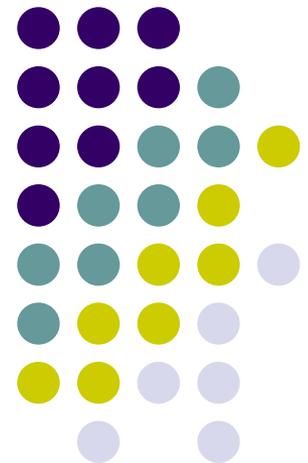


**ПРЕДСТАВЛЕНИЕ
ЗНАНИЙ:**

**ЛИНГВИСТИЧЕСКИЕ
ОНТОЛОГИИ**



СОДЕРЖАНИЕ



1. Подходы к созданию онтологий
2. Виды лингвистических онтологий
3. -ономии и рубрикаторы
4. Тезаурусы:
 - Информационно-поисковые тезаурусы
 - РуТез
5. Тезаурус WordNet
6. Заключение
7. Домашнее задание

ФОРМАЛЬНЫЙ ПОДХОД К СОЗДАНИЮ ОНТОЛОГИЙ



- Методология построения: от понятий
- Как правило, опирается на формальный язык и логику (например, предикатов первого порядка)
- На выходе – формальная онтология, включающая математически точные аксиомы и **языковонезависимая**:
 - ✓ не содержит единиц, специфичных для какого-либо естественного языка (ЕЯ)
 - ✓ в общем случае ее понятия не имеют взаимнооднозначного соответствия со значениями языковых единиц

При этом для названий понятий, отношений и пр. все равно используются слова и словосочетания ЕЯ

ЛИНГВИСТИЧЕСКИЙ ПОДХОД К СОЗДАНИЮ ОНТОЛОГИЙ



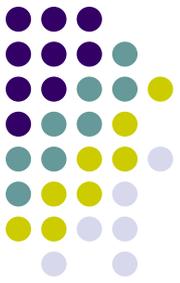
- Методология построения: от слов к понятиям
- Предполагает изучение ЕЯ и привлечение корпусов текстов
- На выходе – менее формальный ресурс онтологического характера
(*лингвистическая онтология*)
- Поскольку понятия связаны со значениями языковых единиц, для переноса на другой язык требуется серьезная адаптации
- Для более строгого описания знаний о мире необходимо сопоставлять лингвистические онтологии с формальными

ЛИНГВИСТИЧЕСКИЕ ОНТОЛОГИИ: ВИДЫ

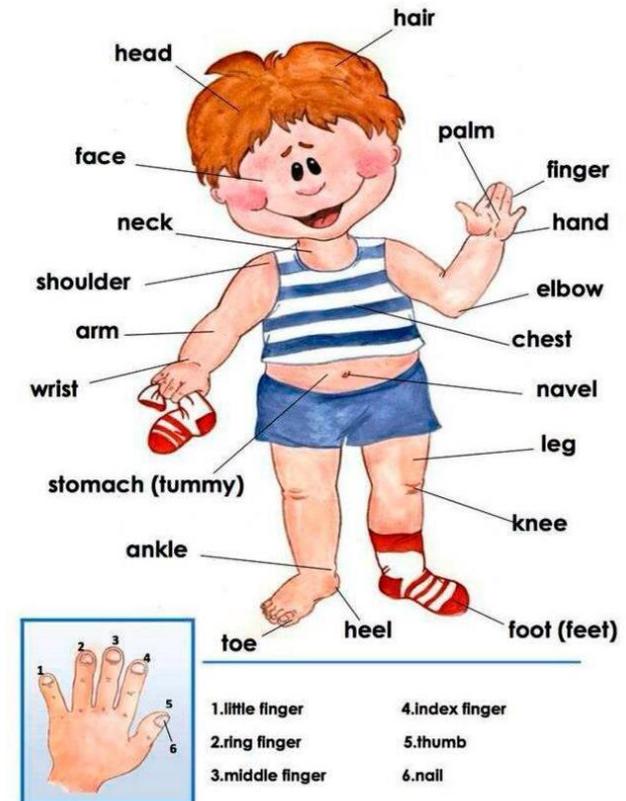
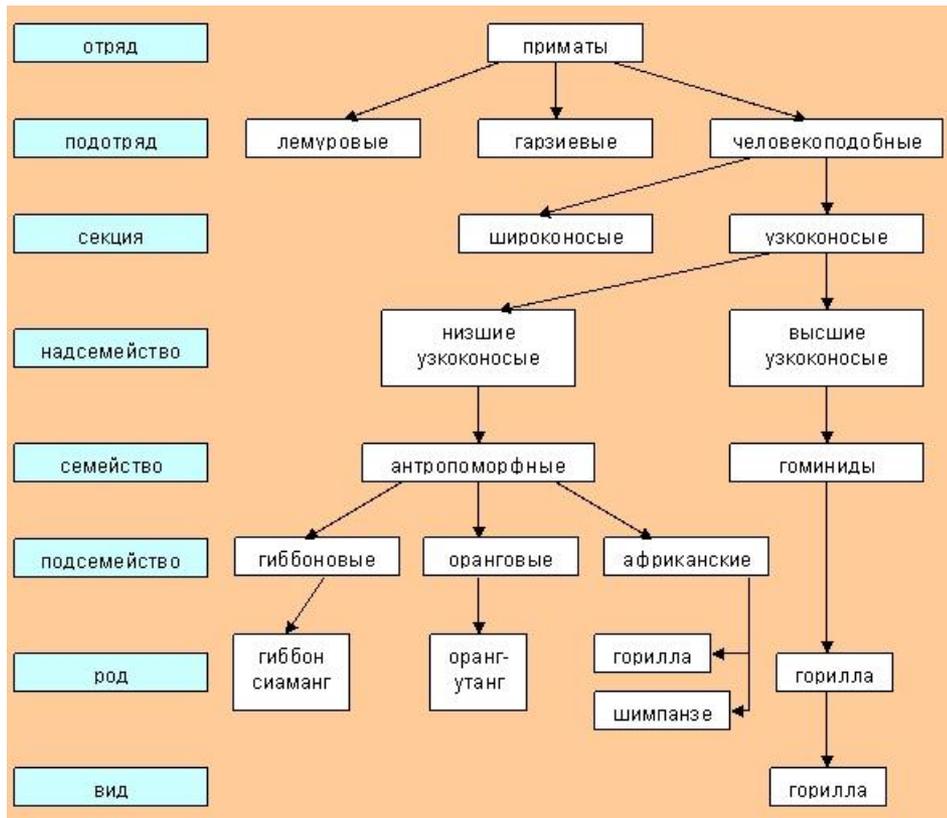


- ◆ *-ономия* – понятия узкой области в одном определенном отношении
- ◆ *Рубрикатор (классификатор, каталог)* – иерархическая/фасетная классификация понятий
- ◆ *Тезаурус* – понятия предметной области (ПО) в отношениях синонимии, иерархии, ассоциации
- Различие: создаются для разных целей
- Сходство: понятиям сопоставлены текстовые/языковые (лексические) единицы

-ОНОМИИ



- ◆ Таксономия – отношение класс-подкласс
- ◆ Партономия – отношение часть-целое



РУБРИКАТОРЫ



- Предназначены для **систематизации** информации, знаний, документов
- Содержат взаимно непересекающиеся *рубрики*, покрывающие всю ПО
- Назначение рубрики – четко определить отдельную концептуальную категорию
- Зачастую названия рубрик длинные, «неуклюжи» и не встречаются в текстах: *прочие отрасли экономики, художественная литература на балтийских языках*
- Отношения между рубриками не сводятся к одному конкретному типу
- Для автоматического использования каждой рубрике сопоставляются текстовые единицы

РУБРИКАТОРЫ: КАТАЛОГ ЯНДЕКС.МАРКЕТ



Витрина

Новинки, скидки, рекомендации

Каталог

Более 1500 категорий товаров



Скидки

Товары для дачи · Гардероб ·
Техника для дома · Мебель ·
Электроника · Ремонт · Спорт · Детям ·
Косметика и уход

[Все скидки >](#)



Электроника

Телевизоры · Фотокамеры · Объективы ·
Видеокамеры · Игровые приставки ·
Наушники · Мобильные телефоны ·
Чехлы · Умные часы

[Вся электроника >](#)



Компьютеры

Планшеты · Ноутбуки · Настольные ·
Комплектующие · Аксессуары ·
Мониторы · Wi-Fi · Принтеры ·
Клавиатуры и мыши

[Все компьютеры >](#)



Авто

Шины · Диски · Регистраторы ·
Магнитолы · Акустика · GPS ·
Детские автокресла · Запчасти ·
Аксессуары · Моторные масла

[Всё для авто >](#)



Одежда и обувь

Платья · Женская одежда ·
Мужская одежда · Обувь · Сумки ·
Наручные часы · Аксессуары ·
Детская одежда

[Вся одежда и обувь >](#)



Детские товары

Для малышей · Коляски · Кроватки ·
Школа и творчество · Игрушки ·
Конструкторы · Велосипеды · Самокаты

[Все детские товары >](#)

ТЕЗАУРУСЫ



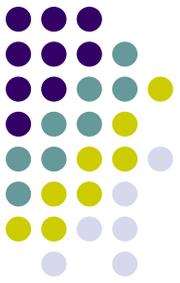
- Тезаурус (или *идеографический* словарь):
 - ✓ статьи упорядочены по смыслу, а не по алфавиту
 - ✓ смысл понятия выражается посредством соотнесения его с другими понятиями и их группами, а не через определение
- Описывает понятия определенной ПО, **облегчает коммуникацию** людей
- Понятиям сопоставлены языковые единицы
- Единицы с близкими значениями сгруппированы в концепты (понятия/*дескрипторы*), семантические отношения между концептами явно указаны
- Для удобства пользования снабжается алфавитным указателем

РАЗНОВИДНОСТИ ТЕЗАУРУСОВ



- Типа Тезауруса Роже:
 - цель – классификация слов ЕЯ для подбора синонимов и близких по смыслу слов, например, при написании текстов
 - отношения, как правило, не специфицированы
- Информационно-поисковые тезаурусы:
 - используются при информационном поиске (ИП)
 - цель – унифицированное описание содержания документов и/или поисковых запросов к автоматизированным ИП системам
- Типа WordNet:
 - цель – подробное описание лексической системы конкретного ЕЯ

ТЕЗАУРУС РОЖЕ



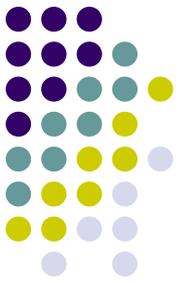
- Автор – Peter Mark Roget (1779–1869)
- Принцип классификации:
 - на верхних уровнях – объекты/понятия действительности
 - на нижнем – слова
- Слова организованы в пронумерованные группы. Название группы – слово с широкой семантикой
- В первом издании 1852 г. – 1000 групп
- В группах слова разбиты по частям речи, расположение – по степени близости значений
- От некоторых слов есть отсылки к другим словам, возможны пометы, ограничивающие употребление
- Противоположные по смыслу группы слов следуют друг за другом (*connect* – *disconnect*)

ТЕЗАУРУС РОЖЕ. СХЕМА КЛАССИФИКАЦИИ (ФРАГМЕНТ)



| Классы | Категории | Секции (+Понятийные группы) |
|---------------------------------|------------------------|--------------------------------|
| I. Абстрактные отношения | 1. Бытие | |
| | 2. Отношение | |
| | 3. Количество | |
| | 4. Порядок | |
| | 5. Число | |
| | 6. Время | |
| | 7. Изменение | |
| | 8. Возможность | |
| | 9. Причинность | |
| | 10. Сила | |
| II. Пространство | 1. Пространство вообще | A. Абстрактное пространство |
| | | 185. Присутствие |
| | | 186. Отсутствие |
| | | 187. Проживание |
| | | 188. Рождение |
| | | 189. Житель |
| | | 190. Местопребывание |
| | | 191. Помещение |
| | | 192. Вместительность |
| | | 193. Содержимое |
| B. Ограниченное пространство | | |
| C. Существование в пространстве | | |
| D. Относительное пространство | | |
| 2. Мера | | |
| 3. Структура, форма | | |
| 4. Движение | | |

ТЕЗАУРУС РОЖЕ. ГРУППА СЛОВ (ФРАГМЕНТ)



Land 342

N. land, earth, ground, dry land, terra firma

continent, mainland, peninsula; tongue of land, neck of land; isthmus, oasis; highland &c. (height) 206.

riverbank, river bank, levee

...

V. land, come to land, set foot on the soil; come ashore, go ashore, debark

Adj. earthy, continental, midland, coastal, littoral, riparian,; alluvial; terrene &c. (world) 318; landed, predial, territorial

Adv. ashore; on shore, on land

ИНФОРМАЦИОННО- ПОИСКОВЫЕ ТЕЗАУРУСЫ



- Информационно-поисковый тезаурус (ИПТ) – нормативный словарь терминов ПО, явно указывающий отношения между ними
- Термины – слова и словосочетания ЕЯ, называющие понятия ПО: *абсолютно черное тело, молодняк животных, торговля на вынос*
- Основное развитие ИПТ – 60-80 гг. XX века:
 - сначала: поиск по библиографическим данным и реферату
 - затем: сопоставление документу терминов ПО, характеризующих его содержание
- Существуют стандарты ISO и ГОСТ на разработку ИПТ

СОСТАВ ИПТ



- ◆ *Дескрипторы и аскрипторы:*
 - понятие ПО обычно имеет синонимы
 - среди них выбирается термин, используемый как основной способ ссылки на понятие – *дескриптор*
 - другие термины из синонимического ряда – *аскрипторы* (А) – используются для поиска подходящих дескрипторов
языкознание – лингвистика – языковедение (А)
- ◆ Отношения между дескрипторами:
 - ✓ родовидовые отношения:
мебель: вид стулья, прятки: род игра
 - ✓ отношение ассоциации (все остальное)

ПРИМЕРЫ АССОЦИАТИВНЫХ ОТНОШЕНИЙ (ГОСТ 7.25-2001)



- часть-целое: *комната, лестница – дом*
- причина-следствие: *смерть – похороны*
- сырье-продукт: *сахарная свекла – сахар*
- административная иерархия: *заведующий кафедрой – декан – ректор*
- процесс-объект: *рубка леса – дерево*
- процесс-субъект: *рубка леса – лесоруб*
- функциональное сходство: *гнездо – нора*
- свойство-носитель свойства: *огнедышащий – дракон, хрупкий – стекло*
- антонимия: *революция – контрреволюция*

ПРИМЕНЕНИЕ ИПТ: РУЧНОЕ ИНДЕКСИРОВАНИЕ



- Процедура ручного индексирования экспертом:
 - определяется содержание документа
 - для его описания подбирается подходящий набор дескрипторов
 - для уточнения набора дескрипторов документа используются их отношения
- Применяется для обработки документов в государственных и международных организациях:
 - ✓ тезаурус AGROVOC продовольственной и сельскохозяйственной организация при ООН
 - ✓ тезаурус EUROVOC законодательных служб ЕС
 - ✓ тезаурус научно-технических терминов , например, Ю.А. Шемакин, 1972 г.

АВТОМАТИЧЕСКОЕ ИНДЕКСИРОВАНИЕ



- При применении ИПТ, созданного для ручного индексирования, для автоматического индексирования:
 - важные термины документа могут быть выражены иначе, чем в тезаурусе (всего 31% документов явно содержит дескрипторы, приписанные экспертом)
 - менее значимые термины могут найтись в документе (90% дескрипторов, найденных в тексте, не приписаны документу экспертом)
- Для эффективной работы ИП приложений необходимо построение специального ИПТ, в котором для понятия фиксируются всевозможные текстовые выражения
прикладное программное обеспечение – ППО – прикладное ПО

ПРИМЕР ИПТ: РУТЕЗ



- ◆ Разрабатывается в НИВЦ МГУ с 1994 г.
- ◆ РуТез – ИПТ для текстов русском языке.
Состав:
 - **Общественно-политический тезаурус:**
экономика, политика, спорт, искусство и др.
 - **Общий лексикон**, независящий от ПО
- ◆ Объем:
 - ✓ 55 тыс. понятий
 - ✓ 158 тыс. текстовых выражений
 - ✓ 210 тыс. отношений (установлены вручную)
 - ✓ 125 тыс. текстовых входов на английском

ОТНОШЕНИЯ В РУТЕЗ: НИЖЕ-ВЫШЕ И ЧАСТЬ-ЦЕЛОЕ



- ❑ Отношение НИЖЕ-ВЫШЕ: *щенок: выше собака*
- ❑ Отношение ЧАСТЬ-ЦЕЛОЕ (части жестко и всегда связаны с целым): *глаз: часть веко*

➤ Существуют модификаторы, показывающие, что отношение слабое:

- В («возможно») – отношение выполняется не для всех примеров
- А («аспект») – отношение существует не все время

пенсионер:

выше_в старый человек

целое_а пенсионная система

ОТНОШЕНИЯ В РУТЕЗ: АССОЦИАЦИЯ



- Устанавливается, если для связанных по смыслу понятий не подошли предыдущие отношения
- Виды ассоциации:
 - ❖ несимметричная ассоциация – X не может существовать без Y: $\text{ассоц1}(X, Y) = \text{ассоц2}(Y, X)$
*саммит **ассоц1** глава государства*
*жидкость **ассоц2** кипение жидкости*
 - ❖ симметричная ассоциация – при поиске текстов о X могут пригодиться тексты о Y:
*яблоко **ассоц** яблоня, родители **ассоц** дети*
- Составляют около половины отношений РуТез
- Другое название: отношение **онтологической зависимости**

ТЕКСТОВЫЙ ВХОД: УРОК



ДОМАШНЕЕ ЗАДАНИЕ

(ДОМАШНЕЕ ЗАДАНИЕ, ДОМАШНЯЯ РАБОТА, ЗАДАНИЕ НА ДОМ, УРОК, УРОКИ НА ДОМ)

ВЫШЕ УЧЕБНОЕ ЗАДАНИЕ
ЦЕЛОЕ ЗАДАТЬ УРОК НА ДОМ
ЦЕЛОЕ ОБУЧЕНИЕ, УЧЕБНАЯ ДЕЯТЕЛЬНОСТЬ

ШКОЛЬНЫЙ УРОК

(УРОК, УРОК В ШКОЛЕ, УРОЧНЫЙ, ШКОЛЬНЫЕ ЗАНЯТИЯ, ШКОЛЬНЫЙ УРОК)

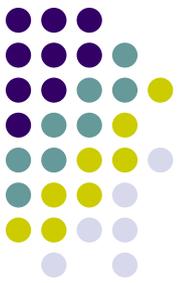
ВЫШЕ УЧЕБНЫЕ ЗАНЯТИЯ
ВЫШЕ УЧЕБНЫЙ ЧАС
ЦЕЛОЕ СРЕДНЕЕ УЧЕБНОЕ ЗАВЕДЕНИЕ
АССОЦ₂ ШКОЛЬНАЯ ПЕРЕМЕНА

УРОК НА БУДУЩЕЕ

(НАУКА, НАУКА НА БУДУЩЕЕ, ПОСЛУЖИТЬ НАУКОЙ, ПОСЛУЖИТЬ УРОКОМ, ПОУЧИТЕЛЬНЫЙ УРОК, УРОК, УРОК НА БУДУЩЕЕ)

ВЫШЕ РЕАЛЬНЫЙ ПРИМЕР
ЦЕЛОЕ ОПЫТ (СОВОКУПНОСТЬ ЗНАНИЙ И УМЕНИЙ)
ЦЕЛОЕ_В ПОУЧЕНИЕ, НАСТАВЛЕНИЕ
ЧАСТЬ ПОУЧИТЕЛЬНЫЙ
АССОЦ₂ ПРОУЧИТЬ (НАКАЗАТЬ)

ТЕКСТОВЫЙ ВХОД: *ВЕТЕР В ГОЛОВЕ*

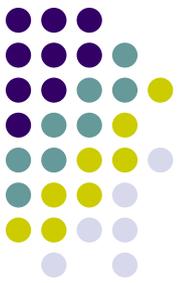


ЛЕГКОМЫСЛЕННЫЙ, БЕЗДУМНЫЙ

(БЕЗДУМНОСТЬ, БЕЗДУМНЫЙ, БЕЗЗАБОТНОСТЬ, БЕЗЗАБОТНЫЙ, БЕСПЕЧНОСТЬ, БЕСПЕЧНЫЙ, ВЕТЕР В ГОЛОВЕ, ВЕТРЕННОСТЬ, ВЕТРЕННЫЙ, ЛЕГКОМЫСЛЕННОСТЬ, ЛЕГКОМЫСЛЕННЫЙ, ЛЕГКОМЫСЛИЕ, НЕ СЕРЬЕЗНОСТЬ, НЕ СЕРЬЕЗНЫЙ)

| | |
|--------------------|---|
| ВЫШЕ | <u>ПОВЕДЕНИЕ</u> |
| ВЫШЕ | <u>ПУСТОГОЛОВЫЙ (ГЛУПЫЙ, НЕ СЕРЬЕЗНЫЙ)</u> |
| ЦЕЛОЕ | <u>ЛЕГКОМЫСЛЕННЫЙ ЧЕЛОВЕК</u> |
| НИЖЕ | <u>БЕЗАЛАБЕРНЫЙ (БЕСТОЛКОВЫЙ И БЕСПОРЯДОЧНЫЙ)</u> |
| НИЖЕ | <u>БЕЗОТВЕТСТВЕННЫЙ</u> |
| НИЖЕ | <u>БЕСШАБАШНЫЙ</u> |
| НИЖЕ | <u>ИГРИВЫЙ, ШАЛОВЛИВЫЙ</u> |
| НИЖЕ | <u>ЛЕГКИЙ (ПОВЕРХНОСТНЫЙ, ЛЕГКОМЫСЛЕННЫЙ)</u> |
| НИЖЕ | <u>РАСТОЧИТЕЛЬНЫЙ</u> |
| АССОЦ | <u>СЕРЬЕЗНЫЙ, СЕРЬЕЗНО НАСТРОЕННЫЙ</u> |
| АССОЦ ₁ | <u>БЕЗДУМНЫЙ (НЕРАЗУМНЫЙ)</u> |
| АССОЦ ₁ | <u>НЕРАЗУМНЫЙ (НЕУМНЫЙ, НЕЦЕЛЕСООБРАЗНЫЙ)</u> |
| АССОЦ ₂ | <u>ИГРАТЬ, ОБРАЩАТЬСЯ ЛЕГКОМЫСЛЕННО</u> |
| АССОЦ ₂ | <u>РАСТОЧИТЕЛЬСТВО</u> |

ТЕЗАУРУС *WORDNET*: ОБЩАЯ ХАРАКТЕРИСТИКА



- **WordNet** – большой лексический ресурс для английского языка
- Разрабатывался с 1985 г. в Принстонском университете, у истоков – психолог Джордж Миллер
- Изначальная цель – воссоздать модель хранения слов в человеческой памяти
- Основное – отношение синонимии и понятие синсета (*synset* – synonym set)
- Две единицы синонимичны, если замена одной на другую в определенном контексте не меняет истинности высказывания:
 - square* – *second power* (в алгебре)
 - Earth* – *earth* – *world* (про нашу планету)

WORDNET: СИНСЕТ



- ❑ Набор синонимов одной части речи – *синсет* – представляет понятие (концепт) английского языка
- ❑ Большинство синсетов снабжены толкованиями
- ❑ Каждое слово относится к одному или нескольким синсетам
- ❑ Различные значения слов входят в разные синсеты
- ❑ Разные части речи (существительные, глаголы, прилагательные, наречия) сгруппированы отдельно
- ❑ Синсеты существительных и глаголов образуют иерархическую сеть.
Сеть организована на основе разных отношений:
 - ❖ для существительных – родовидового
 - ❖ для глаголов – тропонимии

СИНСЕТЫ СЛОВА *TABLE*



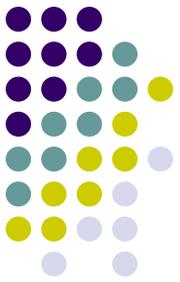
Noun

- **S: (n) table, tabular array** (a set of data arranged in rows and columns) "*see table 1*"
- **S: (n) table** (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) "*it was a sturdy table*"
- **S: (n) table** (a piece of furniture with tableware for a meal laid out on it) "*I reserved a table at my favorite restaurant*"
- **S: (n) mesa, table** (flat tableland with steep edges) "*the tribe was relatively safe on the mesa but they had to descend into the valley for water*"
- **S: (n) table** (a company of people assembled at a table for a meal or game) "*he entertained the whole table with his witty remarks*"
- **S: (n) board, table** (food or meals in general) "*she sets a fine table*"; "*room and board*"

Verb

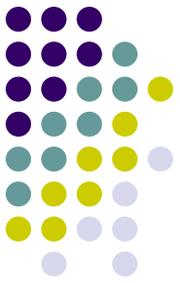
- **S: (v) postpone, prorogue, hold over, put over, table, shelve, set back, defer, remit, put off** (hold back to a later time) "*let's postpone the exam*"
- **S: (v) table, tabularize, tabularise, tabulate** (arrange or enter in tabular form)

WORDNET: ОТНОШЕНИЯ



- ◆ Для существительных:
 - основные: родовидовое и *меронимия*
 - другие: антонимия, экземпляр, иметь общий гипероним
- ◆ Для прилагательных и наречий:
 - антонимия, синонимия
- ◆ Для глаголов:
 - основное: *тропонимия*
 - другие: следование, причина-результат, антонимия
- ◆ Есть отношения между разными частями речи
- С 1985 г. набор отношений существенно расширился

СУЩЕСТВИТЕЛЬНЫЕ: РОДОВИДОВОЕ ОТНОШЕНИЕ



- Видовой синсет называется *гипонимом*,
родовой – *гиперонимом*
- Синсет X является гипонимом синсета Y,
если истинно высказывание типа
An X is a (kind of) Y
- Самое многочисленное отношение

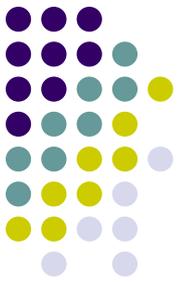
Верхний уровень иерархии существительных:

entity: hyponym physical entity

abstraction, abstraction entity

thing

ГИПОНИМЫ И ГИПЕРОНИМЫ СИНСЕТА *TABLE* (ТАБЛИЦА)



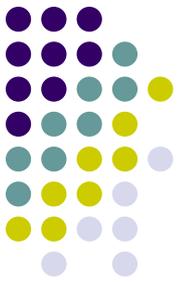
Гипонимы

- S: (n) table, tabular array
 - direct hyponym / full hyponym
 - S: (n) correlation table
 - S: (n) contents, table of contents
 - S: (n) actuarial table, statistical table
 - S: (n) calendar
 - S: (n) file allocation table
 - S: (n) periodic table

Гиперонимы

- direct hypernym / inherited hypernym / sister term
 - S: (n) array
 - S: (n) arrangement
 - S: (n) group, grouping
 - S: (n) abstraction, abstract entity
 - S: (n) entity

СУЩЕСТВИТЕЛЬНЫЕ: ОТНОШЕНИЕ ЧАСТЬ-ЦЕЛОЕ



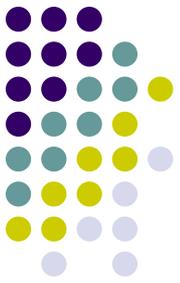
- Синсет-часть называется *меронимом*, синсет-целое – *ХОЛОНИМОМ*
- Внутри меронимии выделяют отношения:
 - ✓ быть частью (**part**)
bird: part meronym feather, wing
point: part holonym icepeak, knife
 - ✓ быть элементом (**member**)
table: member meronym row, column
human: member holonym genus Homo
 - ✓ быть сделанным из (**substance**)
wood: substance meronym lignin
glass: substance holonym glasswork

СУЩЕСТВИТЕЛЬНЫЕ: ДРУГИЕ ОТНОШЕНИЯ



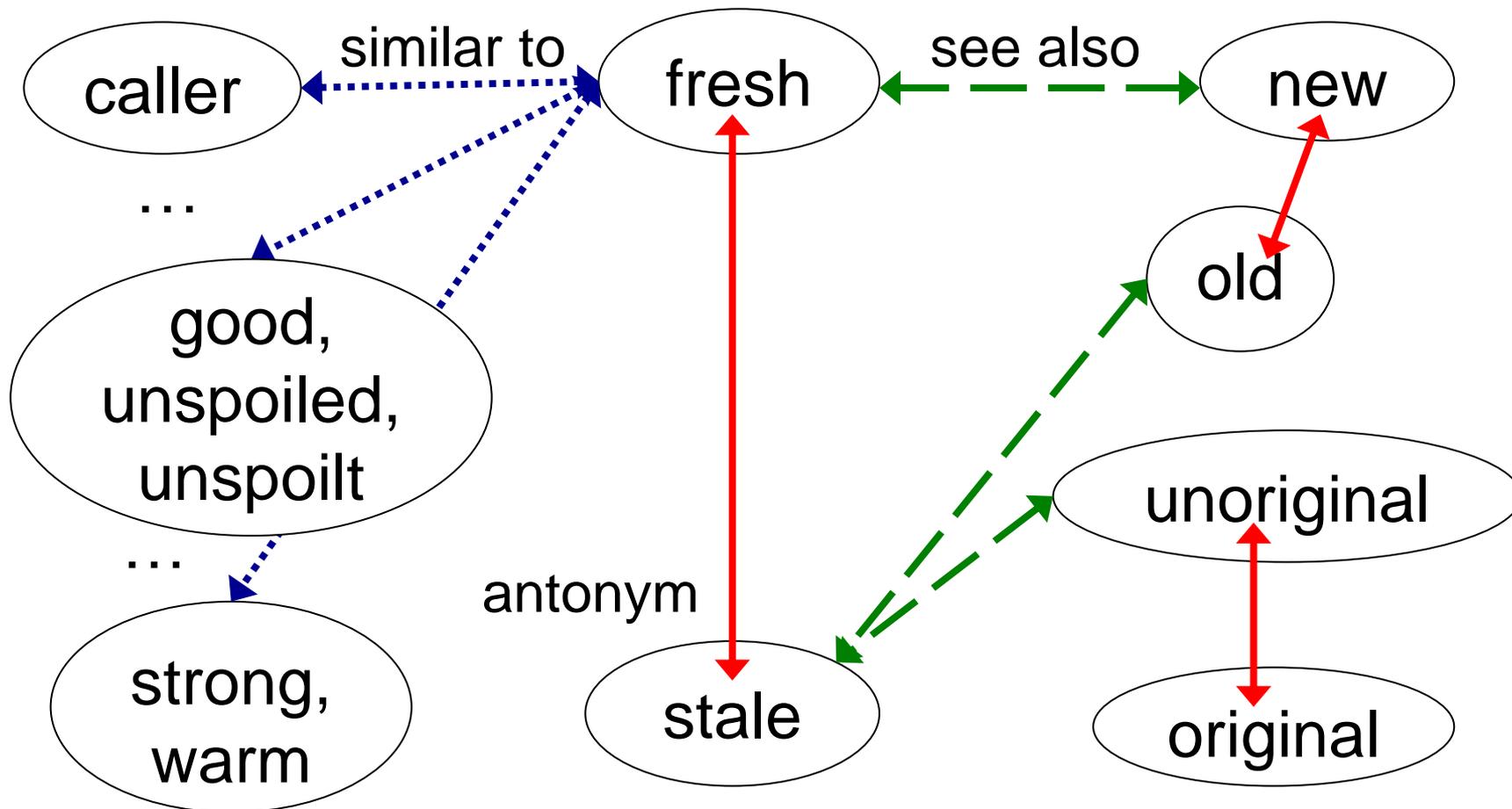
- Антонимия (**antonym**): между конкретными словами, не синсетами
*dark: **antonym** light*
*woman: **antonym** man*
- Экземпляр:
 - ❖ иметь экземпляр (**has instance**)
*table: **has instance** King Arthur's Round Table*
 - ❖ быть экземпляром (**instance**)
*Russia: **instance** country*
- Иметь общий гипероним (**sister term**)
*outdoor sport: **direct hyponym** hunting, fishing*
*fishing: **sister term** hunting*

КАЧЕСТВЕННЫЕ ПРИЛАГАТЕЛЬНЫЕ



- ❑ Качественные прилагательные (*fresh, slow*) выражают величины атрибутов, которые обычно биполярны
- ❑ Основное отношение – антонимия
- ❑ Для отсылки к близким по смыслу прилагательным, которые имеют свои антонимы, введено отношение между синсетами **see also**
- ❑ Поскольку не у всех прилагательных есть антонимы, они присоединяются к близкими по смыслу прилагательным (у которых есть антонимы) с помощью отношения **similar to**

ОТНОШЕНИЯ МЕЖДУ *FRESH И STALE*



ПАРА ПРИЛАГАТЕЛЬНЫХ

FRESH – *STALE*



fresh – свежий,
stale – несвежий

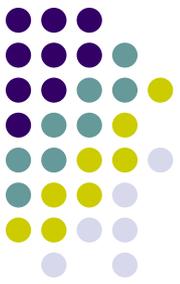
- **S:** (adj) **fresh**
 - see also
 - **S:** (adj) new
 - similar to
 - **S:** (adj) caller
 - **S:** (adj) crisp
 - **S:** (adj) fresh-cut
 - **S:** (adj) good, unspoiled, unspoil
 - **S:** (adj) undecomposed, incorrupt
 - **S:** (adj) hot
 - **S:** (adj) new-made
 - **S:** (adj) strong, warm
 - attribute
 - antonym
 - **W:** (adj) stale [Opposed to: fresh]

- **S:** (adj) **stale**
 - see also
 - **S:** (adj) old
 - **S:** (adj) unoriginal
 - similar to
 - **S:** (adj) addled
 - **S:** (adj) bad, spoiled, spoilt
 - **S:** (adj) cold
 - **S:** (adj) day-old
 - **S:** (adj) hard
 - **S:** (adj) flyblown, maggoty
 - **S:** (adj) moldy, mouldy, musty
 - **S:** (adj) rancid
 - **S:** (adj) rotten
 - **S:** (adj) corrupt, tainted
 - **S:** (adj) putrid
 - **S:** (adj) putrescent
 - attribute
 - antonym
 - **W:** (adj) fresh [Opposed to: stale]

S – отсылка к синсету (семантическое отношение)

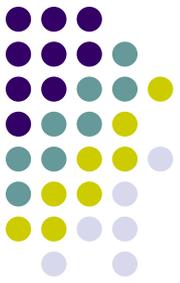
W – отсылка к слову (лексическое отношение)

ОТНОСИТЕЛЬНЫЕ ПРИЛАГАТЕЛЬНЫЕ



- Значение относительного прилагательного (*metal, home*) поясняется через отсылку к существительному (**pertainym**)
- Для некоторых прилагательных одно из значений представлено как качественное прилагательное, а второе – как относительное
 - **S: (adj) agricultural**
 - **pertainym**
 - **W: (n) agriculture** [Related to: **agricultural**]
 - **S: (adj) agrarian, agricultural, farming**
 - **similar to**
 - **S: (adj) rural**
 - **antonym**
 - **W: (adj) urban** [Indirect via **rural**]

ГЛАГОЛЫ: ОТНОШЕНИЯ



- Следование (**entailment**):
из someone V1 логически следует someone V2
walk: entailment step
- Причина (**cause**): *V1 – причина, V2 – результат*
give: cause have, have got, hold
- Антонимия:
shout: antonym whisper
- Тропонимия (**troponym**) – задает иерархию:
to V1 это to V2 в особой манере
sing: troponym hymn, hymn: hypernym sing
- На верхнем уровне иерархии глаголы разделены на семантические поля: глаголы движения, восприятия, изменения, эмоций и пр.

СИНСЕТ ГЛАГОЛА *SIT* (ПРИСЕСТЬ)



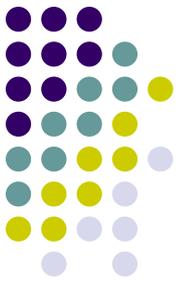
- **S: (v) sit, sit down** (be seated)
 - *direct troponym / full troponym*
 - **S: (v) lounge** (sit or recline comfortably)
 - **S: (v) sprawl** (sit or lie with one's limbs spread out)
 - **S: (v) spread-eagle** (stand with arms and legs spread out)
 - **S: (v) perch, roost, rest** (sit, as on a branch)
 - **S: (v) squat, crouch, scrunch, scrunch up, hunker, hunker down** (sit on one's heels)
 - *verb group*
 - **S: (v) sit down, sit** (take a seat)
 - **S: (v) seat, sit, sit down** (show to a seat; assign a seat for)
 - *entailment*
 - **S: (v) sit down, sit** (take a seat)
 - *phrasal verb*
 - **W: (v) sit out** [Related to: **sit**] (endure to the end)
 - **W: (v) sit down** [Related to: **sit**] (take a seat)
 - *antonym*
 - **W: (v) stand** [Opposed to: **sit**] (be standing; be upright)
 - **W: (v) lie** [Opposed to: **sit**] (be lying, be prostrate; be in a horizontal position)
 - *derivationally related form*
 - *sentence frame*
 - Something ----s
 - Somebody ----s
 - Something is ----ing PP
 - Somebody ----s PP
 - [Applies to **sit**] The children sit in the rocking chair
 - [Applies to **sit**] There sit some children in the rocking chair

ОТНОШЕНИЯ МЕЖДУ РАЗНЫМИ ЧАСТЯМИ РЕЧИ



- Дериваты (**derivationally related form**) – производные слова разных частей речи
table: derivationally related form
tabulate (глагол)
tabular (прилагательное)
- Атрибуты (**attribute**)
fresh: attribute freshness
pleasantness: attribute
pleasant
unpleasant

WORDNET: ТЕКУЩЕЕ СОСТОЯНИЕ



- С 1995 года в Интернет свободном доступе
- Версия 3.1 (2012 г.): 155 287 слов, 117 659 синсетов
- Существует множество аналогичных ресурсов:
 - ❖ точно воспроизводящих структуру и состав оригинального WordNet
 - ❖ учитывающих специфику языка, критику и проблемы WordNet
- Для русского языка были попытки создать аналог WordNet обоими способами
- Проект EuroWordNet, 1996-1999 гг.: учет специфики языка (голландский, итальянский, испанский, немецкий, французский, чешский, эстонский), установление связей между словами разных языков

ЗАКЛЮЧЕНИЕ



- ◆ Существуют разные виды лингвистических онтологий
- ◆ Различие: назначение и типы/количество связей между понятиями
- ◆ Общее: тесная связь с языковыми/текстовыми единицами – структура понятий «подгоняется» под них
- ◆ Лингвистическая онтология – ресурс, в котором адекватное представление лексической системы важнее адекватного представления понятийной



СПАСИБО ЗА ВНИМАНИЕ!

ДОМАШНЕЕ ЗАДАНИЕ № 8



Изучить два лексических ресурса –

выбрав по одному из указанных ниже двух групп (I,II)

- 1) кратко охарактеризовать особенности каждого лексического ресурса (не более 1 стр. для каждого) и
- 2) для ресурса из первой группы указать вид онтологии
- 3) для ресурса второй группы – ответить на вопрос, является ли он лингвистической онтологией, и если да, то какого вида? Ответ обосновать.

I. Идеографический словарь О.С. Баранова

Викисловарь

АвиаОнтология

<http://ipm.lviv.ua/library/0/004/004.8/aviaontology.pdf>

II. УДК

Фолксономия (любая), напр.: <http://www.last.fm/music>

Контрольный срок сдачи задания – 5 мая

УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ (УДК)



Из-за длинных названий рубрикам могут присваиваться классификационные коды

Пример. **УДК** – систематизация произведений науки, литературы и искусства, периодической печати, различных видов документов и картотек

8 Язык. Языкознание. Лингвистика. Литература

81 Языкознание и языки. Лингвистика

811 Языки

811.1/.8 Естественные языки

811.11 Германские языки

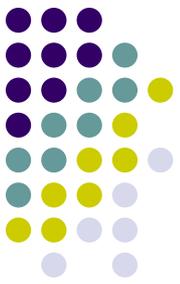
811.111 Английский язык

811.112 Другие западногерманские языки

811.112.2 Немецкий язык

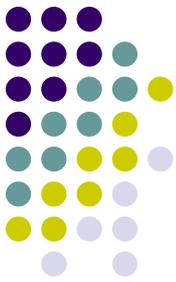
ИДЕОГРАФИЧЕСКИЙ СЛОВАРЬ

О. С. БАРАНОВА



- Разрабатывается с 1990 г., электронное издание 2012 г. – <http://www.thesaurus1.narod.ru>
- Назначение – дать смысловую картину окружения конкретного понятия и картину всей лексики в целом
- Слова собраны в гнезда по смыслу и группируются вокруг некоторого понятия. Гнезда группируются в подразделы и т.д.
- В гнезде – определение основного понятия, слова разных частей речи с синонимами
- Могут указываться ссылки на родовидовые и противоположные понятия, примеры употребления
- Размер: 100 000 лексем, объединенных в 7800 гнезд

СЛОВАРЬ БАРАНОВА. КЛАССИФИКАЦИЯ И ПРИМЕР ГНЕЗДА



| | | | | | |
|-----------------------|-----------------|---------------------|--------------|-------------------------|----------------|
| ПОРЯДОК | ПРИРОДА | ЧЕЛОВЕК | ДЕЯТЕЛЬНОСТЬ | ОБЩЕСТВО | КУЛЬТУРА |
| ДЕЙСТВИЯ ЧЕЛОВЕКА | | | ДЕЯТЕЛЬНОСТЬ | | |
| ДЕЯТЕЛЬНОСТЬ | ДОСТИЖЕНИЕ ЦЕЛИ | | ТРУДНОСТИ | УСПЕШНОСТЬ ДЕЯТЕЛЬНОСТИ | |
| ДЕЯТЕЛЬНОСТЬ ЧЕЛОВЕКА | | СТАДИИ ДЕЯТЕЛЬНОСТИ | | ТРУД | СОТРУДНИЧЕСТВО |

193 ДЕЯТЕЛЬНОСТЬ ЧЕЛОВЕКА

193.1 ДЕЯТЕЛЬНОСТЬ ЧЕЛОВЕКА

193.1.1 деятельность человека

▲ деятельность

↑ преднамеренный

деятельность человека - *взаимодействие человека с окружающей средой; действия, посвященные какой-л. цели; процесс осуществления крупной цели; последовательность действий; решение стоящих перед субъектом задач; включает в себя предметную составляющую (заниматься какой-л. деятельностью. посвятить себя какой-л. деятельности).*

развернуть (# деятельность. # борьбу).

развить (# бурную деятельность).

работа. вести [проводить] работу.

политика. вести политику.

кампания (*избирательная #*).

давать [дать] себе труд.

вдаваться (*не будем # в рассуждения*).

ковать (*# победу*).

ставить (*# спектакль*).

↓ интересоваться

▼ искусство промышленность предпринимательство

действовать интенсивно

см. действовать медленно

свободное время

характер *чего*

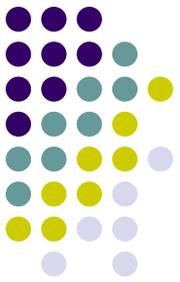
ВИКИСЛОВАРЬ



Структура словарной статьи:

- ❑ Морфологические и синтаксические свойства
- ❑ Произношение
- ❑ Семантические свойства
 - Значение
 - Синонимы, антонимы, гиперонимы, гипонимы
- ❑ Родственные слова
- ❑ Этимология
- ❑ Фразеологизмы и устойчивые сочетания
- ❑ Перевод
- ❑ Библиография
- ✓ Категории

АВИАОНТОЛОГИЯ



- Построен на базе RuTез
- Описывает процессы функционирования бортовой аппаратуры, а также экипажа в различных полетных режимах
- Создавалась для использования в системе АЛОТ – автоматического анализа технических текстов типа «Логика работы...»
- Включает около 1600 понятий и 4700 терминов
- Содержит термины и общезначимые слова (*необходимость, вероятность, условие*)
- Использовались отношения синонимии, ВЫШЕ-НИЖЕ, ЧАСТЬ-ЦЕЛОЕ и ассоциации

АВИАОНТОЛОГИЯ: ПРИМЕР ОПИСАНИЯ ТЕРМИНА



| Пуск ракеты | |
|--------------------|--|
| СИН | запуск ракеты, запустить ракету, отстрел ракеты, применение ракет, применение ракетного вооружения, пуск ракеты, пустить ракету, ракетный удар, ракетная атака |
| ВЫШЕ | применение оружия |
| НИЖЕ | эффективный пуск ракеты |
| ЧАСТЬ | максимально допустимая перегрузка при пуске, дальность пуска ракеты, зона возможного пуска |
| АСЦ1 | ракета |
| АСЦ2 | выход в зону возможного пуска, ошибка пуска, команда «пуск разрешен», управление ракетой |

ФОЛКСОНОМИЯ



Фолксономия – отношения устанавливаются «народом» (пользователями системы)

На сайте <http://www.last.fm/music>

- ❑ пользователи добавляют произвольные теги к информации об исполнителях и композициях
- ❑ из наиболее частых тегов создается навигационная строка

Например, для группы The Beatles:

classic rock • rock • 60s • the beatles • british • pop

- ❑ при нажатии на тег получаем список исполнителей/композиций с таким же тегом