

Научная статья

УДК 004.85:004.89

DOI 10.25205/1818-7900-2025-23-3-44-56

Разработка прототипа информационно-аналитического интернет-ресурса по машинному обучению

Ирина Олеговна Плотникова

Новосибирский государственный университет

Новосибирск, Россия

i.plotnikova1@g.nsu.ru

Аннотация

В условиях стремительного развития технологий искусственного интеллекта и машинного обучения доступ к структурированной и понятной информации в данной области становится критически важен как для специалистов, так и для широкого круга пользователей. Существующие ресурсы часто страдают от фрагментарности данных, отсутствия систематизации и сложности навигации. Данная статья посвящена разработке прототипа интернет-ресурса [1], который бы решал эти проблемы и предоставлял пользователям доступ к обширной базе знаний по машинному обучению, включая описания алгоритмов, методов, инструментов, а также примеры их применения.

Ключевые слова

онтология, информационно-аналитический интернет-ресурс, машинное обучение, Protégé

Для цитирования

Плотникова И. О. Разработка прототипа информационно-аналитического интернет-ресурса по машинному обучению // Вестник НГУ. Серия: Информационные технологии. 2025. Т. 23, № 3. С. 44–56. DOI 10.25205/1818-7900-2025-23-3-44-56

Development of a prototype of an informational and analytical Internet resource on machine learning

Irina O. Plotnikova

Novosibirsk State University

Novosibirsk, Russian Federation

i.plotnikova1@g.nsu.ru

Abstract

With the rapid development of artificial intelligence and machine learning technologies, access to structured and understandable information in the field is becoming critical for both specialists and a wide range of users. Existing resources often suffer from data fragmentation, lack of systematization, and difficulty in navigation. This paper is devoted to the development of a prototype Internet resource [1] that would solve these problems and provide users with access to an extensive knowledge base on machine learning, including descriptions of algorithms, methods, tools, and examples of their application.

© Плотникова И. О., 2025

ISSN 1818-7900 (Print). ISSN 2410-0420 (Online)

Вестник НГУ. Серия: Информационные технологии. 2025. Том 23, № 3

Vestnik NSU. Series: Information Technologies, 2025, vol. 23, no. 3

Keywords

ontology, informational and analytical Internet resource, machine learning, Protégé

For citation

Plotnikova I. O. Development of a prototype of an informational and analytical Internet resource on machine learning. Vestnik NSU. Series: Information Technologies, 2025, vol. 23, no. 2, pp. 44–56 (in Russ.) DOI 10.25205/1818-7900-2025-23-3-44-56

Введение

Машинное обучение (МО) претерпевает стремительное развитие, оказывая все более значительное влияние на различные отрасли науки и техники. Его применение уже сегодня решает сложные задачи в областях медицины, финансов, автоматизированного управления и многих других, а перспективы дальнейшего развития обещают еще более масштабные преобразования [2]. Этот бурный рост, однако, сопровождается быстрым увеличением объема информации, связанной с новыми алгоритмами, библиотеками, наборами данных и результатами научных исследований.

В настоящее время существует множество информационных ресурсов, посвященных машинному обучению. Некоторые из них предлагают обширные базы данных научных статей, другие – доступ к открытым наборам данных, третьи – обучающие материалы и примеры кода. Однако многие из этих ресурсов имеют существенные недостатки: отсутствие единой системы классификации и поиска информации, недостаточная актуальность данных, сложная навигация и отсутствие инструментов для анализа информации. Это создает значительные трудности для исследователей, разработчиков и студентов, которые вынуждены тратить значительное время на поиск и обработку разрозненных данных из различных, часто несовместимых источников.

В связи с этим необходимость систематизации и интеграции информации в этой области является актуальной задачей.

Целью данного исследования является разработка прототипа информационно-аналитического интернет-ресурса, призванного улучшить доступ к информации по машинному обучению.

В рамках исследования планируется решение следующих задач: создание онтологии [3], отражающей основные понятия и связи между ними в сфере машинного обучения; разработка архитектуры информационного ресурса, основанного на данной онтологии; расширение функциональных возможностей ресурса путем интеграции рекомендательных систем.

Предполагается, что онтология будет динамически развиваться и пополняться по мере возникновения новых концепций и методологий в области машинного обучения.

В результате работы будет создан прототип, демонстрирующий возможность решения проблемы фрагментарности информации в области МО.

Теория онтологии и средства ее описания

Онтология – один из наиболее популярных способов формализации знаний [4]. Она представляет собой структурированную систему понятий (или классов) определенной предметной области, где сущности связаны различными отношениями.

С формальной точки зрения онтология представляется как $O = (C, R, A)$ [5], где C обозначает совокупность концептов (понятий), формирующих фундамент предметной области, R – это множество связей (отношений) между этими концептами, а A – набор аксиом, формализующих семантические и структурные характеристики понятий и отношений. Аксиоматическая составляющая играет важнейшую роль в поддержании логической целостности онтологии, давая возможность осуществлять логический вывод, опираясь на имеющуюся базу знаний.

В сущности, онтологию можно представить как сложную, многокомпонентную систему, которая интегрирует сущности (понятия и отношения) и утверждения (аксиомы и правила), регулирующие и конкретизирующие семантику сущностей в пределах определенной предметной области. Основное предназначение онтологии состоит в формировании строго определенного глоссария (словаря терминов) для отображения и передачи знаний о предметной области, а также в установлении взаимосвязей между этими терминами, что способствует недвусмысленному толкованию информации.

В наиболее строгом формальном представлении онтологии выступают как логические теории, базирующиеся на аксиомах – утверждениях, определяющих взаимосвязи и свойства понятий. Для конструирования и описания онтологий применяется разнообразный инструментарий, включающий логические формализмы, специализированные языки и программное обеспечение.

Среди наиболее востребованных подходов [3] выделяются **дескриптивные логики (ДЛ)** [6]. Это семейство логических систем, целенаправленно созданных для моделирования знаний о конкретной предметной области. Появление ДЛ было обусловлено недостаточной формализацией более ранних методов представления знаний, таких как семантические сети и фреймы, а также сложностями применения исчисления предикатов для описания специфических знаний. Дескриптивные логики предоставляют механизмы для строгого определения понятий (формирования терминологической базы) и последующего использования этих понятий для точной характеристики свойств объектов (индивидуумов, экземпляров) в рамках рассматриваемой предметной области.

Технологии Semantic Web. Технология Semantic Web предлагает набор инструментов для представления данных в машиночитаемом формате. Центральное место в этой технологии занимает RDF (Resource Description Framework), формализм для описания свойств ресурсов через триплеты «субъект – предикат – объект», записываемые в XML. RDF обеспечивает базовую структуру для представления знаний в вебе.

Помимо RDF, в семантической паутине (Semantic Web) существенную роль играет язык OWL (Web Ontology Language) [7], предоставляющий расширенные возможности для представления знаний и определения более выразительной семантики при построении онтологий. OWL представлен тремя версиями: OWL Full, OWL DL и OWL Lite. Каждая версия предлагает компромисс между выразительностью языка и вычислительной сложностью работы с ним. Запись OWL-онтологий осуществляется с использованием различных форматов представления, таких как RDF/XML, OWL/XML и Syntax. Protégé [8] – один из самых популярных редакторов онтологий, поддерживающий экспорт в различные форматы. Таким образом, Semantic Web предоставляет мощный инструментарий для формализации знаний, сочетающий выразительность с вычислительной разрешимостью, что делает его применимым на практике.

Формализация предметной области машинного обучения. Формализация осуществляется с применением технологий семантической паутины, в частности, языка OWL для описания онтологий. Однако использование OWL может вызывать трудности как у экспертов в машинном обучении, так и у специалистов по онтологическому инжинирингу ввиду его ориентации на представление относительно простых сущностей, тогда как онтологии машинного обучения зачастую требуют моделирования сложных понятий и взаимосвязей.

Для преодоления этой проблемы применяются шаблоны (паттерны) онтологического проектирования [3]. Так, для отображения бинарных отношений с атрибутами и доменов этих атрибутов разработаны специальные структурные шаблоны и введены вспомогательные классы типа «Отношение с атрибутами» и «Домен». Более того, для повышения эффективности процесса разработки онтологий используются паттерны содержания, которые предопределяют способы представления типичных фрагментов базовых онтологий, таких как «Виды машинного обучения», «Задачи машинного обучения», «Метрики» и т. д., включая соответствующие отношения, например, «включает в себя алгоритм», «включает в себя модель». Этот подход

не только расширяет выразительные возможности OWL, позволяя представлять сложные понятия и отношения, но и упрощает и ускоряет процесс разработки онтологий.

Основные понятия машинного обучения

Машинное обучение является подобластью искусственного интеллекта, фокусирующейся на разработке алгоритмов, обучающихся на данных без необходимости явного программирования. В отличие от традиционного подхода к созданию программ, основанного на жестких правилах, системы МО строят гибкие модели, позволяющие делать прогнозы и принимать решения, адаптируясь к новым данным. В рамках данной работы будут рассмотрены ключевые концепции и принципы МО.

В первую очередь, необходимо проанализировать существующие типы МО [9] и определить их ключевые различия.

Первой категорией является **обучение с учителем** (supervised learning). Эта методика подразумевает использование размеченных данных, где для каждого примера данных на вход есть соответствующие целевые показатели. Основная цель данного подхода – научить модель предсказывать выход для новых, невиданных ранее входных данных. Примеры задач включают классификацию (фильтрация нежелательных писем) и регрессию (прогнозирование стоимости жилья). Примеры алгоритмов: статистические методы, такие как наивный Байес, логистическая, линейная и полиномиальная регрессии, метод опорных векторов, k-ближайших соседей, дерево решений.

Второй тип МО – **обучение без учителя** (unsupervised learning) – предполагает анализ неразмеченных данных без заданного желаемого результата. Цель – обнаружение неочевидных взаимосвязей и закономерностей данных. В качестве примеров задач можно привести кластеризацию (например, сегментирование пользователей по поведенческим признакам) и понижение размерности (например, выделение главных компонент). Для решения подобных задач применяются такие алгоритмы, как итеративные методы типа k-средних, латентное размещение Дирихле, сингулярное разложение, DBSCAN, а также анализ главных компонент как метод сокращения данных;

Обучение с подкреплением (reinforcement learning) – интерактивный формат, где система обучается при взаимодействии со средой, получая положительную/отрицательную обратную связь. Целью является разработка стратегии, которая максимизирует общую награду. Примеры включают обучение автономных роботов двигательным навыкам. Примеры алгоритмов: SARSA, генетический алгоритм, Q-обучение, глубокие нейросетевые подходы вроде DDPG и DQN;

Обучение с частичным привлечением учителя (semi-supervised learning) – гибридный подход, совмещающий элементы контролируемого и неконтролируемого обучения, где используется небольшое количество размеченных данных и в несколько раз больше набор неразмеченных. Цель – построить более точную, обобщающую модель, чем та, которая была бы обучена только на размеченных данных. Примеры задач включают классификацию текстов, распознавание изображений с частичной разметкой, анализ медицинских данных и т. д.

Ансамблевое обучение основано на комбинировании предсказаний нескольких базовых классификаторов, обученных на одном и том же наборе данных, для повышения точности классификации новых объектов [10]. Рассмотрим основные методы ансамблевого обучения:

Бэггинг. В основе метода лежит генерация множества случайных подвыборок из исходной обучающей выборки с помощью выборки с возвращением. Несмотря на возможное дублирование элементов, агрегирование результатов, полученных на различных подвыборках, зачастую обеспечивает более высокую точность по сравнению с использованием единственной выбор-

ки. Эффективность бэггинга проявляется при использовании разнообразных и нестабильных классификаторов, чувствительных к изменениям в обучающих данных [11].

Бустинг. Этот метод представляет собой итеративный процесс построения композиции алгоритмов, где каждый последующий алгоритм корректирует ошибки предыдущих. Бустинг является жадным алгоритмом, изначально разработанным в контексте вероятностного почти корректного обучения. Основная идея заключается в возможности получения высокоточного классификатора путем комбинирования множества слабых алгоритмов [12].

Стэкинг. Данный метод также комбинирует несколько классификаторов, но, в отличие от бэггинга и бустинга, использует метаалгоритм. Стэкинг предполагает использование классификаторов различной природы и включает следующие шаги [13]:

1. Разделение обучающей выборки на два непересекающихся подмножества;
2. Обучение базовых классификаторов на первом подмножестве;
3. Тестирование базовых классификаторов на втором подмножестве;
4. Обучение метаалгоритма на основе предсказаний базовых классификаторов (входные данные) и истинных меток классов (выходные данные) для второго подмножества.

Нейронные сети – это вычислительные модели, вдохновленные структурой и функцией человеческого мозга. Они состоят из взаимосвязанных узлов, называемых нейронами, организованных в слои. Каждый нейрон получает входные данные, обрабатывает их с помощью функции активации и передает результат следующему слою. Сеть «обучается», корректируя веса связей между нейронами, чтобы минимизировать ошибку предсказаний на обучающих данных.

Глубокое обучение – это подмножество машинного обучения, использующее нейронные сети с множеством слоев (отсюда «глубокие»). Большое количество слоев позволяет сети извлекать сложные признаки и иерархии из данных, что делает ее эффективной для решения таких задач, как распознавание изображений, обработка естественного языка и другие. Ключевым фактором успеха глубокого обучения является наличие больших объемов данных и мощных вычислительных ресурсов.

Машинное обучение предлагает решения для разнообразных задач, варьирующихся в зависимости от целей анализа и типа данных. Выбор подходящего метода напрямую зависит от ожидаемого результата и характера входной информации. Основные типы задач машинного обучения включают следующие методы:

- **Регрессия.** Этот метод фокусируется на предсказании непрерывных значений, опираясь на набор входных признаков. Примерами задач регрессии могут служить прогнозирование цены квартиры на основе ее характеристик, оценка стоимости ценной бумаги с учетом рыночных факторов или прогнозирование ожидаемого дохода магазина, исходя из исторических данных и текущих трендов. Регрессионные модели строят функцию, аппроксимирующую зависимость целевой переменной от предикторов.
- **Классификация.** Классификация занимается отнесением объектов к одной из predeterminedных категорий. Результатом работы классификатора является категориальный ответ, часто представленный в бинарной форме (да/нет). Примеры задач классификации включают определение наличия кота на фотографии, распознавание лиц на изображениях или диагностику заболеваний на основе медицинских данных. Точность классификации зависит от качества данных и выбранного алгоритма.
- **Кластеризация.** Этот метод направлен на группировку данных на основе сходства между объектами. Кластеризация позволяет выявлять скрытые структуры в данных без предварительного определения категорий. Примеры применения кластеризации включают сегментацию клиентов мобильного оператора по уровню платежеспособности для персонализированных предложений или классификацию космических объектов по их характеристикам (планета, звезда, черная дыра и т. п.). Выбор метрики сходства играет ключевую роль в результатах кластеризации.

- **Понижение размерности.** Данный метод предназначен для сокращения количества признаков, описывающих данные, с сохранением при этом наиболее существенной информации. Понижение размерности полезно для сжатия данных, уменьшения вычислительной сложности моделей и устранения избыточности. Это может улучшить производительность алгоритмов машинного обучения и интерпретируемость результатов. Примерами методов понижения размерности являются метод главных компонент (PCA) и t-SNE.
- **Выявление аномалий.** Эта задача фокусируется на идентификации объектов, которые отклоняются от нормы или ожидаемого поведения. В отличие от классификации, аномалии, как правило, являются редкими событиями, и обучающих примеров для их обнаружения может быть недостаточно. Это требует специализированных алгоритмов, способных работать с несбалансированными данными. Типичным примером задачи выявления аномалий является обнаружение мошеннических операций с банковскими картами или выявление дефектов на производстве.

Разработка онтологии

Разработка онтологии началась с определения ее предметной области. Вместо стремления к созданию всеобъемлющей онтологии машинного обучения, было принято решение сфокусироваться на ключевых и наиболее часто используемых концепциях. Такой подход позволил разработать более компактную, прагматичную и целенаправленную онтологию, непосредственно применимую к решению конкретных задач, избегая излишней сложности и абстракции. Это решение обусловлено практической необходимостью создания инструмента, готового к использованию в реальных проектах, а не теоретической конструкции, охватывающей все возможные аспекты, но сложной в применении.

В качестве языка представления знаний был выбран OWL (Web Ontology Language). Этот выбор обоснован выразительностью OWL, позволяющей моделировать сложные взаимосвязи между концепциями, а также его широкой поддержкой различными инструментами и технологиями Semantic Web. Совместимость с существующими инструментами упрощает интеграцию онтологии в различные приложения. Разработка и редактирование онтологии производились в среде Protégé, которая предлагает удобный графический интерфейс и богатый набор функций для моделирования знаний, включая поддержку OWL, визуализацию онтологии и проверку ее на корректность.

Онтология включает 11 основных классов, каждый из которых детально описывает определенный, но взаимосвязанный аспект машинного обучения. Эти классы, вместе с определенными между ними отношениями, формируют фундамент онтологии и обеспечивают ее способность представлять знания о машинном обучении в структурированном и логически связанном виде. Взаимодействие этих классов позволяет моделировать различные сценарии и процессы машинного обучения. Рассмотрим их подробнее.

- **Модели.** Этот класс описывает различные модели машинного обучения, такие как линейные модели, деревья решений, нейронные сети (с указанием типов архитектур, например, CNN, RNN) и др. Включает в себя атрибуты, характеризующие модель: параметры, точность, сложность и т. д.
- **Публикации.** Класс содержит информацию о научных публикациях, связанных с машинными методами. Атрибуты включают библиографические данные, аннотацию, ключевые слова и ссылки на соответствующие модели, алгоритмы или наборы данных.
- **Алгоритмы.** Описывает конкретные алгоритмы, используемые для построения моделей машинного обучения. Для каждого алгоритма указываются его параметры, свойства

и особенности. Связь с классами «Модели» и «Задачи» определяет использование алгоритма для построения конкретных моделей и решения конкретных задач.

- **Виды МО.** Класс систематизирует различные подходы к машинному обучению, такие как контролируемое, неконтролируемое, полуконтролируемое, обучение с подкреплением [14; 15] и т. д. Этот класс обеспечивает иерархическую классификацию алгоритмов и моделей.
- **Датасеты.** Этот класс описывает наборы данных, играющие ключевую роль в обучении и тестировании моделей. Он содержит важную информацию о характеристиках набора данных, включая их размер, тип данных, источник происхождения, а также другие параметры, необходимые для анализа и сравнения результатов.
- **Задачи.** Определяет спектр задач машинного обучения, таких как классификация, регрессия, кластеризация и другие [16]. Он устанавливает важную связь с классами «Алгоритмы» и «Модели», определяя, какие алгоритмы и модели подходят для решения конкретных задач.
- **Метрики.** Описывает метрики, используемые для оценки качества и эффективности моделей машинного обучения. Включает в себя такие распространенные метрики, как точность, полнота, F1-мера, AUC-ROC и др., предоставляя информацию об их применении и интерпретации в контексте различных задач.
- **Обученные модели.** Этот класс представляет собой конкретные экземпляры моделей машинного обучения, полученные после процесса обучения на определенных наборах данных с использованием выбранных алгоритмов. Он хранит информацию о параметрах обученной модели, ее производительности на тестовых данных, а также ссылки на использованные наборы данных и алгоритмы.
- **Персоны.** Включает информацию об исследователях и разработчиках, внесших вклад в область машинного обучения. Атрибуты класса включают имя, аффилиацию и список публикаций.
- **Тезаурус.** Класс содержит структурированный словарь терминов, связанных с машинным обучением, обеспечивая единообразие терминологии и улучшая поиск информации.
- **Фреймворки.** Описывает различные программные инструментарии и библиотеки, используемые для разработки и реализации моделей машинного обучения (например, TensorFlow, PyTorch, Scikit-learn). Включает информацию о функциональности и возможностях каждого инструментария.

Взаимосвязи между этими классами позволяют создавать сложные запросы и получать интегрированную информацию о различных аспектах машинного обучения.

Особое внимание уделялось моделированию отношений между классами. Вот некоторые из них:

- «*используется для решения задач*» связывает класс «Алгоритмы» с классом «Задачи», указывая, какие алгоритмы применяются для решения конкретных задач;
- «*включает в себя алгоритм*» связывает конкретный алгоритм с классом «Виды МО» (например, «логистическая регрессия» является типом «обучения с учителем»);
- «*обучена на*» связывает класс «Модели» с классом «Датасеты», отражая процесс обучения моделей;
- «*автор*» связывает «Публикации» с «Персонами» (персона является автором публикации);
- «*определяет*» связывает «Тезаурус» со всеми остальными классами (тезаурус определяет терминологию для каждого класса);
- «*использует датасет*» связывает «Алгоритмы» с «Датасетами» (алгоритм использует датасет для обучения);
- «*использует алгоритм*» связывает «Модели» с «Алгоритмами» (модель использует алгоритм) и т. д.

Эти отношения позволяют не только структурировать информацию, но и выводить новые знания на основе существующих.

Разработанная онтология является гибкой и может быть расширена за счет добавления новых классов и свойств, что делает ее ценным инструментом для организации и анализа знаний в этой динамично развивающейся области.

Пример онтологии представлен на рис. 1.

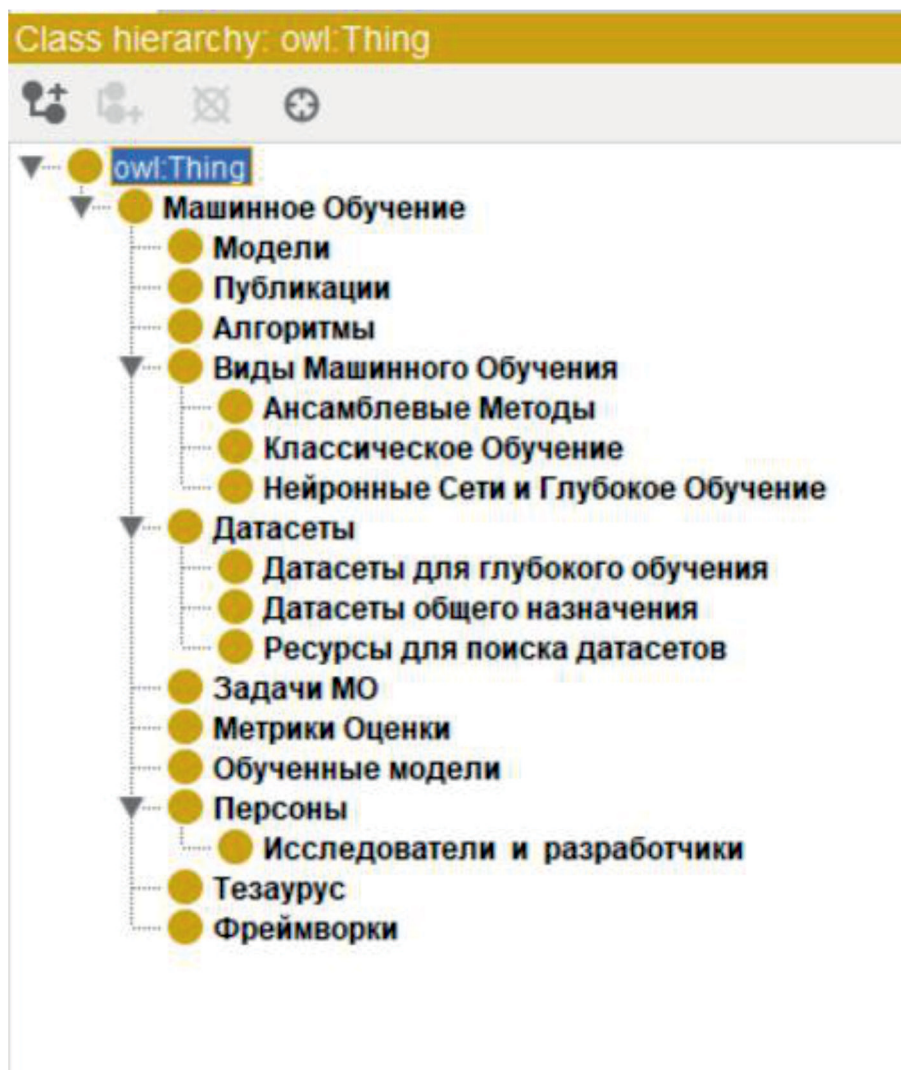


Рис. 1. Верхний уровень онтологии машинного обучения

Fig. 1. The top level of the machine learning ontology

В рамках расширения и обогащения онтологии была проведена масштабная работа по ее семантическому заполнению. Для каждого концепта, включенного в онтологию, было разработано подробное определение, отражающее его сущность и взаимосвязи с другими понятиями. Эти определения были тщательно проработаны с целью обеспечить точность и непротиворечивость представленной информации, а также исключить неоднозначные интерпретации. В процессе разработки определений использовались как общепринятые в научной литературе термины, так и специально подобранные описания, обеспечивающие доступность информации для широкого круга пользователей, включая как специалистов в области машинного об-

учения, так и людей, только начинающих изучать эту область. На рис. 2 представлен пример описания одного из понятий.

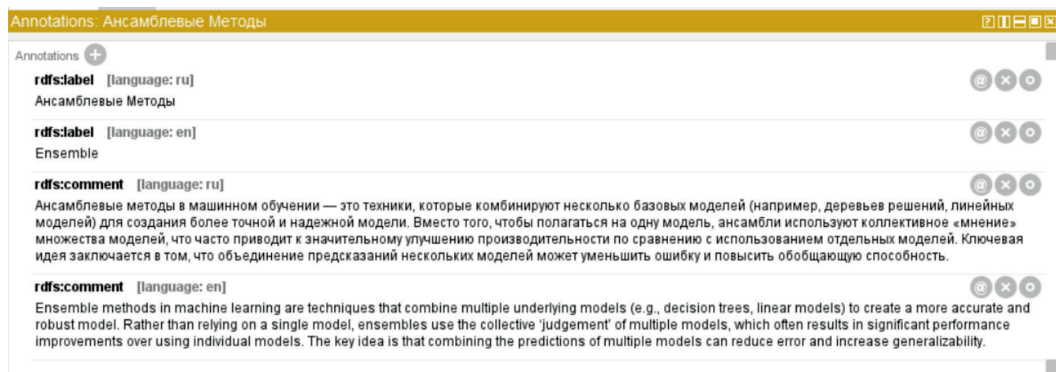


Рис. 2. Описание «Ансамблевые методы»
Fig. 2. The description of “Ensemble Methods”

Для обеспечения международной доступности и расширения аудитории ресурса вся онтология была переведена на английский язык. Процесс перевода включал не только прямой перевод определений, но и верификацию терминологической согласованности с английской научной литературой в области машинного обучения. Это гарантирует высокое качество перевода и исключает возможность неточностей и несоответствий в переведенной версии онтологии. Таким образом, созданный ресурс обеспечивает двуязычность (русский и английский языки), что значительно расширяет его потенциальную аудиторию и способствует международному обмену знаниями в области машинного обучения.

Разработка информационно-аналитического интернет-ресурса

Разработанная онтология машинного обучения, созданная в среде Protégé, изначально проектировалась с учетом требований интеграции в информационно-аналитический интернет-ресурс. Это обеспечило совместимость онтологии с целевой веб-системой, исключив необходимость последующих преобразований формата. В процессе разработки был проведен тщательный анализ требований к представлению онтологий в веб-среде, учитывающий особенности целевой информационной системы, а также критерии производительности и масштабируемости.

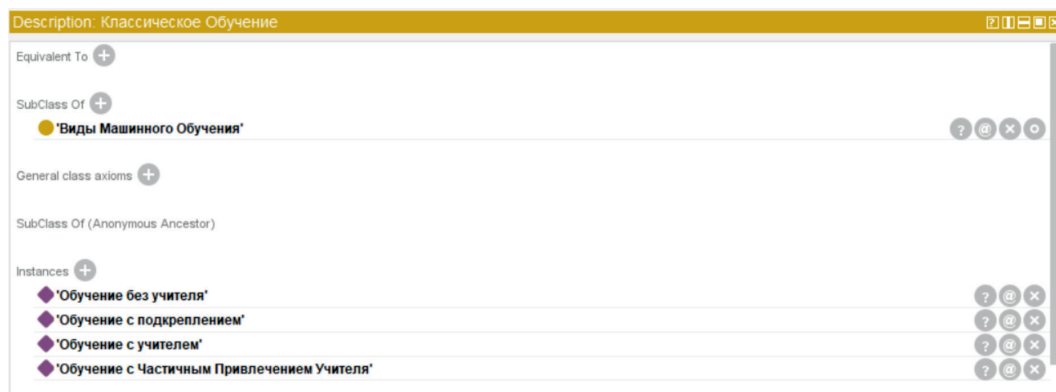


Рис. 3. Описание одного из видов МО
Fig. 3. Description of one of the types of ML

Проектирование онтологии в Protégé проводилось с акцентом на семантическую непротиворечивость и целостность знаний. Для обеспечения качества и минимизации ошибок была использована методология разработки, включающая поэтапный контроль соответствия целевому формату и многократную проверку на наличие ошибок и несоответствий. Примеры описания понятий онтологии представлены на рис. 3 и 4.

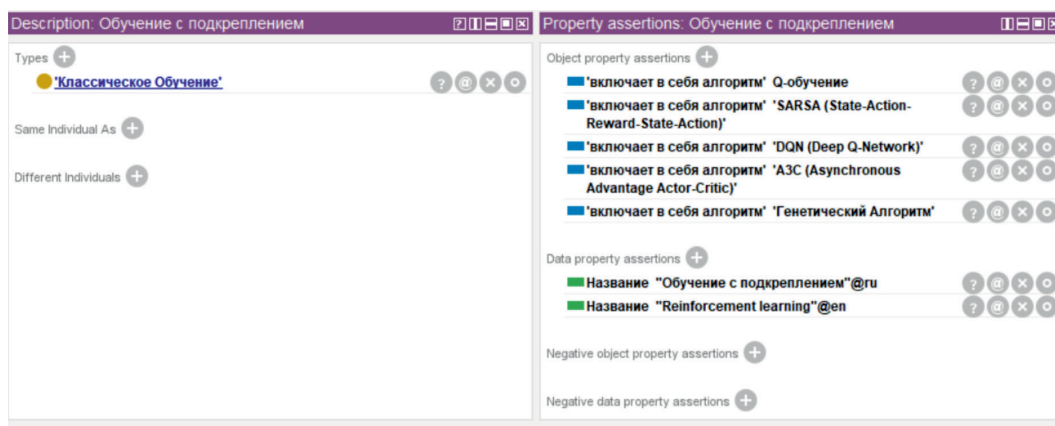


Рис. 4. Описание обучения с подкреплением

Fig. 4. Description of reinforcement learning

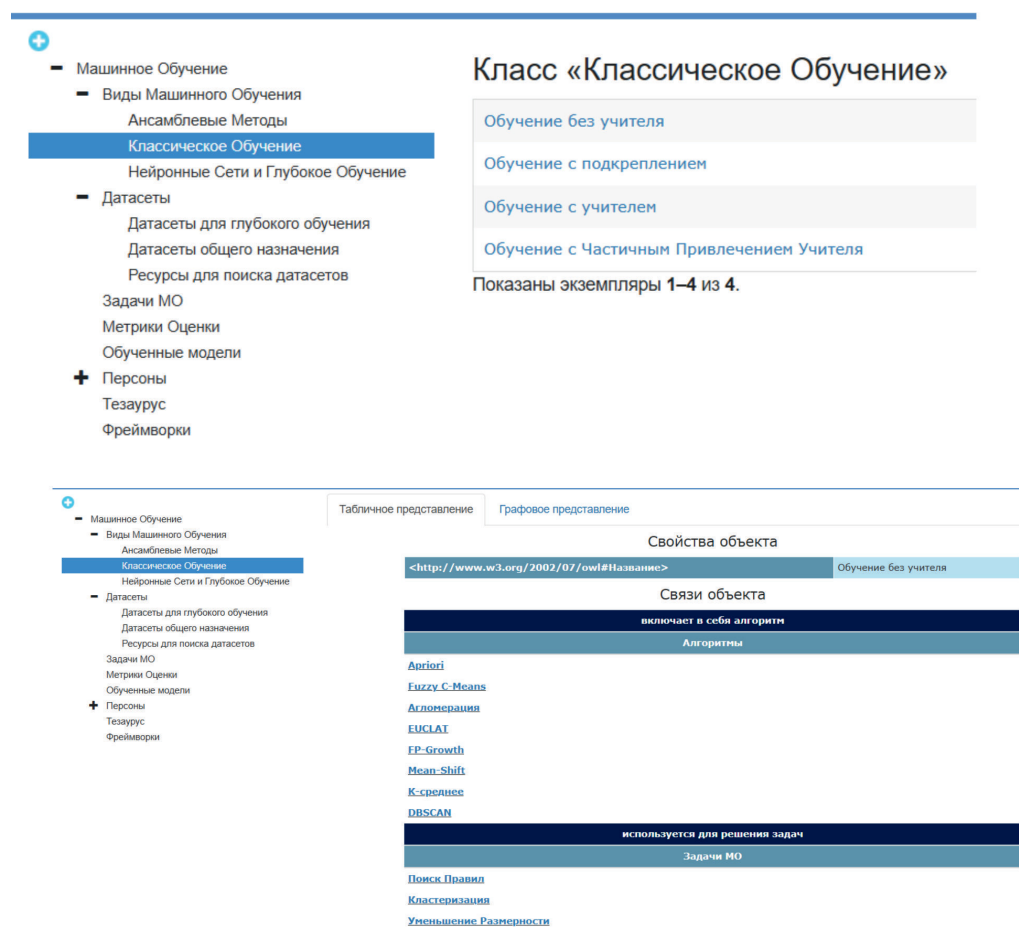


Рис. 5. Прототип информационно-аналитического интернет-ресурса

Fig. 5. Prototype of an informational and analytical Internet resource

Успешная интеграция разработанной онтологии машинного обучения в аналитический интернет-ресурс позволила создать полностью функционирующий прототип информационной системы. Этот прототип демонстрирует значительный потенциал для предоставления пользователям удобного и эффективного доступа к структурированной информации об этой быстро развивающейся области. Интеграция обеспечила не только хранение и поиск данных, но и возможность установления сложных семантических связей между различными концептами МО, такими как алгоритмы, модели, наборы данных и метрики. Благодаря структурированному представлению знаний, пользователи могут проводить углубленный анализ, легко находить необходимую информацию и эффективно изучать взаимосвязи между различными аспектами машинного обучения. Развертывание прототипа подтверждает практическую ценность разработанной онтологии и ее способность служить надежной основой для построения масштабируемых и интуитивно понятных информационных систем в области машинного обучения.

Дальнейшее развитие прототипа направлено на расширение его функциональности и интеграцию с другими информационными ресурсами для создания еще более полной и удобной экосистемы для исследователей и практиков машинного обучения.

Разработанный прототип был подвергнут всестороннему тестированию. Результаты тестирования подтвердили успешность преобразования онтологии и работоспособность созданного прототипа интернет-ресурса (рис. 5).

Заключение

Разработка информационно-аналитического интернет-ресурса (ИАИР), базирующегося на методах машинного обучения, представляет собой актуальную научную задачу, решение которой способствует углублению понимания и интерпретации результатов, получаемых в процессе применения алгоритмов машинного обучения. Формализацию предметной области обеспечивает разработанная онтология, структурирующая знания о машинном обучении посредством создания системы классификации концепций, методов и ресурсов. Интеграция онтологии в ИАИР повышает его эргономичность и доступность для пользователей, предоставляя удобный инструмент для работы с информацией. Полученные результаты могут быть применены в будущих исследованиях и разработках в сфере машинного обучения. Дальнейшая работа заключается в разработке сервиса, реализующего сценарий подбора представленных в контенте прототипа ИАИР МО инструментальных средств и информационных ресурсов, необходимых для решения заданной задачи методами машинного обучения.

Список литературы

1. Загорюлько Ю. А., Загорюлько Г. Б., Боровикова О. И. Технология создания тематических интеллектуальных научных интернет-ресурсов, базирующаяся на онтологии // Программная инженерия, 2016. Т. 7. № 2. С. 51–60.
2. Шумский С. А. Машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта. М.: РИОР, 2019. 340 с.
3. Загорюлько Ю. А. Обзор современных средств формализации семантики областей знаний на основе онтологий // Информационные и математические технологии в науке и управлении. 2018. № 3 (11). С. 27–36. DOI 10.25729/2413-0133-2018-3-03
4. Лукашевич Н. В., Добров Б. В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования. 2015. № 1 (15). С. 47–69.
5. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем: учебник. Питер: СПб., 2000. 384 с.

6. **Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F.** The Description Logic Handbook: Theory, Implementation, Applications. Cambridge, 2003. 574 p.
7. **Antoniou G., Harmelen F.** Web Ontology Language: OWL // Handbook on Ontologies. Berlin: Springer Verlag, 2004. P. 67–92.
8. Protégé. A free, open-source ontology editor and framework for building intelligent systems. URL: <http://protege.stanford.edu/>
9. **Sighn A.** Machine Learning with Python, 2019. С. 7–10.
10. **Izenman A. J.** Modern Multivariate Statistical Techniques, Springer Texts in Statistics. Springer Science+Business Media New York, 2013.
11. **Breiman L.** Bagging predictors // Machine Learning. 1996. 24. P. 123–140.
12. **Freund Y., Schapire R. E.** A Short Introduction to Boosting // Journal of Japanese Society for Artificial Intelligence. 1999. Vol. 14 (5). P. 771–780 .
13. **Wolpert D. H.** Stacked Generalization // Neural Networks. 1992. Vol. 5. P. 241–259.
14. **Goodfellow I., Bengio Y., Courville A.** Deep Learning. MIT Press, 2016.
15. **Mitchell T. M.** Machine Learning. McGraw-Hill Education, 1997.
16. **Флах П.** Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. М.: ДМК Пресс, 2023. 401 с.

References

1. **Zagorulko Yu. A., Zagorulko G. B., Borovikova O. I.** Technology for building subject based intelligent scientific internet resources based on ontology. *Software Engineering*, 2016, vol. 7, no. 2, pp. 51–60. (in Russ.)
2. **Shumskiy S. A.** Mashinnyy intellekt. Ocherki po teorii mashinnogo obucheniya i iskusstvennogo intellekta [Machine Intelligence. Essays on the Theory of Machine Learning and Artificial Intelligence]. Moscow, RIOR, 2019, 340 p. (in Russ.)
3. **Zagorulko Yu. A.** Obzor sovremennykh sredstv formalizatsii semantiki oblastey znaniy na osnove ontologiy [Modern means for formalizing the semantics of knowledge areas on the basis of ontologies]. *Informatsionnye i matematicheskiye tekhnologii v nauke i upravlenii [Information and Mathematical Technologies in Science and Management]*, 2018, no. 3 (11), pp. 27–36. DOI: 10.25729/2413-0133-2018-3-03 (in Russ.)
4. **Lukashevich N. V., Dobrov B. V.** Proyektirovaniye lingvisticheskikh ontologiy dlya informatsionnykh sistem v shirokikh predmetnykh oblastyakh [Developing linguistic ontologies in broad domains]. *Ontologiya proyektirovaniya [Ontology of Design]*, 2015, no. 1 (15), pp. 47–69. (in Russ.)
5. **Gavrilova T. A., Horoshevskiy V. F.** Bazy znaniy intellektualnykh sistem. Uchebnik [Knowledge Bases of Intelligent Systems. Textbook]. St. Petersburg, Piter, 2000, 384 p. (in Russ.)
6. **Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F.** The Description Logic Handbook: Theory, Implementation, Applications. Cambridge, 2003. 574 p.
7. **Antoniou G., Harmelen F.** Web Ontology Language: OWL. *Handbook on Ontologies. Berlin: Springer Verlag*, 2004. pp. 67–92.
8. Protégé. A free, open-source ontology editor and framework for building intelligent systems. URL: <http://protege.stanford.edu/>
9. **Sighn A.** Machine Learning with Python, 2019, pp. 7–10.
10. **Izenman A. J.** Modern Multivariate Statistical Techniques, Springer Texts in Statistics. Springer Science+Business Media New York, 2013.
11. **Breiman L.** Bagging predictors. Machine Learning, 1996, 24, pp. 123–140.
12. **Freund Y., Schapire R. E.** A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 1999, no. 14 (5), pp. 771–780.

13. **Wolpert D. H.** Stacked Generalization. *Neural Networks*, 1992, no. 5, pp. 241–259.
14. **Goodfellow I., Bengio Y., Courville A.** Deep Learning. MIT Press, 2016.
15. **Mitchell T. M.** Machine Learning. McGraw-Hill Education, 1997.
16. **Flakh P.** Mashinnoye obucheniye. Nauka i iskusstvo postroyeniya algoritmov, kotoryye izvlekayut znaniya iz dannykh [Machine learning. The science and art of building algorithms that extract knowledge from data]. Moscow, DMK Press, 2023. 401 p. (in Russ.)

Информация об авторе

Плотникова Ирина Олеговна, магистрант

Information about the Author

Irina O. Plotnikova, Master Student

*Статья поступила в редакцию 31.03.2025;
одобрена после рецензирования 29.04.2025; принята к публикации 29.04.2025*
*The article was submitted 31.03.2025;
approved after reviewing 29.04.2025; accepted for publication 29.04.2025*