

# **Онтологии и Представление Знаний**

**Борис Конев**

**Department of Computer Science**

**Liverpool University**

# Дескрипционные логики

## Комбинирование определений

Коллекция медицинских терминов SNOMED допускала порождение новых терминов путем комбинирования известных

**T-1X500 = bone**

**T-1X501 = long bone (kind-of)**

**T-1X505 = shaft of bone (part-of)**

**T-1X520 = cortex of bone (constituent-of)**

## Комбинирование определений

Коллекция медицинских терминов SNOMED допускала порождение новых терминов путем комбинирования известных

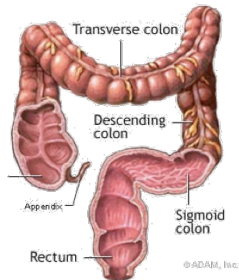
**T-1X500 = bone**

**T-1X501 = long bone (kind-of)**

**T-1X505 = shaft of bone (part-of)**

**T-1X520 = cortex of bone (constituent-of)**

**T-67000 + M-12000 + E-4986 + F-90000 = fracture in colon caused by donkey together with emotional state**



## Что такое дескрипционные логики?

- **семейство** логических формализмов для **представления знаний**
  - происходят от of **семантических сетей** и системы **KL-ONE**
  - описывают домен в терминах **концептов** (классов, понятий), **ролей** (отношений) и **индивидов**
  - определение сложных понятий как комбинации простых

Отличаются

- наличием **формальной семантики** (обычно, стандартная семантика Тарского)
  - более выразительны, чем пропозициональная логика
  - обычно разрешимы
  - родственны модальными и динамическим логикам
- наличие **средств логического анализа**
  - полные и корректные процедуры
  - (оптимизированные) реализации

## Архитектура ДЛ

### База знаний (КВ)

**ТВох** (терминология, схема)

$\text{Man} \equiv \text{Human} \sqcap \text{Male}$

$\text{HappyFather} \equiv \text{Man} \sqcap \exists \text{hasChild}$

...

**АВох** (assertion box, данные)

$\text{john} : \text{Man}$

$(\text{john}, \text{mary}) : \text{hasChild}$

...

Система анализа

Интерфейс

## Краткая история

- 1-й этап (1975-1990):
  - Системы логического анализа не соответствовали формальной семантике (т.е., системы были не полные и/или не корректные) (Back, Classic, Loom, ...)
  - основаны на **структурном анализе**
- 2-й этап (1990-1995):
  - развитие полных и корректных **табличных алгоритмов** (Kris, Crack)
  - анализ **вычислительной сложности** задач логического анализа
  - первые реализации табличных систем систем
  - развитие **методов оптимизации**
- 3-й этап (from 1995):
  - табличные алгоритмы для **ДЛ большой выразительной силы**
  - оптимизированные реализации, “успешно” решающие (N)EXPTIME-полные задачи (FaCT, DLP, Racer)
  - понимание связей с **модальными логиками** и разрешимыми фрагментами **логики первого порядка**

## Последние результаты

- Зрелые реализации (например, FaCT++, RacerPro, Pellet, KAON2, CEL )
- Новые задачи логического анализа (ответ на запросы, поиск объяснений, поддержка версий)
- Полиномиальная разрешимость некоторых ДЛ (EL, DL-Lite)
- **Базы данных**
  - доступ к данным с использование онтологий
  - непротиворечивость модели данных (ER, UML, и т.п.)
  - интеграция схем
- **Онтологии** и **Семантическая Паутина**
  - методологии дизайна, поддержки, интеграции онтологий
  - логический анализ и разметка на основе онтологий (метаданные)
  - поиск и описание сервисов



# Дескрипционная логика $\mathcal{EL}$ : терминологическая часть

## База знаний (KB)

**TBox** (терминология, схема)

$\text{Man} \equiv \text{Human} \sqcap \text{Male}$

$\text{HappyFather} \equiv \text{Man} \sqcap \exists \text{hasChild}$

...

**ABox** (assertion box, данные)

$\text{john} : \text{Man}$

$(\text{john}, \text{mary}) : \text{hasChild}$

...

Система анализа

Интерфейс

## Синтаксис $\mathcal{EL}$

- **Язык  $\mathcal{EL}$  концептов (классов):**

- имена концептов  $A_0, A_1, \dots$  (e.g., Person, Female, ...)
- имена ролей  $r_0, r_1, \dots$  (e.g., hasChild, loves, ...)
- концепт  $\top$  (иногда называют "thing" или "вещь")
- операция  $\sqcap$  (пересечение, конъюнкция, или просто "и").
- квантор  $\exists$  (отверждение о существовании).

### Простейшая дескрипционная логика

## Синтаксис $\mathcal{EL}$

- **Язык  $\mathcal{EL}$  концептов (классов):**

- имена концептов  $A_0, A_1, \dots$  (e.g., Person, Female, ...)
- имена ролей  $r_0, r_1, \dots$  (e.g., hasChild, loves, ...)
- концепт  $\top$  (иногда называют "thing" или "вещь")
- операция  $\sqcap$  (пересечение, конъюнкция, или просто "и").
- квантор  $\exists$  (отверждение о существовании).

### Простейшая дескрипционная логика

- **Множество  $\mathcal{EL}$  концептов определяется индуктивно:**

- все имена концептов —  $\mathcal{EL}$  концепты
- $\top$  является  $\mathcal{EL}$ -концептом
- Если  $C$  и  $D$  являются  $\mathcal{EL}$ -концептами, а  $r$  — имя роли, то

$$(C \sqcap D), \exists r.C$$

являются  $\mathcal{EL}$ -концептами

## Примеры

- **Person**  $\sqcap$  **Female** (женщина),
- **Person**  $\sqcap$   $\exists$ **hasChild.Person** (человек, у которого есть ребенок),
- **Person**  $\sqcap$   $\exists$ **hasChild.Person**  $\sqcap$   $\exists$ **hasParent.Person** (человек, у которого есть ребенок и родитель),
- **Person**  $\sqcap$   $\exists$ **hasChild.(Person**  $\sqcap$  **Female)**,
- **Person**  $\sqcap$   $\exists$ **hasChild.Person**  $\sqcap$  **Female**,
- **Person**  $\sqcap$   $\exists$ **hasChild.T**,
- **Person**  $\sqcap$   $\exists$ **hasChild.** $\exists$ **hasChild.T**.

## Определения в $\mathcal{EL}$

Если  $A$  имя концепта, а  $C$  —  $\mathcal{EL}$ -концепт, то

- $A \equiv C$  — **определение**  $\mathcal{EL}$ -концепта.  $C$  описывает необходимое и достаточное условие, чтобы объект принадлежал классу  $A$ . "A эквивалентен  $C$ ".
- $A \sqsubseteq C$  — **примитивное определение**  $\mathcal{EL}$ -концепта.  $C$  описывает необходимое условие, чтобы объект принадлежал классу  $A$ . "A поглощен  $C$ ".

### Примеры:

- **Father** = **Person**  $\sqcap$  **Male**  $\sqcap$   $\exists$ **hasChild**.**T**.
- **Student** = **Person**  $\sqcap$   $\exists$ **is\_registered\_at**.**University**.
- **Father**  $\sqsubseteq$  **Person**.
- **Father**  $\sqsubseteq$   $\exists$ **hasChild**.**T**.

## $\mathcal{EL}$ -терминология (TBox)

$\mathcal{EL}$  терминология  $T$  это конечное множество определения вида

$$A \equiv C, \quad A \sqsubseteq C$$

таких что никакое имя концепта не определяется более одного раза.  
Таким образом, в терминологиях не бывает одновременно, например,

- **University**  $\equiv$  **Institution**  $\sqcap$   $\exists$ **grants.academicdegree**
- **University**  $\equiv$  **Institution**  $\sqcap$   $\exists$ **supplies.higher\_education**

## $\mathcal{EL}$ -терминология (TBox)

$\mathcal{EL}$  терминология  $T$  это конечное множество определения вида

$$A \equiv C, \quad A \sqsubseteq C$$

таких что никакое имя концепта не определяется более одного раза.  
Таким образом, в терминологиях не бывает одновременно, например,

- **University**  $\equiv$  **Institution**  $\sqcap$   $\exists$ **grants.academicdegree**
- **University**  $\equiv$  **Institution**  $\sqcap$   $\exists$ **supplies.higher\_education**

Циклические определения допускаются

$$\mathbf{Human\_being} \equiv \exists \mathbf{has\_parent.Human\_being}$$

## Ациклические терминологии

**Ациклическая  $\mathcal{EL}$ -терминология  $T$**  это  $\mathcal{EL}$ -терминология в которой нет (явных или неявных) циклических определений.



## Ациклические терминологии

**Ациклическая  $\mathcal{EL}$ -терминология  $T$**  это  $\mathcal{EL}$ -терминология в которой нет (явных или неявных) циклических определений.

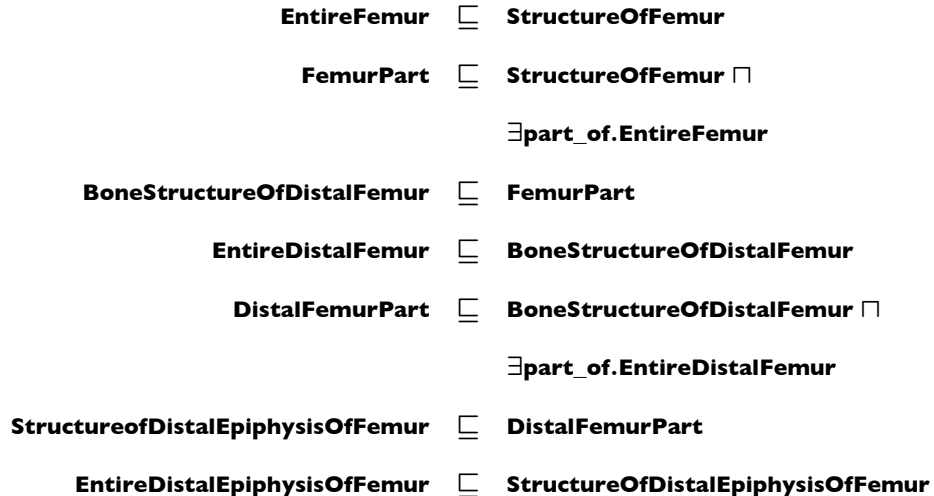
$T$  — терминология

- Определим отношение  $\prec_T$  на парах имен концептов:  $A \prec_T B$  т. и т.т., когда в  $T$  найдется  $A \equiv C$  или  $A \sqsubseteq C$  т.ч.  $B$  входит в  $C$
- Терминология  $T$  **ациклична** если транзитивное замыкание  $\prec_T$  иррефлексивно.

## Пример: SNOMED CT

- Терминология медицинских терминов насчитывающая около 400 000 определений (400 000 имен концептов в 60 имен ролей)
- Ацикличная *ЭЛ*-терминология (почти)
- Принадлежит (и поддерживается) некоммерческой организации IHSTDO (International Health terminology Standards Development Organisation).
- IHSTDO имеет 9 членов (бесплатное членство для 49 развивающихся стран).
- Цель: обеспечение обмена информацией между медиками, исследователями и пациентами по всему миру.

## Фрагмент SNOMED CT



## Импликация $\mathcal{EL}$ -концептов (concept inclusion, CI)

$C$  и  $D$  – произвольные  $\mathcal{EL}$  концепты.

- $C \sqsubseteq D$  называется **импликацией  $\mathcal{EL}$  концептов**. Означает, что каждый  $C$  **является**  $D$ . Также говорят, что  $C$  поглощен  $D$  или что  $D$  поглощает  $C$ . Иногда говорят, что  $C$  включается в  $D$ .
- $C \equiv D$  является формой записи двух импликаций,  $C \sqsubseteq D$  и  $D \sqsubseteq C$ . " $C$  и  $D$  эквивалентны.

Например:

- **Disease**  $\sqcap$   $\exists$ has\_location.Heart  $\sqsubseteq$  **NeedsTreatment**
- $\exists$ student\_of.ComputerScience  $\sqsubseteq$  **Human\_being**  $\sqcap$   $\exists$ knows.Programming\_Language

## $\mathcal{EL}$ -TBox

$\mathcal{EL}$ -TBox это конечное множество импликаций  $\mathcal{EL}$ -концептов вида  $C \sqsubseteq D$  (также используем сокращение  $C \equiv D$ ).

Пример:

**Pericardium**  $\sqsubseteq$  **Tissue**  $\sqcap \exists \text{cont\_in.Heart}$

**Pericarditis**  $\sqsubseteq$  **Inflammation**  $\sqcap \exists \text{has\_loc.Pericardium}$

**Inflammation**  $\sqsubseteq$  **Disease**  $\sqcap \exists \text{acts\_on.Tissue}$

**Disease**  $\sqcap \exists \text{has\_loc.}\exists \text{cont\_in.Heart}$   $\sqsubseteq$  **Heartdisease**  $\sqcap$  **NeedsTreatment**

## $\mathcal{EL}$ -TBox

$\mathcal{EL}$ -TBox это конечное множество импликаций  $\mathcal{EL}$ -концептов вида  $C \sqsubseteq D$  (также используем сокращение  $C \equiv D$ ).

Пример:

**Pericardium**  $\sqsubseteq$  **Tissue**  $\sqcap \exists \text{cont\_in.Heart}$

**Pericarditis**  $\sqsubseteq$  **Inflammation**  $\sqcap \exists \text{has\_loc.Pericardium}$

**Inflammation**  $\sqsubseteq$  **Disease**  $\sqcap \exists \text{acts\_on.Tissue}$

**Disease**  $\sqcap \exists \text{has\_loc.}\exists \text{cont\_in.Heart}$   $\sqsubseteq$  **Heartdisease**  $\sqcap$  **NeedsTreatment**

Т.о.,

ациклические терминологии  $\Rightarrow$  терминологии  $\Rightarrow$  TBox

## Как используют TBox?

**Иерархия концептов** задаваемая  $T$  определяется как

$$\{A \sqsubseteq B \mid A, B \text{ имена концептов в } T \text{ и } T \text{ влечет } A \sqsubseteq B\}$$

например, рассмотренный фрагмент SNOMED CT влечет

**EntireDistalEpiphysisOfFemur**

⊆

**StructureOfDistalEpiphysisOfFemur**

⊆

**DistalFemurPart**

⊆

**BoneStructureOfDistalFemur**

⊆

**FemurPart**

## Стандартное использование SNOMED CT на основе иерархии

- SNOMED CT используется, чтобы породить иерархию медицинских терминов. Каждый термин ассоциирован с числовым кодом и определением.
- Врачи используют иерархию чтобы
  - порождать,
  - обрабатывать, и
  - хранить

электронные медицинские карточки пациента, содержащие диагноз, рекомендованное лечение, назначенные лекарства, результаты исследований и т.п.



## Семантика $\mathcal{EL}$

- **интерпретация** это структура  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$  в которой
  - $\Delta^{\mathcal{I}}$  это **носитель, домен** (непустое множество)
  - **интерпретация**  $\cdot^{\mathcal{I}}$  отображает:
    - \* имя концепта  $A$  в  $A^{\mathcal{I}}$ , подмножество  $\Delta^{\mathcal{I}}$   $(A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}})$
    - \* имя роли  $r$  в бинарное отношение  $r^{\mathcal{I}}$  на  $\Delta^{\mathcal{I}}$   $(r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}})$
- Интерпретация  $C^{\mathcal{I}}$  произвольного концепта  $C$  определяется индуктивно:
  - $(\top)^{\mathcal{I}} = \Delta^{\mathcal{I}}$
  - $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
  - $(\exists r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \text{существует } y \in \Delta^{\mathcal{I}} \text{ т.ч. } (x, y) \in r^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}}\}$

## Пример

Пусть  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where

- $\Delta^{\mathcal{I}} = \{a, b, c, d, e, f\}$ ;
- $\mathbf{Person}^{\mathcal{I}} = \{a, b, c, d, f\}$ ;  $\mathbf{Female}^{\mathcal{I}} = \{a, b, c, e\}$ ;
- $\mathbf{hasChild}^{\mathcal{I}} = \{(a, b), (b, c), (d, e), (f, f)\}$ .

Вычислить

- $(\mathbf{Person} \sqcap \mathbf{Female})^{\mathcal{I}}$ ,
- $(\mathbf{Person} \sqcap \exists \mathbf{hasChild}.\mathbf{Person})^{\mathcal{I}}$ ,
- $(\mathbf{Person} \sqcap \exists \mathbf{hasChild}.\mathbf{Female})^{\mathcal{I}}$ ,
- $(\mathbf{Person} \sqcap \exists \mathbf{hasChild}.\mathbf{Person} \sqcap \mathbf{Female})^{\mathcal{I}}$ ,
- $(\mathbf{Person} \sqcap \exists \mathbf{hasChild}.\top)^{\mathcal{I}}$ ,
- $(\mathbf{Person} \sqcap \exists \mathbf{hasChild}.\exists \mathbf{hasChild}.\top)^{\mathcal{I}}$ .

## Интерпретация импликаций концептов

$\mathcal{I}$  интерпретация,  $C \sqsubseteq D$  импликация концептов, и  $T$  – TBox.

- $\mathcal{I} \models C \sqsubseteq D$  т. и т.т, когда  $C^{\mathcal{I}} \sqsubseteq D^{\mathcal{I}}$ . Говорят, что
  - $\mathcal{I}$  выполняет  $C \sqsubseteq D$  или
  - $C \sqsubseteq D$  истинна в  $\mathcal{I}$  или
  - $\mathcal{I}$  является моделью для  $C \sqsubseteq D$ .
- $\mathcal{I} \models C \equiv D$  т. и т.т., когда  $C^{\mathcal{I}} = D^{\mathcal{I}}$
- $\mathcal{I} \models T$  т. и т.т., когда  $\mathcal{I} \models E \sqsubseteq F$  для всех  $E \sqsubseteq F$  в  $T$ .
  - $\mathcal{I}$  выполняет  $T$  или
  - $\mathcal{I}$  является моделью для  $T$ .

## ТВох и импликации концептов

$T$  – ТВох и  $C \sqsubseteq D$  – импликация концептов.  $T$  влечет  $C \sqsubseteq D$  т. и т.т., когда каждая модель  $T$  является моделью  $C \sqsubseteq D$ .

- $T \models C \sqsubseteq D$  или
- $C \sqsubseteq_T D$ .

Пример: пусть MED является следующим  $\mathcal{EL}$ -ТВох'ом

**Pericardium**  $\sqsubseteq$  **Tissue**  $\sqcap \exists \text{cont\_in.Heart}$

**Pericarditis**  $\sqsubseteq$  **Inflammation**  $\sqcap \exists \text{has\_loc.Pericardium}$

**Inflammation**  $\sqsubseteq$  **Disease**  $\sqcap \exists \text{acts\_on.Tissue}$

**Disease**  $\sqcap \exists \text{has\_loc.}\exists \text{cont\_in.Heart}$   $\sqsubseteq$  **Heartdisease**  $\sqcap$  **NeedsTreatment**

Можно убедиться что **Pericarditis**  $\sqsubseteq_{\text{MED}}$  **NeedsTreatment**.