УДК 004.822 *Научная статья* DOI: 10.18287/2223-9537-2026-16-3-491-501

Доменно-независимый онтологический паттерн проектирования интерпретируемых нейросимвольных систем

© 2026, А.Н. Жданова

Самарский национальный исследовательский университет им. академика С.П. Королева, Самара, Россия

Аннотация

Интерпретируемые нейросимвольные системы анализа данных, объединяющие графы знаний и графовые нейронные сети, разрабатываются для конкретных предметных областей. В работе предложен доменно-независимый онтологический паттерн проектирования «Объект анализа – Признак – Целевая характеристика», задающий структуру интерпретируемой нейросимвольной системы, в которой формирование объяснений определяется онтологической организацией знаний. Паттерн включает классы концептов, предикаты, правила преобразования онтологии в граф знаний и построения графовой нейронной сети. На его основе разработан метод проектирования, включающий переход от концептуализации предметной области к построению графа знаний, обучению модели и формированию семантических путей объяснения. Предложенный паттерн объединяет структуру знаний, механизмы вывода и формирования объяснений в единой концептуальной схеме. Применение паттерна показано на примере двух предметных областей: психологического профилирования пользователей социальной сети; дерматоскопической диагностики новообразований кожи. Выделены доменно-инвариантные и доменно-специфичные элементы онтологии и выполнено отображение онтологической структуры на граф знаний. Показано, что структура паттерна, механизмы вывода и формирования объяснений сохраняются для различных предметных областей. Интерпретация решения определяется онтологической структурой нейросимвольной системы и формируется посредством семантических связей между объектом анализа, его признаками и целевой характеристикой.

Ключевые слова: *онтологический паттерн, нейросимвольная система, граф знаний, графовые нейронные сети, интерпретируемость, объяснимый искусственный интеллект.*

Цитирование: *Жданова А.Н.* Доменно-независимый онтологический паттерн проектирования интерпретируемых нейросимвольных систем. *Онтология проектирования*. 2026. Т.16, №3(61). С.491-501. DOI: 10.18287/2223-9537-2026-16-3-491-501.

Благодарности: автор выражает благодарность д.т.н., доценту А.В. Куприянову за обсуждение результатов исследования и замечания при подготовке статьи.

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Введение

Интерпретируемость выводов в системах поддержки принятия решений – важное практическое требование, в том числе в таких областях, как медицина, психологическая диагностика и безопасность. Распространённым способом повышения интерпретируемости являются апостериорные методы атрибуции, формирующие объяснения для уже обученной модели. Объяснения формируются независимо от процесса вывода и не опираются на формализованное представление предметного знания, их связь с используемыми понятиями предметной области (ПрО) остаётся неявной.

В нейросимвольных системах (НСС) графы знаний (ГЗ) используются для представления предметного знания, а графовые нейронные сети (ГНС) – для анализа данных и формирования предсказаний. Интерпретация результата в этом случае может быть представлена семантическим путём в ГЗ. Подобные системы проектируются независимо для каждой ПрО.

Предметом настоящей статьи является онтологический паттерн (ОП) проектирования и его применение в различных ПрО. Предлагаемый ОП основан на разработанном автором подходе к построению интерпретируемых НСС [1].

1 Источниковая база для исследования

Онтология в инженерии знаний определяется как формальная спецификация концептуализации ПрО [2]. Реализация онтологий обычно основана на языках *RDF* и *OWL* [3, 4].

ОП проектирования представляют собой переиспользуемые решения типовых задач построения онтологий [5, 6]. ОП фиксируют устойчивые структуры классов и отношений, применимые в различных ПрО. Выделяются содержательные, логические и другие виды ОП, рассматриваются методы их каталогизации и повторного использования [7].

В онтологии проектирования рассматриваются знания о процессах создания и модификации артефактов, включая описание объектов и субъектов проектирования, используемых понятий, ограничений и процедур [8]. Вопросы инженерии знаний, онтологического моделирования и построения порталов знаний исследованы в работах [4, 9].

В нейросимвольном искусственном интеллекте объединены методы машинного обучения и формализованное представление знаний [10]. В таких системах предметное знание представлено ГЗ, а обработка данных выполняется ГНС [11].

Для обработки графовых структур используются ГНС, в том числе графовые сети внимания [12], включающие структуры ГЗ и выявленные зависимости при формировании предсказаний. Существующие НСС ориентированы преимущественно на решение задач в отдельных ПрО. Инвариантные компоненты их онтологической организации и механизмы междоменного переноса не формализованы.

В обзоре [13] систематизированы методы нейросимвольных рассуждений с помощью ГЗ; в [14] причинно-ориентированный механизм объяснения встроен непосредственно в ГНС; в [15] в сеть встраиваются обучаемые нечёткие правила для интерпретируемой классификации медицинских изображений; в [16] интерпретируемость графовой сети усиливается интеграцией внешних предметных знаний; в [17] предложены объяснения для задач дополнения ГЗ. Онтологическая составляющая объяснимого искусственного интеллекта рассматривается в обзоре [18].

В настоящей работе под *интерпретируемостью* понимается возможность проследить формирование предсказания через семантические связи ГЗ: от наблюдаемого признака к целевой характеристике [1]; интерпретируемость определяется структурой онтологии и связями между её концептами. *Объяснение* рассматривается как представление этого процесса в форме подграфа вывода, набора семантических путей или их текстового описания.

2 Паттерн «Объект анализа – Признак – Целевая характеристика»

ОП задаёт доменно-инвариантную структуру онтологии интерпретируемой НСС. В его состав входят классы концептов, предикаты и ограничения, не зависящие от ПрО. Наполнение признаков и целевых характеристик определяется применением ОП в конкретной ПрО.

ОП включает следующие классы сущностей.

Объект анализа (*AnalysisObject*) – сущность ПрО, для которой выполняется анализ и формируется предсказание. Объект описывается набором атрибутов, получаемых из исходных данных и производных признаков.

Признак (*Feature*) – характеристика объекта анализа, используемая при построении предсказания. Состав признаков определяется ПрО.

Целевая характеристика (*TargetCharacteristic*) – концепт ПрО, отражающий интерпретируемое свойство объекта анализа и являющийся объектом предсказания или объяснения.

Значение целевой характеристики (*TargetMeasurement*) – значение целевой характеристики, полученное в результате наблюдения, измерения или экспертной оценки и используемое при обучении и валидации модели.

Разделение классов *TargetCharacteristic* и *TargetMeasurement* отражает различие между концептами ПрО и результатами их измерения. Наличие этих классов упрощает расширение онтологии и формулирование зависимостей между признаками и характеристиками.

Семантические связи ОП задаются бинарными предикатами.

- *has_feature* (имеет_признак): домен – *AnalysisObject*, ко-домен – *Feature*. Связывает объект анализа с наблюдаемым признаком.
- *associated_with* (ассоциируется_с): домен – *Feature*, ко-домен – *TargetCharacteristic*. Отражает связь признака с целевой характеристикой. Количественная оценка этой связи определяется при построении ГЗ для конкретной ПрО.
- *correlates_with* (коррелирует_с): домен и ко-домен – *TargetCharacteristic*. Описывает взаимосвязи между целевыми характеристиками и используется при задании предметных ограничений.

Композиция предикатов *has_feature* и *associated_with* образует элементарный семантический путь вывода вида «Объект анализа – Признак – Целевая характеристика». Данный путь является базовой структурой интерпретации результатов в рамках предлагаемого ОП. Для описания ОП используются языки *OWL* и *RDF*. Фрагмент спецификации инвариантного ядра в синтаксисе *Turtle* имеет вид:

```
:AnalysisObject          a owl:Class.
:Feature                 a owl:Class.
:TargetCharacteristic    a owl:Class.
:TargetMeasurement       a owl:Class.
:has_feature a owl:ObjectProperty;
   rdfs:domain :AnalysisObject;
   rdfs:range :Feature.
:associated_with a owl:ObjectProperty;
   rdfs:domain :Feature;
   rdfs:range :TargetCharacteristic.
```

Спецификация задаёт структуру ОП и служит основой для построения ГЗ, используемого в НСС. Схема ОП представлена на рисунке 1.

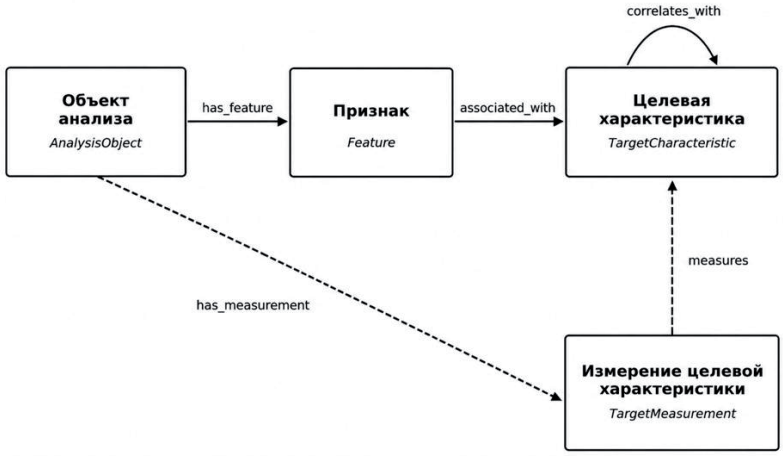

Рисунок 1 – Схема онтологического паттерна

ОП содержит набор аксиом, определяющих структуру онтологии и правила формирования интерпретации.

Для предикатов *has_feature* и *associated_with* задаются ограничения домена и ко-домена. Предикат *has_feature* связывает экземпляры классов *AnalysisObject* и *Feature*, а предикат *associated_with* – экземпляры классов *Feature* и *TargetCharacteristic*. Ограничения предотвращают формирование семантически некорректных связей.

Классы *AnalysisObject, Feature, TargetCharacteristic* и *TargetMeasurement* являются попарно дизъюнктными. Один и тот же индивид не может одновременно принадлежать нескольким классам ОП.

Предикат *correlates_with* задаётся как симметричный. Если одна целевая характеристика связана отношением корреляции с другой, то справедливо и обратное отношение.

Для отношения *associated_with* может задаваться характеристика *associationStrength*, отражающая силу связи между признаком и целевой характеристикой. Способ определения данного значения зависит от конкретной реализации и может основываться на статистических оценках, экспертных знаниях или параметрах обучаемой модели.

3 Метод проектирования нейросимвольной системы на основе паттерна

Проектирование начинается с выделения объекта анализа и целевых характеристик ПрО. Далее выполняется выделение признаков объекта анализа на основе знания ПрО. Каждый признак отображается на концепт класса *Feature* и включается в онтологическую модель.

Устанавливаются связи между признаками и целевыми характеристиками. Отношения *associated_with* задаются на основе статистических зависимостей, экспертных знаний или их комбинации.

Построенная онтология отображается в ГЗ. На следующем этапе обучается ГНС. В настоящей работе используется ГНС внимания, но допускается применение других моделей обработки графов.

После обучения модели формируются объясняющие пути, связывающие объект анализа, признаки и целевые характеристики. Интерпретация результата представляется в виде подграфа вывода или набора семантических путей, а её качество может оцениваться специализированными метриками на основе онтологии.

Схема метода проектирования НСС представлена на рисунке 2. Этапы концептуализации ПрО, выделения признаков и построения онтологии зависят от рассматриваемого домена. Построение ГЗ, обучение графовой модели и формирование объясняющих путей определяются инвариантной структурой паттерна и воспроизводятся при его переносе между ПрО.

Рисунок 2 – Схема метода проектирования нейросимвольной системы

4 Объяснимость как следствие онтологической модели

Объяснение строится на основе онтологии. Элементарное объяснение предсказания есть семантический путь, составленный из ребра *has_feature* и ребра *associated_with*. Онтология фиксирует допустимые типы связей и их семантику. Специалист может сопоставить путь с предметным знанием и принять либо опровергнуть полученную интерпретацию.

Применение ОП можно показать на примере дерматоскопической диагностики. Пусть для изображения *o* графовая модель сформировала предсказание класса «меланома» *t*. Объясняющий путь строится из двух рёбер: ребро *has_feature* связывает *o* с присутствующим на изображении признаком «структурная гетерогенность» *f*, а ребро *associated_with* связывает этот признак с целевой характеристикой «меланома», поскольку в ГЗ задана соответствую-

щая значимая ассоциация. Полученный путь «*o* – *has_feature* – структурная гетерогенность – *associated_with* – меланома» – допускает проверку специалистом: структурная неоднородность относится к признанным дерматоскопическим индикаторам злокачественности [19]. Если модель опирается на признак, не имеющий значимого ребра к предсказанному классу, путь не строится. Таким образом, наличие и состав объяснения определяются структурой онтологии, а не постобработкой.

Для количественной оценки качества объяснений используется структурная метрика, анализирующая граф вывода. Метрика учитывает наличие значимого объясняющего пути и степень его использования моделью при формировании решения. Поскольку метрика определяется на уровне онтологической модели ГЗ, она сохраняет применимость при переносе ОП на различные ПрО.

Объясняющим путём для предсказания целевой характеристики *t* объекта *o* называется путь $\pi = (o \to f \to t)$, состоящий из ребра *has_feature* (объект обладает признаком *f*) и ребра *associated_with* (признак ассоциирован с целевой характеристикой *t*). Путь считается значимым, если одновременно выполняются два условия: нормированная выраженность признака *c* объекта *o* удовлетворяет условию $c(f, o) > \tau,\ \tau = 0{,}2$, а связь *associated_with* становится статистически значимой после поправки на множественные сравнения $q < 0{,}05$.

Основной характеристикой интерпретируемости является покрытие значимыми объясняющими путями. Для каждого объекта *o* вводится индикатор

$$I_n = \begin{cases} 1, & \text{если существует хотя бы один значимый путь } \pi, \\ 0, & \text{иначе.} \end{cases}$$

Тогда показатель покрытия $\mathrm{GPI}_{cov} = \frac{1}{N}\sum_{n=1}^{N} I_n$, где *N* – количество объектов выборки.

Дополнительно оценивается взвешенная согласованность объясняющих путей, отражающая долю вклада значимых путей в итоговое решение:

$$\mathrm{GPI}_w = \frac{\sum_{\pi \in \Pi_{sig}} w(\pi)}{\sum_{\pi \in \Pi} w(\pi)},$$

где Π – множество всех путей от объекта *o* к предсказанной целевой характеристике *t*, $\Pi_{sig} \subseteq \Pi$ – множество значимых путей.

Вес отдельного пути вычисляется по формуле $w(\pi) = c(f, o) \cdot \alpha_e$, где $c(f, o)$ – нормированная выраженность признака объекта, а α_e – коэффициент внимания ребра *associated_with* между признаком *f* и целевой характеристикой *t*, полученный в процессе обучения ГНС внимания. Значения GPI_{cov} и GPI_w вычисляются для каждого объекта и затем усредняются по всей выборке. В качестве дополнительной проверки объяснений используется тест верности. Узлы графа, соответствующие признакам, последовательно удаляются в порядке убывания их вклада в предсказание (например, по величине коэффициента внимания α_e). Корректное объяснение должно приводить к монотонному снижению уверенности модели при удалении наиболее значимых признаков.

5 Отображение онтологии на графовую нейронную сеть

Связь онтологии с ГНС устанавливается на двух уровнях.

На структурном уровне классы ОП отображаются в типы узлов ГЗ, а предикаты – в типизированные рёбра. Ограничения онтологии фиксируют допустимую топологию графа, поэтому механизм передачи сообщений графовой сети внимания (*GAT*) действует только вдоль разрешённых типов связей, и любой путь вывода основан на онтологии.

На параметрическом уровне атрибут *associationStrength* инициализируется статистически значимой мерой связи между признаком и целевой характеристикой. В процессе обуче-

ния этот атрибут уточняется как коэффициент внимания соответствующего ребра в последнем слое *GAT*. Таким образом, структура определяет, что может быть связано, а данные – что действительно связано и насколько сильно.

Онтологические ограничения учитываются при обучении. Функция потерь включает основной член ошибки предсказания и регуляризирующий член, штрафующий нарушение структуры взаимосвязей целевых характеристик, заданной предикатом *correlates_with*:

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda \cdot \mathcal{R},$$

где $\mathcal{L}_{\text{pred}}$ – стандартная функция потерь (например, кросс-энтропия для классификации или среднеквадратичная ошибка для регрессии), λ – гиперпараметр, управляющий силой регуляризации, $\mathcal{R}$ - штраф за несоответствие между предсказанными значениями целевых характеристик и их известными корреляционными связями:

$$\mathcal{R} = \sum_{(i,j)\in E_{\text{corr}}} \left(\text{sim}(\hat{y}_i, \hat{y}_j) - \rho_{ij}\right)^2,$$

где E_{corr} – множество рёбер *correlates_with* между целевыми характеристиками i и j; $\hat{y}_i$ и $\hat{y}_j$ – предсказания модели для соответствующих характеристик; $\text{sim}(\cdot)$ – функция сходства (например, косинусное сходство или коэффициент корреляции Пирсона); ρ_{ij} – заданная в онтологии эталонная мера связи (например, экспертно определённый коэффициент корреляции). Регуляризатор способствует тому, чтобы предсказания модели согласовывались с известной структурой зависимостей между целевыми характеристиками.

Обучению подлежит только графовый модуль (слой внимания и классификатор), тогда как энкодеры признаков используются в режиме фиксированных параметров. Это предотвращает искажение семантики признаков в процессе обучения и сохраняет прослеживаемость связи «онтология – ГЗ – архитектура – обучение»: узлы и рёбра сети сохраняют предметную семантику на всех этапах.

Отображение онтологии на ГНС$_a$ и интерпретация объяснения могут быть представлены следующим псевдокодом.

Вход: онтология O (классы, предикаты, ограничения); обучающая выборка; извлечённые признаки объектов.
Выход: обученная ГНС внимания; объясняющие пути; значение метрики интерпретируемости.

1) Для каждого класса O создать тип узла ГЗ; для каждого предиката – тип ребра.
2) Для каждого объекта o добавить узел и рёбра *has_feature* к присутствующим признакам (вес – нормированная выраженность признака $c(f, o)$).
3) Для каждой пары «признак – целевая характеристика» вычислить статистическую меру связи (например, корреляцию Пирсона); включить ребро *associated_with* только при статистической значимости после поправки на множественность ($q < 0{,}05$); вес ребра инициализируется значением *associationStrength*.
4) Инициализировать ГНС внимания топологией ГЗ; механизм передачи сообщений ограничен типами рёбер, заданными онтологией.
5) Обучить графовый модуль, минимизируя $\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda\mathcal{R}$ с онтологической регуляризацией; веса рёбер *associated_with* уточняются механизмом внимания, т.е. финальный вес ребра становится равным коэффициенту внимания α_e из последнего слоя *GAT*.
6) Извлечь объясняющие пути (композиции рёбер *has_feature* и *associated_with*) и вычислить компоненты метрики интерпретируемости (GPI_cov, GPI_w).

6 Перенос паттерна между предметными областями

Возможность переноса предложенного ОП показана на двух ПрО. В обоих случаях используется одно и то же инвариантное ядро, включающее классы и предикаты, описанные в разделе 2. Различается наполнение сущностей и отношений ПрО.

Задача 1 – Применение ОП для психологического профилирования. В задаче анализа пользователей социальных сетей класс «Объект анализа» соответствует профилям пользова-

телей. Признаки формируются на основе поведенческих и визуальных модальностей цифровой активности, а целевые характеристики соответствуют психометрическим факторам личности. Измерения целевых характеристик получены с использованием известных психологических методик. Отношения *associated_with* связывают поведенческие признаки с факторами личности, тогда как отношения *correlates_with* отражают взаимосвязи между факторами [1].

Задача 2 – Применение ОП для дерматоскопической диагностики. В задаче анализа изображений новообразований кожи объектом анализа является дерматоскопическое изображение. Признаки соответствуют дескрипторам правила *ABCD* (асимметрия, нерегулярность границы, вариативность цвета и диаметр) [19], а также показателям структурной гетерогенности. Целевыми характеристиками являются диагностические категории, измерения которых определяются клиническим диагнозом. Отношения *associated_with* связывают выявленные дерматоскопические признаки с диагностическими категориями. Постановка задачи основана на правиле [19] и наборе данных [20].

Сопоставление двух реализаций ОП (таблица 1) показывает, что инвариантными остаются классы и предикаты, способ построения ГЗ, тип обучаемой модели и механизм формирования объяснений.

Таблица 1 – Сопоставление доменно-инвариантных и доменно-специфичных элементов паттерна

Элемент проектирования	Инвариантно (ядро паттерна)	Домен 1 – Психология	Домен 2 – Дерматоскопия
Объект анализа	класс *AnalysisObject*	профиль соцсети	дерматоскопическое изображение
Признаки	класс *Feature*, предикат *has_feature*	данные цифровой активности и визуальные	дескрипторы *ABCD* и структуры
Целевая характеристика	класс *TargetCharacteristic*	факторы личности	диагностические категории
Измерение	класс *TargetMeasurement*	психологическая методика	клинический диагноз
Ассоциации	предикат *associated_with*	связь данных цифровой активности с факторами	связь дескрипторов с диагнозом
Обучаемая модель	графовая нейронная сеть	графовые сети внимания	графовые сети внимания
Механизм объяснения	семантический путь в графе	идентичен	идентичен

Доменно-зависимыми являются состав признаков и целевых характеристик, источники ассоциаций и используемые измерительные методики. Сопоставление двух онтологий с выделением общего ядра представлено на рисунке 3.

Для формализации переносимости ОП разграничиваются терминологический и фактологический компоненты онтологии. ОП как терминологический компонент (*TBox*) задаётся четвёркой $P = \langle C, R, A, M \rangle$, где C – множество классов, R – множество типов предикатов, A – онтологические аксиомы, M – механизм вывода интерпретации (семантический путь). Реализация ОП в конкретной ПрО представляет собой отображение, которое наполняет фактологический компонент (*ABox*) экземплярами классов «Признак» и «Целевая характеристика», а также определяет источник весов ассоциаций.

Для рассмотренных ПрО без изменений сохраняются элементы *TBox* онтологии. Доменная адаптация сводится к пополнению *ABox*. Сохранение единого онтологического ядра при существенном различии ПрО подтверждает доменную независимость предложенного ОП и общность метода проектирования НСС.

Работоспособность ОП подтверждена количественно в двух рассмотренных ПрО.

Результаты в психологическом домене представлены в [1].

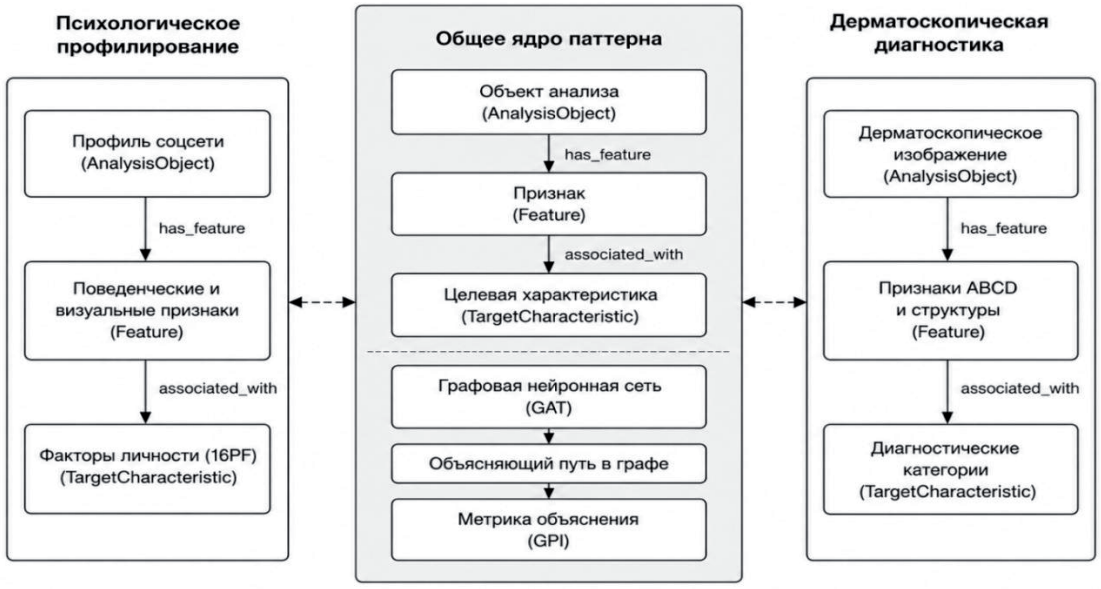

Рисунок 3 – Психологическая и дерматоскопическая реализации паттерна с выделением общего онтологического ядра

В дерматоскопическом домене (HAM10000 [20], 10 015 изображений, 7 диагностических классов) получены следующие значения показателей работы модели: извлечение признаков *Accuracy* = 0,807 и *AUC* = 0,955; при равной точности рассуждающих модулей (≈0,67) отделено графовое рассуждение, опирающееся на концепты, от моделей типа «чёрный ящик» без онтологической опоры GPI_w = 0,42 (GAT) при покрытии GPI_cov ≈ 1,0. Переносимость подтверждена внешней валидацией на независимом наборе ISIC 2018[1] (1 511 изображений), где воспроизведены показатели классификатора и значимые рёбра ГЗ. ОП апробирован на задачах интерпретируемой классификации токсичных текстов и классификации эмоций.

ОП предполагает наличие полной и непротиворечивой предметной онтологии. При неполной онтологии часть решений модели не покрывается объясняющими путями; это состояние диагностируется метрикой (снижение GPI_cov) и указывает на необходимость расширения словаря признаков. Противоречивые или ложные экспертные ассоциации отсеиваются на этапе статистической валидации рёбер: связи, не подтверждённые данными после поправки на множественность проверок, в ГЗ не включаются. Качество объяснений ограничено качеством исходной концептуализации: ОП структурирует и верифицирует предметные знания, но не заменяет доменную экспертизу, а при смене ПрО требуется повторная реализация словаря признаков и целевых характеристик.

ОП реализован в программном комплексе, включающем модули: извлечения признаков; наполнения ГЗ по правилам ОП; нейросимвольного вывода и генерации объяснений с ранжированием семантических путей. Веб-интерфейс визуализирует объясняющие пути и вклад признаков, обеспечивая возможность экспертной верификации результатов анализа.

Заключение

Предложен доменно-независимый ОП проектирования интерпретируемых НСС анализа данных «Объект анализа – Признак – Целевая характеристика» и метод его применения. ОП включает классы, типы отношений, способ построения ГЗ, тип обучаемой модели и механизм формирования объяснений.

[1] *ISIC 2018 Challenge* (*International Skin Imaging Collaboration*) – международный конкурс по анализу изображений кожных поражений. https://challenge.isic-archive.com/landing/2018/.

Применение ОП в задачах психологического профилирования и дерматоскопической диагностики показало, что при различии ПрО сохраняются структура онтологии, механизм вывода и способ формирования объяснений. Новым является выделение доменно-инвариантного ядра проектирования, сохраняемого при изменении ПрО, а также метод построения объясняющих путей вывода на основе онтологической структуры ГЗ.

Список литературы

[1] ***Жданова А.Н.*** Онтологический подход к предсказанию психологических характеристик пользователей соцсетей на основе графов знаний. *Искусственный интеллект и принятие решений*. 2025. №3. С.40-59. DOI: 10.14357/20718594250304.

[2] ***Studer R., Benjamins V.R., Fensel D.*** Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*. 1998. Vol.25(1-2). P.161-197.

[3] ***Hitzler P., Krötzsch M., Rudolph S.*** Foundations of Semantic Web Technologies. Boca Raton: CRC Press, 2009. 455 p.

[4] ***Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И.*** Инженерия знаний. Модели и методы. СПб.: Лань, 2016. 324 с.

[5] ***Gangemi A.*** Ontology Design Patterns for Semantic Web Content. *The Semantic Web – ISWC 2005*. Berlin: Springer, 2005. P.262-276.

[6] ***Presutti V., Gangemi A.*** Pattern-Based Ontology Design. *Ontology Engineering in a Networked World*. Berlin: Springer, 2012. P.35-64. DOI: 10.1007/978-3-642-24794-1_3.

[7] ***Blomqvist E., Hammar K., Presutti V.*** Engineering Ontologies with Patterns – The eXtreme Design Methodology. *Ontology Engineering with Ontology Design Patterns*. Amsterdam: IOS Press, 2016. P.23-50. DOI: 10.3233/978-1-61499-676-7-23.

[8] ***Боргест Н.М.*** Научный базис онтологии проектирования. *Онтология проектирования*. 2013. Т.3. №1. С.7-25.

[9] ***Загорулько Ю.А.*** Построение порталов научных знаний на основе онтологии. *Вычислительные технологии*. 2007. Т.12. Спец. вып. С.169-177.

[10] ***d'Avila Garcez A., Lamb L.C.*** Neurosymbolic AI: The 3rd Wave. *Artificial Intelligence Review*. 2023. Vol.56. P.12387-12406. DOI: 10.48550/arXiv.2012.05876.

[11] ***Hogan A., Blomqvist E., Cochez M.*** et al. Knowledge Graphs. *ACM Computing Surveys*. 2021. Vol.54(4). Art.71. DOI: 10.1145/3447772.

[12] ***Veličković P., Cucurull G., Casanova A.*** et al. Graph Attention Networks. *International Conference on Learning Representations* (ICLR). 2018. 12 p. DOI: 10.48550/arXiv.1710.10903.

[13] ***Liu L., Wang Z., Tong H.*** Neural-Symbolic Reasoning over Knowledge Graphs: A Survey from a Query Perspective. *ACM SIGKDD Explorations Newsletter.* 2025. DOI: 10.48550/arXiv.2412.10390.

[14] ***Zheng K., Yu S., Chen B.*** CI-GNN: A Granger Causality-Inspired Graph Neural Network for Interpretable Brain Network-Based Psychiatric Diagnosis. *Neural Networks.* 2024. Vol.172. Art.106147. arXiv:2301.01642v3 [stat.ML] 28 Jan 2024.

[15] ***Sengupta P., Rekik I.*** FireGNN: Neuro-Symbolic Graph Neural Networks with Trainable Fuzzy Rules for Interpretable Medical Image Classification. arXiv preprint arXiv:2509.10510. 2025.

[16] ***Raj K. A*** Neuro-symbolic Approach to Enhance Interpretability of Graph Neural Network through the Integration of External Knowledge. *Proc. of the 32nd ACM Int. Conf. on Information and Knowledge Management* (CIKM). 2023. DOI: 10.1145/3583780.3616008.

[17] ***Chang H., Ye J., Lopez Avila A., Du J., Li J.*** Path-based Explanation for Knowledge Graph Completion. *Proc. of the 30th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining* (KDD). 2024. DOI: 10.1145/3637528.3671683.

[18] ***Barredo Arrieta A***. et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 2020. Vol.58. P.82-115. DOI: 10.48550/arXiv.1910.10045.

[19] ***Nachbar F., Stolz W., Merkle T.*** et al. The ABCD Rule of Dermatoscopy: High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin Lesions. *Journal of the American Academy of Dermatology*. 1994. Vol.30(4). P.551-559.

[20] ***Tschandl P., Rosendahl C., Kittler H.*** The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*. 2018. Vol.5. Art.180161. DOI: 10.1038/sdata.2018.161.

Сведения об авторе

Жданова Александра Николаевна, 1986 г. рождения. Окончила Самарский государственный аэрокосмический университет им. С.П. Королёва в 2009 г., к.т.н. (2012). Доцент кафедры программных систем Самарского университета. В списке научных трудов более 90 работ в области ИИ и систем поддержки принятия решений. Author ID (РИНЦ): 762284; Author ID (Scopus): 57076115300; Researcher ID (WoS): QMP-9127-2026; ORCID: 0009-0002-6370-1894. *aleksandra.zhd@yandex.ru.*

Поступила в редакцию 19.06.2026, после рецензирования 19.07.2026. Принята к публикации 31.07.2026.

Scientific article DOI: 10.18287/2223-9537-2026-16-3-491-501

Domain-independent ontology design pattern for interpretable neuro-symbolic systems

© 2026, A.N. Zhdanova

Samara University (Samara National Research University named after academician S.P. Korolev), Samara, Russia

Abstract

Interpretable neuro-symbolic data analysis systems integrating knowledge graphs and graph neural networks are typically developed for specific application domains. This paper proposes a domain-independent ontology design pattern, "Analysis Object – Feature – Target Characteristic," which defines the structure of an interpretable neuro-symbolic system in which explanation generation is governed by the ontological organization of knowledge. The pattern includes concept classes, predicates, rules for transforming the ontology into a knowledge graph, and rules for constructing a graph neural network. Based on this pattern, a design method is developed that covers the transition from domain conceptualization to knowledge graph construction, model training, and the generation of semantic explanation paths. The proposed pattern integrates knowledge structure, inference mechanisms, and explanation generation within a unified conceptual framework. The application of the pattern is demonstrated in two domains: psychological profiling of social network users and dermoscopic diagnosis of skin lesions. Domain-invariant and domain-specific ontology elements are identified, and the ontological structure is mapped onto a knowledge graph. It is shown that the pattern structure, inference mechanisms, and explanation-generation procedures are preserved across various application domains. The interpretation of a decision is determined by the ontological structure of the neuro-symbolic system and is generated through semantic relationships between the analysis object, its features, and the target characteristic.

Keywords: *ontology design pattern, neuro-symbolic system, knowledge graph, graph neural networks, interpretability, explainable artificial intelligence.*

For citation: *Zhdanova A.N.* Domain-independent ontology design pattern for interpretable neuro-symbolic systems. [In Russian]. *Ontology of Designing*. 2026; 16(3): 491-501. DOI: 10.18287/2223-9537-2026-16-3-491-501.

Acknowledgment: The author expresses gratitude to A.V. Kupriyanov, Doctor of Technical Sciences, Associate Professor, for valuable discussions and providing valuable comments during the preparation of this paper.

Conflict of interest: The author declares no conflict of interest.

List of figures and table

References

[1] ***Zhdanova AN.*** Ontological Approach to Predicting Psychological Characteristics of Social Network Users Based on Knowledge Graphs [In Russian]. *Artificial Intelligence and Decision Making*, 2025; 3: 40–59. DOI: 10.14357/20718594250304.

[2] ***Studer R, Benjamins VR, Fensel D.*** Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering*. 1998; 25(1-2): 161-197.

[3] ***Hitzler P, Krötzsch M, Rudolph S.*** Foundations of Semantic Web Technologies. Boca Raton: CRC Press, 2009. 455 p.

[4] ***Gavrilova TA, Kudryavtsev DV, Muromtsev DI.*** Knowledge Engineering: Models and Methods [In Russian]. Saint Petersburg: Lan Publishing House, 2016. 324 p.

[5] ***Gangemi A.*** Ontology Design Patterns for Semantic Web Content. *The Semantic Web – ISWC 2005*. Berlin: Springer, 2005. P.262-276.

[6] ***Presutti V, Gangemi A.*** Pattern-Based Ontology Design. *Ontology Engineering in a Networked World*. Berlin: Springer, 2012. P.35-64. DOI: 10.1007/978-3-642-24794-1_3.

[7] ***Blomqvist E, Hammar K, Presutti V.*** Engineering Ontologies with Patterns – The eXtreme Design Methodology. *Ontology Engineering with Ontology Design Patterns*. Amsterdam: IOS Press, 2016. P.23-50. DOI: 10.3233/978-1-61499-676-7-23.

[8] ***Borgest NM.*** Scientific Basis of the Ontology of Designing [In Russian]. *Ontology of Designing*, 2013; 1(7): 7–25.

[9] ***Zagorulko YuA.*** Development of Scientific Knowledge Portals Based on Ontologies [In Russian]. *Computational Technologies*, 2007; 12(2): 169–177.

[10] ***d'Avila Garcez A, Lamb LC.*** Neurosymbolic AI: The 3rd Wave. *Artificial Intelligence Review*. 2023; 56: 12387-12406. DOI: 10.48550/arXiv.2012.05876.

[11] ***Hogan A, Blomqvist E, Cochez M.*** et al. Knowledge Graphs. *ACM Computing Surveys*. 2021; 54(4): Art.71. DOI: 10.1145/3447772.

[12] ***Veličković P, Cucurull G, Casanova A.*** *et al*. Graph Attention Networks. International Conference on Learning Representations (ICLR). 2018. 12 p. DOI: 10.48550/arXiv.1710.10903.

[13] ***Liu L, Wang Z, Tong H.*** Neural-Symbolic Reasoning over Knowledge Graphs: A Survey from a Query Perspective. *ACM SIGKDD Explorations Newsletter.* 2025. DOI: 10.48550/arXiv.2412.10390.

[14] ***Zheng K, Yu S, Chen B.*** CI-GNN: A Granger Causality-Inspired Graph Neural Network for Interpretable Brain Network-Based Psychiatric Diagnosis. *Neural Networks.* 2024; 172: Art.106147. arXiv:2301.01642v3 [stat.ML] 28 Jan 2024.

[15] ***Sengupta P, Rekik I.*** FireGNN: Neuro-Symbolic Graph Neural Networks with Trainable Fuzzy Rules for Interpretable Medical Image Classification. arXiv preprint arXiv:2509.10510. 2025.

[16] ***Raj K.*** A Neuro-symbolic Approach to Enhance Interpretability of Graph Neural Network through the Integration of External Knowledge. *Proc. of the 32nd ACM Int. Conf. on Information and Knowledge Management* (CIKM). 2023. DOI: 10.1145/3583780.3616008.

[17] ***Chang H, Ye J, Lopez Avila A, Du J, Li J.*** Path-based Explanation for Knowledge Graph Completion. *Proc. of the 30th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining* (KDD). 2024. DOI: 10.1145/3637528.3671683.

[18] ***Barredo Arrieta A***. et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Information Fusion*. 2020; 58: 82-115. DOI: 10.48550/arXiv.1910.10045.

[19] ***Nachbar F, Stolz W, Merkle T.*** et al. The ABCD Rule of Dermatoscopy: High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin Lesions. *Journal of the American Academy of Dermatology*. 1994; 30(4): 551-559.

[20] ***Tschandl P, Rosendahl C, Kittler H.*** The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Scientific Data*. 2018. Vol.5. Art.180161. DOI: 10.1038/sdata.2018.161.

About the author

Aleksandra Nikolaevna Zhdanova (b.1986) graduated from the S.P. Korolev Samara State Aerospace University in 2009, Candidate of Technical Sciences (2012). Associate Professor at the Department of Software Systems, Samara University. She is the author of more than 90 scientific publications in the fields of artificial intelligence and decision support systems. Author ID (RSCI): 762284; Scopus Author ID: 57076115300; Researcher ID (WoS): QMP-9127-2026; ORCID: 0009-0002-6370-1894. *aleksandra.zhd@yandex.ru*.

Received June 19, 2026. Revised July 19, 2026. Accepted July 31, 2026.