

Таксономия: Что это? Зачем это?

- ▶ Д.т.н. Миркин Борис Григорьевич
- Зам. Председателя Классификационной Комиссии ВСНТО СССР (1984 –1991)
- Профессор, ФКН НИУ Высшая школа экономики, Москва РФ
- В.н.с., Международный Центр анализа и принятия решений
- Prof. Emeritus, University of London UK
- Member, Academia Europaea
- Fellow, CORE Academy

Биологическая таксономия: Классификация всего живого

К. Линней (1707–78) Удачная
таксономия / по сходству
органов размножения /

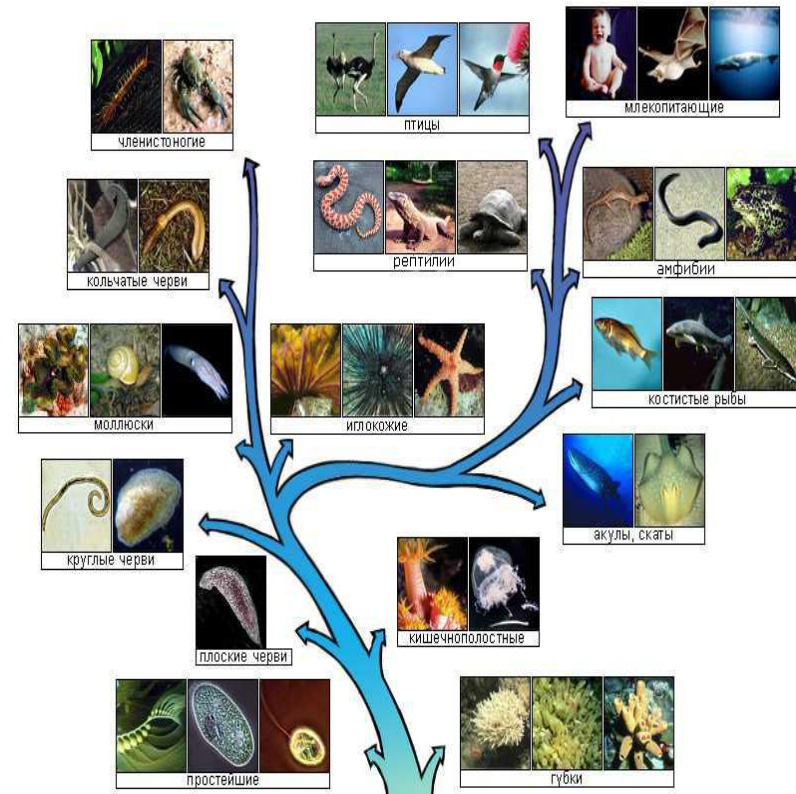
Ч. Дарвин (1809–82)

Удачное осмысление / как схемы
биологической **эволюции** /

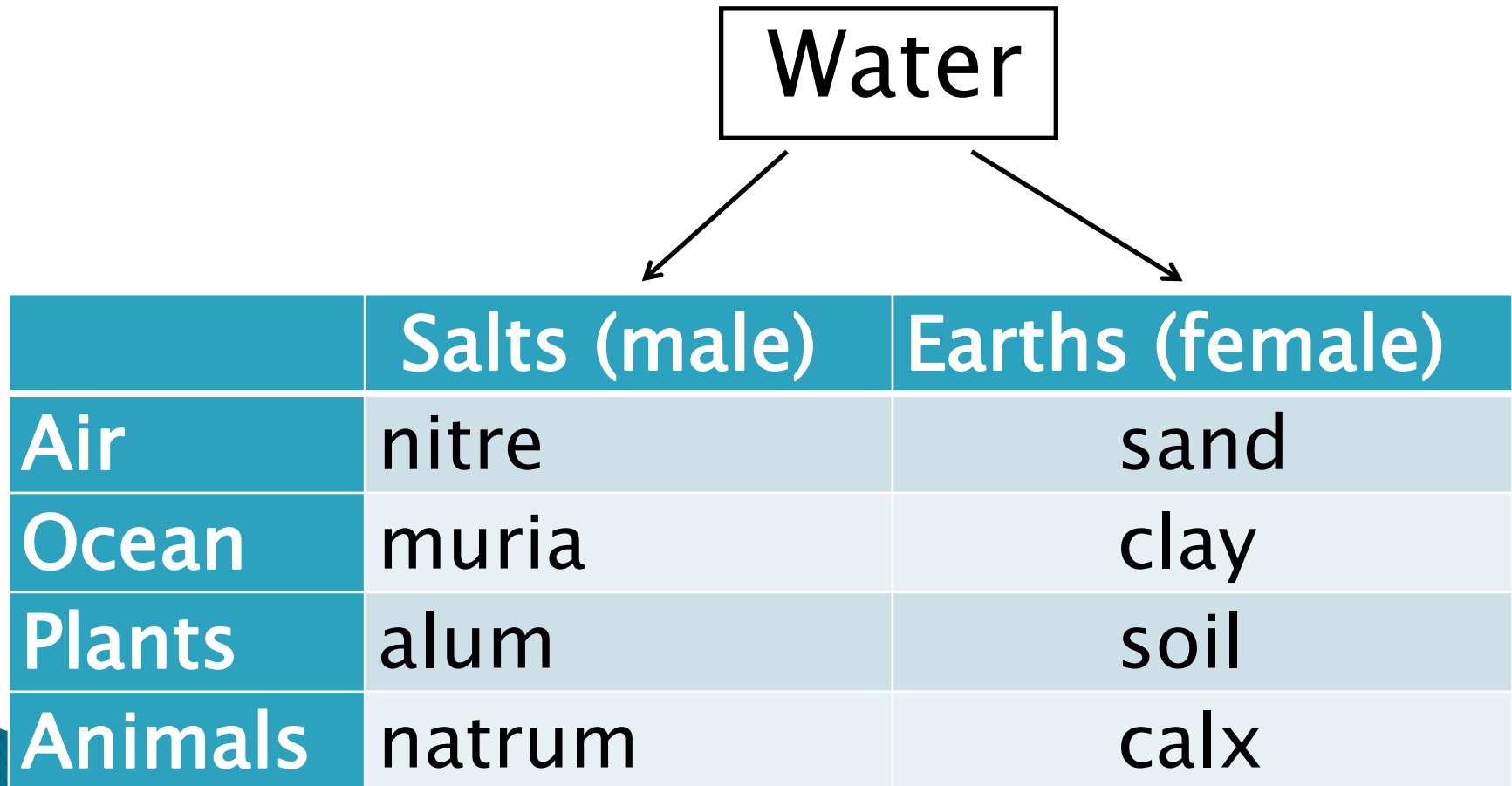
20 век: Две интерпретации для
Дарвина:

– **Кладистика** (дивергенция по
одному признаку)

– **Нумерическая таксономия**
(дивергенция по многим
признакам)



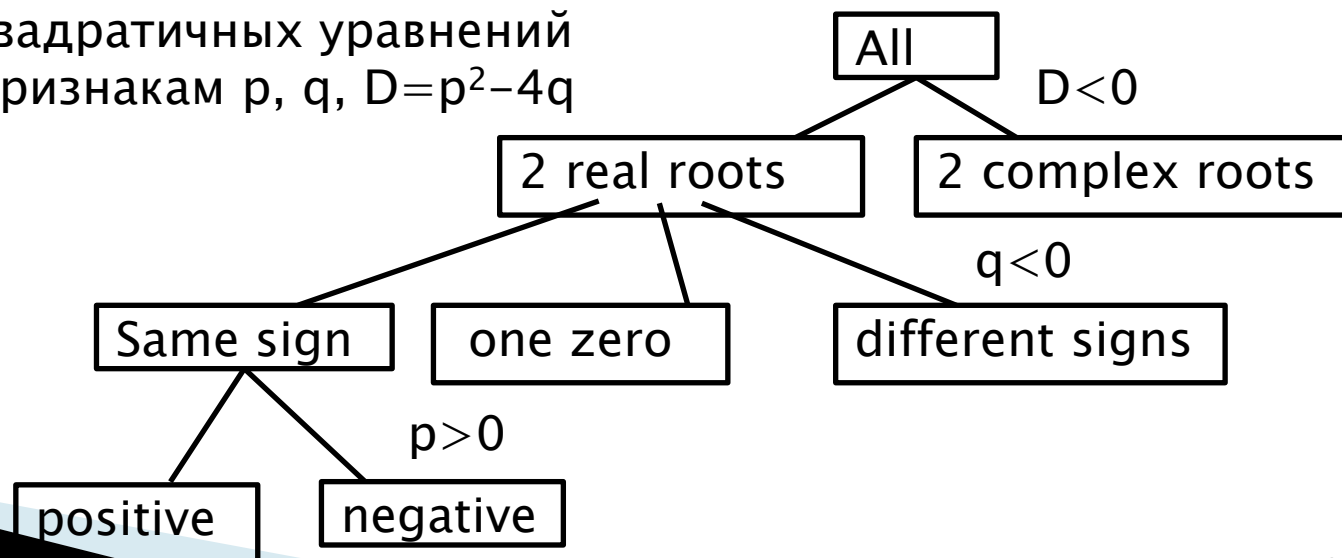
К. Линней: Неудачная таксономия минералов
/по сходству «органов размножения»/, т.е.
чушь – верхние уровни



Таксономия:

- Форма классификации понятий в виде корневого дерева, в котором отношение между вершинами А и Б разного уровня (А – «родитель» Б) описывается условием вида «Б это такая часть А, которая...» или «Б это такое А, у которого...»

Пример: 25 веков развития математики,
Классификация квадратичных уравнений
 $x^2+px+q=0$ по признакам p , q , $D=p^2-4q$



Другие формы классификации

Писатели Википедии не подозревают о существовании других форм классификации:

- ▶ – **Относительная группировка** / весовые категории в боксе, Таблица Менделеева/
- ▶ – **Стратификация** (ранжирование университетов, стратиграфия)
- ▶ – **Типология** (литературные типы, виды почв, темперамент)
- ▶ – **Выделенный класс** (G7, аутизм и др.)

Классификация – это

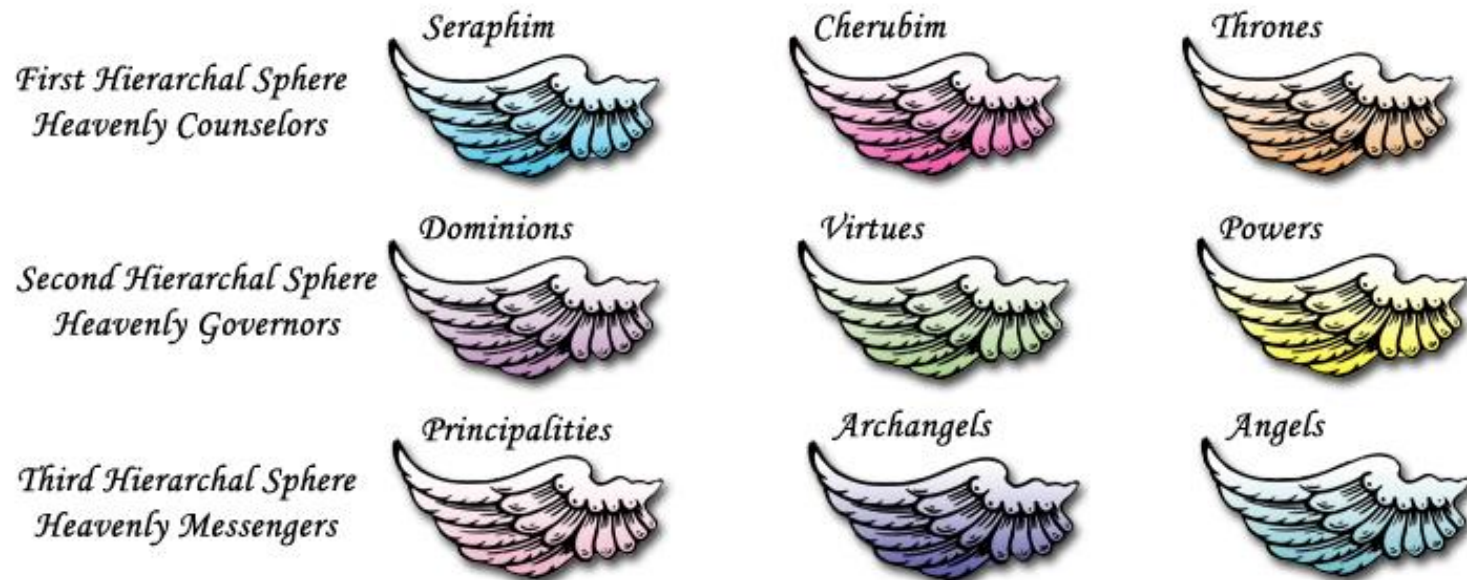
- ▶ Разделение «универсума» на группы той или иной степени общности с целью
 - – структуризации,
 - – выявления связи между разными аспектами,
 - – а также для организации и хранения знаний

Классификация – не наука

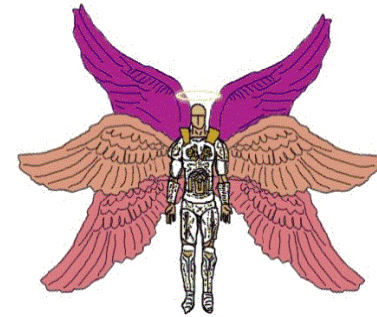
- В настоящее время не существует как научное или инженерное понятие
- Нет ни одной монографии, не говоря уж об учебниках, посвященных классификации
- Выброшена из научного знания в 17 веке вместе с многими другими понятиями, введенными или рассматривавшимися Аристотелем
- Пример: не присуждение Нобелевской премии Менделееву (периодическая таблица элементов) и Woese (выделил и включил Archaea в древо жизни)
- Таксономия – то же самое

Не наука: Классификация ангелов в христианстве

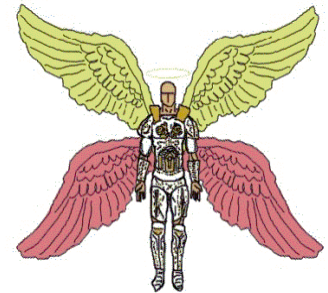
Hierarchy of Angels



Seraphim



Cherubim



Angel



Формы классификации

▶ Иерархия

- Биологическая таксономия
- Библиотечная классификация

▶ Типология

- Классификация почв
- Классификация литературных произведений

▶ Стратификация

- Социальная

▶ Отдельная группа

- «Большая семерка» стран
- «Злостные неплательщики»

Психология: Типы темпераментов

- ▶ Гиппократ (460–370 до Р.Х.): Четыре сока в организме человека:
- ▶ Кровь (сангуа) ++ **сангвиник**
- ▶ Желчь (холе) ++ **холерик**
- ▶ Лимфа (флегма) ++ **флегматик**
- ▶ Чёрная желчь (меланха) ++ **меланхолик**



Психология: Типы темпераментов

- ▶ В. Д. Небылицын (1930–72): Типов больше, но эти четыре могут быть определены силой и скоростью нервных процессов:
- ▶ Скорость +, Сила – **сангвиник**
- ▶ Скорость +, Сила + **холерик**
- ▶ Скорость –, Сила + **флегматик**
- ▶ Скорость –, Сила – **меланхолик**



Роль классификаций огромна в любой области теории и практики

- **Технология**
 - Стандарты
- **Образование**
 - Дифференциация по уровню и области знаний
 - Университетская табель о рангах
- **Полития**
 - Разделение стран на блоки
 - Партийная система
- **Интернет**
 - Рубрикация
 - Страницы

Три примера таксономии в порядке нарастания сложности работы с ней

- ▶ Таксономия товаров и услуг: просто
- ▶ Таксономия области науки : сложнее
- ▶ Таксономия серьезной реальной проблематики (здоровье населения) : усилие национального уровня

Пример 1

Таксономия товаров и услуг

- Товарная номенклатура внешнеэкономической деятельности ЕАЭС – три уровня (21/97/...)

https://base.garant.ru/402888401/#block_100000

– Таксономия товаров и услуг IAB (The Interactive Advertising Bureau) Content Taxonomy

<https://www.iab.com/guidelines/content-taxonomy/>

Они удобны для решения всевозможных задач планирования, управления и контроля, потому что каждый товар нетрудно отнести к единственному узлу таксономии (мясо/рыба/платье/обувь/фильм/картина...)

IAB (The Interactive Advertising Bureau) Content Taxonomy

Fragment <https://www.iab.com/>

```
graph TD; Root[ ] --- SF[Style & Fashion]; Root -.- PC[Personal Care]; Root -.- DC[Designer Clothing]; SF --- MF[Men's Fashion]; SF -.- PC; SF -.- DC; MF --- MOW[Men's Outerwear]; MF --- MW[Men's Watches]; MF --- MS[Men's Shoes];
```

Style & Fashion

**Men's
Fashion**

**Personal
Care**

**Designer
Clothing**

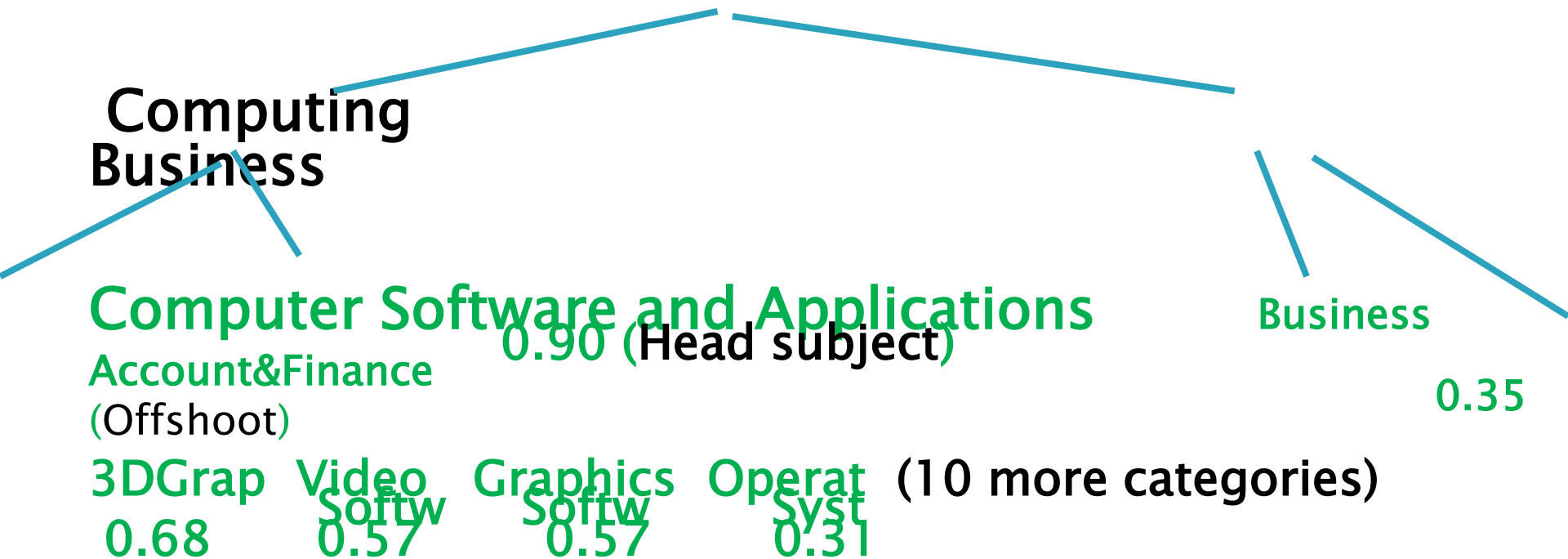
...

**Men's
Outerwear**

**Men's
Watches**

**Men's
Shoes**

Lifting IAB user profile to a Head Subject: **with our Generalization Model**



Applying to Targeted adverts

USER'S TARGET: an advert on
*Daycare and Pre-School; Internet Safety; Parenting
Children Aged 4-11; Parenting Teens; Antivirus
Software*

To be shown to a million visitors

Our method

Impressions	378933	942104	1017598
Clicks	1061	2544	1526
	Standard	LIFT	Solution 2

Пример 2

Таксономия «Науки данных»

- ↪ На основе Классификации Компьютерных Наук, разработанной Всемирной Ассоциацией Компьютерной Техники ACM
- ↪ The 2012 ACM Computing Classification System
- ↪ <https://www.acm.org/publications/class-2012>
- ↪ Предыдущая классификация 1998.
Следующая – 2026 (?)

Upper layers of Taxonomy for “Data analysis” extracted from ACM-CCS 2012

Subject index	Subject name
1.	Theory of computation
1.1.	Theory and algorithms for application domains
2.	Mathematics of computing
2.1.	Probability and statistics
3.	Information systems
3.1.	Data management systems
3.2.	Information systems applications
3.3.	World Wide Web
3.4.	Information retrieval
4.	Human-centered computing
4.1.	Visualization
5.	Computing methodologies
5.1.	Artificial intelligence
5.2.	Machine learning

Куда отнести нашу статью?

Nascimento, S., Martins, A., Relvas, P., Luís, J. F., & Mirkin, B. (2023). Core–shell clustering approach for detection and analysis of coastal upwelling.

Computers & Geosciences, 179, 105421.

1.1. или 3.2. или 4.1 или 5.2 ?

Каждая рубрика имеет отличающегося предка!

**Ограниченность такой таксономии:
несовместимое разделение возможностей**

Пример 3

SNOMED CT – A set of bio–medical hierarchical ontologies and semantic mappings among them

Поддерживается головной научной организацией США в области здоровья – Национальной медицинской библиотекой с начала 1999 для компьютеризации здравоохранения в США

▶ **Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT)**

- ▶ <https://www.nlm.nih.gov/healthit/snomedct/index.html>
- ▶ – a multinational effort in computerization of all things related to health and medicine,
- ▶ ~360 000 concepts so far

Hierarchical taxonomies SNOMED CT (currently, 19 of them)

Clinical finding/disorder
Procedure/intervention
Observable entity
Body structure
Organism
Substance
Pharmaceutical/biologic product
Specimen
Special concept
Physical object
Physical force
Event
Environment or geographical location
Social context
Staging and scales

SNOMED CT

Документ2 - Microsoft Word некоммерческое использование

Файл Главная Вставка Разметка страницы Ссылки Рассылки Рецензирование Вид Настройки

Вырезать Копировать Вставить Формат по образцу Буфер обмена

Calibri (Основно) 11 A A Aa Шрифт

Абзац Стили

АаБбВвГг, АаБбВвГг, АаБбВв АаБбВв АаБ АаБбВв. Изменить стили

1 Обычный 1 Без интер... Заголовок 1 Заголовок 2 Название Подзаголо...

Найти Заменить Выделить Редактирование

Parent(s): body structure
(Select a parent to make it the "Current Concept".)
Action (qualifier value)

Current Concept:
Evaluation - action (qualifier value)

Child(ren):
(N=5) (Select a child to make it the "Current Concept".)
Examination - action (qualifier value)
Imaging - action (qualifier value)
Measurement - action (qualifier value)
Monitoring - action (qualifier value)
Spectroscopy - action (qualifier value)

Current Concept:

Fully Specified Name:	Evaluation - action (qualifier value)
ConceptId:	129265001

Defining Relationships:

Is a Action (qualifier value)
This concept is primitive.

Descriptions (Synonyms):

Fully Specified Name:	Evaluation - action (qualifier value)
Preferred:	Evaluation - action [208001015]
Synonym:	Patient evaluation - action[1490072019]

Страница: 1 из 1 Число слов: 104 английский (США) 170%

Проблемы внедрения

- ▶ Сами таксономии достаточно просты и адекватны, их обилие и сложности связи между ними вполне преодолимы
- ▶ Внедрение тормозится отсутствием электронных записей того же качества в больницах

Заключение

- ▶ Таксономия – удобный инструмент структурирования, хранения и использования знания
- ▶ Таксономию удобно применять в натуральных системах и, вероятно, в системах банковских услуг
- ▶ Трудности могут быть связаны
 - (а) с наличием многих аспектов – для каждого своя таксономия
 - (б) с появлением сложных продуктов, которые могут принадлежать сразу нескольким узлам таксономии