экскурсии по москве бюро переводов готовые предприятия недвижимость в Испании

Визуальный словарь (<http://vslovar.org.ru>)

СЛОВАРИК визуальный

О проекте | Помощь

Найти

Энциклопедия | Компьютеры | Финансы | Психология | Право | Философия | Культура | Медицина

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я

Ба Бб Бв Бг Бд Бе Бж Бз Би Бй Бк Бл Бм Бн Бо Бп Бр Бс Бт Бу Бф Бх Бц Бч Бш Бщ Бь Бы Бь Бю Бя

БАЗА ДАННЫХ

БАЗА ДАННЫХ (БД) (database (DB)). Один или несколько специальным образом организованных файлов, хранящих систематизированную информацию, для доступа к которой используются программные средства системы управления базой данных (СУБД). Б. д. представляет собой хранилище связанных между собой данных, не зависящих от программ, которые к ним обращаются. Например, в Б. д. предприятия хранятся конструкционные, технологические, производственные, эксплуатационные, экономические и т. п. характеристики изделий предприятия и их аналогов, выпускаемых конкурирующими фирмами. Такой Б. д. могут пользоваться специалисты из отдела маркетинга (для прогноза конкурентоспособности изделий на рынке), плановики (для определения необходимых объемов выпуска продукции), конструкторы (для анализа конструктивных особенностей изделий конкурентов) и т. д. Б. д. должна быть достаточно полной, правильно организованной, всегда соответствовать текущему моменту времени и удобной для использования. Содержащиеся в ней данные не должны противоречить друг другу. Редактирование, пополнение и исключение данных, их поиск и сортировка осуществляются под управлением СУБД. Б. д. могут быть персональными и коллективного пользования. Сопровождение крупных Б. д. коллективного пользования осуществляет администратор базы данных. Ср. банк данных

```
graph TD;
  A[БАЗА ДАННЫХ] --- B[ПРОГРАММНЫЕ СРЕДСТВА];
  A --- C[СУБД];
  A --- D[РЕДАКТИРОВАНИЕ];
  A --- E[СОРТИРОВКА];
  A --- F[БАНК ДАННЫХ];
  B --- G[ДОКУМЕНТ];
  B --- H[ДААННЫЕ];
  C --- I[СОРТИРОВКА ДАННЫХ];
  C --- J[СОРТИРОВКА];
  D --- K[ТЕКСТ];
  D --- L[ФОРМАТ];
  D --- M[УДАЛЕНИЕ];
  F --- N[ФАЙЛ];
  F --- O[ДОСТУП];
```

Грамматический словарь

<http://www.solarix.ru/>

Компьютерная лингвистика

Грамматический словарь

- ▶ Русский, английский, французский и др. языки
- ▶ Хранение в реляционной БД, доступ средствами SQL
- ▶ Пользовательские словари
- ▶ Склонение существительных и прилагательных
- ▶ Спряжение глаголов

поиск по сайту:



Тезаурус

- ▶ Переводы, синонимы, антонимы, гипернимы, гипонимы
- ▶ Редактор тезауруса

Морфологический и синтаксический анализ текста

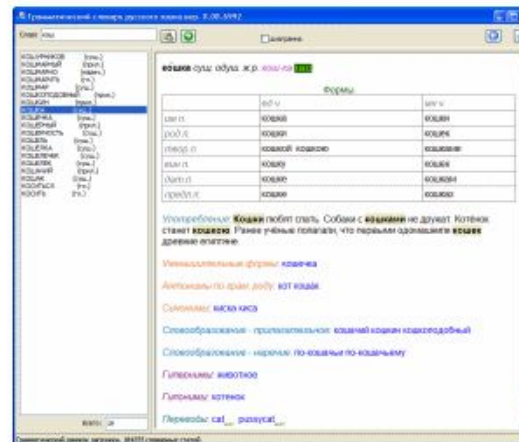
- ▶ Сегментация и токенизация текста
- ▶ Морфологический разбор слов и словосочетаний
- ▶ Синтаксический анализ предложений

Статистическая обработка текста

- ▶ Частотный анализ
- ▶ Работа с N-граммами - накопление и доступ к базе данных
- ▶ Полнотекстовая поисковая система с индексированием

Преобразование текста

- ▶ Полнотекстовый переводчик
- ▶ Перефразировщик текста



Новости:

- 21-02-2011 Грамматический Словарь в.9.11
- 07-02-2011 Обновление Грамматического Словаря
- 05-02-2011 Исходные тексты C API
- 01-02-2011 Обновление SDK поисковой системы
- 18-01-2011 Обновление Dictionary ORM
- 14-01-2011 Быстрый поиск словарных статей
- 09-01-2011 Алфавиты
- 22-12-2010 Тестовая база N-грамм в SDK
- 12-12-2010 Консольный поиск под Linux
- 09-12-2010 Релиз версии 9

Текущая версия: 9.11

о проекте

документы

online demo

скачать

купить

контакты

новости

Тезаурус грамматического словаря

<http://www.solarix.ru/>

Состав тезауруса

Синонимы (19 тыс.), антонимы (более 2500), уменьшительные формы (более 1150) и словообразовательные связи (более 220 тыс.), переводы, гипонимы и гиперонимы

Назначение тезауруса в составе словаря

Тезаурус - это база данных, хранящая информацию о различных связях (отношениях) между словами (словарными статьями) и словосочетаниями (фразовыми статьями).

Визуализация тезауруса

Тезаурус разработан так, чтобы максимально облегчить использование хранимой информации в алгоритмах компьютерной лингвистики, например при машинном переводе.

Тезаурус грамматического словаря

ThesaurusBrowser v.7.00.6072 Professional Win32 (build date Aug 21 2009)

Word: кош поиск

КОШМАРИТЬ (инф.)
КОШЕЛЕЧЕК (сущ.)
КОШЕЛЕК (сущ.)
КОШЕЛЬ (сущ.)
КОШМАР (сущ.)
КОШУРНИКОВ (сущ.)
КОШАК (сущ.)
КОШЕРНОСТЬ (сущ.)
КОШЕРНЫЙ (прил.)
КОШКА (сущ.)
КОШЕЧКА (сущ.)
КОШЕЛКА (сущ.)
КОШАЧИЙ (прил.)
КОШКОПОДОБНЫЙ (прил.)
КОШМАРНЫЙ (прил.)
КОШМАРНО (нареч.)

ко́шка сущ. одуш. ж.р.

Формы

	ед.ч.	мн.ч.
<i>им.п.</i>	кошка	кошки
<i>род.п.</i>	кошки	кошек
<i>твор.п.</i>	кошкой	кошками
<i>вин.п.</i>	кошку	кошек
<i>дат.п.</i>	кошке	кошкам
<i>предл.п.</i>	кошке	кошках

Синонимы: кот кошак киска киса

Антонимы по грам. роду: кот кошак

Грам. связи - прилагательное: кошачий кошкоподобный

Гипернимы: животное

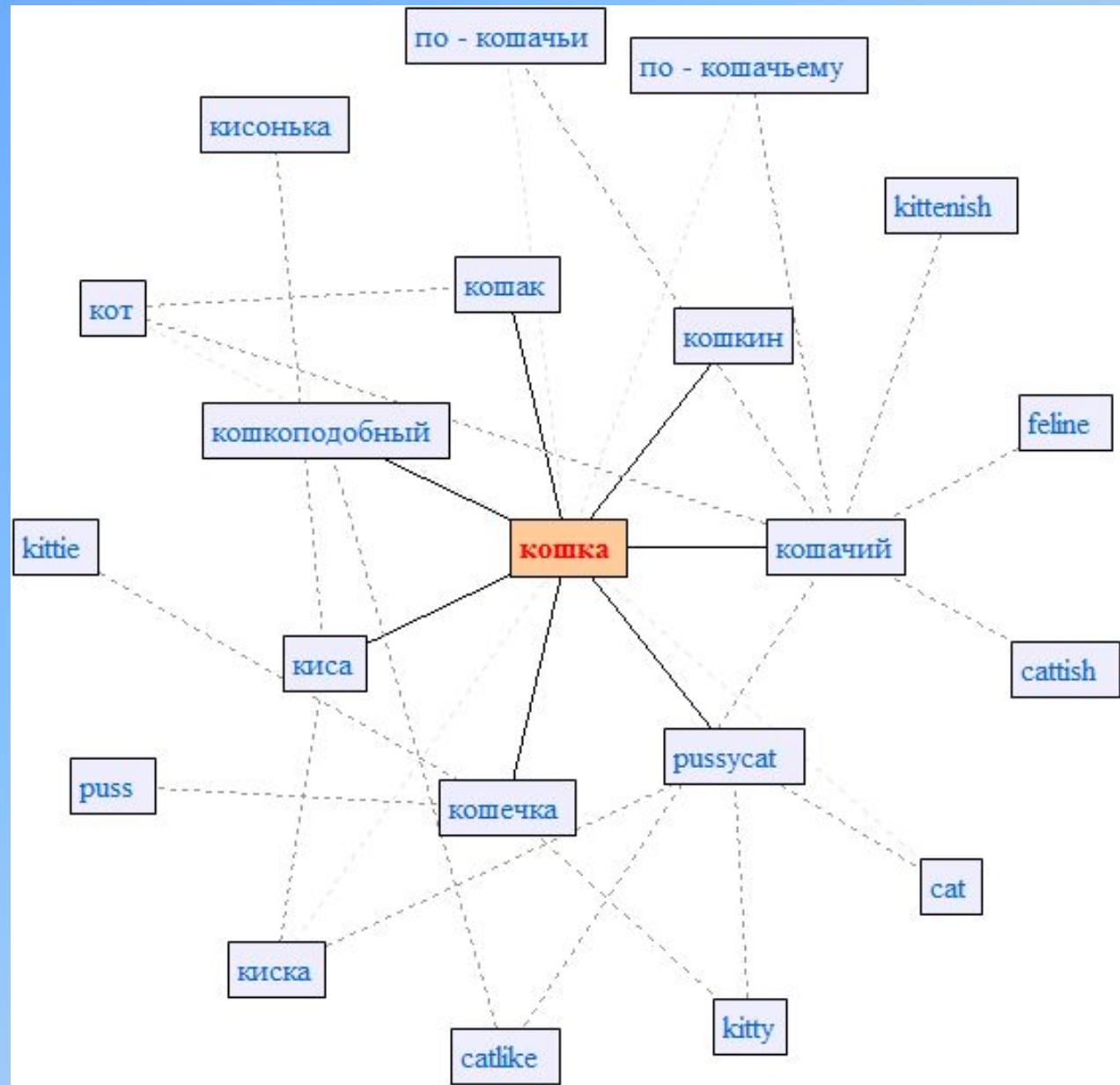
Гипонимы: котенок

Викисловарь: ссылка

Total: 16

SOLARIX loaded, 112476 entries

Визуализация тезауруса



Функции информационно-поисковых тезаурусов

- ✓ **Терминологическая** – отражение терминологии той или иной предметной области (отрасли знаний или их совокупности);
- ✓ **Систематизирующая** – представление системы понятий и терминов определенной предметной области и показ их взаимосвязей через систему помет (синоним, выше, ниже, ассоциация и др.);
- ✓ **Навигационная** – оперативная ориентация в терминологии даже незнакомой предметной области за счет явного выражения парадигматических отношений в структуре дескрипторной статьи;
- ✓ **Регламентирующая** (нормативная) – представление нормализованной лексики за счет устранения синонимии и многозначности;
- ✓ **Технологическая** – обеспечение технологии индексирования документов и запросов с целью поиска информации в АИС.