

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

# **Создание лексико- семантической основы ИПЯ**

# План

1. Отбор лексических единиц
2. Нормализация лексики
3. Систематизация и группировка лексических единиц
4. Построение классификационных схем понятий
5. Организационное оформление лексики ИПЯ

# Список литературы

1. Антопольский А. Б. Лингвистическое обеспечение электронных библиотек : учеб. пособие / А. Б. Антопольский. – Москва : Московский гос. ун-т культуры и искусств, 2002. – 229 с.
2. Гендина Н. И. Лингвистические основы информатики: гипертекст. учеб. терминолог. слов.-справ. : [рек. М-вом культуры в качестве учеб. пособия для вузов искусств и культуры] / Н. И. Гендина, И. Л. Скипор ; Кемеров. гос. акад. культуры и искусств. – Кемерово : Кузбассвузиздат, 2002. – 124 с.
3. ГОСТ 7.74-96. Информационно-поисковые языки. Термины и определения. – Введ. 1997-07-01. – Минск : Межгос. совет по стандартизации, метрологии и сертификации, 1997. – 34 с.

# Отбор лексических единиц

**Ключевое слово** – это полнозначное слово или словосочетание, являющееся носителем существенной в данном тексте (документе или запросе) информации с точки зрения информационного поиска.

**Неключевые слова** – служебные слова (частицы, предлоги, союзы, междометия); часто употребляемые общенаучные термины (*проблема, актуальность, значение* и т. п.); термины, не относящиеся к понятийному аппарату того раздела науки или техники, для которого создается конкретный ИПЯ.

**Термин** – слово специальной сферы употребления (науки, техники, ремесла), имеющее четко определенный смысл.

**Различают:**

- 1. Специальные термины**, употребляемые только в данной отрасли знания. Например: *«диагноз, гастрит, анестезия, переливание крови»;*  
*«антрацит, угольный карьер, проходка, метаноопасность»*
- 2. Термины смежных отраслей знания.** Например, в медицинской терминологии используются биологические, химические термины; терминология горного дела активно привлекает лексику геологии, кристаллографии.
- 3. Общенаучные термины**, бытующие в любой отрасли знания, например: *метод, анализ, оборудование, дефект, исследование, объект* и т. п.

# Источники отбора лексических единиц

- Ранее разработанные ИПЯ, располагающие фиксированным словарем: таблицы УДК, ББК, списки предметных рубрик, классификаторы, тезаурусы и т. п.;
- Терминологические ГОСТы, словари, справочники, энциклопедии;
- Массив документов по тематике создаваемого ИПЯ за последние 5-6 лет, объемом 3-4 тыс. документов;
- Массив запросов, полученных от потребителей информации, отражающий новые термины.

# Нормализация лексики

**Нормализация лексики** – это обеспечение однозначности лексических единиц ИПЯ. Достигается посредством двух операций:

1. Представлением ключевых слов (КС) в единообразной грамматической форме (морфологический уровень нормализации лексики);
2. Устранением синонимии и полисемии (семантический уровень).

# Нормализация лексики: представление КС в единообразной грамматической форме

- Установление унифицированной формы записи лексических единиц (род, число, падеж). Исчисляемые существительные рекомендуется записывать в именительном падеже множественного числа, неисчисляемые – в именительном падеже единств. числа: *автомобили, радиодетали, документы*, но *информация, реле, резонанс*.
- Решение вопроса о единообразной формулировке ключевого слова: запись в виде отдельного слова (унитерм) или словосочетания.

## Проблема устойчивого словосочетания:

- «Поиск», «Информация» или «Поиск информации»?  
«Комплектование фондов» или «Фонды», «Комплектование»?

## Проблема инверсии:

- «Гусеничные тракторы» или «Тракторы гусеничные»?

## Эвристическая (поисковая) функция инверсии:

- Тракторы болотоходные
- Тракторы гусеничные
- Тракторы колесные
- Тракторы трелевочные

# Нормализация лексики.

## Два способа устранения полисемии и омонимии

### 1. Развертывание многозначного слова до словосочетания.

Например: *Сети: сети рыбацкие; сети информационные; сети коммуникационные.*

### 2. Использование специальных ограничительных или пояснительных помет, помещаемых в скобках.

Например:

*Ребро (геометрия), ребро (авиация), ребро (анатомия);*

*Емкость (предмет), емкость (характеристика),*

*Смазка (процесс), смазка (материал);*

*Танк (резервуар), танк (вооружение);*

*Обработка (пищевых продуктов), обработка (информации).*

# Систематизация лексики

**Систематизация лексики** – это группировка и упорядочение ЛЕ по степени смысловой близости.

**Цель систематизации** – установление парадигматических отношений между отобранными лексическими единицами и построение классов условной эквивалентности.

**Класс условной эквивалентности (КУЭ)** – это совокупность ЛЕ, которые считаются условно синонимичными в данной области знания и используются в данном ИПЯ как равнозначные с точки зрения поиска информации.

**Виды парадигматических отношений между ЛЕ в КУЭ:**

1. Равнозначности, тождества, выражаемые синонимами;
2. Иерархии, подчинения, т. е. отношения типа «Род–Вид», «Целое–Часть», «Выше–Ниже»;
3. Ассоциативные.

**Пример класса условной эквивалентности понятия «ФБД»:**

- Фактографические базы данных; Фактографические БД; ФБД;
- Базы данных; информационные массивы; массивы данных;
- Системы управления базами данных; СУБД.

# Систематизация лексики.

## Построение классификационных схем понятий

**Классификационная схема понятия** – это графическое средство фиксирования парадигматических связей между понятиями.

**Ее основное назначение** – обозначение границы того или иного понятия, определение объема, содержания и структуры понятия, установление его связей с другими понятиями.

**Разработка классификационной схемы включает:**

- Выделение наиболее существенных признаков понятия в качестве основания деления;
- Построение частных схем, отражающих взаимосвязь понятий и использующих выделенные признаки.

# Логические правила деления понятий (классификации)

**1. Единство основания деления:** основание деления должно оставаться неизменным.

**Основание деления** – это признак, по которому производится разделение (классификация) объектов.

**Неправильно:** Издания делятся на периодические, непериодические и художественные. (**Смешаны два основания деления - периодичность и целевое назначение**).

**Правильно:** Издания делятся на периодические, непериодические и продолжающиеся (**Основание деления – периодичность**).

**Правильно:** Издания делятся на официальные, научные, научно-популярные, производственно-практические, нормативные производственно-практические, учебные, массово-политические, справочные, для досуга, литературно-художественные (**Основание деления – целевое назначение**).

# Логические правила деления понятий (классификации)

**2. Взаимоисключение видовых понятий:** объемы видовых понятий должны исключать друг друга.

**Неправильно:** Издания делятся на периодические, справочные, технические журналы, массовые газеты.

**Правильно:** периодические издания делятся на газеты и журналы.

# Логические правила деления понятий (классификации)

**3. Адекватность по объему:** объемы видовых понятий должны исчерпывать объем родового понятия.

**Неправильно:** Издания делятся на официальные, научные, научно-популярные, производственно-практические, нормативные, учебные, массово-политические, литературно-художественные.

**Правильно:** Издания делятся на официальные, научные, научно-популярные, производственно-практические, нормативные, производственно-практические, учебные, массово-политические, справочные, для досуга, литературно-художественные.

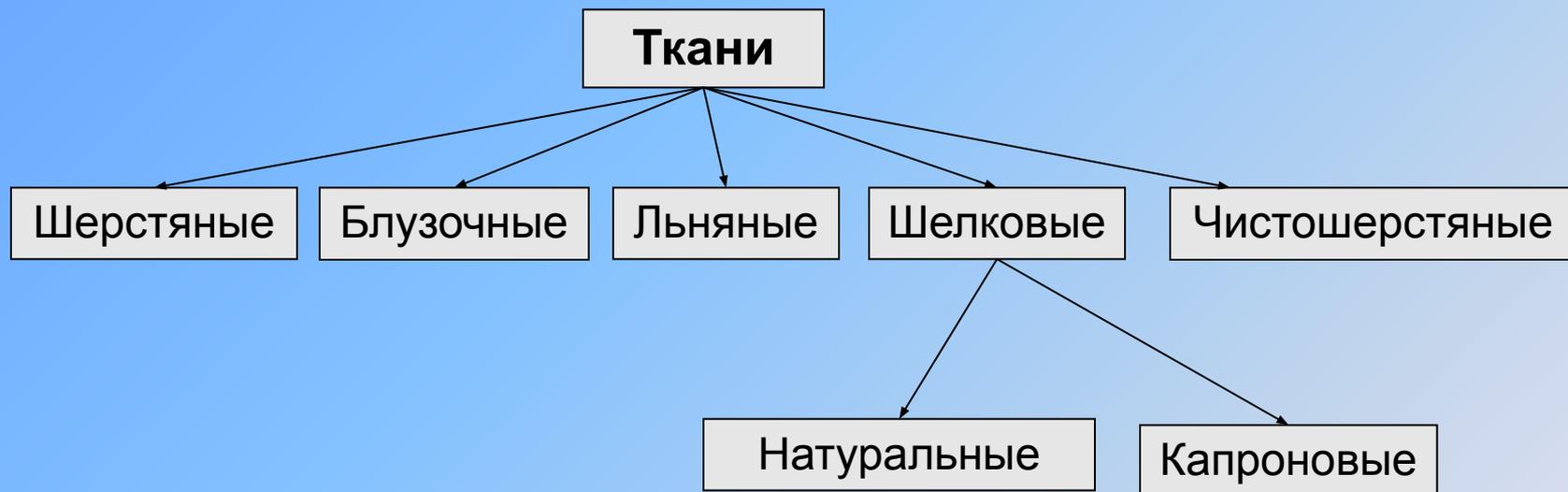
# Логические правила деления понятий (классификации)

**4. Непрерывность, отсутствие «логических скачков»:** родовое (делимое) понятие должно быть ближайшим видом для членов деления.

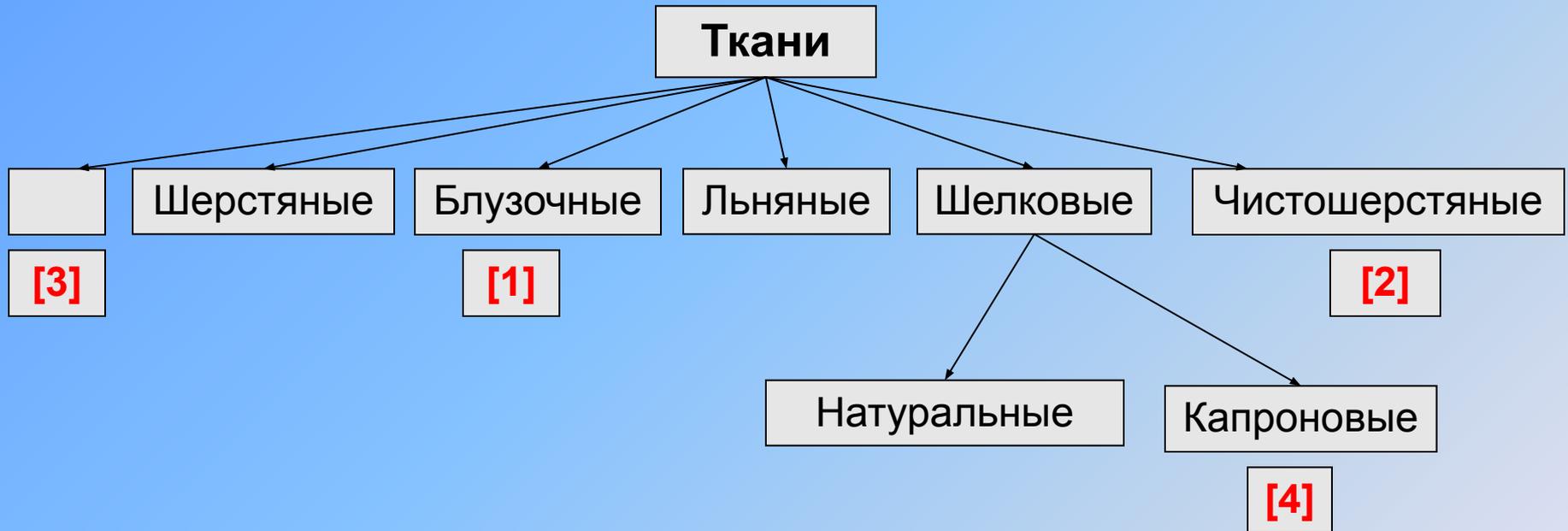
**Неправильно:** Издания делятся на периодические издания, книги, газеты, журналы.

**Правильно:** Издания делятся на периодические и непериодические. Периодические издания делятся на газеты и журналы, а непериодические - на книги и брошюры.

# Найдите ошибки в классификационной схеме понятия «Ткани»



# Классификационная схема понятия «Ткани»: анализ нарушений требований логики



## Ошибки (нарушения требований логики):

- 1 – смешаны разные основания деления (существенные признаки)
- 2 – нарушена взаимоисключаемость видовых понятий
- 3 – нарушена адекватность деления родового понятия «Ткани» по объему – не указаны хлопчатобумажные ткани
- 4 – нарушена непрерывность деления объема понятия, допущен логический «скачок»

# Организационное оформление лексики ИПЯ

Организационное оформление лексики – это ее графическое, знаковое воплощение, которое целиком зависит от типа разрабатываемого ИПЯ.

Классификационные ИПЯ оформляются в виде таблиц классификации, классификаторов; дескрипторные ИПЯ – в виде словарей (тезаурусов); язык предметных рубрик – в виде списка предметных рубрик (авторитетных файлов) и т.п.



# Фрагмент основных таблиц ДКД (780 Музыка. Основные деления)

780 Музыка

781 Основные принципы и музыкальные формы

782 Вокальная музыка

.1 Театральные вокальные формы. Оперы

.2 Нетеатральные вокальные формы

.3 Службы (Литургия и ритуал)

.4 Светские формы

.5 Смешанные голоса

.6 Женские голоса

.7 Детские голоса

.8 Мужские голоса

.9 Другие типы голосов

783 Музыка для отдельных голосов Голос

784 Инструменты и инструментальные ансамбли

785 Камерные ансамбли

786 Клавишные и другие инструменты

787 Струнные инструменты (Хордофоны)

788 Духовые инструменты (Аэрофоны)

# Общероссийский классификатор профессий рабочих, должностей служащих и тарифных разрядов



- ✓ Введение
- ✓ Профессии рабочих
- ✓ Информационный блок классификатора профессий рабочих
- ✓ Фасет 01. Виды производств и работ
- ✓ Фасет 02. Тарифные разряды
- ✓ Фасет 03. Классы (категории) квалификации
- ✓ Фасет 04. Формы и системы оплаты труда
- ✓ Фасет 05. Условия труда
- ✓ Фасет 06. Степень механизации труда
- ✓ Фасет 07. Производные профессии
- ✓ Должности служащих
- ✓ Информационный блок классификатора должностей служащих
- ✓ Фасет 10. Категории должностей
- ✓ Фасет 11. Производные должности
- ✓ Фасет 12. Категории (классы) квалификации

# Профессии рабочих (фрагмент общероссийского классификатора)

Код	КЧ	Наименование профессии	Диапазон тарифных разрядов	Код выпуска ЕТКС	Код по ОКЗ
10003	5	Авербандщик	2-5	47	7332
10005	4	Авиационный механик (техник) по планеру и двигателям	2-6	57	7515
10007	3	Авиационный механик (техник) по приборам и электрооборудованию	2-6	57	7515
10008	8	Авиационный механик (техник) по радиооборудованию	2-6	57	7515
10010	5	Авиационный техник (механик) по парашютным и аварийно-спасательным средствам	4-6	57	7515
10012	4	Авиационный техник по горюче-смазочным материалам	3-5	57	7515
10014	3	Автоклавщик	4	43	8139
10015	8	Автоклавщик	3-4	44	8139
30018	3	Автоклавщик		73	8122
10013	9	Автоклавщик	3-4	30	8152
10016	2	Автоклавщик литья под давлением	3	2	8122
10017	7	Автоклавщик на запарке брикетов	3	4	8112
10019	6	Автоклавщик-сушильщик аккумуляторных пластин в производстве свинцовых аккумуляторов	2-3	19	8290
10021	3	Автоматчик	2-4	2	8211

# Фрагмент АПУ книги

## Методы

- активного обучения 272; 279
- дидактические игры 282-283
- имитационные 280-281
- имитационные игровые 281
- имитационные неигровые 281
  - - действия по инструкции 281
  - - ситуация выбора 281
  - - ситуация выработки решения 281
  - - ситуация доказательства 281
  - - ситуация конфликта 281
  - - ситуация несоответствия 281
- исследовательские 270, 279
- наглядные 269; 275-276
  - - демонстрация 276
  - - иллюстрация 276

# Фрагмент информационно-поискового тезауруса по технологии машиностроения

## **ГОРНЫ 2703, 2709**

в Оборудование термическое

н Горны высокие

Горны кузнечные

Горны низкие

Горны плавильные

а Вагранки

Печи (оборудование)

Топки

## **ГОРНЫ ВЫСОКИЕ 2709**

в Горны

Горючее

см. Топливо

ГОСТ

См.Стандарт

## **ГОФРИРОВАНИЕ 2703**

в Обработка механическая

а Накатка

Рифление

## **ГРАВИРОВАНИЕ 2709**

с Гравировка

а Нанесение рисунка

Обработка электроискровая

Обработка электрохимическая

Чеканка

Гравировка

См. Гравирование