

Лекция 12

**Лингвистическое обеспечение ИС.
Знаковые системы.**

**Кодификаторы, классификаторы,
тезаурусы, онтологии.**

**Частотные словари, словари предметной
области.**

Информационно-поисковые языки.

**ИПЯ, основанные на естественных
языках**

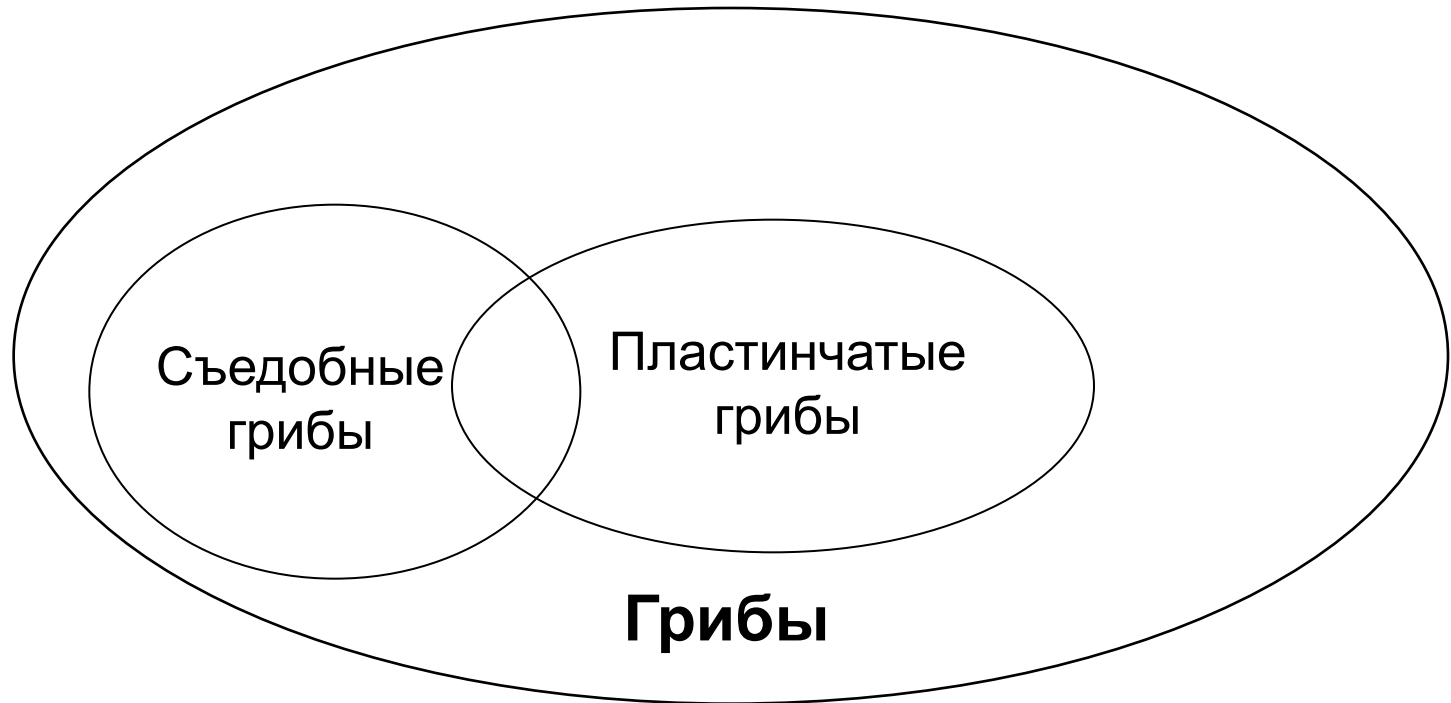
Базовые понятия лингвистического обеспечения ИС

- *Лингвистическое обеспечение* — языки представления и управления информацией
- *Язык* — знаковая система любой физической природы, выполняющая познавательную и коммуникативную функцию в процессе человеческой деятельности
- *Естественный язык* — основное средство общения и выражения мысли
- *Искусственный язык* — специализированный язык, основное назначение которого состоит в устранении многозначности слов естественного языка
- *Информационный язык* — искусственный язык, созданный в целях лучшей реализации основной функции языка — коммуникативной, то есть функции передачи информации

Внутренняя структура языка

- **Лексика** (словарный состав) — вся совокупность слов, входящих в состав языка
- **Грамматика** — система способов и средств построения слов и предложений в рассматриваемом языке
- **Морфология** — совокупность действующих в языке способов и средств построения слов
- **Морфема** — наименьшая неделимая без потери данного качества единица системы выражения, непосредственно соотносимая с соответствующим ей элементом содержания
- **Аффикс** (грамматическая морфема) — предназначен для видоизменения значения основной части слова (базы)
- **Лексема** (лексическая морфема) — корневая или словообразовательная морфема, выражающая основное значение слова
- **Синтаксис** — совокупность действующих в языке способов и средств построения предложений
- **Словообразование** — совокупность способов и средств образования слов на базе уже существующих
- **Парадигма** — лексико-семантическая группа слов, построенная на основе некоторого семантического признака

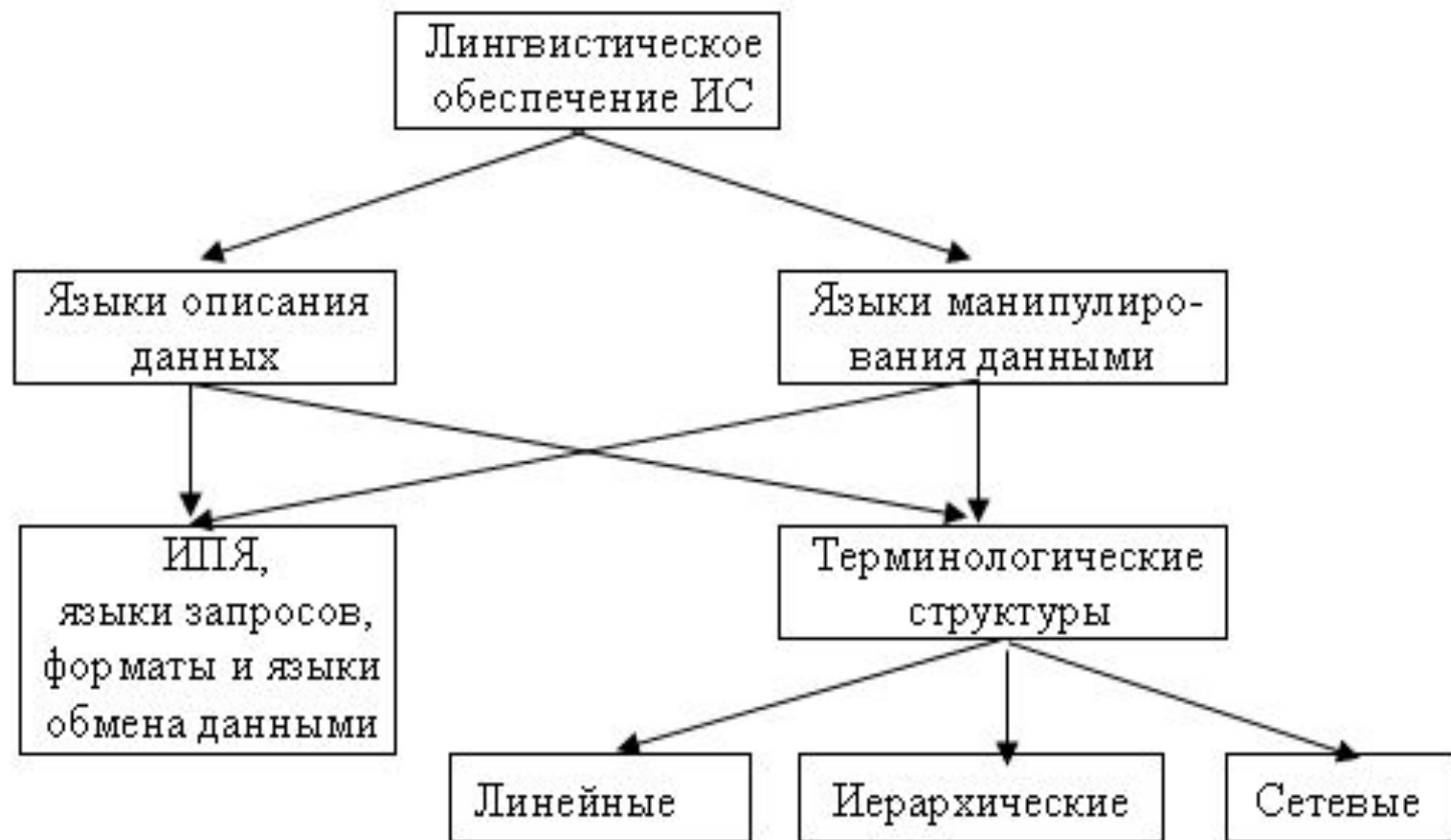
Пример лексико-семантической парадигмы



Типы отношений между словами

- *Парадигматические отношения* (или - аналитические, базисные отношения) – логические отношения, существующие между лексическими единицами языка независимо от контекста их конкретного употребления
- *Синтагматические отношения* (или - текстуальные, синтетические, синтаксические отношения) – это связи между лексическими единицами языка, выражающие действительные конкретные логические связи между соответствующими понятиями в тексте сообщения, которые, в свою очередь, отражают конкретные ситуативные отношения между обозначаемыми ими объектами, описываемыми в сообщении

Состав лингвистического обеспечения ИС



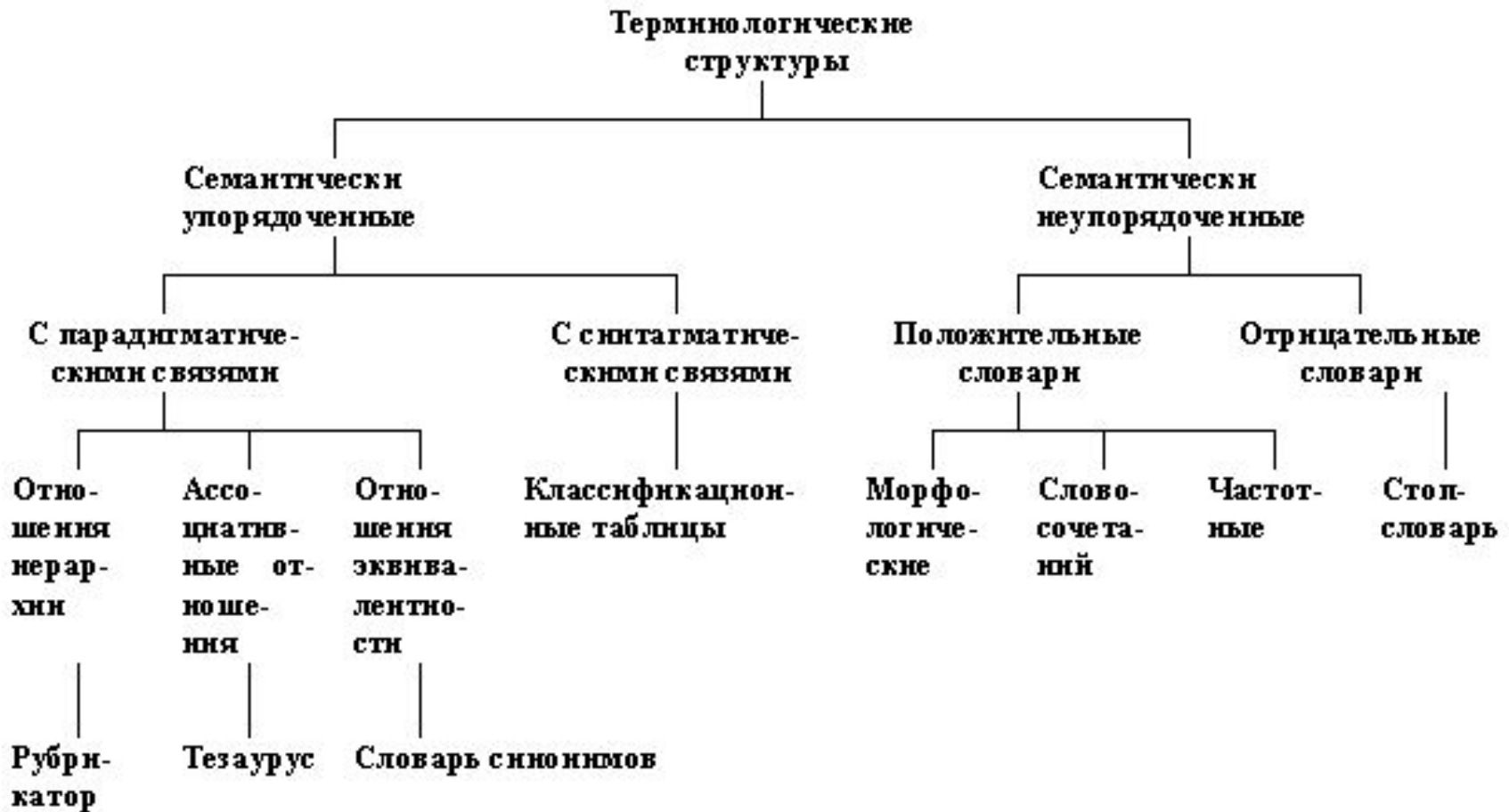
Лингвистическое обеспечение ИС - совокупность языковых средств, позволяющих представить информационную составляющую ИС на различных этапах внутрисистемной обработки и взаимодействия с пользователем

Информационно-поисковый язык (ИПЯ) - искусственный язык, специально сконструированный для удобного (человеку) выражения основного смыслового содержания документов и запросов с целью последующего их сопоставления (системой)

Типы знаковых систем

- ✓ Знаковые системы, построенные на «естественных знаках»
- ✓ Образные знаковые системы
- ✓ Языковые системы
- ✓ Знаковые системы записи
- ✓ Математико-формализованные (кодовые) знаковые системы

Типология терминологических структур по семантическому признаку



Фрагмент частотного словаря

51 ИНФОРМАЦИОННАЯ ГРАМОТНОСТЬ

1 ИНФОРМАЦИОННАЯ ГРАНИЦА
ВСЕЛЕННОЙ

1 ИНФОРМАЦИОННАЯ ДЕМОКРАТИЯ

1563 ИНФОРМАЦИОННАЯ ДЕЯТЕЛЬНОСТЬ

2 ИНФОРМАЦИОННАЯ ДИАГНОСТИКА

2 ИНФОРМАЦИОННАЯ ДОКТРИНА

1 ИНФОРМАЦИОННАЯ ДОКУМЕНТАЦИЯ

Фрагмент Рубрикатора ВИНТИ

201.01 Общие вопросы информатики

201.01.04 Информатизация общества. Информационная политика

201.01.07 Теория и методология информатики

201.01.07.03 Предмет и объекты исследования информатики

201.01.07.03.03 Информация как знание

201.01.07.03.07 Научная коммуникация

201.01.07.03.11 Информационная деятельность

201.01.07.06 Методы информатики

201.01.07.06.03 Наукометрический анализ

201.01.07.06.05 Семиотический подход

201.01.07.06.07 Кластерный анализ

201.01.07.06.09 Теория информации

201.01.07.06.11 Эвристические методы

201.01.07.06.13 Системный подход

201.01.07.06.15 Моделирование

201.01.07.06.99 Другие методы

201.01.07.08 Проблемы информатики

201.01.07.08.03 Взаимосвязи данных, информации и знания

201.01.07.08.05 Представление знаний

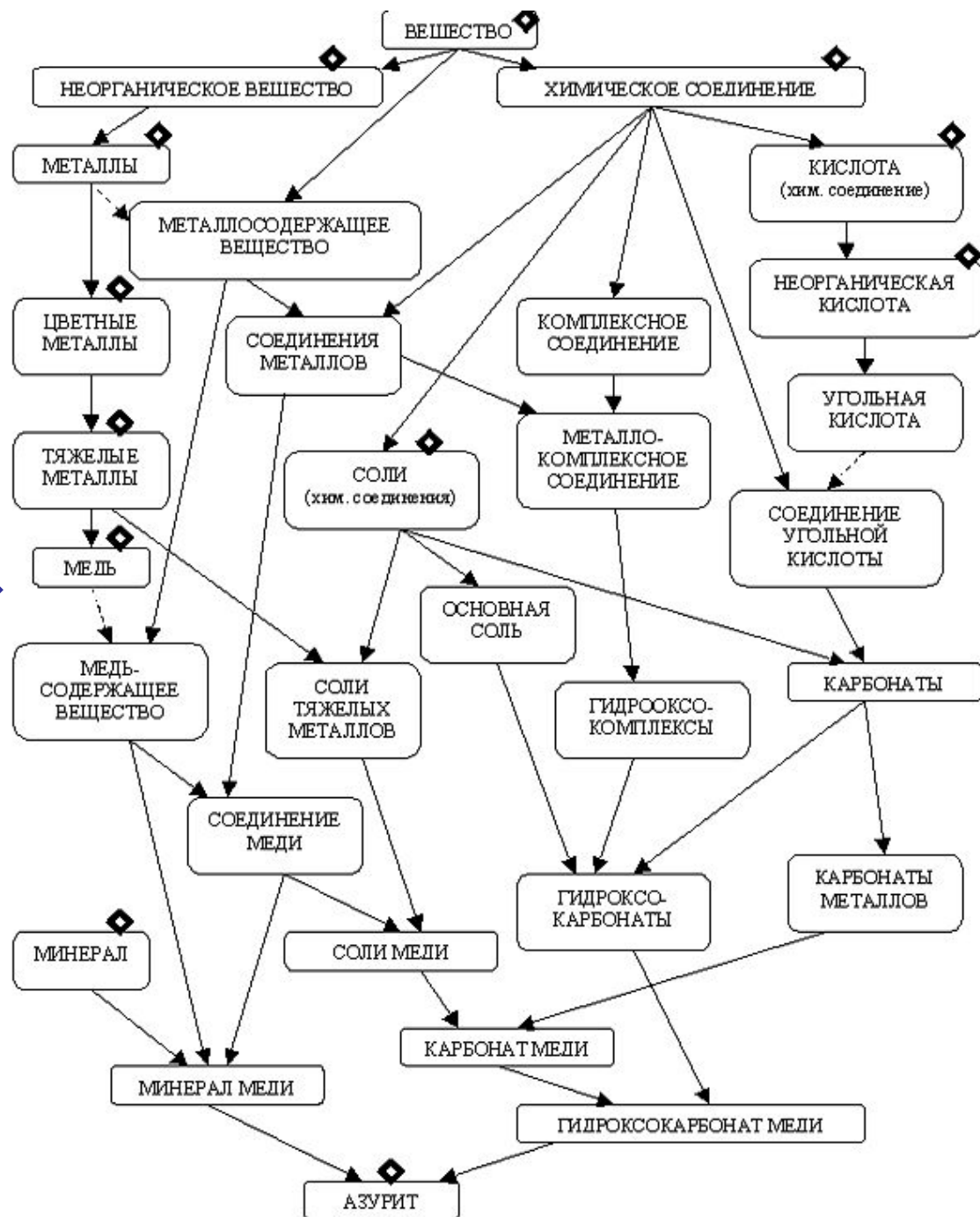
201.01.07.08.07 Разработка классификаций

Пример словарной статьи дескриптора Тезауруса по информатике

ИПЯ /информационно-поисковые языки /

- с** информационно-поисковые языки
поисковые языки
- ср** информационные языки
- в** ИСКУССТВЕННЫЕ ЯЗЫКИ
ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ
- н** ДЕСКРИПТОРНЫЕ ИПЯ
КЛАССИФИКАЦИИ
- а** ВХОДНЫЕ ЯЗЫКИ
ВЫХОДНЫЕ ЯЗЫКИ
ЕСТЕСТВЕННЫЕ ЯЗЫКИ
ИНДЕКСАЦИОННЫЕ ТЕРМИНЫ
ИНДЕКСИРОВАНИЕ
ИНФОРМАЦИОННО-ЛОГИЧЕСКИЕ ЯЗЫКИ
ИНФОРМАЦИОННЫЙ ПОИСК
ИПС
ИПТ
ПАРАДИГМАТИЧЕСКИЕ ОТНОШЕНИЯ
ПОСТКООРДИНАЦИЯ
ПРЕДКООРДИНАЦИЯ
ФОРМАЛИЗОВАННЫЕ ЯЗЫКИ

**Фрагмент
Онтологии по
естественным
наукам и
технологиям**



Типология методов описания содержания документов

Описание содержания документа

Классификация

1 документ → 1 класс
1 класс → N документов

ИПЯ
классификационного
типа

Индексирование

1 документ → N классов
1 класс → N документов

Рубрицирование
(предметизация)

Описание
понятиями

Описание
словами

Тезаурусное
(фиксированное)

Свободное

ИПЯ дескрипторного типа

Основные разделы ББК

Индекс для массовых библиотек	Разделы (таблицы) ББК	Индекс для научных библиотек
1	Общенаучное и междисциплинарное знание	А
2	Естественные науки	
20	Естественные науки в целом	Б
22	Физико-математические науки	В
24	Химические науки	Г
26	Науки о Земле (геодезические, геофизические, геологические и географические науки)	Д
28	Биологические науки	Е
3	Техника. Технические науки	Ж/О
4	Сельское и лесное хозяйство. Сельскохозяйственные и лесохозяйственные науки	П
5	Здравоохранение. Медицинские науки	Р
6/8	Общественные и гуманитарные науки	
60	Общественные науки в целом	С
63	История. Исторические науки	Т
65	Экономика. Экономические науки	У
66	Политика. Политическая наука	Ф
67	Право. Юридические науки	Х
68	Военное дело. Военная наука	Ц
70/79	Культура. Наука. Просвещение	Ч
80/84	Филологические науки. Художественная литература	Ш
85	Искусство	Щ
86	Религия. Мистика. Свободомыслие	Э
87	Философия	Ю
88	Психология	Ю
9	Литература универсального содержания	Я

Типовые деления ББК

Таблица	Кодирование	Назначение	Примеры
<i>Основные типовые деления</i>			
ОТД	Строчные буквы русского алфавита (первый уровень деления) Арабские цифры (второй, третий и последующие уровни деления)	Тематические и формальные деления	22.141я72 - Учебники для средней школы по алгебре. <i>Здесь:</i> 22.141 - Элементарная алгебра, я72 - Учебные издания для средней школы.
ТТД	Цифры и буквы русского алфавита, прописные и строчные, заключенные в круглые скобки	Деление по региональному признаку	<i>Страны мира:</i> (4) - Европа (4Авс) - Австрия (4Алб) - Албания (5Япо) - Япония 60.7(4) - Народонаселение Европы, 30г(2) - История техники в России и СССР <i>Крупные физико-географические регионы:</i> (53) - Западная (Передняя) Азия (531) - Кавказ
ТДС	Арабские цифры с опознавательным знаком апостроф (')	Деление по признаку социальной системы	66.2(0)'6 - Политика и современное политическое положение развитых стран. <i>Здесь:</i> 66.2 - Политика и современное политическое положение; (0) - деление ТТД, обозначающее весь мир, все страны; '6 - деление ТДС, обозначающее систему развитых стран.
<i>Специальные типовые деления</i>			
СТД	Последовательность арабских цифр и знак «-» (дефис) перед цифровой частью	Используются только в пределах одного или нескольких классов одной отрасли знания	(4Лат-5) - Районы Латвии, (5Гру-6) - Автономные республики Грузии, (4Укр-6Кры-2) - Города Крыма.

Фасеты УДК

№ пп	Наименование фасета	Тип фасета	Нотация фасета
1	Общий предмет	Фасет содержания	индекс
2	Место	Фасет содержания	(индекс)
3	Народность	Фасет содержания	(=индекс)
4	Время	Фасет содержания	"индекс"
5	Язык документа	Фасет формы	= индекс
6	Форма документа	Фасет формы	(0индекс)

**Главные классы
фасета
«Общий предмет»**

- 0 Общий отдел
- 1 Философия. Психология. Логика
- 3 Общественные науки
- 5 Математика. Естественные науки
- 6 Прикладные науки. Медицина. Техника
- 7 Искусство. Декоративно-прикладное искусство. Фотография.
Музыка. Игры Спорт
- 8 Языкознание. Филология. Художественная литература.
Литературоведение
- 9 География. Биографии. История

Фрагмент таблицы УДК

004 Информационные технологии. Вычислительная техника. Теория, технология и применения вычислительных машин и систем.

Все подклассы 004 подразумевают цифровую обработку данных, кроме тех, которые включают индексы 004.386 или 004.387.

=> [621.3.049.77](#) Микроэлектроника. Интегральные схемы

=> [621.39](#) Электросвязь

Основные деления

- 004.2 Архитектура вычислительных машин
- 004.3 Аппаратные средства. Техническое обеспечение
- 004.4 Программные средства
- 004.5 Человеко-машинное взаимодействие. Пользовательский интерфейс
- 004.6 Данные
- 004.7 Связь компьютеров. Сети ЭВМ. Вычислительные сети
- 004.8 Искусственный интеллект
- 004.9 Прикладные информационные (компьютерные) технологии

Специальные определители

- 004.01 Документация
- 004.02 Методы решения задач
 - 004.021 Алгоритмы
 - => [004.421](#) Алгоритмы составления программ
 - 004.023 Эвристические методы
- 004.03 Типы и характеристики систем
 - 004.031 Типы систем
 - 004.031.2 Автономные системы. Системы с пакетной обработкой
 - 004.031.4 Неавтономные системы. Онлайн-овые системы
 - 004.031.42 Интерактивные системы
 - 004.031.43 Системы реального времени. Системы обработки транзакций
 - 004.031.6 Встроенные системы
 - 004.032 Характеристики систем
 - 004.032.2 Режим обработки данных
 - 004.032.22 последовательный
 - 004.032.24 параллельный, одновременный
 - => [004.272](#) Архитектуры параллельной обработки
 - 004.032.26 Нейронные сети
 - 004.032.3 Согласование по времени. Задание времени цикла
 - => [004.074.34](#) Время цикла памяти
 - 004.032.32 Синхронные процессы
 - => [004.451.23](#) Синхронизация
 - 004.032.322 Синхронизирующая частота. Тактовая частота

Международная патентная классификация

- A - удовлетворение жизненных потребностей человека;
- B - различные технологические процессы, транспортирование;
- C - химия, металлургия;
- D - текстиль, бумага;
- E - строительство, горное дело;
- F - механика, освещение, отопление, двигатели и насосы, оружие, боеприпасы, взрывные работы;
- G - физика;
- H - электричество.

Уровень	Наименование	Индекс	Примеры
1	Раздел	латинская заглавная буква от А до H	H – Электричество
2	Класс	двузначное число	H 04 Техника электрической связи H 05 Специальные области техники, не отнесенные к другим классам.
3	Подкласс	заглавная буква латинского алфавита	H 04 J Многоканальные системы связи H 05 K Печатные схемы, корпуса или конструктивные элементы электрических приборов; изготовление блоков электронной аппаратуры
4	Группа	одно-, двух- или трехзначное число, наклонная черта и два нуля	H05 K 1/00 Печатные схемы A 63 F 13/00 Аспекты игр с использованием двух- и более размерного дисплея, работающего на электронных средствах, например, телевизионного экрана, и показывающего изображения, относящиеся к игре A 63 F 9/00 Различные игры
5	Подгруппа	одно-, двух- или трехзначное число основной группы, наклонная черта и, по крайней мере, две цифры, кроме 00. Каждая следующая цифра обозначает дальнейшее десятичное деление предыдущей цифры	A 63 F 13/10 Управление ходом игры, например началом, развитием, концом A 63 F 9/28 Игры с цепной реакцией опрокидывания частей

Разделы МПК

Дескрипторные ИПЯ

- *Дескрипторы* — это предназначенные для координатного индексирования документов и информационных запросов нормативные ключевые слова, которые по определенным правилам отобраны из основного словарного состава того или иного естественного языка
- *Дескрипторный язык* — специальный ИПЯ, словарный состав которого состоит из дескрипторов, а грамматика — по крайней мере, из способа построения поисковых образов документов и поисковых предписаний путем координации соответствующих дескрипторов.

Типология ИШЯ с точки зрения семантической силы

