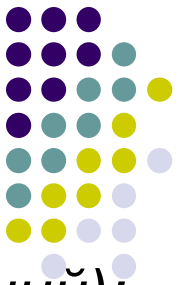




# **ЛЕКСИЧЕСКАЯ СЕМАНТИКА: ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ**

Большакова Елена Игоревна

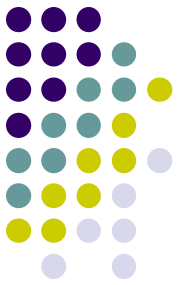
# СОДЕРЖАНИЕ



*Лексическая семантика связана со смыслом лексических единиц языка (слов, словосочетаний):  
и отображена в языке / тексте*

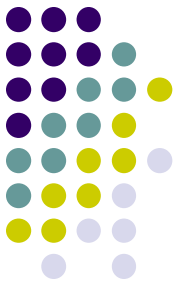
1. Семантические классы слов
2. Лексическая неоднозначность
3. Лексические (*парадигматические*) отношения:  
синонимия, род-вид, часть-целое, антонимия
4. Лексические ресурсы: тезаурус *WordNet*
  - общая характеристика, понятие *сунсета*
  - лексические отношения и иерархии
  - аналоги *WordNet* для русского языка
5. Лексические ресурсы: тезаурус *PyТез*, сеть *BabelNet*
6. Заключение, Упражнения

# СЕМАНТИЧЕСКИЕ ХАРАКТЕРИСТИКИ



- Семантические характеристики/категории слов позволяют учесть **семантическую сочетаемость**:
  - *Поет* – кто? Категория/характеристика **ОДУШ**
- Слову может быть приписано несколько семантических признаков, они представлены в семантич. словарях
- Каждый признак образует семантический класс слов
- Основные классы: **Объекты**(оду/неодуш), **Отношения Действия/Процессы**, **Свойства** (качества, количества),
- Нет универсальной системы семантич. характеристик
- Системы семантических классов/характеристик в разных подходах/системах различаются
- Примеры семантических характеристик: **ДЕЙСТВИЕ, СОСТОЯНИЕ, СВОЙСТВО, ВРЕМЯ, ЛИЦО, ПРЕДМЕТ**

# СЕМАНТИЧЕСКИЕ КЛАССЫ: ДИАЛИНГ (*aot.ru*)



Н.Н. Леонтьева, семантический словарь РУСЛАН:  
большой набор семантических характеристик

- АБСТР: *модель, план, тенденция, обстоятельство*
- ВЕЛИЧ: *высокий, мощный*
- ВЕЩВО: *вода, уголь*
- УСТР (устройство): *компьютер, принтер*
- ВМЕСТЛ: *мешок, сейф, гараж*
- ВРЕД: *катастрофа, война, перегрузка*
- ГЕОГР (любой географический объект): *Минск*
- ОРГ (организация): *Дума, спецшкола, ООН*
- ДВИЖ: *поездка, бежать*
- ДОЛЖ (профессия, социальный статус): *врач*
- ВОСПР – *слышать, видеть*

# СЕМАНТИЧЕСКИЕ ХАРАКТЕРИСТИКИ: ДИАЛИНГ (продолжение)



- ИНФ (информация): *утверждение, новость*
- НОСИНФ (носитель информации): *книга, газета*
- ОДЕЯТ (область деятельности): *художник*
- ПРЕДМЕТ (объект по размерам меньше среднего человека)
- ПРОТЯЖ (протяженные географические объекты): *река*
- ВЛАСТЬ: *парламент*
- СОБИР: *библиотека*
- ФИН (все, что связано с финансами): *деньги, банк, долг*
- ЭМОЦ: *страх, радость*
- АРТ (артефакты): *дом, автомобиль*
- ЯВЛЕНИЯ: *смерч, терроризм*
- ИЗМ (действия изменения): *переключить*

*Какое свойство языка делать классы пересекающимися?*

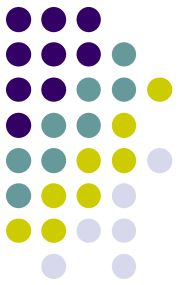
# ЛЕКСИЧЕСКАЯ МНОГОЗНАЧНОСТЬ



Многозначность (неоднозначность) –  
особенность ЕЯ на всех его уровнях

- **Омонимия**
    - Лексическая: *коса, лук*
    - Морфологическая: *зала*
    - Синтаксическая: *отец любит дочь*
  - Лексическая многозначность словосочетаний
    - *катать шарики* (делать или перемещать)
    - *вытравить рисунок* (сделать или уничтожить)
    - *проехать остановку* (проехать расстояние или мимо)
  - **Полисемия** – важная особенность ЕЯ, одно из средств концептуализации/понимания нового опыта
- Чай*: 1) кустарник 2) высушенные листья этого растения  
3) напиток, настоянный на этих листьях  
4) высушенные травы для заваривания

# ЛЕКСИЧЕСКАЯ МНОГОЗНАЧНОСТЬ: ПОЛИСЕМИЯ И ОМОНИМИЯ



- **Омонимия** – случайное внешнее совпадение двух разных слов, нет общих элементов смысла:
    - *лук* (оружие) – *лук* (растение)
  - **Полисемия** – имеется общий элемент смысла:
    - *клапан мотора* – *клапан кармана* – *клапан фагота* – *сердечный клапан* –  
часть предмета, закрывающая отверстие в нем
    - *класс* (разряд) – *класс* (общественная группа) –  
*класс* (группа учащихся по году обучения) – *другие?*
- Но не всегда есть четкая граница между значениями
- ❖ Причины полисемии (экономичность языка):
    - метафоричность (сходство явлений) : *клапан*
    - метонимия (перенос на смежное явление):  
*город* (населенный пункт и городское население)

# ПОЛИСЕМИЯ: СЛОВАРИ и КЛ



- Регулярная многозначность
  - сосуд – количество: *ведро, стакан*
  - животное – мех: *белка, норка, лиса*
  - и многое др.
- Словари: дискретная организация смысла слов, но между значениями часто нет четкой границы (диффузность)
- Разные словари – разное количество значений многозначных слов, например: *Идти*
  - 30 значений в Большом толковом словаре
  - 25 значений в словаре Ожегова
- А в словаре для автоматической обработки текстов?
  - Лучше меньше значений, но значения могут сильно различаться
  - Выбор зависит от приложения: машинный перевод – смыслы не разделяются, если в целевом языке также есть подобная многозначность



# КРИТЕРИИ ДЛЯ РАЗДЕЛЕНИЯ ЗНАЧЕНИЙ ПРИ ПОЛИСЕМИИ



- Различия в морфологии, например, *appendix*:  
мн.ч. *appendices* (приложения) или *appendixes* (части тела)
- Синтаксические различия, например, счетное – несчетное:  
*a forest of oaks / a table made of oak*
- Семантические ограничения модели управления:  
*crawl* (ползать, пресмыкаться): insects, snakes, people, babies
- Лексическая сочетаемость (устойчивые словосочетания):  
*fully absorbed* – *полностью поглощенный* (into the bloodstream) и *completely/totally absorbed* (in her book).
- Лексические отношения: синонимы, антонимы, гипонимы:  
*земля* – *земельный, земляной*
- Предметная область (терминологичность):  
*bonds* : химия – связи и финансы – облигации
- Наличие эквивалентов на других языках

# ЗАДАЧА РАЗРЕШЕНИЯ ЛЕКСИЧЕСКОЙ МНОГОЗНАЧНОСТИ



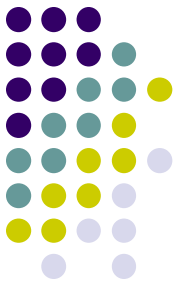
- При автоматическом анализе текста требуется выбор правильного значения из нескольких возможных (*lexical disambiguation*)
- Два основных пути решения, оба используют **контекст** употребления слова
  - ❖ методы, основанные на таких лингвистических ресурсах, как **словари, тезаурусы, онтологии**
  - ❖ методы, основанные на машинном обучении и размеченных корпусах текстов
- Качество разрешения многозначности оценивается на основе «золотого стандарта», т.е. экспертной разметки

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ



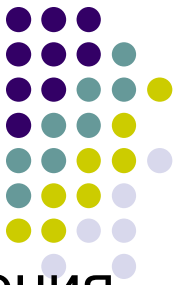
- Лексические отношения – семантические связи лексических единиц (слов, словосочетаний) в системе языка:
  - Синонимия
  - Антонимия
  - Деривационные отношения (*дериваты*)
  - Родо-видовое отношение (Род – Вид)
  - Отношение Часть – Целое
- Синонимы – слова, имеющие одно и то же лексическое значение, но различающиеся его оттенками:
  - стилистическими: *жена – супруга*
  - эмоционально-экспрессивными: *лошадь – кляча*
  - профессиональными: *воспаление легких – пневмония*

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ: СИНОНИМИЯ



- Критерий синонимии – взаимозаменяемость в одном и том же контексте без (заметного) различия по смыслу, но с возможными стилистич. Различиями
- Взаимозаменяемость:
  - сильная, во всех контекстах: *собака/пес* (таких синонимов очень мало)
  - слабая, в некоторых контекстах – **квазисинонимы**
- *Синонимы / Квазисинонимы ? Или другое?*
  - *самолет, авиалайнер, аэробус, аэроплан*
  - *здание, корпус, строение, постройка, помещение*
  - *палка, трость, посох*
  - *преступник, убийца, злодей, душегуб*
  - *ранец, рюкзак, сумка*

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ: СВЯЗЬ РОД-ВИД



- Общелогическая связь, базовая для челов. мышления
- Термины: *гипоним – гипероним*  
*медведь – животное, саднить – болеть*
- В области ИИ обозначается как *IS-A*
- Свойства отношения Род-Вид:
  - транзитивность отношения  
*грач – птица – животное – живое существо*
  - наследование свойств родовых понятий на видовые
  - заменимы в части контекстов
- Критерии выявления: лингвистические тесты
  - *X – это Y*
  - *X, Z и другие Y*
  - *к числу Y относятся X*

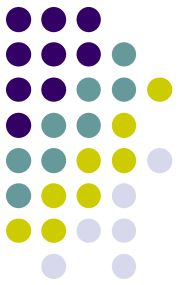
Могут быть записаны как шаблоны для автом. анализа

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ: СВЯЗЬ ЧАСТЬ-ЦЕЛОЕ



- Термины: *мероним* (часть) и *холоним* (целое)
- Критерии выявления – лингвистические тесты:
  - *У имеет X* (или *Xы*)
  - *X – это часть Y*
  - *Части Y включают X, Z и ...*
  - *X и другие части Y*(X и Y – родовые, обобщенные понятия)
- Проблемы формального описания отношения:
  - Наследование свойств?
  - «Части» могут принадлежать многим «целым» :  
*ножка* – стул, стол, кресло...
  - Факультативность/обязательность «части» ?  
*этаж* обязательно находится в *здании*
- Важный вопрос: одно или несколько разных отношений?

# ВИДЫ ОТНОШЕНИЯ ЧАСТЬ-ЦЕЛОЕ



- Физические объекты (центральный тип) – для них понятие части определяется четко.
- Сущности, длящиеся во времени, тоже могут иметь части: стадии, фазы, этапы: *матч*
- Абстрактные сущности, такие как *опера, симфония, пьеса, балет* тоже могут иметь четкие части.
- Такие сущности, как группы, классы и коллекции состоят в отношении часть-целое со своими элементами: *племя, команда, куча, лес*
- Ингредиенты: *заварной крем и молоко*
- Можно считать частью – материал предмета:  
*бокал – стекло*

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ: ДЕРИВАТЫ



- **Синтаксические дериваты** — производные слова другой части речи, полностью сохраняющие лексическое значение исходного слова
  - *двигаться — движение*
  - *рожать — роды, белый — белизна*
  - *строить — строительный*
  - *республика — республиканский*
  - *карты — карточный; быстрый — быстро*
  - *строительный — строить*
  - *любовь — любить*
- **Семантические дериваты** — только смысловая связь:  
*свадьба, невеста, жених*

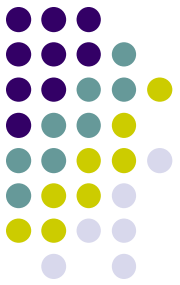


# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ: АНТОНИМЫ



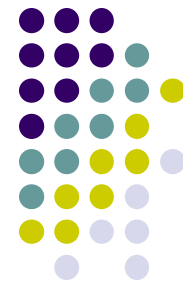
- Антонимы имеют противоположные или обратные, но не противоречащие значения (в структуре их значений может быть одна противоположная сема)
- Основные типы:
  - $P$  – не  $P$ : *влажный – сухой*
  - больше – меньше: *узкий – широкий*
  - корреляты: *вектор – скаляр*
  - противоположные действия:  
*спрыгнуть – запрыгнуть*
  - действие – уничтожение результата действия:  
*приклеить – отклеить*

# ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ И ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ



- Лексические отношения = Семантические отношения
- Лексические отношения учитываются в задачах КЛ, а также в смежных областях ИИ и информационного поиска: построение тезаурусов и онтологий
- **Тезаурусы и онтологии** – лингвистические ресурсы, в которых описываются понятия (значения слов) и их взаимосвязи (семантические отношения)
- Применение лексических отношений в задачах КЛ:
  - анализ и синтез связного текста
  - распознавание перифразирования:  
*цены контролируются правительством –  
правительство осуществляет контроль цен*
  - Информ. поиск: расширение запроса (гиперонимом)

# ПРИМЕНЕНИЕ ЛЕКСИЧЕСКИХ ОТНОШЕНИЙ : ПРИМЕРЫ ЗАДАЧ КЛ

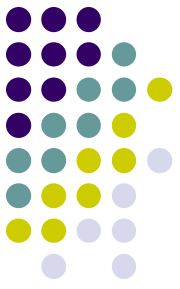


- Упрощение текстов (*Lexical text simplification*): замена сложных слова на их синонимы (*ринит* – *насморк*)
- Выводы по тексту (*Textual entailment*)
  - Мальчик играет => Ребенок играет (гипероним),  
≠ Девочка играет (*Contradiction, ко-гипоним*)
  - Запрещены домашние животные => Собаки запрещены (гипоним)
  - Ресторан в Японии => ресторан в Азии (холоним)
- Диалог в вопросно-ответных системах  
*Когда Дональд Трамп посещал Францию? => Трамп посещал Париж в сентябре* (географическая часть)
- Информационный поиск: при запросе *дешевый бар на востоке* не должна выдаваться информация о дорогом ресторане на западе

# ЛЕКСИЧЕСКИЕ РЕСУРСЫ: ТЕЗАУРУСЫ И ОНТОЛОГИИ



- ❖ **Тезаурус** – семантический словарь ( слова + указание их лексических связей : род-вид, часть-целое и др.)
- ❖ **Онтология** – формальное описание понятий/сущностей некоторой предметной области и их семантических связей, а также ограничений (аксиом) их использования
- **WordNet** – тезаурус (лингвистическая онтология) : систематизация общей английской лексики; связи: синонимия, род-вид, антонимия, часть-целое и др.
- **EuroNet** – аналогичные (*WordNet*) лексические ресурсы для других европейских языков
- **PyTез** – информационно-поисковый тезаурус для РЯ: 52 тыс. понятий из общественно-политической области; связи: род-вид (выше-ниже), синонимия, ассоциация



# WordNet: ОБЩАЯ ХАРАКТЕРИСТИКА

Истоки – 1985 г., Принстонский университет,  
психолог Дж. Миллер (автор «магического числа 7»)

- Цель: реляционное описание лексики английского языка, как некоторой модели хранения слов в человеч. памяти
- Ключевые решения:
  - Отдельное рассмотрение слов разных частей речи (существительные, глаголы, прилагательные, наречия)
  - Семантические отношения устанавливаются между *синсетами* – понятиями, объединяющими синонимы

Ныне большой лексический ресурс, в свободном доступе с 1995 г.: <http://wordnetweb.princeton.edu/perl/webwn>

- Версия 1.6: 95 тысяч понятий, около 130 тысяч слов
- Версия 3.0 (≈ 2011 г.): 117 тыс. понятий, ≈ 155 тыс. лексем (слов и словосочетаний), 200 тыс. связей

# WordNet: СИНСЕТЫ И СВЯЗИ

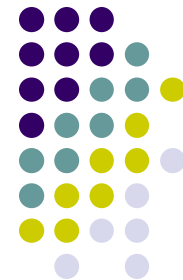


## *Synset* (*synonym set*)

- Синсет – единица словаря, объединяющая слова-синонимы одной части речи с одинаковым лексическим значением (толкованием)
- Синсет соответствует некоторому понятию (концепту) английского языка
- Большинство синсетов снабжены толкованиями
- Каждое слово относится к одному/нескольким понятиям, различные значения слова входят в разные синсеты
- Семантические (лексические) отношения (связи) устанавливаются между синсетами (концептами)
- На основе определенного лексического отношения строится иерархия синсетов для выбранной части речи (разные иерархии для существительных и глаголов)

# WordNet:

## ОПИСАНИЕ ЗНАЧЕНИЙ СЛОВ



Пример:

The noun *forest* has 2 senses:

1. forest, wood, woods

-- the trees and other plants  
in a large densely wooded area

2. forest, woodland, timberland, timber

-- land that is covered with trees and shrubs

The verb *forest* has 1 sense:

1. afforest, forest

-- establish a forest on previously unforested land  
("afforest the mountains")

# WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

## Noun

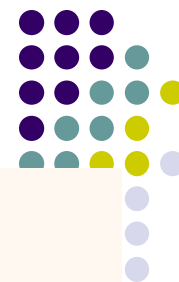
- [S:](#) [\(n\)](#) **forest**, [wood](#), [woods](#) (the trees and other plants in a large densely wooded area)
- [S:](#) [\(n\)](#) **forest**, [woodland](#), [timberland](#), [timber](#) (land that is covered with trees and shrubs)

## Verb

- [S:](#) [\(v\)](#) [afforest](#), **forest** (establish a forest on previously unforested land) "*afforest the mountains*"



# 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:

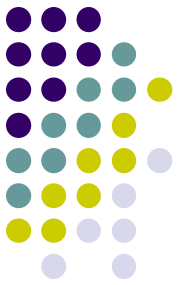
## ЛЕКСИЧЕСКИЕ ОТНОШЕНИЯ



Связи значений слов (синсетов), зависят отчасти речи

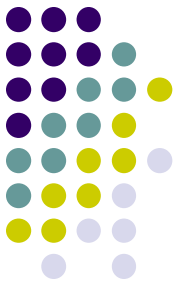
- ◆ Для существительных:
  - Род-Вид (*гипероним – гипоним*) – самое многочисленное, основа лексической иерархии  
*птица* – гипероним для гипонима *грач*
  - Часть-Целое (*мероним – холоним*)
  - другие: антонимия, экземпляр, ко-гипонимия
- ◆ Для прилагательных:
  - антонимия, синонимия
- ◆ Для глаголов: *лексическое следование (lexical entailment)*
  - основное: *тропонимия*
  - другие: причина-результат, антонимия
- ◆ Есть отношения между разными частями речи

# **WordNet: СУЩЕСТВИТЕЛЬНЫЕ ОПИСАНИЕ ГИПЕРОНИМОВ**



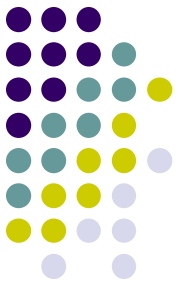
1. forest, wood, woods
  - => vegetation, flora --
  - => collection, aggregation, accumulation, assemblage
  - => group, grouping
  
2. forest, woodland, timberland, timber
  - => land, dry land, earth, ground, solid ground, terra firma
  - => object, physical object
  - => entity, something

# WordNet: ОПИСАНИЕ ГИПОНИМОВ



1. forest, wood, woods --
  - => grove -- (a small growth of trees without underbrush) -- (роща)
  - => old growth, virgin forest -- (forest or woodland having a mature or overmature ecosystem more or less uninfluenced by human activity)
  - => second growth -- (a second growth of trees covering an area where the original stand was destroyed by fire or cutting)
2. forest, woodland, timberland, timber --
  - => Black Forest, Schwarzwald -- (a hilly forest region in southwestern Germany)
  - => greenwood -- (woodlands in full leaf)
  - => jungle -- (an impenetrable equatorial forest)
  - => rain forest, tropical rain forest -- (a tropical forest with heavy annual rainfall)
  - => riparian forest -- (woodlands along the banks of stream or river)
  - => Sherwood Forest -- (an ancient forest in central England)
  - => silva, sylva -- (the forest trees growing in a country or region)...

# ГИПОНИМЫ И ГИПЕРОНИМЫ СИНСЕТА *TABLE* (ТАБЛИЦА)



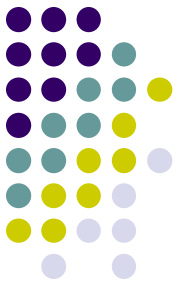
## Гипонимы

- 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

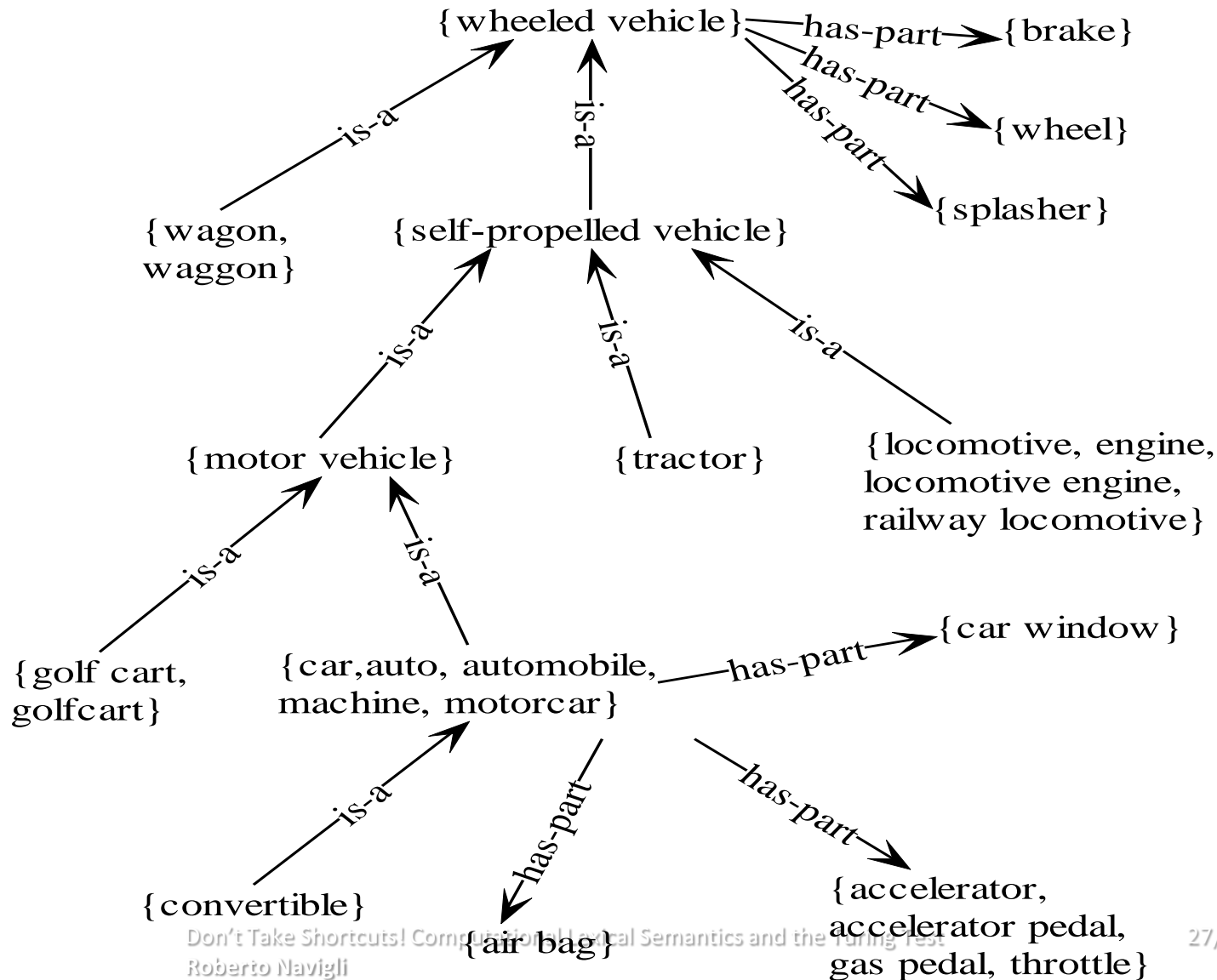
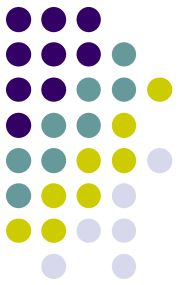
- direct hypernym / inherited hypernym / sister term
  - S: (n) array
    - S: (n) arrangement
      - S: (n) group, grouping
        - S: (n) abstraction, abstract entity
          - S: (n) entity

# ЛЕКСИЧЕСКАЯ ИЕРАРХИЯ СУЩЕСТВИТЕЛЬНЫХ



- В основе— родовидовое отношение, оно образует цепочки вида: *грач — птица — животное — живое сущ-во*
- Концепты на верхнем уровне иерархии:
  - Entity : *hyponym Physical entity , Abstraction , Abstraction Entity , Thing*
  - Psychological Feature
  - Abstraction
  - State
  - Event
  - Human Action
  - Group
  - Possession
  - Phenomenon и др.
- Онтологичность: верхние уровни иерархии близки к описанию общего устройства мира (общей онтологии )

# WordNet: ФРАГМЕНТ ИЕРАРХИИ



Miller et al.  
1990;  
Fellbaum,  
1998

# WordNet: СУЩЕСТВИТЕЛЬНЫЕ ОТНОШЕНИЕ ЧАСТЬ-ЦЕЛОЕ



- *Мероним* – часть, *холоним* – целое
- Внутри *меронимии* выделяют подвиды отношения Часть-Целое:
  - ❖ быть частью (*part\_of*) – типичное физическое  
*bird: part meronym feather, wing*  
*point: part holonym icepeak, knife*
  - ❖ быть элементом (*member\_of*)  
*table: member meronym row, column*  
*tree: member holonym forest*
  - ❖ быть сделанным из (*substance\_of*)  
*wood: substance meronym lignin*  
*glass: substance holonym glasswork*



Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

## Noun

- **S: (n) forest**, [wood](#), [woods](#) (the trees and other plants in a large densely wooded area)
  - [direct hyponym](#) / [full hyponym](#)
  - [member meronym](#)
    - **S: (n) underbrush**, [undergrowth](#), [underwood](#) (the brush (small trees and bushes and ferns etc.) growing beneath taller trees in a wood or forest)
    - **S: (n) tree** (a tall perennial woody plant having a main trunk and branches forming a distinct elevated crown; includes both gymnosperms and angiosperms)
  - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
  - [derivationally related form](#)
- **S: (n) forest**, [woodland](#), [timberland](#), [timber](#) (land that is covered with trees and shrubs)

## Verb

- **S: (v) afforest**, **forest** (establish a forest on previously unforested land) "*afforest the mountains*"

# WordNet: 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, когипоним*)  
*outdoor sport: direct hyponym hunting, fishing*  
*fishing: sister term hunting*

# WordNet: ПРИЛАГАТЕЛЬНЫЕ



- Виды прилагательных:
  - ❖ качественные (*descriptive, qualitative*) – *fresh*
  - ❖ относительные (*relational, relative*) – *wood*
- Описание качественных прилагательных – на основе отношений синонимии и антонимии
  - Основное отношение – антонимия
  - Для отсылки к близким по смыслу прилагательным, которые имеют свои антонимы, введено отношение *see also* между синсетами
  - Поскольку не у всех прилагательных есть антонимы, они присоединяются к близким по смыслу (у которых есть антонимы) с помощью отношения *similar to*

# АНТОНИМИЯ ПРИ ОПИСАНИИ ПРИЛАГАТЕЛЬНЫХ

Качественные прилагательные *Heavy Vs. Light*



Heavy

Dense  
Doughy,  
Soggy  
Heavyweight  
Hefty  
Massive  
Middleweight  
Ponderous  
Welterweight

Light

Airy  
Bantamweight  
Buoyant,  
Floaty  
Chiffon  
Feathery,  
Featherlike  
Low-density  
Lightweight  
Weightless

# ПАРА *FRESH* – *STALE*

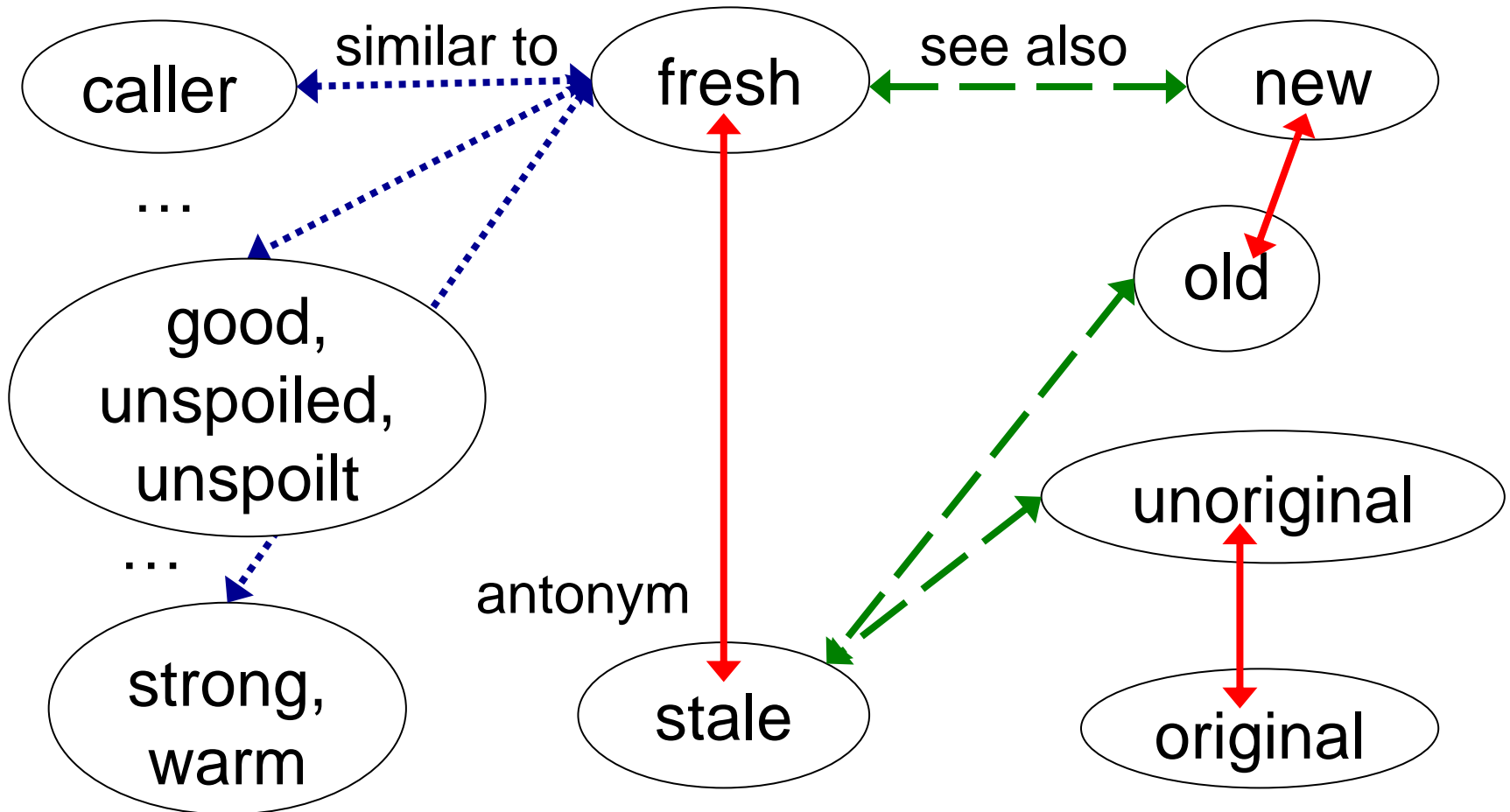
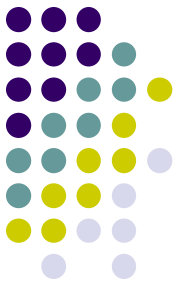
*fresh* – свежий,  
*stale* – несвежий

- S: (adj) **fresh**
  - see also
    - S: (adj) new
  - similar to
    - S: (adj) caller
    - S: (adj) crisp
    - S: (adj) fresh-cut
    - S: (adj) good, unspoiled, unspoilt
    - 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  
W – отсылка к слову

# ОТНОШЕНИЯ МЕЖДУ *FRESH И STALE*



## WordNet:

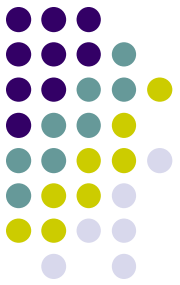
# ОТНОСИТЕЛЬНЫЕ ПРИЛАГАТЕЛЬНЫЕ



- Значение относительного прилагательного (пример: *metal*, *home*) поясняется через отсылку к соответствующему 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]

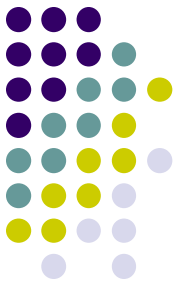
# WordNet: ГЛАГОЛЫ



- Лексические отношения:
  - Лексическое следование: *lexical entailment*
  - Каузальное (причинно-следственное): *causal relation*  
*give: cause have, have got, hold*
  - Антонимия: *shout antonym whisper*  
*Appear – disappear, Lengthen – shorten, Tie – untie,*  
*Emigrate – immigrate, Exclude – include,*  
*Fall / descend – Rise / ascend*
- Разбиение на *семантические поля*:  
Motion, perception, contact, communication,  
competition, change, creation, emotion, possession,  
social behavior, bodily care and functions, cognition,  
weather verbs ...  
*possessions: (give, transfer), (take, receive), (have, hold)*



# ГЛАГОЛЫ: *LEXICAL ENTAILMENT*



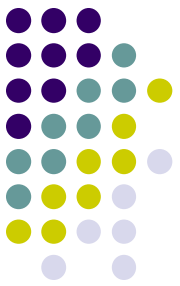
- ❖ Логическое следование для глаголов  $V1$ ,  $V2$ :  
 $V1 \Rightarrow V2$  – Если некто  $V1$ , то значит, некто  $V2$   
*walk: entailment step*, Например:
  - если человек храпит (*snore*), то он спит (*sleep*)
  - если человек купил, значит, он заплатил
  - если он ведет (*drive*) транспорт. средство, то едет (*ride*)
- ❖ Виды лексического следования:
  - Совпадение по времени  $V1$  и  $V2$  (temporal inclusion)
    - Тропонимия, *troponymy*: *march* – *walk*
    - Собственно *entailment*: *walk* – *step*, *snore* – *sleep*
  - Несовпадение по времени  $V1$  и  $V2$ 
    - Каузальность: если  $V1$  является причиной  $V2$ , то  $V2$  следует из  $V1$ : *show* – *see*, *give* – *have*
    - Обратный вывод, пресуппозиция:  
*forget* – *know*, *unwrap* – *wrap*

# ГЛАГОЛЫ: ТРОПОНИМИЯ



- Лексическое отношение *troponymy* :  
делать V1 означает  
делать V2 в некоторой особой манере  
*Fight: fistfight, battle, war, attack, feud, box ...*  
*Walk: stumble, toe, limp, stroll, march*  
*sing: troponym hymn, hymn: hypernym sing*
- Рассматривается как вид отношения лексического следования (при совпадении по времени)
- Тропонимия – основа иерархии глаголов (4 уровня)
  - есть уровни с большим количеством глаголов:  
*communicate – talk – babble / mumble / murmur...*
  - сужение потенциальных аргументов
  - наследование аргументной структуры глагола (модели управления)

# 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

# WordNet: ОТНОШЕНИЯ МЕЖДУ РАЗНЫМИ ЧАСТЯМИ РЕЧИ



- За время развития ресурса с 1985 г. набор отношений существенно расширился, в частности, были добавлены:
- Дериваты (*derivationally related forms*) – производные слова разных частей речи  
*table: derivationally related form*  
*tabulate* (глагол)  
*tabular* (прилагательное)
- Атрибуты (*attribute*)  
*fresh: attribute freshness*  
*pleasantness: attribute pleasant*  
*unpleasant*

# WordNet: ПРОБЛЕМЫ РАЗРАБОТКИ



- Дублирование понятий в большой системе
- Выявление отношений: шаги в цепочке похожи на истину, но вся цепочка проблемна:

*spider web, spider's web* (паутина)

--(a web spun by spiders to trap insect prey )

=> *web, entanglement* -- (запутанность)

=> *trap* -- (ловушка)

=> *device* --

=> *instrumentality, instrumentation* --

=> *artifact, artefact* --(a man-made object)

- Большое количество значений слов (критично для автоматической обработки текстов)
  - 41 значение глагола *cut* , 25 значений *good*
  - Среднее количество значений больше, чем в традиц. словарях, есть плохо различимые значения – почему?

# ЗНАЧЕНИЕ *WordNet*

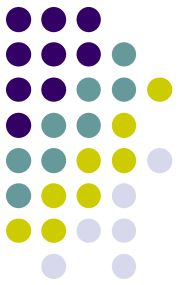


*WordNet* – первый большой лексический ресурс, создававшийся годами и до сих пор развивающийся

Стимулировал создание других ресурсов :

- Аналоги для европейских языков – *EuroWordNet*(1996-99 гг.)  
Первоначально:  
испанский, итальянский, голландский  
Далее: немецкий, французский, чешский, эстонский  
Установление связей слов разных языков
- Попытки создать свои структурные ресурсы на базе *WordNet*: японский, болгарский, румынский, шведский  
русский
- Лингвистический ресурс *SentiWordNet* – оценочная лексика
- *ImageNet*, используемый для обучения систем компьютерного зрения, основан на *WordNet*  
(присоединение картинок к синсетам)

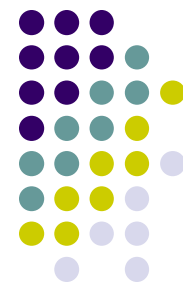
# WordNet: ПРИКЛАДНОЕ ЗНАЧЕНИЕ



Применение *WordNet* в задачах КЛ и АТ:

- Информационный поиск: расширение поискового запроса (синонимы, гипонимы, гиперонимы)
- Алгоритмы на правилах: например, разрешение омонимии – выбор правильного значения слова из нескольких возможных на базе его контекста (путем проверки соответствия слов контекста тезаурусным данным WordNet)
- Тестирование и валидация разных моделей, созданных на базе машинного обучения
- Создание конкретных словарей, ориентированных на прикладную задачу

# WordNet для РУССКОГО ЯЗЫКА



- Несколько попыток создать аналог *WordNet* :
  - Разработка лингвистами по словарям и корпусу, филфак СПбГУ: <http://project.phil.spbu.ru/RussNet/>
  - Автоматич.перевод исходного WordNet: <http://wordnet.ru>  
Проблемы: неправильные синонимы, лишние значения  
*Почему попытка неудачна?*
  - *YARN – Yet Another RussNet*: Краудсорсинг + Wiktionary
- Помимо попыток создания прямого аналога *WordNet* разработаны лингвист. ресурсы с исходно иными целями, но в которых также представлены лексические отношения:
  - Электронная версия тезауруса Баранова  
<http://www.thesaurus1.narod.ru>
  - Информационно-поисковый тезаурус *PyТез*  
<http://www.labinform.ru/pub/ruthes/index.htm>  
некоторая часть трансформирована в формат *WordNet*:  
*RuWordNet* [ruwordnet.ru/ru](http://ruwordnet.ru/ru)





Введите слово



Главная страница

Публикации

Данные

Технологии

Все слова

Утверждённые  
слова

Синсеты

## Yet Another RussNet

**Целью проекта** является создание большого открытого электронного тезауруса русского языка.

Работа над проектом началась в 2012 году.

**Основная идея проекта** — эксперимент по комбинированию традиционных принципов создания ворднетов и вики-подхода к наполнению и редактированию лингвистических ресурсов.

Любой желающий может присоединиться к обсуждению проекта в [рассылке](#) и непосредственной работе по наполнению тезауруса.

Проект осуществляется при участии [УрФУ](#), [ИММ УрО РАН](#), [ВШЭ](#) и выполняется при финансовой поддержке [РГНФ](#), грант №13-04-12020 («Новый открытый электронный тезаурус русского языка»).



[Главная страница](#)

[Публикации](#)

[Данные](#)

[Технологии](#)

[Инструкция](#)

[Новости](#)

[Словник](#)

[Редактор  
связей](#)

# Слово №5130: «закон»

Связанные синсеты 12

## Синсет

[54](#): закон, постановление, распоряжение, указ

[55](#): закон

[85](#): закон, указ, распоряжение, повеление

[821](#): право, закон

[1640](#): норма, правило, порядок, закон

[1863](#): правило, норма, принцип, закон, обычай, устой, традиция, строй, канон

[6910](#): закон, законодательство, законы, право, свод законов

[9718](#): термин, закон

[10186](#): догмат, аксиома, закон, основа, положение, постулат, правило, принцип, пункт

[10804](#): кредо, взгляд, закон, принцип, убеждение, уверение, воззрение

Версия 1

Автор [Викисловарь](#)

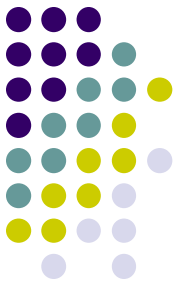
Обновлено 23:04:37 08.07.2013

# РУТЕЗ: ОБЩАЯ ХАРАКТЕРИСТИКА



- Тезаурус разрабатывается в НИВЦ МГУ с 1994 г. для задач информационного поиска в текстах общественно-политической тематики
- Тексты – официальные документы, междунар. договоры, новости (экономика, политика, спорт, искусство и др.)
- Представление понятия:
  - имя понятия
  - набор выражающих его текстовых выражений
- Отличие от *WordNet* – нет разделения по части речи:  
РЕГИСТРАЦИЯ: *Зарегистрировать, регистрационный, регистрация, регистрирование, регистрировать*
- Семантические связи понятий: синонимия,  
ВЫШЕ-НИЖЕ (Род-Вид ), Ассоциация  
Часть-Целое (части жестко и всегда связаны с целым)
- Объем РуТез: 55 тыс. понятий,  
158 тыс. слов и выражений, 210 тыс. отношений

# РУТЕЗ: ОТНОШЕНИЯ ВЫШЕ-НИЖЕ И АССОЦИАЦИИ



- Отношение НИЖЕ-ВЫШЕ (аналог Род-Вид)
  - есть модификаторы слабости отношения, в частности, *В* – «возможно», т.е. выполняется не всегда  
*пенсионер: выше\_В старый человек*
- Отношение онтологической зависимости: *авто – гараж*
- Ассоциация (около половины связей) – устанавливается, если для связанных по смыслу понятий не подходит другое
- Виды ассоциации *X* и *Y*:
  - ❖ симметричная ассоциация – если при поиске текстов о *X* могут пригодиться тексты о *Y*: *родители ассоц дети*
  - ❖ несимметричная ассоциация (онтологич. зависимость) – *X* не может существовать без *Y*:  
*саммит ассоц1 глава государства*  
*жидкость ассоц2 кипение жидкости*  
 $\text{ассоц1}(X, Y) = \text{ассоц2}(Y, X)$

# РУТЕЗ: ПОНЯТИЕ 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

АССОЦ<sub>2</sub> ШКОЛЬНАЯ ПЕРЕМЕНА

## УРОК НА БУДУЩЕЕ

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

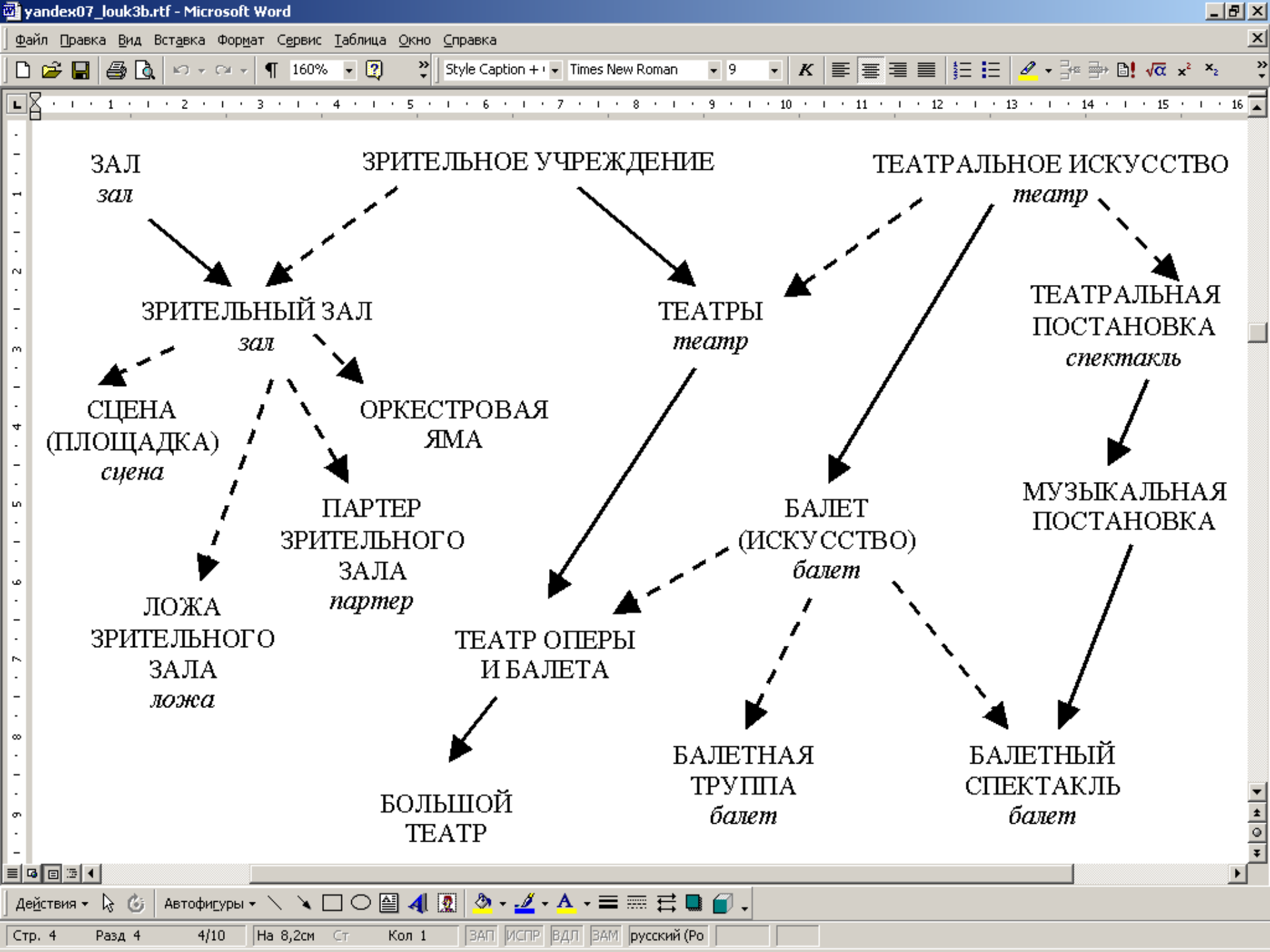
ВЫШЕ РЕАЛЬНЫЙ ПРИМЕР

ЦЕЛОЕ ОПЫТ (СОВОКУПНОСТЬ ЗНАНИЙ И УМЕНИЙ)

ЦЕЛОЕ<sub>В</sub> ПОУЧЕНИЕ, НАСТАВЛЕНИЕ

ЧАСТЬ ПОУЧИТЕЛЬНЫЙ

АССОЦ<sub>2</sub> ПРОУЧИТЬ (НАКАЗАТЬ)



# *BabelNet*: МНОГОЯЗЫЧНАЯ СЕМАНТИЧЕСКАЯ СЕТЬ

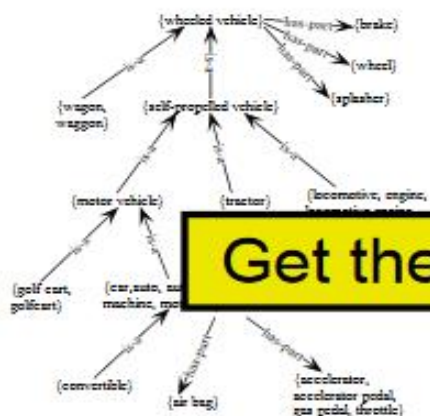


*Семантическая сеть* – модель представления знаний в ИИ: ориентированный граф из поименованных вершин и ребер

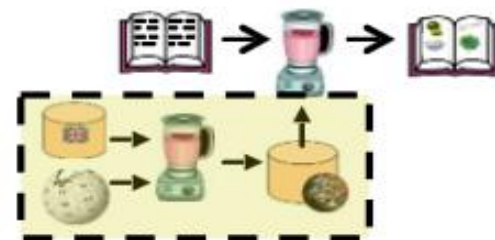
Start from two large **complementary** resources:

WordNet: full-fledged taxonomy

Wikipedia: multilingual and continuously updated

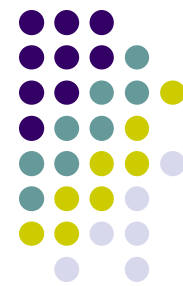


Get the best from both worlds





# РАЗРАБОТКА *BabelNet*



Проект склеивания Википедии и *WordNet*  
(словарных и энциклопедических знаний):

- WordNet: структурированные словарные данные, но не хватает именованных сущностей и нет многоязычия
- Wikipedia: мультязычная, обширная и подробная, но слабоструктурированная

Автоматически решались задачи:

- Объединение этих 2 ресурсов только для английского: выбор лучшего синсета для статьи в Википедии (? 87% ассигурация для 50 тысяч пересекающихся понятий)  
Узлы сети – понятия, ребра – отношения между ними:  
*IS\_A, INSTANCE\_OF, PART\_OF* и др.
- Перевод на другие языки английских статей из Википедии

*BabelNet 2.0* – использованы также *WikiData*, *Wiktionary*,  
*OmegaWiki* – 50 языков, 9 млн. понятий, 250 млн. связей





# ЗАКЛЮЧЕНИЕ

- Рассмотрены вопросы лексической семантики, заключенной в системе языка
- Идет развитие и объединение разных видов ресурсов
- *RuWordNet* (133 тыс. слов и выражений) [ruwordnet.ru/ru](http://ruwordnet.ru/ru) получен из *PyТез*: разделение понятий по 3 частям речи, перенесена часть отношений, добавлены новые
- *BabelNet* – мультилингвистический (271 язык), лексический и энциклопедический ресурс на базе нескольких источников [babelnet.org](http://babelnet.org)
- Подход к семантике со стороны речи/текстов – *дистрибутивная лексическая семантика*: векторные пространства слов, включая модели *Word2Vec* и более сложные.

## СПАСИБО ЗА ВНИМАНИЕ

# УПРАЖНЕНИЯ



- Для следующих предложений нарисовать размеченные деревья составляющих и деревья зависимостей:
  - *Серая птичка весело поет незатейливую песенку.*
  - *Зеленая лягушка поймала быстро летящую муху в саду.*КС-грамматика:  $S \rightarrow NP VP$      $NP \rightarrow N | DT N | DT AN$   
 $DT \rightarrow a | the$      $AN \rightarrow A N | A AN$      $VP \rightarrow V | V NP$   
что в ней надо подправить?
- Определить проективность фраз: *Признания нам его не нужно.*  
*Этому человеку мы будем обязаны всю жизнь.*
- Сколько различных деревьев зависимостей возможно для фраз?
  - *Эти школьники скоро напишут диктант по русскому языку*
  - *Продолжается прием в колледж на территории Англии.*
  - *Мы увидели больного врача Сидорова.*
  - *Привет освободителям Минска от немецких захватчиков.*
  - *Миронов вернулся из своей командировки на север в Москву.*
  - *Я видел его молодым.*