

ОГЛАВЛЕНИЕ

- М. В. Бобырь, Н. И. Храпова**
ДВУХУРОВНЕВАЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА УПРАВЛЕНИЯ
ИНТЕЛЛЕКТУАЛЬНЫМ СВЕТОФОРОМ 696–717
- А. К. Журавлёв, К. А. Григорян**
АВТОМАТИЧЕСКАЯ РАЗМЕТКА ОБУЧАЮЩИХ ВЫБОРОК В КОМПЬЮТЕРНОМ
ЗРЕНИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ 718–729
- Т. Ф. Ибрагимов, А. А. Ференец**
АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ HTML-ДОКУМЕНТОВ ПО СТАНДАРТУ
MICRODATA 730–744
- В. В. Наумова, А. А. Загумённых, В. С. Ерёмченко**
ЧИСЛОВЫЕ ИНСТРУМЕНТЫ И ВИРТУАЛЬНЫЕ АССИСТЕНТЫ ДЛЯ ПОДДЕРЖКИ
НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ 745–757
- М. И. Патук, В. В. Наумова**
ИСПОЛЬЗОВАНИЕ СЕМАНТИЧЕСКОГО ПОИСКА ДЛЯ ВЫБОРА И РАНЖИРОВАНИЯ
НАУЧНЫХ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ 758–773
- А. В. Шубин, В. В. Кугуракова**
ПРОЕКТИРОВАНИЕ ИНСТРУМЕНТАРИЯ ДЛЯ СОЗДАНИЯ ИГРОВОГО ПРОЦЕССА
ЧЕРЕЗ СИСТЕМАТИЗАЦИЮ ИГРОВЫХ МЕХАНИК 774–795

ДВУХУРОВНЕВАЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМ СВЕТОФОРОМ

М. В. Бобырь¹ [0000-0002-5400-6817], **Н. И. Храпова**³ [0000-0001-7947-1427]

^{1,2}*Юго-Западный государственный университет, Курская обл., Курск, 305040
Россия*

¹maxbobyrg@gmail.com, ²KhrapovaNI@yandex.ru

Аннотация

В современном мире проблемы, возникающие в сфере дорожного движения, имеют большую значимость. С целью решения существующих задач разрабатываются различные интеллектуальные системы, одной из которых является система «Умный город». Данная работа посвящена разработке информационно-аналитической системы (ИАС) для управления интеллектуальным светофором. Представленная система состоит из двух уровней, каждый из которых реализует набор определенных операций. Первый уровень отвечает за обнаружение объектов, в частности, пешеходов и автомобилей, находящихся на перекрестке, а второй уровень осуществляет расчёт времени работы сигналов светофора для управляющего сигнала, который передаётся на устройство. Для сравнительного анализа выбран комбинационный метод (HOG+SVM) Histogram of Oriented Gradients, основанный на подсчёте числа направлений градиента на отдельных областях изображения и Support Vector Machines, с помощью которого строятся гиперплоскости в n -мерном пространстве с целью разделения объектов, относящихся к разным классам. Результаты экспериментального исследования, в ходе которого проводилось распознавание объектов на изображениях, показали превосходство разработанной информационно-аналитической системы над существующими. Среднее значение точности выявления пешеходов и автомобилей посредством ИАС составило 69,4%. Кроме того, по результатам проведенного эксперимента сделан вывод, что точность выявления объектов на изображениях прямо пропорциональна расстоянию от видеокамеры до объекта.

Ключевые слова: интеллектуальный светофор, детектирование объектов, машинное обучение, нечётко-логический метод детектирования границ, YOLO, HOG, SVM.

ВВЕДЕНИЕ

Интеллектуальные системы для управления городским потоком имеют значительную актуальность в современном мире [1]. Задача регулирования транспортного и пешеходного потоков возникает из-за изменяющейся интенсивности дорожного движения в зависимости от времени суток, так как в «часы-пик» (утреннее и вечернее время) наблюдается рост загруженности перекрестков, в отличие от ночного времени, когда улицы практически пусты. Подобные системы позволяют повысить безопасность дорожного движения за счёт решения основных транспортных проблем, таких как высокая загруженность улично-дорожной сети, дорожно-транспортные происшествия и регулирование средней скорости городского потока [2].

В настоящей работе рассмотрено управление городским потоком с помощью системы «Умный город», которая содержит информационно-аналитическую систему, состоящую из двух уровней: детекции числа участников дорожного движения (автомобилей и пешеходов) и расчета времени задержки сигналов светофора.

Для распознавания автомобилей и пешеходов использована нейронная сеть YOLO [3, 4]. Для передачи ей информации необходимо выполнить предобработку изображения, которая включает следующие операции: преобразование в градации серого [5], размытие изображения по Гауссу, детектирование границ и построение контуров [6]. Ранее для построения контуров объектов на изображениях нами был использован алгоритм Канни [7], который имеет недостатки, из-за которых снижается точность распознавания. В данной работе с целью повышения точности распознавания объектов предложено использовать нечётко-логический метод детектирования границ.

С целью проведения сравнительного анализа разработанной системы был выбран комбинационный метод HOG+SVM (Histogram of Oriented Gradients +

Support Vector Machines) [8]. Техника метода HOG [9, 10] (гистограмма направленных градиентов) основана на определении числа направлений градиента в отдельных областях изображений и состоит из пяти шагов: нахождение градиента интенсивности для каждого пикселя; разбиение изображения на ячейки; построение гистограммы, которая указывает направление градиента в каждой ячейке; группировка гистограмм в блоки; нормализация полученных блоков с целью снижения воздействия изменений контраста и освещенности. Недостатками HOG являются вычислительная сложность и чувствительность к шуму на изображении.

SVM [11, 12] (метод опорных векторов) используется в задачах классификации и регрессии и основан на поиске оптимальной гиперплоскости, разделяющей данные на классы. Метод состоит из двух основных этапов: обучение алгоритма на наборе данных, когда каждая точка относится к определенному классу; предсказание класса новой точки данных на основании её расположения относительно гиперплоскости. Недостатками SVM являются чувствительность к шуму и масштабированию данных, ресурсоёмкость, а также сложность выделения наиболее важных признаков для классификации объектов.

МЕТОДЫ И МАТЕРИАЛЫ

Была разработана двухуровневая информационно-аналитическая система управления интеллектуальным светофором. Первый уровень ИАС состоит из семи операций:

1. Получение RGB-изображения с видеокамеры, установленной на перекрестке.
2. Преобразование полученного изображения из RGB в градации серого по формуле

$$I = \frac{R + G + B}{3},$$

где I – интенсивность цвета каждого пикселя, R , G , B – значения интенсивности красного, зелёного и синего цветов в пикселе.

3. Применение операции сглаживания по Гауссу [13] с целью уменьшения влияния шума на изображение по формуле

$$K_{x,y} = \sum_{x=-2}^2 \sum_{y=-2}^2 \frac{1}{b} G_m \cdot I_{x,y},$$

где $K_{x,y}$ – значения интенсивности каждого пикселя; b – коэффициент нормировки, равный сумме элементов матрицы G ; G_m – матрица Гаусса.

4. Детектирование границ объектов с помощью нечётко-логического метода. Данный метод содержит несколько шагов [14]:

4.1. Вычисление значений градиента яркости изображения происходит с помощью свёртки изображения матрицами:

- в вертикальном направлении

$$GX_{x,y} = \sum_{y=-1}^1 \sum_{x=-1}^1 Ver \cdot K_{x,y};$$

- в горизонтальном направлении

$$GY_{x,y} = \sum_{y=-1}^1 \sum_{x=-1}^1 Gor \cdot K_{x,y}.$$

Далее вычисляется градиент для каждого пикселя с помощью формулы

$$G_{x,y} = \sqrt{GX_{x,y}^2 + GY_{x,y}^2},$$

4.2. Расчёт угла направления градиента по формуле

$$\Theta_{x,y} = \text{round} \left(\text{atan} \left(\frac{GY_{x,y}}{GX_{x,y}} \right) \right),$$

где round – функция для округления результата до целого числа; atan – функция для определения арктангенса отношения градиентов.

4.3. Фаззификация входных и выходных переменных происходит путём вычисления разности градиентов соседних ячеек относительно центральной:

$$\Delta G_i = G - G_i,$$

где i – номер рассматриваемой ячейки.

4.4 Формирование базы нечётких правил представлено в таблице 1.

Таблица 1 – База нечетких правил

Rule	$\Theta_{x,y}$	$\mu(\Delta G1)$	$\mu(\Delta G2)$	$\mu(\Delta G3)$	$\mu(\Delta G4)$	$\mu(\Delta G5)$	$\mu(\Delta G6)$	$\mu(\Delta G7)$	$\mu(\Delta G8)$	Out
R1	0°		Low					High		Edge
R2			High					Low		Edge
R3			High					High		Edge
R1	45°	Low							High	Edge
R2		High							Low	Edge
R3		High							High	Edge
R1	90°				Low	High				Edge
R2					High	Low				Edge
R3					High	High				Edge
R1	135°			Low			High			Edge
R2				High			Low			Edge
R3				High			High			Edge

Далее с помощью операции нечёткого минимума определяется значение степеней нечётких правил [15]:

$$R_i = \min[\mu(\Delta G_i), \mu(\Delta G_i)].$$

4.5 Дефаззификация чёткого значения осуществляется посредством нечётного α -среза, после чего происходит определение границ объектов на изображении.

5 Выделение контуров объектов на изображении производится с помощью алгоритма Suzuki-Abe [16].

6 Распознавание классов объектов происходит с помощью алгоритма YOLO.

7 Блок принятия решений, выходными данными которого является информация о количестве автомобилей и пешеходов, распознанных на входном изображении.

Выходные переменные с первого уровня передаются на второй уровень информационно-аналитической системы. Второй уровень состоит из пяти операций:

1. Определение входных переменных, которыми являются значения количества автомобилей и пешеходов, полученные на первом уровне ИАС.

2. Фаззификация значений входных переменных включает в себя построение функций принадлежности (ФП) и расчёт их степеней с помощью формулы

$$\mu(x) = \begin{cases} 0, & \text{если } x \notin [a;c]; \\ \frac{x-a}{b-a}, & \text{если } x \in [a;b]; \\ \frac{c-x}{c-b}, & \text{если } x \in [b;c]; \end{cases}$$

где $\mu(x)$ – характеристическая ФП; a, b, c – вершины треугольника ФП.

3. Выборка решений на основе лингвистических переменных и базы нечетких правил:

3.1. Вычисление степеней истинности предпосылок нечётких правил выполняется с помощью формул

$$R_1 = \vartheta_t(A_1; B_1);$$

$$R_2 = \vartheta_t(A_1; B_2);$$

$$R_3 = \vartheta_t(A_1; B_3);$$

$$R_4 = \vartheta_t(A_2; B_1);$$

$$R_5 = \vartheta_t(A_2; B_2);$$

$$R_6 = \vartheta_t(A_2; B_3);$$

$$R_7 = \vartheta_t(A_3; B_1);$$

$$R_8 = \vartheta_t(A_3; B_2);$$

$$R_9 = \vartheta_t(A_3; B_3);$$

где ϑ_t – знак, обозначающий операцию нечеткой импликации с использованием одной из t -норм [17].

3.2. Усечение ФП в зависимости от нечетких правил происходит по формулам

$$\begin{aligned} Y_5 &= R_3; \\ Y_4 &= \theta_s(R_2; R_6); \\ Y_3 &= \theta_s(R_1; \max(R_5; R_9)); \\ Y_2 &= \theta_s(R_8; R_4); \\ Y_1 &= R_7, \end{aligned}$$

где θ_s – знак, обозначающий операцию нечеткой импликации с использованием одной из s -норм.

4. Дефаззификация полученных значений [18] выполняется по формуле

$$d_{sum} = \frac{\sum_{i=1}^{11} M_i Y_i}{\sum_{i=1}^{11} Y_i},$$

где M_i – метки выходной ФП, которые задаются синглтонной ФП.

5. Определение выходной переменной, которой является время для регулирования работы сигналов светофора.

Расчёт данных операций и экспериментальные исследования по моделированию системы управления интеллектуальным светофором представлены в статье [19]. Также была разработана и запатентована специализированная программная модель, реализующая регулирование светофора на основе нечеткой логики [20].

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Разработано специализированное программное обеспечение «Программа детектирования объектов на пешеходном переходе и определения времени задержки управляющих сигналов светофора» [21]. Для экспериментальных исследований взяты изображения с видеокамеры, установленной на пересечении ул. Карла Маркса и ул. Хуторской города Курска (рисунок 1).

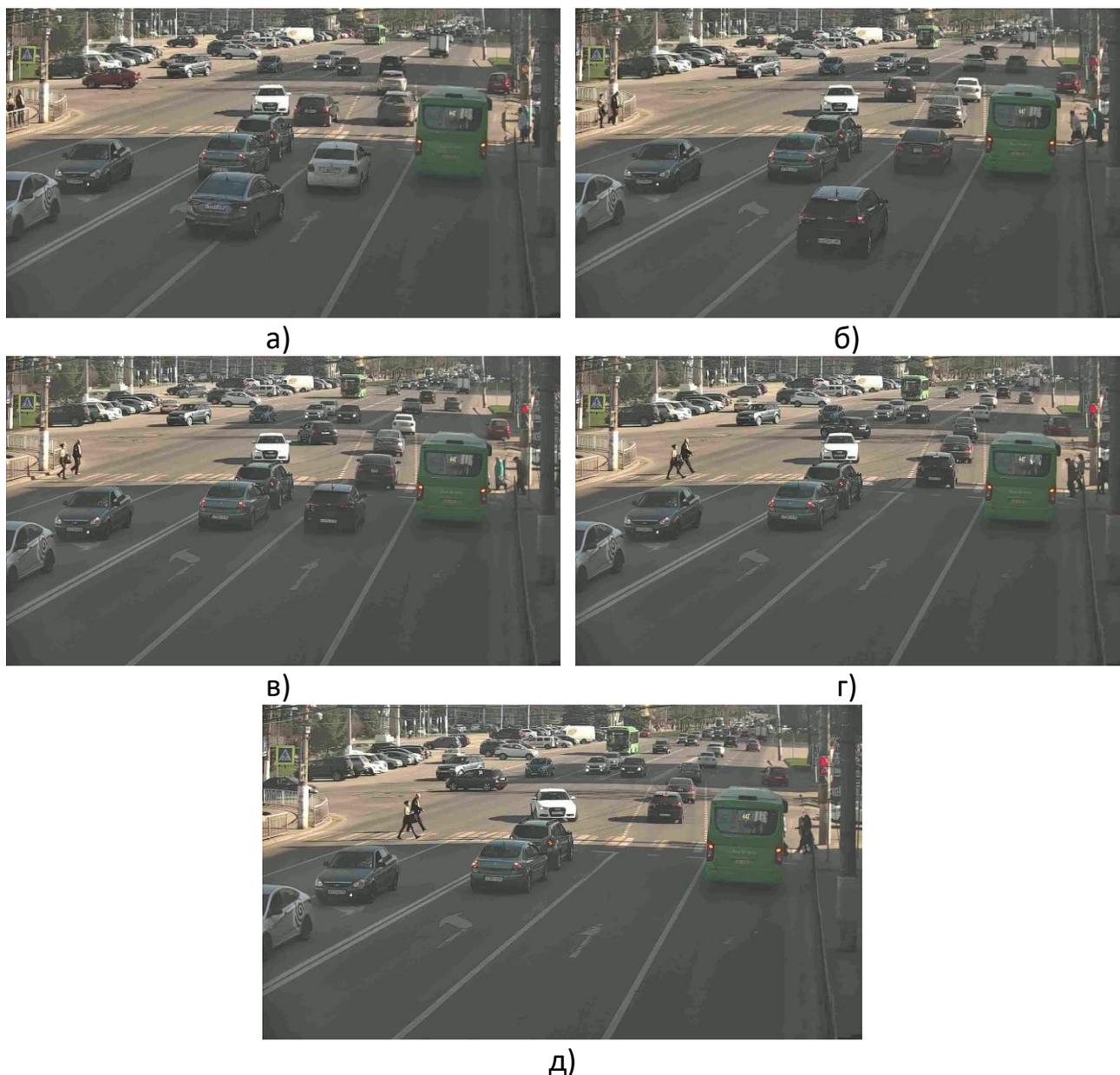
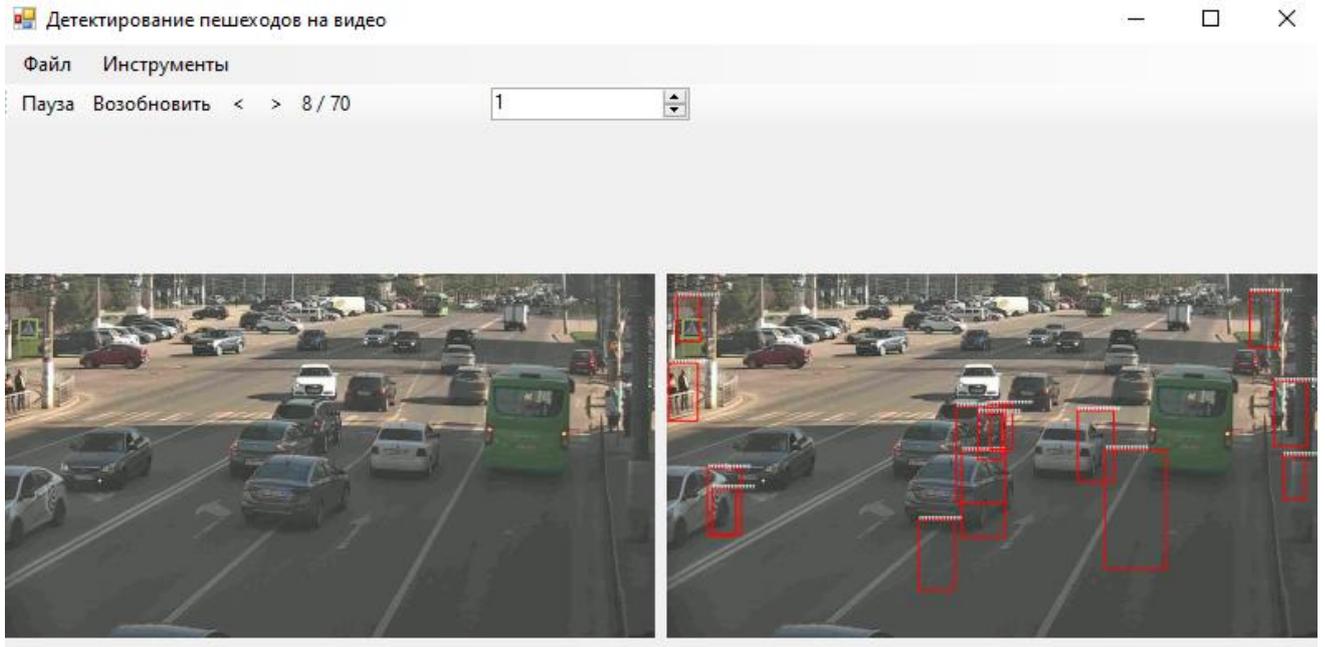


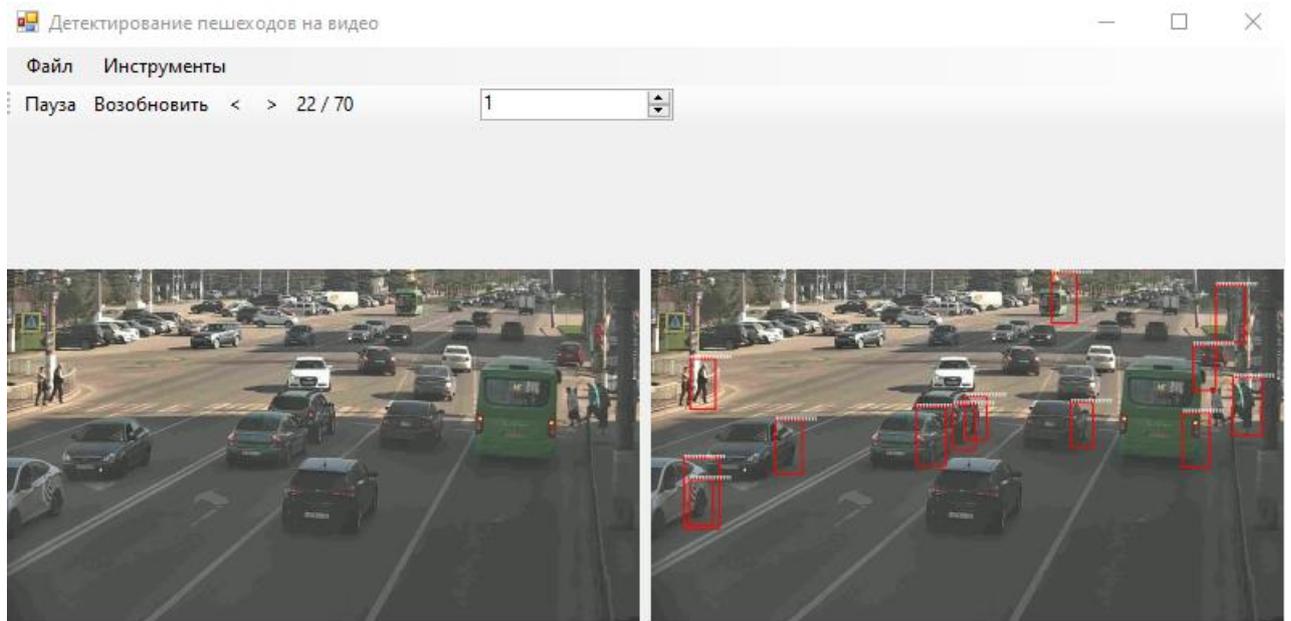
Рис. 1. Детектирование движения людей на пешеходном переходе, исходное изображение: а – кадр 8; б – кадр 22; в – кадр 27; г – кадр 30; д – кадр 34

В ходе экспериментальных исследований распознавание участников дорожного движения производилось с помощью двух методов: HOG+SVM (Histogram of oriented gradients + Support Vector Machines) и разработанной информационно-аналитической системы [22]. Недостатком метода HOG+SVM является возможность распознавания только людей (People), следовательно, эксперимент для автомобилей с помощью данного метода не проводился. Результаты

распознавания пешеходов комбинацией методов HOG+SVM представлены на рисунке 2.



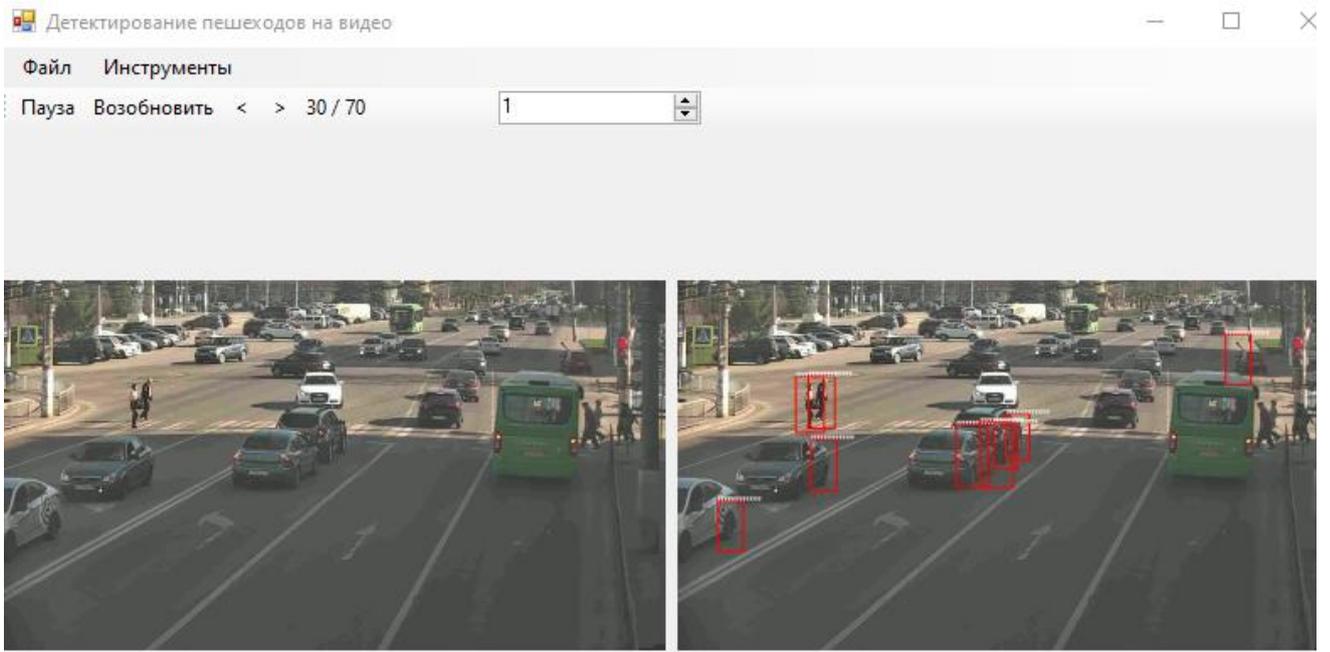
а)



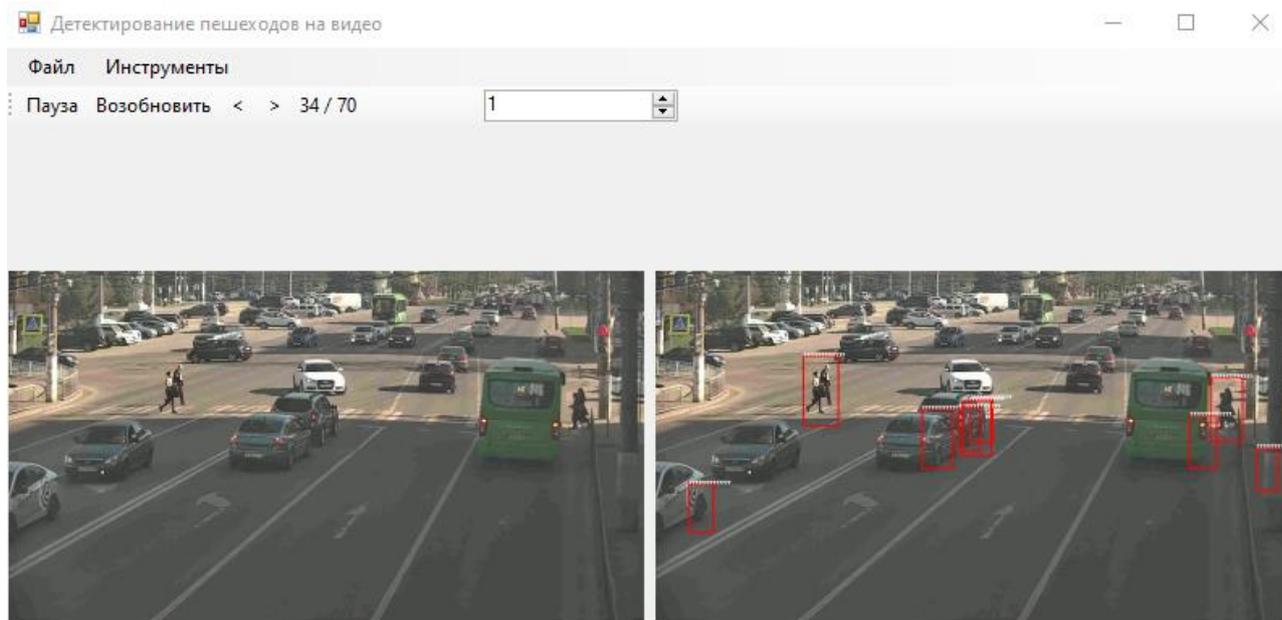
б)



в)



г)



д)

Рис. 2. Детектирование движения людей на пешеходном переходе с помощью комбинации методов HOG+SVM: а – кадр 8; б – кадр 22; в – кадр 27; г – кадр 30; д – кадр 34

Результаты экспериментального исследования сведены в таблицу 2.

Таблица 2 – Результаты экспериментальных исследований детектирования движения людей на пешеходном переходе комбинацией методов HOG+SVM.

Дистанция 51 м.

Номер кадра	Объект	Всего	Количество верно обнаруженных объектов, TP	Количество ложно обнаруженных объектов, FP	Количество пропущенных объектов, FN
8 кадр	Пешеходы	5	2	13	3
22 кадр	Пешеходы	5	2	11	3
27 кадр	Пешеходы	4	2	12	2
30 кадр	Пешеходы	5	2	7	3
34 кадр	Пешеходы	4	2	7	2
Итого			10	50	13

Точность распознавания пешеходов на рассмотренных изображениях рассчитывалась по формуле

$$N = \frac{TP}{TP + FP + FN} \times 100\%,$$

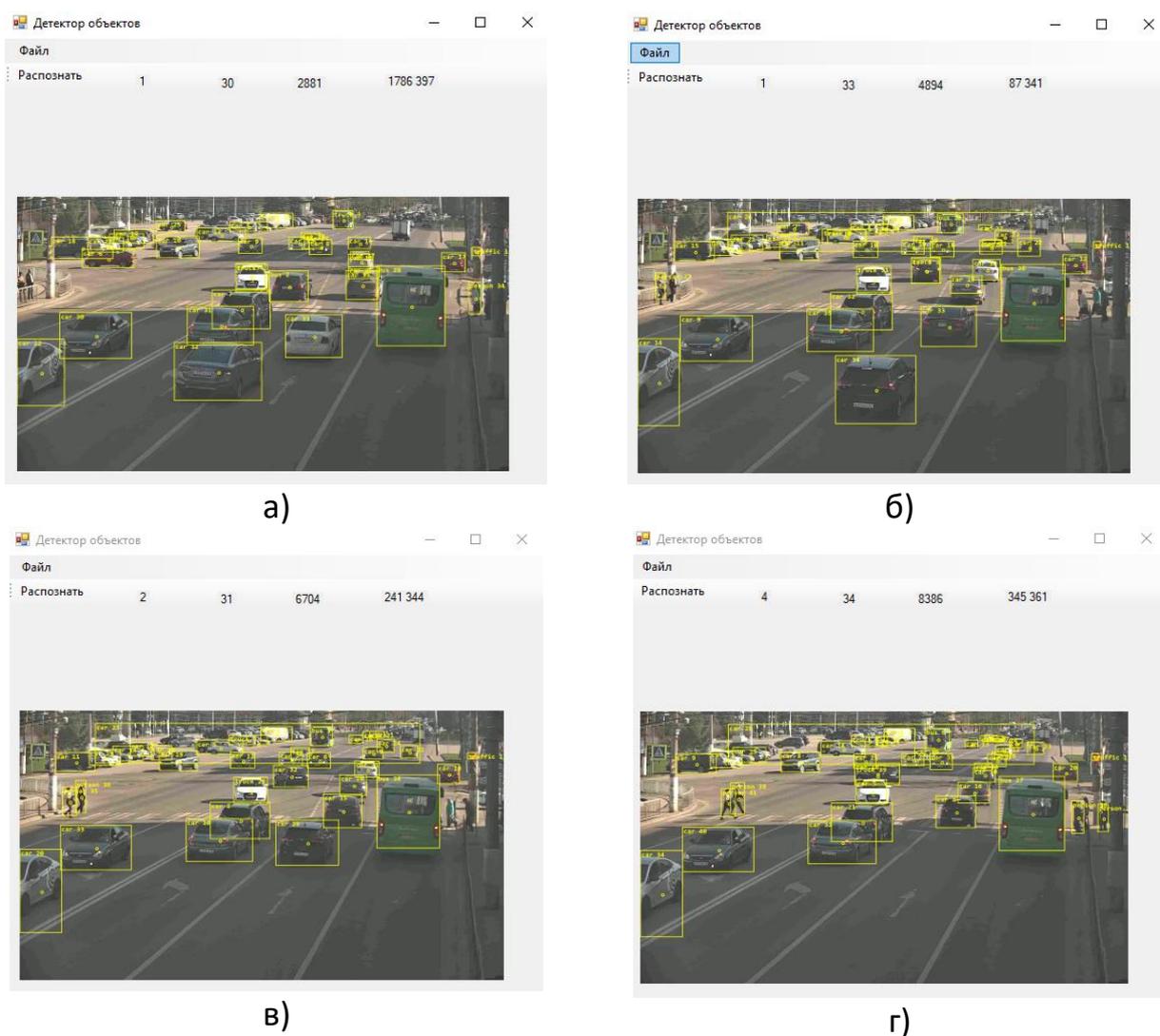
где TP – количество верно обнаруженных объектов; FP – количество ложно обнаруженных объектов; FN – количество пропущенных объектов.

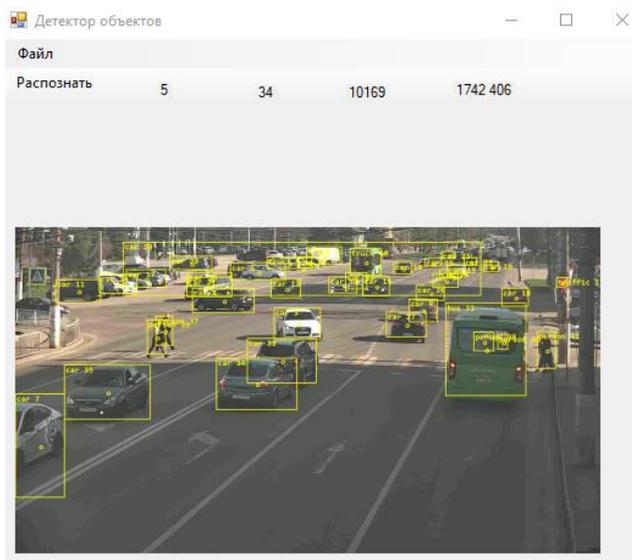
На основе значений из таблицы 1 точность распознавания составила

$$N = \frac{10}{10 + 50 + 13} = \frac{13}{73} = 0,1369 \times 100\%.$$

Следовательно, точность распознавания людей на пешеходном переходе с помощью комбинации методов HOG+SVM составила 13,7%.

Второй эксперимент проводился с помощью разработанной информационно-аналитической системы. Результаты исследования представлены на рисунке 3.





д)

Рис. 3. Детектирование движения людей на пешеходном переходе с помощью ИАС: а – кадр 8; б – кадр 22; в – кадр 27; г – кадр 30; д – кадр 34

Результаты экспериментальных исследований представлены в таблице 3.

Таблица 3 – Результаты экспериментальных исследований детектирования движения людей на пешеходном переходе с помощью ИАС

Номер кадра	Объект	Всего	Количество верно обнаруженных объектов, TP	Количество ложно обнаруженных объектов, FP	Количество пропущенных объектов, FN
8 кадр	Пешеходы	5	1	0	4
22 кадр	Пешеходы	5	1	0	4
27 кадр	Пешеходы	4	2	0	2
30 кадр	Пешеходы	5	4	0	1
34 кадр	Пешеходы	4	3	0	1
Итого			11	0	12

Точность распознавания пешеходов на рассмотренных изображениях составила

$$N = \frac{11}{11 + 0 + 12} = \frac{11}{23} = 0,4782 \times 100\%.$$

Следовательно, точность распознавания людей на пешеходном переходе с помощью метода ИАС составила 47,8%, что выше в 3,5 раза по сравнению с комбинационным методом HOG+SVM.

В таблице 4 представлены результаты распознавания автомобилей с помощью ИАС. При проведении данного эксперимента учитывались автомобили, находящиеся на ближайшем перекрестке.

Таблица 4. Результаты экспериментальных исследований детектирования движения автомобилей на перекрестке с помощью ИАС

Номер кадра	Объект	Всего	Количество верно обнаруженных объектов, <i>TP</i>	Количество ложно обнаруженных объектов, <i>FP</i>	Количество пропущенных объектов, <i>FN</i>
8 кадр	Автомобиль	11	11	0	0
22 кадр	Автомобиль	11	11	1	0
27 кадр	Автомобиль	10	10	1	0
30 кадр	Автомобиль	10	10	1	0
34 кадр	Автомобиль	10	9	1	1
Итого			51	4	1

Точность распознавания автомобилей, находящихся на перекрестке, на рассмотренных изображениях составила

$$N = \frac{51}{51 + 4 + 1} = \frac{51}{56} = 0,9107 \times 100\%.$$

В ходе экспериментальных исследований было выявлено, что автомобили, стоящие перед стоп-линией по направлению движения и находящиеся на перекрестке, распознаются с точностью около 91%. Однако, если рассматривать объекты, находящиеся за перекрестком, точность распознавания значительно снижается, и появляются ложно обнаруженные объекты.

Рассчитаем среднее арифметическое значение точности распознавания объектов с помощью информационно-аналитической системы на основе данных, полученных в ходе экспериментального исследования:

$$\frac{47,8\% + 91\%}{2} = 69,4\%$$

Следовательно, среднее значение точности выявления объектов посредством ИАС составляет 69,4%.

ЗАКЛЮЧЕНИЕ

Таким образом, рассмотрены два метода для распознавания объектов на изображении: существующий метод Histogram of oriented gradients + Support Vector Machines и разработанная информационно-аналитическая система для управления интеллектуальным светофором. Результаты экспериментального исследования, в ходе которого проводилось детектирование объектов на изображениях, показали превосходство разработанной информационно-аналитической системы над существующими методами HOG+SVM. Выявление пешеходов на перекрестке с помощью ИАС выполнено с точностью 47,8%, в то время как комбинационный метод HOG+SVM справился с поставленной задачей с результатом 13,7%, что в 3,5 раза ниже результата ИАС. Среднее значение точности детектирования пешеходов и автомобилей, находящихся на ближайшем перекрестке, с помощью разработанной информационно-аналитической системы равняется 69,4%. Результаты эксперимента также доказывают, что точность распознавания объектов напрямую зависит от расстояния от видеокамеры до объекта.

Благодарности

Работа выполнена при поддержке гранта РФФИ 23-21-00071 «Разработка модели компьютерного зрения для интеллектуальной навигации робототехнических систем, основанной на построении трёхмерных сцен по картам глубин».

СПИСОК ЛИТЕРАТУРЫ

1. *Евстигнеев И.А.* Основы создания интеллектуальных транспортных систем на автомобильных дорогах федерального значения России. М.: Издательство «Перо», 2016. 260 с.
2. ГОСТ Р 50597-93 «Автомобильные дороги и улицы требования к эксплуатационному состоянию, допустимому по условиям обеспечения безопасности дорожного движения».
3. *Муаль М.Н.Б., Козырев Д.В., Уанкпо Г.Ж.К., Нибасумба Э.* Разработка нейросетевого метода в задаче классификации и распознавании изображения //

Современные информационные технологии и ИТ-образование. 2021. Т. 17, № 3. С. 507–518.

4. Брехт Э.А., Коншина В.Н. Применение нейронной сети YOLO для распознавания дефектов // Интеллектуальные технологии на транспорте. 2022. № 2(30). С. 41–47.

5. Sala Filip A. Design of false color palettes for grayscale reproduction // Displays. 2017. №46. P. 9–15.

6. Чернухин Н.А. Комбинированный метод детектирования границ на рентгенографических медицинских изображениях, использующий методику активных контуров // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 88. С. 530–544.

7. Canny J. A computational approach to edge detection // IEEE Trans Pattern Anal Mach Intell. 1986. № 8(6). P. 679–698.

8. Ярошевич П.В., Богуш Р.П. Алгоритм классификации изображений парковочных мест автостоянки на основе гистограмм ориентированных градиентов и метода опорных векторов // Компьютерная оптика. 2017. Т. 41, № 1. С. 110–117.

9. Singh A.K., Shukla V.P., Tiwari Sh., Biradar S.R. Wavelet Based Histogram of Oriented Gradients Feature Descriptors for Classification of Partially Occluded Objects // International Journal of Intelligent Systems and Applications. 2015. Vol. 7, No. 3. P. 54–61.

10. Gornale Sh.S., Patravali P.U., Marathe K.S., Hiremath P.S. Determination of Osteoarthritis Using Histogram of Oriented Gradients and Multiclass SVM // International Journal of Image, Graphics and Signal Processing. 2017. Vol. 9, No. 12. P. 41–49.

11. Михайлов И.С., Зеар Аунг, Йе Тху Аунг. Разработка модификации метода опорных векторов для решения задачи классификации с ограничениями на предметную область // Программные продукты и системы. 2020. № 3. С. 439–448.

12. Зенков В.В. Применение аппроксимации дискриминантной функции Андерсона и метода опорных векторов для решения некоторых задач классификации // Автоматика и телемеханика. 2020. № 1. С. 147–160.

13. *Милостная Н.А.* Методология синтеза интеллектуальных высокопроизводительных нейро-нечётких систем технического зрения: диссертация на соискание ученой степени доктора технических наук. 2023. 350 с.

14. *Бобырь М.В., Архипов А.Е., Горбачев С.В.* Нечетко-логические методы в задаче детектирования границ объектов // Информатика и автоматизация. 2022. Т. 21, № 2. С. 376–404.

15. *Спицын В.Г.* Разработка экспертных систем на основе нечетких правил вывода // Методические указания к лабораторным работам, Томск: Изд-во ТПУ, 2011. 33 с.

16. *Suzuki Satoshi, Keiichi Abe.* Topological structural analysis of digitized binary images by border following // Comput. Vis. Graph. Image Process. 1985. Vol. 30. P. 32–46.

17. *Пегат А.* Нечеткое моделирование и управление; пер. с англ. 4-е изд., электрон. М.: Лаборатория знаний, 2020. 801 с.

18. *Бобырь М.В., Кулабухов С.А.* Моделирование процесса управления температурным режимом в зоне резания на основе нечеткой логики // Проблемы машиностроения и надежности машин. 2017. № 3. С. 76–82.
<https://doi.org/10.3103/S1052618817030049>.

19. *Бобырь М.В., Храпова Н.И., Ламонов М.А.* Система управления интеллектуальным светофором на основе нечеткой логики // Известия Юго-Западного государственного университета. 2021. Т. 25, № 4. С. 162–176.

20. *Бобырь М.В., Милостная Н.А., Храпова Н.И.* Свидетельство о государственной регистрации программы для ЭВМ № 2021661796 Российская Федерация. Программа для регулирования светофора на основе нечёткой логики: № 2021660730: заявл. 08.07.2021: опубл. 15.07.2021; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет».

21. *Бобырь М.В., Храпова Н.И.* Свидетельство о государственной регистрации программы для ЭВМ № 2024662790 Российская Федерация. Программа детектирования объектов на пешеходном переходе и определения времени задержки управляющих сигналов светофора: № 2024661177: заявл. 20.05.2024:

опубл. 30.05.2024; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет».

22. Бобырь М.В., Милостная Н.А., Храпова Н.И. О подходе к детектированию движения пешеходов методом гистограмм направленных градиентов // Электронные библиотеки. 2024. Т. 27, №4. 19 с.

THE TWO-LEVEL INFORMATION AND ANALYTICAL CONTROL SYSTEM FOR INTELLIGENT TRAFFIC LIGHTS

M. V. Bobyr¹ [0000-0002-5400-6817], N. I. Khrapova² [0000-0001-7947-1427]

^{1, 2}Southwest State University, Kursk, Russia

¹fregat_mn@rambler.ru, ²KhrapovaNI@yandex.ru

Abstract

In the modern world, the problems arising in the field of traffic are of great importance. In order to solve existing problems, various intelligent systems are being developed, one of which is the Smart City system. This work is devoted to the development of an information and analytical system (IAS) for controlling an intelligent traffic light. The presented system consists of two levels, each of which contains a set of specific operations. The first level is responsible for detecting objects, in particular pedestrians and cars at the intersection, and the second level calculates the operating time of traffic light signals for the control signal that is transmitted to the device. For comparative analysis, the combined method (HOG+SVM) Histogram of oriented gradients was chosen, based on counting the number of gradient directions on individual image areas and Support Vector Machines, which are used to construct hyperplanes in n-dimensional space in order to separate objects belonging to different classes. The results of an experimental study, during which the recognition of objects in images was carried out, showed the superiority of the developed information and analytical system over existing methods. The average accuracy of detecting pedestrians and cars through the

IAS was 69.4%. In addition, according to the experiment, it was concluded that the accuracy of detecting objects in images is directly proportional to the distance from the video camera to the object.

Keywords: *intelligent traffic light, object detection, machine learning, fuzzy logic boundary detection method, YOLO, HOG, SVM.*

REFERENCES

1. *Evstigneev I.A.* Fundamentals of the creation of intelligent transport systems on highways of federal importance in Russia. M.: Publishing house "Pero", 2016. 260 p.
2. GOST R 50597-93 "Highways and streets requirements for the operational condition permissible under the conditions of ensuring traffic safety".
3. *Mual M.N.B., Kozyrev D.V., Uankpo G.J.K., Nibasumba E.* Development of a neural network method in the problem of image classification and recognition // Modern information technologies and IT education. 2021. Vol. 17, No. 3. P. 507–518.
4. *Brecht E.A., Konshina V.N.* Application of the YOLO neural network for defect detection // Intelligent technologies in transport. 2022. No. 2(30). P. 41–47.
5. *Sala Filip A.* Design of false color palettes for grayscale reproduction // Displays. 2017. №46. P. 9–15.
6. *Chernukhin N.A.* The combined method of detecting boundaries on radiographic medical images using the technique of active contours // Polythematic network electronic scientific journal of the Kuban State Agrarian University. 2013. No. 88. P. 530–544.
7. *Canny J.* A computational approach to edge detection // IEEE Trans Pattern Anal Mach Intell. 1986. № 8(6). P. 679–698.
8. *Yaroshevich P.V., Bogush R.P.* Algorithm for classifying images of parking lots based on histograms of oriented gradients and the method of support vectors // Computer optics. 2017. Vol. 41, No. 1. P. 110–117.
9. *Singh A.K., Shukla V.P., Tiwari Sh., Biradar S.R.* Wavelet Based Histogram of Oriented Gradients Feature Descriptors for Classification of Partially Occluded Objects // International Journal of Intelligent Systems and Applications. 2015. Vol. 7, No. 3. P. 54–61.

10. *Gornale Sh.S., Patravali P.U., Marathe K.S., Hiremath P.S.* Determination of Osteoarthritis Using Histogram of Oriented Gradients and Multiclass SVM // International Journal of Image, Graphics and Signal Processing. 2017. Vol. 9, No. 12. P. 41–49.

11. *Mikhailov I.S., Zear Aung, Ye Thu Aung.* Development of a modification of the method of support vectors for solving the problem of classification with domain constraints // Software products and systems. 2020. No. 3. P. 439–448.

12. *Zenkov V.V.* Application of approximation of the discriminant Anderson function and the support vector machine method for solving some classification problems // Automation and Telemekhanics. 2020. No. 1. P. 147–160.

13. *Milostnaya N.A.* Methodology of synthesis of intelligent highly productive neuro-fuzzy systems of technical vision: dissertation for the degree of Doctor of Technical Sciences. 2023. 350 p.

14. *Bobyry M.V., Arkhipov A.E., Gorbachev S.V.* Fuzzy logic methods in the problem of detecting object boundaries // Informatics and Automation. 2022. Vol. 21, No. 2. P. 376–404.

15. *Spitsyn V.G.* Development of expert systems based on fuzzy inference rules // Methodological guidelines for laboratory work, Tomsk: TPU Publishing House, 2011. 33 p.

16. *Suzuki Satoshi, Keiichi Abe.* Topological structural analysis of digitized binary images by border following // Comput. Vis. Graph. Image Process. 1985. Vol. 30. P. 32–46.

17. *Pegat A.* Fuzzy modeling and control; trans. from English. 4th ed., electron. M.: Laboratory of Knowledge, 2020. 801 p.

18. *Bobyry M.V., Kulabukhov S.A.* Simulation of control of temperature mode in cutting area on the basis of fuzzy logic // Journal of Machinery Manufacture and Reliability. 2017. Vol. 46. P. 288–295. <https://doi.org/10.3103/S1052618817030049>.

19. *Bobyry M.V., Khrapova N.I., Lamonov M.A.* Intelligent traffic light control system based on fuzzy logic // Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. 2021. Vol. 25, No. 4. P. 162–176.

20. *Bobyry M.V., Milostnaya N.A., Khrapova N.I.* Certificate of state registration of the computer program No. 2021661796 Russian Federation. The program for traffic light regulation based on fuzzy logic: No. 2021660730: application 08.07.2021: publ.

15.07.2021; the applicant is the Federal State Budgetary Educational Institution of Higher Education "Southwestern State University".

21. *Bobyry M.V., Khrapova N.I.* Certificate of state registration of the computer program No. 2024662790 Russian Federation. The program for detecting objects at a pedestrian crossing and determining the time of holding traffic light control signals: No. 2024661177: application 20.05.2024: publ. 30.05.2024; applicant Federal State Budgetary Educational Institution of Higher Education "Southwestern State University".

22. *Bobyry M.V., Milostnaya N.A., Khrapova N.I.* On the approach to detecting pedestrian movement by the method of histograms of directional gradients // *Electronic libraries. 2024. Vol. 27, No. 4. 19 p.*

СВЕДЕНИЯ ОБ АВТОРАХ



БОБЫРЬ Максим Владимирович. Доктор технических наук, профессор. Профессор кафедры «Программная инженерия», ФГБОУ ВО «Юго-Западный государственный университет». Области исследований: интеллектуальные системы управления, адаптивные нейро-нечеткие системы вывода, распознавание и обработка изображения, робототехнические системы.

Maxim V. BOBYR. Doctor of technical sciences, professor. Professor of the Department of Software Engineering, Southwest State University. Research areas: intelligent control systems, adaptive neuro-fuzzy inference systems, image recognition and processing, robotic systems.

Email: maxboby@gmail.com

ORCID: 0000-0002-5400-6817



ХРАПОВА Наталья Игоревна. Аспирант кафедры «Программная инженерия», ФГБОУ ВО «Юго-Западный государственный университет». Область исследования: распознавание сложных объектов с использованием нечеткой логики.

Natalia Igorevna KHRAPOVA. Postgraduate student of the Department of Software Engineering at the Southwest State University. Research area: recognition of complex objects using fuzzy logic.

Email: KhrapovaNI@yandex.ru

ORCID: 0000-0001-7947-1427

Материал поступил в редакцию 25 июня 2024 года

УДК 004

АВТОМАТИЧЕСКАЯ РАЗМЕТКА ОБУЧАЮЩИХ ВЫБОРОК В КОМПЬЮТЕРНОМ ЗРЕНИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

А. К. Журавлёв¹ [0009-0001-1265-9916], **К. А. Григорян**² [0000-0001-6470-1832]

^{1, 2}*Казанский (Приволжский) федеральный университет, ул. Кремлевская,
35, г. Казань, 420008*

¹UnMelow@yandex.ru, ²karigri@yandex.ru

Аннотация

Рассмотрена проблема автоматической разметки обучающих выборок в области компьютерного зрения с использованием методов машинного обучения.

Разметка данных является ключевым этапом в разработке и обучении моделей глубокого обучения, однако процесс создания размеченных данных зачастую требует значительных временных и трудовых затрат. В статье предложен механизм автоматической разметки, основанный на использовании сверточных нейронных сетей и методов активного обучения.

Предложенная методология включает анализ и оценку существующих подходов к автоматической разметке. Эффективность предложенных решений оценена на общедоступных наборах данных. Результаты показали, что предложенный метод в значительной мере сокращает время, необходимое для разметки данных, но в любом случае требует вмешательства оператора-разметчика.

Обзор литературы включает анализ современных методов разметки и существующих автоматических систем, что позволяет лучше понять контекст и преимущества предлагаемого подхода. В заключении обсуждены достижения, ограничения и возможные направления для будущих исследований в данной области.

Ключевые слова: компьютерное зрение, машинное обучение, автоматическая разметка данных, обучающая выборка, сегментация изображений.

ВВЕДЕНИЕ

С быстрым развитием технологий в области искусственного интеллекта и машинного обучения компьютерное зрение стало одной из самых активных и быстрорастущих областей исследований. Эта область включает в себя методы и технологии для автоматического извлечения, анализа и интерпретации значимой информации из изображений и видео. Основными задачами компьютерного зрения являются классификация, сегментация и обнаружение объектов.

Одним из ключевых аспектов успешного обучения моделей глубокого обучения является наличие большого объема качественно размеченных данных. Разметка данных – критически важный этап в создании высококачественных обучающих выборок. Однако процесс разметки является трудоемким и требует значительных ресурсов, что служит основным препятствием для быстрой разработки новых моделей и приложений. Согласно исследованию, на разметку крупных наборов данных уходит до 80% времени всего проекта по разработке модели машинного обучения [1].

Для облегчения процесса разметки используют различные инструменты и библиотеки, такие как LabelImg и VGG Image Annotator (VIA). LabelImg привлекает своей простотой и доступностью [2, 3], поддерживает популярные форматы аннотаций, такие как Pascal VOC, и обеспечивает интуитивно понятный интерфейс для пользователей [4]. Однако, несмотря на его достоинства, LabelImg не предоставляет продвинутых возможностей автоматизации и интеграции. VIA – инструмент с открытым исходным кодом и поддержкой различных типов аннотаций (для аннотирования доступны ограничивающие рамки, ломаные линии, многоугольники, круги, овалы, точки и эллипсы). Работать с этим инструментом можно через веб-браузер, что делает его удобным для удаленной работы и совместного использования.

В последние годы были разработаны различные методы автоматической разметки данных, которые могут значительно сократить время и усилия, необходимые для этого процесса. Одним из таких методов является использование сверточных нейронных сетей (Convolutional Neural Network, CNN), которые показали высокую эффективность в задачах классификации и сегментации изображений.

Активное обучение, другой важный метод, также используется для автоматической разметки данных. Активное обучение позволяет модели выбирать наиболее информативные примеры для разметки, что значительно улучшает эффективность процесса разметки и снижает затраты на разметку больших объемов данных. Установлено, что этот метод при разметке данных на видео снижает рабочую нагрузку на человека-оператора примерно на 78%, что значительно превосходит полную ручную разметку [5].

В настоящей работе проведены результаты исследования методов автоматической разметки обучающих выборок для задач компьютерного зрения с использованием передовых методов машинного обучения. Рассмотрены основные подходы к автоматической разметке данных, а также представлены результаты экспериментов по оценке эффективности предложенных методов.

ОБЗОР ЛИТЕРАТУРЫ

Современные методы автоматической разметки данных в задачах компьютерного зрения демонстрируют значительный прогресс благодаря использованию сверточных нейронных сетей (CNN) и активного обучения. В работе [6] рассмотрены различные инструменты и методы разметки изображений, систематизированы существующие программные решения и оценены по таким параметрам, как степень автоматизации и удобство пользовательского интерфейса. Большинство современных инструментов стремится минимизировать ручной труд и повысить точность разметки за счет внедрения методов машинного обучения и активного обучения.

В исследовании [7] описан один из методов автоматической аннотации изображений, основанный на сверточных нейронных сетях и алгоритме оптимизации порога (threshold optimization, THOP), который значительно улучшает

точность разметки по сравнению с традиционными методами и демонстрирует на 11% более высокую точность на нескольких наборах данных. В названном методе CNN превосходит традиционные методы по извлечению признаков и точности аннотации.

Модель автоматической аннотации изображений на основе модели внимания и равновесия данных, описанная в обзоре [8], использует глубокие нейронные сети для извлечения признаков и автоэнкодер для балансировки словаря аннотированных ключевых слов. Эксперименты показали, что предложенная модель превосходит существующие модели по критериям производительности на двух контрольных наборах данных.

Обзор [9] современных методов сегментации изображений на основе глубокого обучения подчеркивает важность сверточных нейронных сетей, таких как U-Net и Fully Convolutional Networks (FCN). Эти архитектуры показали высокую производительность в задачах сегментации естественных и медицинских изображений.

Исследование [10] также представляет интерактивные системы самоаннотации для детекции объектов на видео, сочетающие автоматические и интерактивные процессы, что позволяет ускорить аннотацию и снизить затраты на ручную работу. Эксперименты показали, что такие системы могут эффективно аннотировать видео и снижать затраты времени на аннотацию без потери качества.

Кроме того, статьи [11, 12] описывают модель Segment Anything (SAM), которая использует подход "promptable segmentation", позволяя модели эффективно справляться с задачами сегментации при различных входных данных, таких как точки, боксы или текст. Это значительно улучшает гибкость и применимость модели для автоматической аннотации данных в различных сценариях.

МЕТОДОЛОГИЯ

Для оценки различных методов автоматической разметки данных мы рассмотрели несколько современных подходов, включая использование мо-

делей сверточных нейронных сетей, активного обучения, интерактивной аннотации, а также методов, реализованных в модели YOLOv8 и в возможностях платформы Roboflow [13]. Основное внимание было уделено следующим аспектам: эффективность, точность на различных наборах данных и применимость в различных ситуациях.

Эффективность. Оценено время, необходимое для разметки одного изображения. Оценка включает измерение средней продолжительности аннотации для каждого метода (рис. 1).

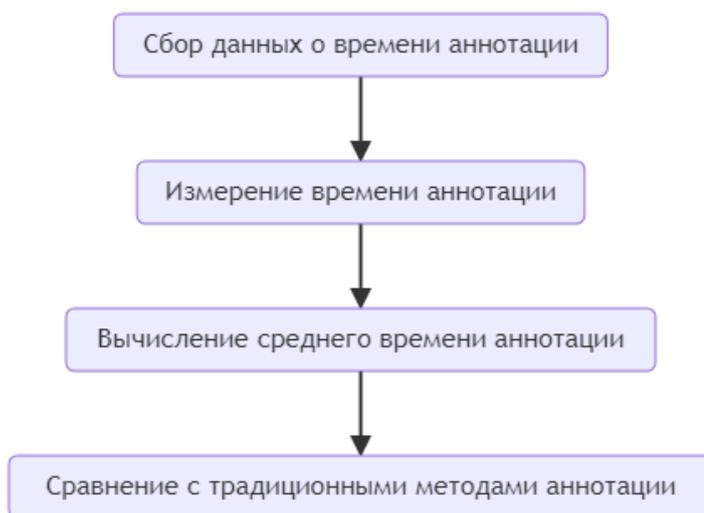


Рис. 1. Алгоритм оценки эффективности отдельно взятого метода

Точность. Определяется средней точностью (mean average precision, mAP) на стандартных наборах данных. Оценивалась на различных уровнях точности, таких как AP50 и AP75 (формулы (1) и (2) ниже), средняя точность при уровне уверенности 50% и 75% соответственно (рис. 2).

$$AP_{50} = 1/N \sum_{i=1}^N 1(IoU_i \geq 0.50), (1)$$

где (AP_{50}) – средняя точность при пороге IoU 50%; (N) – общее количество предсказаний; $(1(\cdot))$ – индикаторная функция, принимающая значение 1, если условие выполняется, и 0 – если нет; (IoU_i) – метрика степени пересечения между двумя ограничивающими рамками для (i) -го предсказания.

$$AP_{75} = 1/N \sum_{i=1}^N 1(IoU_i \geq 0.75), (2)$$

где (AP_{75}) – средняя точность при пороге IoU 75%; (N) – общее количество предсказаний; $(1(\cdot))$ – индикаторная функция, принимающая значение 1, если условие выполняется, и 0 – если нет; (IoU_i) – метрика степени пересечения между двумя ограничивающими рамками для (i) -го предсказания.

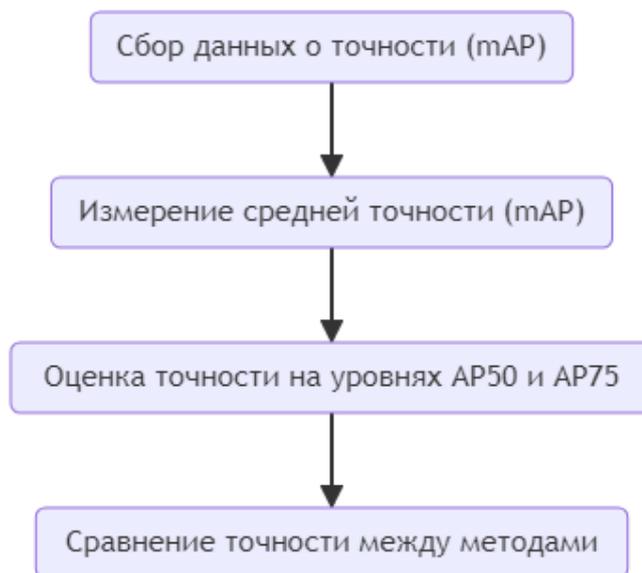


Рис. 2. Алгоритм оценки точности отдельно взятого метода

Применимость. Оценивалась способность метода адаптироваться к различным задачам и сценариям, включая аннотацию в реальном времени, медицинскую аннотацию и сложные сцены (рис. 3).



Рис. 3. Алгоритм оценки применимости отдельно взятого метода

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Анализ показал, что различные методы автоматической разметки данных имеют свои уникальные преимущества и ограничения (таблица 1). Для оценки были использованы стандартные наборы данных COCO [14] и Cityscapes [15], которые позволяют объективно сравнить производительность методов в различных сценариях. Основные метрики включают эффективность (время аннотации одного изображения), точность (средняя точность, mAP) и применимость в реальных задачах.

Таблица 1. Результаты сравнения подходов к автоматической разметке

Метод	Эффективность (время/изобр.)	Точность (mAP, COCO)	Точность (mAP, Cityscapes)	Применимость
Mask RCNN	1–2 секунды	37,1%	57,58%	Городские сцены, медицинские изображения
Интер-активная аннотация	Варьируется (в среднем 5 секунд)	45,0%	60,0%	Сложные задачи, требующие ручной доработки
YOLOv8	<1 секунды	50,2%	63,0%	Обнаружение объектов в реальном времени
Roboflow Auto Label [16]	0,5–1 секунд	55,6%	64,2%	Медицинская аннотация, автономные транспортные средства

В результате анализа были получены следующие результаты. Mask RCNN отличается высокой точностью в сегментации сложных сцен и медицинских изображений, однако время аннотации может быть выше по сравнению с другими методами. Показатель mAP для COCO составляет 37,1%, а для Cityscapes – 57,58%.

Интерактивная аннотация, в свою очередь, эффективно снижает затраты на ручную разметку и достигает высокой точности благодаря ручной корректировке. Среднее время аннотации повышается в связи с разметкой вручную и составляет около 5 секунд, точность (mAP) для COCO – 45,0%, для Cityscapes – 60,0%. Подходит для сложных задач, требующих ручной доработки.

Наилучшие результаты по скорости и точности для задач реального времени дает YOLOv8. Показатель mAP для COCO составляет 50,2%, а для Cityscapes – 63,0%. Время обработки одного изображения занимает менее 1 секунды.

Инструмент Roboflow Auto Label демонстрирует наибольшую гибкость, обеспечивая быстрое и точное аннотирование с возможностью ручной корректировки. Показатель mAP для COCO составляет 55,6%, а для Cityscapes – 64,2%. Время разметки одного изображения занимает 0,5–1 секунды. Указанный инструмент подходит для множества сценариев, включая медицинскую аннотацию и аннотацию для автономных транспортных средств.

ЗАКЛЮЧЕНИЕ

Использование современных методов автоматической разметки данных позволяет значительно сократить время и усилия, затрачиваемые на аннотацию изображений, обеспечивая при этом высокую точность. Наиболее эффективными и универсальными методами являются те, которые используют современные модели глубокого обучения и активное обучение. Метод Roboflow Auto Label выделяется как лучший вариант, предлагая высокую точность и гибкость благодаря использованию текстовых подсказок и предобученных моделей. Этот метод обеспечивает оптимальное сочетание автоматической разметки и возможности ручной корректировки, что делает его пригодным для широкого спектра задач в области компьютерного зрения.

СПИСОК ЛИТЕРАТУРЫ

1. Council J. Data challenges are halting AI projects, IBM executive says // The Wall Street Journal. 2019. Vol. 28.
2. Labellmg for Image Annotation.
URL: <https://viso.ai/computer-vision/labelimg-for-image-annotation/>.
3. VGG Image Annotator.
URL: https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html.
4. Everingham M. et al. The pascal visual object classes challenge: A retrospective // International Journal of Computer Vision. 2015. Vol. 111. P. 98–136.
5. Berg A. et al. Semi-automatic annotation of objects in visual-thermal video // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. <https://doi.org/10.1109/ICCVW.2019.00277>
6. Sager C., Janiesch C., Zschech P. A survey of image labelling for computer vision applications // Journal of Business Analytics. 2021. Vol. 4, No. 2. P. 91–110.
7. Cao J., Zhao A., Zhang Z. Automatic image annotation method based on a convolutional neural network with threshold optimization // Plos one. 2020. V. 15, No. 9. e0238956. <https://doi.org/10.1371/journal.pone.0238956>
8. Vatani A., Ahvanooy M.T., Rahimi M. An effective automatic image annotation model via attention model and data equilibrium // arXiv preprint arXiv:2001.10590. 2020.
9. Gu Y. et al. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs // Computers in Biology and Medicine. 2018. Vol. 103. P. 220–231.
10. Levine S. et al. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection // The International Journal of Robotics Research. 2018. Vpl. 37, No. 4-5. P. 421–436.
11. Kirillov A. et al. Segment anything // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. P. 4015–4026.

12. Zou X. et al. Segment everything everywhere all at once // Advances in Neural Information Processing Systems. NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. Article No. 868. P.19769–19782. <https://dl.acm.org/doi/10.5555/3666122.3666990>
 13. Ultralytics YOLOv8 Docs. URL: <https://docs.ultralytics.com/ru>.
 14. COCO Dataset. URL: <https://cocodataset.org/#home>.
 15. Cityscapes Dataset. URL: <https://www.cityscapes-dataset.com/>.
 16. Auto-Label. URL: <https://roboflow.com/auto-label>.
-

AUTOMATIC ANNOTATION OF TRAINING DATASETS IN COMPUTER VISION USING MACHINE LEARNING METHODS

A.K. Zhuravlyov¹ [0009-0001-1265-9916], **K.A. Grigorian**² [0000-0001-6470-1832]

^{1, 2}*Kazan (Volga Region) Federal University, 35 Kremlevskaya str., Kazan, 420008*

¹UnMelow@yandex.ru, ²karigri@yandex.ru

Abstract

This paper addresses the issue of automatic annotation of training datasets in the field of computer vision using machine learning methods. Data annotation is a key stage in the development and training of deep learning models, yet the process of creating labeled data often requires significant time and labor. This paper proposes a mechanism for automatic annotation based on the use of convolutional neural networks (CNN) and active learning methods.

The proposed methodology includes the analysis and evaluation of existing approaches to automatic annotation. The effectiveness of the proposed solutions is assessed on publicly available datasets. The results demonstrate that the proposed method significantly reduces the time required for data annotation, although operator intervention is still necessary.

The literature review includes an analysis of modern annotation methods and existing automatic systems, providing a better understanding of the context and advantages of the proposed approach. The conclusion discusses achievements, limitations, and possible directions for future research in this field.

Keywords: *computer vision, machine learning, automatic data annotation, training datasets, image segmentation.*

REFERENCES

1. *Council J.* Data challenges are halting AI projects, IBM executive says // The Wall Street Journal. 2019. Vol. 28.
2. *LabelImg for Image Annotation.*
URL: <https://visio.ai/computer-vision/labelimg-for-image-annotation/>.
3. *VGG Image Annotator.*
URL: https://www.robots.ox.ac.uk/~vgg/software/via/via_demo.html.
4. *Everingham M. et al.* The pascal visual object classes challenge: A retrospective // International Journal of Computer Vision. 2015. Vol. 111. P. 98–136.
5. *Berg A. et al.* Semi-automatic annotation of objects in visual-thermal video // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. <https://doi.org/10.1109/ICCVW.2019.00277>
6. *Sager C., Janiesch C., Zschech P.* A survey of image labelling for computer vision applications // Journal of Business Analytics. 2021. Vol. 4, No. 2. P. 91–110.
7. *Cao J., Zhao A., Zhang Z.* Automatic image annotation method based on a convolutional neural network with threshold optimization // Plos one. 2020. V. 15, No. 9. e0238956. <https://doi.org/10.1371/journal.pone.0238956>
8. *Vatani A., Ahvanooy M.T., Rahimi M.* An effective automatic image annotation model via attention model and data equilibrium // arXiv preprint arXiv:2001.10590. 2020.
9. *Gu Y. et al.* Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs // Computers in Biology and Medicine. 2018. Vol. 103. P. 220–231.
10. *Levine S. et al.* Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection // The International Journal of Robotics Research. 2018. Vpl. 37, No. 4-5. P. 421–436.
11. *Kirillov A. et al.* Segment anything // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. P. 4015–4026.

12. Zou X. *et al.* Segment everything everywhere all at once // Advances in Neural Information Processing Systems. NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. Article No. 868. P.19769–19782. <https://dl.acm.org/doi/10.5555/3666122.3666990>
 13. Ultralytics YOLOv8 Docs. URL: <https://docs.ultralytics.com/ru>.
 14. COCO Dataset. URL: <https://cocodataset.org/#home>.
 15. Cityscapes Dataset. URL: <https://www.cityscapes-dataset.com/>.
 16. Auto-Label. URL: <https://roboflow.com/auto-label>.
-

СВЕДЕНИЯ ОБ АВТОРАХ



ЖУРАВЛЁВ Алексей Константинович – выпускник бакалавриата Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

Aleksey Konstantinovich ZHURAVLEV – graduate student of the Institute of Information Technologies and Intelligent Systems, Kazan (Volga region) Federal University, Kazan.

Email: UnMelow@yandex.ru

ORCID: 0009-0001-1265-9916



ГРИГОРЯН Карен Альбертович – кандидат экономических наук, доцент, Казанский (Приволжский) федеральный университет, г. Казань.

Karen Albertovich GRIGORIAN – Candidate of Economics, Associate Professor, Kazan (Volga region) Federal University, Kazan.

Email: karigri@yandex.ru

ORCID: 0000-0001-6470-1832

Материал поступил в редакцию 25 сентября 2024 года

АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ HTML-ДОКУМЕНТОВ ПО СТАНДАРТУ MICRODATA

Т. Ф. Ибрагимов¹ [0009-0008-4959-5259], А. А. Ференец² [0000-0002-7859-9901]

^{1, 2}Институт информационных технологий и интеллектуальных систем
Казанского (Приволжского) федерального университета, ул. Кремлевская, 35,
г. Казань, 420008

¹i.timur0701@gmail.com, ²ist.kazan@gmail.com

Аннотация

Описана разработка на основе методов машинного обучения приложения для автоматического аннотирования веб-страниц по стандарту Microdata с возможностью расширения для других стандартов и с внедрением данных в JSX-файлы. Собраны и подготовлены датасеты для обучения моделей Machine Learning (ML). Собраны и проанализированы метрики модели ML.

Ключевые слова: *Microdata, семантическая разметка, HTML5, поисковая оптимизация (SEO), поисковые системы, машинное обучение, schema.org, семантический веб, стандарты разметки, автоматизация SEO.*

ВВЕДЕНИЕ

Современным стандартом для разметки веб-страниц является HTML5 (HyperText Markup Language) [1]. Хотя HTML-элементы предоставляют полезную дополнительную информацию о разделах документа, они описывают лишь общие понятия: навигация, заголовок, ссылка и подобные. Однако в HTML5 отсутствует конкретизация для таких элементов, как фильмы, товары и другие. Этот недостаток в языке разметки преодолевается с помощью описания элементов согласно стандарту «Микроданные», который позволяет встраивать структурированные метаданные в веб-страницы. Авторы вставляют машиночитаемую информацию в свои HTML веб-страницы, которая помогает определить, какой текст является названием фильма, какой – именем режиссера и т. д.

Семантически аннотированные данные на веб-сайте улучшают доступность информации при анализе ботами, в том числе поисковыми, лучше индексируются

и поднимаются в поисковой выдаче. Различные исследования показывают, что все больше и больше ресурсов внедряют семантическую разметку, потому что такая оптимизация (далее SEO) может значительно увеличить пользовательский трафик ресурса.

Семантическую разметку можно внедрить на всех ресурсах, использующих HTML-страницы для представления информации в интернете. Однако сейчас разработчики предпочитают использовать различные технологии, упрощающие процесс создания веб-страниц, которые в свою очередь формируют HTML. Среди таких технологий можно выделить различные шаблонизаторы, которые описывают шаблоны формирования динамических страниц, или в том числе синтаксис JSX [2], который был принят для использования множеством фреймворков. Разнообразие инструментария, используемого при разработке, вносит дополнительную сложность в процесс аннотирования. При этом процесс аннотирования может дополнительно усложняться для больших веб-сайтов из-за роста количества и объема страниц. Поэтому разработка инструмента, автоматизирующего процесс внедрения семантической разметки, важна для обеспечения простоты и скорости разработки веб-приложений.

Существует несколько способов внедрить семантическую разметку: с помощью стандартов Microdata [3] и JSON-LD [4]. Последнее исследование [5] показало, что применение JSON-LD более распространено, однако в данной работе рассматривается только Microdata, потому что не все популярные поисковые системы на данный момент поддерживают формат JSON-LD.

Обзор предметной области

Семантическая разметка используется для аннотирования различных типов данных, таких как товары, события, статьи, организации, фильмы и другие. Например, веб-страница, описывающая фильм, может содержать семантические метаданные, указывающие название фильма, его режиссера, год производства и рейтинг.

Все классы и атрибуты онтологии инициативы Schema.org [6] используются совместно с одним из форматов, с помощью которого эти метаданные внедряются в страницу. Среди используемых форматов: Microdata, JSON-LD, RDFa [7],

Microformats [8]. Наиболее популярными форматами сейчас являются Microdata и JSON-LD.

Для разработки инструмента семантической разметки были определены следующие ключевые критерии:

- Необходима поддержка стандарта Microdata.
- Инструмент должен автоматически добавлять аннотации в HTML-код.
- Инструмент должен иметь возможность расширения набора поддерживаемых понятий из онтологии Schema.org.
- Инструмент должен позволять переносить аннотации в исходный код шаблонов представления данных.

Нами были проведены поиск и анализ существующих решений. В результате установлено, что:

- Microdata Generator [9] предоставляет возможность создания JSON-LD-объекта или HTML-фрагмента для внедрения в веб-страницу. Процесс обработки данных не автоматизирован.

- Отличительной особенностью мастера разметки структурированных данных Google [10] является возможность аннотировать элементы страницы в интерактивном режиме. Этот инструмент предоставляет два способа, с помощью которых можно аннотировать страницу: используя ссылку на веб-страницу либо передав HTML, который необходимо аннотировать.

- InLinks [11], в отличие от предыдущих инструментов, автоматически определяет тематику страницы на основе ее содержимого и генерирует объекты JSON-LD для внедрения. Однако он не поддерживает стандарт Microdata, и все созданные объекты имеют тип Thing.

- Web-segment [12] также автоматизирует процесс обработки данных с поддержкой стандарта Microdata, но поддерживает всего лишь 1 класс из онтологии Schema.org.

Таким образом, в большинстве инструментов нет возможности обработки данных в автоматическом режиме, в некоторых отсутствует поддержка достаточного количества классов из онтологии Schema.org, а InLinks вообще не поддерживает стандарт Microdata.

Проектирование программного решения

Внедрению семантической разметки предшествует несколько этапов: определение страниц, которые необходимо аннотировать, определение типов объектов, находящихся на странице, и выбор соответствующих схем из онтологии Schema.org. В данной работе предложен подход, основанный на машинном обучении для определения типов объектов, выбора соответствующих схем и последующего внедрения семантической разметки на страницу в автоматическом режиме.

Для генерации семантической разметки необходимо получить HTML-документ, для которого будет сгенерирована семантическая разметка. Затем, используя полученный документ и найденные типы объектов на странице, необходимо выбрать соответствующие схемы объектов из онтологии Schema.org. При необходимости, для переноса данных из сгенерированного HTML-документа в файлы, использующие технологии создания веб-страниц, отличные от HTML, мы предполагаем использование независимого инструмента. Такая модульность позволит в будущем, при необходимости поддержки той или иной технологии создания веб-страниц, разработать отдельный механизм, учитывая специфику соответствующей технологии.

Концепция работы инструмента предполагает последовательное выполнение нескольких этапов обработки данных.

1. Изначально необходимо получить документ, в который предполагается внедрить семантическую разметку. Документ можно получить путем передачи URL-адреса страницы в интернете, чтобы она впоследствии была загружена и обработана, либо передав локальный путь доступа к документу.

2. На следующем этапе происходит предобработка документа. Страницу необходимо очистить от элементов, которые не несут семантически значимую информацию. Такие элементы впоследствии могут замедлить выполнение программы и ухудшить результаты работы всего алгоритма, именно поэтому на этапе предобработки документа эти элементы должны быть удалены.

3. На следующем этапе блоки данных сегментируются для их последующей классификации путем поиска в документе шаблонов. При генерации данных на серверной или клиентской частях используются подготовленные шаблоны. Полученная информация из базы данных или с сервера применяется к подготовленным

шаблонам, группируется, и в результате генерируется HTML-страница. Поэтому, используя алгоритм, основанный на сопоставлении структурно схожих элементов, можно сегментировать эти блоки данных для их последующей классификации.

4. На этапе классификации необходимо определить классы объектов онтологии Schema.org, найденных на предыдущем шаге, и сгенерировать css-селекторы для получения элементов.

5. На следующем этапе на основе css-селекторов, сгенерированных на предыдущем шаге, генерируется разметка Microdata. Каждый класс онтологии определяет соответствующий набор атрибутов класса. Для выявленного класса с помощью регулярных выражений и по определенным правилам генерируются соответствующие атрибуты.

6. После создания разметки алгоритм прекращает свою работу и генерирует HTML-документ.

7. С использованием полученных данных, семантическая разметка должна быть внедрена в соответствующие элементы в файле с синтаксисом JSX.

Реализация инструмента

Задача предобработки документа решается нами методами машинного обучения с помощью бинарной классификации. Таким образом, один из классов представляет собой элемент, несущий семантическую информацию, которая впоследствии может быть использована при обработке данных, а другой класс – декоративный элемент либо элемент, не несущий полезной информации, он может быть удален из документа без последствий для работы алгоритма.

При разработке модели машинного обучения было использовано консольное приложение, разработанное Nichita Utii, Vlad-Sebastian Ionescu [13].

Модель была обучена на данных, извлеченных из датасета Dragnet [14]. Он состоит из HTML-страниц, которые были собраны Matthew E. Peters, Dan Lescocq из следующих источников:

- 999 страниц было выбрано через RSS-каналы из ресурсов с большим количеством подписчиков;
- 204 страницы были собраны из 23 популярных новостных ресурсов;
- 178 страниц были случайно выбраны из списка различных блогов.

Каждый тег в датасете представлен следующим набором признаков:

- глубина вложенности;
- позиция среди соседних элементов;
- количество дочерних узлов;
- длина текста;
- длина атрибута класса;
- длина атрибута id;
- категориальный признак, содержащий тип тега.

Также в набор данных включены признаки элементов-предков и дочерних элементов.

Для элементов-предков использованы те же характеристики, что и для рассматриваемого тега: извлечена информация для 5 элементов, являющихся предками тега. Различные исследования (например, [15, 16]) показали, что путь от корня до элемента может нести полезную информацию и положительно влияет на работу подобного решения. Для дочерних элементов упомянутые выше характеристики усреднялись и группировались вместе. Эти усредненные значения вместе с уровнем глубины вложенности извлекались для 5 уровней вложенности в поддереве рассматриваемого элемента. Если этих данных недостаточно, то в случае элементов-предков они заполнялись значениями самого верхнего элемента-предка, а в случае дочерних элементов набор заполнялся значениями самого нижнего уровня.

Для задачи классификации была использована модель, основанная на алгоритме случайного леса, поскольку эта модель, исходя из заключений в ранее упомянутой статье [13], показывает себя наилучшим образом.

Основной задачей алгоритма сегментации является поиск повторяющихся шаблонов для оптимизации последующей классификации. Когда пользователь заходит на веб-сайт в браузере, серверный или клиентский скрипт вставляет результаты, полученные из базы данных, в заранее определенный фрагмент шаблона HTML-кода. Затем эти данные группируются и встраиваются в HTML-документ. Используя алгоритм, основанный на сопоставлении блоков страницы и поиске структурно схожих элементов, можно сегментировать эти блоки данных.

Например, в онлайн-магазинах часто можно увидеть карточки товаров, которые содержат изображение, название товара и цену. Эти элементы имеют схожую DOM-структуру, поэтому, сопоставив эти элементы, можно выделить эти блоки данных.

Сегментация происходит в несколько этапов:

1. На первом этапе происходит обход дерева в ширину, где каждый элемент получает два атрибута: уникальный идентификатор и количество дочерних элементов родителя элемента;

2. Следующий этап – сравнение элементов дерева. На этом этапе генерируется значение динамического окна от 1 до максимального значения, равного половине количества дочерних элементов родителя. Выбор такого максимального значения объясняется тем, что как минимум 2 элемента могут быть сопоставлены. Затем выбирается опорный элемент, на основе которого заполняются окна. Первое окно содержит элементы, идущие перед опорным, второе – элементы, включая опорный элемент и после него, третье – элементы, идущие после опорного. Затем структуры элементов второго и первого, второго и третьего окон сопоставляются путем приведения последовательно идущих тегов к строке вида `tagAtagB`. Если структура элементов второго окна совпала со структурой элементов другого окна, то элементы второго окна выводятся в качестве блоков.

3. На последнем этапе происходит генерация `css`-селекторов блоков, которые были получены на предыдущем шаге. Рекурсивно до тега `html` включительно для каждого элемента генерируется строка вида `nodeName:nth-child(nodePosition)`, где `nodeName` – тип тега, `nodePosition` – позиция среди соседних элементов, либо строка вида `nodeName`, где `nodeName` – название тега, для случаев, когда у узла нет соседних элементов. Затем эти строки конкатенируются между собой с помощью символа `">"`, указывающего на вложенность последующего элемента в предыдущий.

Для обучения модели обработки естественного языка было необходимо подготовить данные, содержащие семантически размеченные текстовые данные. В качестве источника был использован ресурс, подготовленный в рамках проекта `Web Data Commons` [17] и содержащий извлеченные семантически размеченные сущности из корпуса `Common Crawl 2022` [18].

Для извлечения данных, хранящихся в формате n-quad rdf, был использован язык SPARQL [19]. Для получения необходимых данных были составлены запросы для извлечения выделенных сущностей: Event, Book, Movie, Product, Hotel, Restaurant, JobPosting, Recipe.

Для каждой сущности было извлечено по 10 000 записей. Такой объем выбран с целью обеспечить достаточное разнообразие элементов для последующего обучения модели обработки естественного языка.

Для обработки естественного языка была выбрана мультиязычная BERT [20]. Для обучения модели был использован набор данных объемом 70 000 объектов онтологии Schema.org, извлеченный ранее из корпуса Common Crawl 2022 [17]. Набор данных был разделен на обучающий и валидационный наборы данных в соотношении 80% и 20% соответственно.

После этапа обучения был произведен замер показателей производительности модели (рис. 1). Для оценки качества модели работы были использованы следующие известные метрики: точность (precision), полнота (recall) и F-мера (f1-score).

	precision	recall	f1-score	support
Product	0.94	0.92	0.93	1001
Book	0.81	0.70	0.75	981
Event	0.84	0.88	0.86	962
Hotel	0.95	0.95	0.95	1015
JobPosting	1.00	0.98	0.99	972
Movie	0.73	0.83	0.78	1048
Recipe	0.94	0.89	0.91	1000
Restaurant	0.83	0.87	0.85	1021
accuracy			0.88	8000
macro avg	0.88	0.88	0.88	8000
weighted avg	0.88	0.88	0.88	8000

Рис. 1. Значения метрик для оценки производительности и качества предсказаний модели

Также для более детального анализа качества модели была построена матрица ошибок (confusion matrix). На пересечении строки и столбца с одним классом указан процент правильно идентифицированных объектов, а в прочих ячейках указаны проценты ошибочных классификаций (рис. 2).

Из полученной матрицы видно, что модель часто ошибочно идентифицирует класс Book как класс Movie и, наоборот, класс Movie как класс Book. Скорее всего, это связано с семантической близостью классов этих объектов. Часто могут встречаться фильмы и книги с одинаковыми или похожими названиями, при этом модель не учитывает контекст появления этих объектов и опирается только на название и описание объектов, из-за чего модель может ошибочно классифицировать эти объекты.

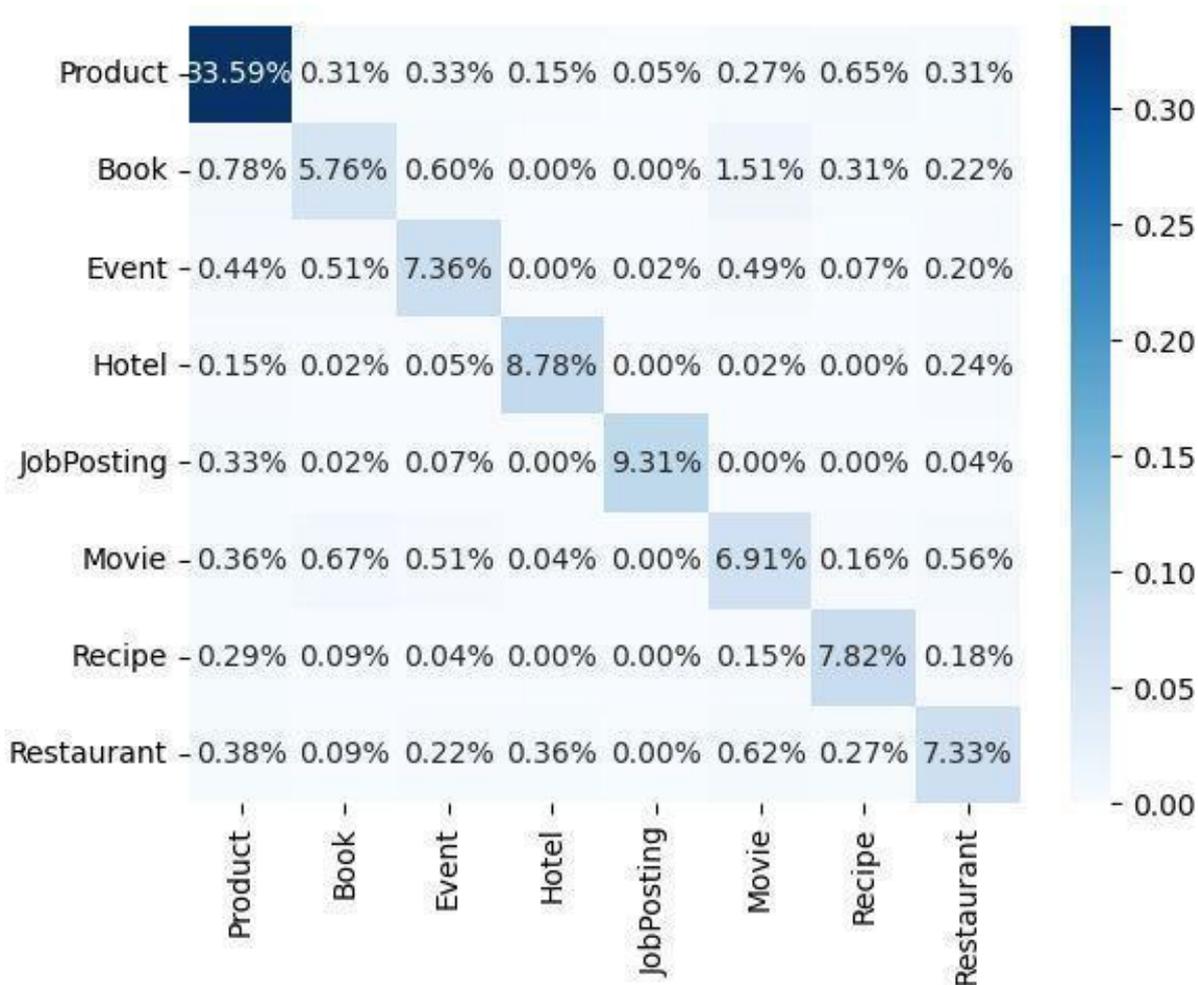


Рис. 2. Матрица ошибок (confusion matrix) модели

Существует большая вариативность выбора технологий для создания веб-страниц, включая различные шаблонизаторы, синтаксис JSX и т. д. Несмотря на то, что возможности этих технологий во многом совпадают, чтобы перенести данные из страницы с автоматически сгенерированной семантической разметкой, используя ту или иную технологию, необходимо учитывать синтаксис, специфический для этой технологии. Поэтому этап автоматического аннотирования был разделен на два независимых механизма: генерация семантической разметки страницы и перенос данных, реализация которого может отличаться для различных технологий. Поскольку JSX уже стал стандартом в современной разработке интерфейсов, эта технология была выбрана для реализации переноса данных.

Для переноса семантической разметки из исходной страницы в JSX сначала необходимо построить абстрактное синтаксическое дерево (AST-дерево) и сопоставить его с исходной страницей, и в случае, если они совпадают, аннотировать соответствующие узлы в JSX.

В качестве парсера для получения AST-дерева были использованы компилятор Babel [21] и язык TypeScript [22]. Обязательным условием для парсинга является наличие экспорта по умолчанию для функции, содержащей JSX. В AST-дереве, помимо обычных тегов, могут встречаться выражения, отличные от объявления тегов, например, объявление переменной. Но поскольку HTML-дерево не содержит этих объявлений, необходимо построить преобразованное дерево, полностью копирующее структуру AST-дерева без выражений, отличных от объявлений HTML-элементов. Для этого была использована рекурсивная функция, которая на вход получает текущий узел и массив, агрегирующий дочерние элементы, в который необходимо передать AST-узел с названием тега, номером строки, на которой появляется тег, номером первого символа в строке и набором семантических атрибутов. Таким образом, на выходе получается структура, полностью копирующая HTML-дерево. Если структуры деревьев совпадут, то происходит генерация групп, содержащих семантическую разметку, которую необходимо внедрить. Группа характеризуется местоположением для внедрения семантических атрибутов и набором связанных с этим узлом семантических атрибутов. Для генерации групп производится последовательный обход HTML-дерева. Если при обходе узла встречаются один или несколько семантических атрибутов, то в

качестве номера строки сохраняется строка, в которой находится этот узел в исходном файле, а в качестве номера символа в строке указывается номер символ в строке, находящийся перед закрывающим символом тега ">" На следующем этапе происходит внедрение разметки на основе групп, сгенерированных на предыдущем этапе. По номеру строки и номеру символа в строке происходит вставка подготовленной строки, содержащей необходимую разметку.

ЗАКЛЮЧЕНИЕ

Результатом работы является программное приложение для автоматического аннотирования по стандарту Microdata, построенное с применением технологий машинного обучения. Им могут пользоваться авторы веб-страниц для улучшения позиций при ранжировании страниц поисковыми системами. Приложение можно доработать, добавив поддержку формата JSON-LD и расширив список технологий, позволяющих его использовать.

СПИСОК ЛИТЕРАТУРЫ

1. HTML5 (HyperText Markup Language).
URL: <https://html.spec.whatwg.org/multipage/introduction.html>.
 2. JSX. URL: <https://www.typescriptlang.org/docs/handbook/jsx.html>.
 3. Microdata.
URL: <https://html.spec.whatwg.org/multipage/microdata.html>.
 4. JSON-LD. URL: <https://json-ld.org>.
 5. *Brinkmann A., Primpeli A., Bizer Ch.* The Web Data Commons Schema.org Data Set Series.
URL: https://www.uni-mannheim.de/media/Einrichtungen/dws/Files_Research/Web-based_Systems/pub/Brinkmann-etal-TheWDCSchemaorgDataSetSeries-WWW2023.pdf.
 6. Schemas. URL: <https://schema.org/docs/schemas.html>.
 7. RDFa. URL: <https://www.w3.org/TR/html-rdfa>.
 8. Microformats. URL: <https://microformats.org>.
 9. Local Business Schema Generator – MicroData & JSON-LD.
URL: <https://microdatagenerator.org/localbusiness-microdata-generator>.
-

10. Structured Data Markup Helper. URL: <https://www.google.com/webmasters/markup-helper/u/0>.
11. Entity SEO Tools. URL: <https://inlinks.com/>.
12. Web-segment. URL: <https://github.com/liaocyintl/web-segment>.
13. *Utiu N., Ionescu V.-S.* Learning Web Content Extraction with DOM Features. URL: <http://dx.doi.org/10.1109/ICCP.2018.8516632>.
14. *Peters M. E., Lecocq D.* Content extraction using diverse feature sets // WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, May 13–17, 2013. Association for Computing Machinery, New York, NY, United States: 2013, pages 89–90.
15. *Gongqing Wu, Li Li, Xuegang Hu, Xindong Wu* Web news extraction via path ratios. URL: <https://dl.acm.org/doi/abs/10.1145/2505515.2505558>.
16. *Vadrevu S., Gelgi F., Davulcu H.* Semantic partitioning of web pages // Web Information Systems Engineering–WISE 2005: 6th International Conference on Web Information Systems Engineering, New York, NY, USA, November 20–22, 2005. Proceedings 6. – Springer Berlin Heidelberg, 2005. P. 107–118.
17. Extraction Results from the October 2022 Common Crawl Corpus. URL: <https://webdatacommons.org/structureddata/#results-2022-1>.
18. Common Crawl September/October 2022 Crawl Archive (CC-MAIN-2022-40). URL: <https://data.commoncrawl.org/crawl-data/CC-MAIN-2022-40/index.html>.
19. SPARQL Query Language. URL: <https://www.w3.org/TR/sparql11-query>.
20. BERT: <https://arxiv.org/abs/1810.04805>.
21. Babel. URL: <https://babeljs.io/>.
22. TypeScript. URL: <https://www.typescriptlang.org/>.

AUTOMATIC ANNOTATION OF HTML DOCUMENTS USING THE MICRO-DATA STANDARD

T. F. Ibragimov¹ [0009-0008-4959-5259], A. A. Ferenets² [0000-0002-7859-9901]

^{1,2}Institute of Information Technology and Intelligent Systems, Kazan Federal University

¹i.timur0701@gmail.com, ²ist.kazan@gmail.com

Abstract

The development of an application based on machine learning methods for automatic annotation of web pages according to the Microdata standard is described, with the possibility of extension to other standards and injecting data to JSX files. Datasets were collected and prepared for training Machine Learning (ML) models. The ML model metrics were collected and analyzed.

Keywords: *Microdata, semantic markup, HTML5, search engine optimization (SEO), search engines, machine learning, schema.org, semantic web, markup standards, SEO automation.*

REFERENCES

1. HTML5 (HyperText Markup Language).
URL: <https://html.spec.whatwg.org/multipage/introduction.html>.
2. JSX. URL: <https://www.typescriptlang.org/docs/handbook/jsx.html>.
3. Microdata.
URL: <https://html.spec.whatwg.org/multipage/microdata.html>.
4. JSON-LD. URL: <https://json-ld.org>.
5. *Brinkmann A., Primpeli A., Bizer Ch.* The Web Data Commons Schema.org Data Set Series.
URL: https://www.uni-mannheim.de/media/Einrichtungen/dws/Files_Research/Web-based_Systems/pub/Brinkmann-et-al-TheWDCSchemaorgDataSetSeries-WWW2023.pdf.
6. Schemas. URL: <https://schema.org/docs/schemas.html>.
7. RDFa. URL: <https://www.w3.org/TR/html-rdfa>.
8. Microformats. URL: <https://microformats.org>.

9. Local Business Schema Generator – MicroData & JSON-LD.
URL: <https://microdatagenerator.org/localbusiness-microdata-generator>.
10. Structured Data Markup Helper. URL: <https://www.google.com/webmasters/markup-helper/u/0>.
11. Entity SEO Tools. URL: <https://inlinks.com/>.
12. Web-segment. URL: <https://github.com/liaocyintl/web-segment>.
13. *Utiu N., Ionescu V.-S.* Learning Web Content Extraction with DOM Features. URL: <http://dx.doi.org/10.1109/ICCP.2018.8516632>.
14. *Peters M. E., Lecocq D.* Content extraction using diverse feature sets // WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, May 13–17, 2013. Association for Computing Machinery, New York, NY, United States: 2013, pages 89–90.
15. *Gongqing Wu, Li Li, Xuegang Hu, Xindong Wu* Web news extraction via path ratios. URL: <https://dl.acm.org/doi/abs/10.1145/2505515.2505558>.
16. *Vadrevu S., Gelgi F., Davulcu H.* Semantic partitioning of web pages // Web Information Systems Engineering–WISE 2005: 6th International Conference on Web Information Systems Engineering, New York, NY, USA, November 20–22, 2005. Proceedings 6. – Springer Berlin Heidelberg, 2005. P. 107–118.
17. Extraction Results from the October 2022 Common Crawl Corpus.
URL: <https://webdatacommons.org/structureddata/#results-2022-1>.
18. Common Crawl September/October 2022 Crawl Archive (CC-MAIN-2022-40). URL: <https://data.commoncrawl.org/crawl-data/CC-MAIN-2022-40/index.html>.
19. SPARQL Query Language. URL: <https://www.w3.org/TR/sparql11-query>.
20. BERT: <https://arxiv.org/abs/1810.04805>.
21. Babel. URL: <https://babeljs.io/>.
22. TypeScript. URL: <https://www.typescriptlang.org/>.

СВЕДЕНИЯ ОБ АВТОРАХ



ИБРАГИМОВ Тимур Фердинандович – выпускник бакалавриата Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

Timur Ferdinandovich IBRAGIMOV – graduate student of the Institute of Information Technologies and Intelligent Systems, Kazan (Volga region) Federal University, Kazan.

Email: i.timur0701@gmail.com

ORCID: 0009-0008-4959-5259



ФЕРЕНЕЦ Александр Андреевич – старший преподаватель кафедры программной инженерии Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

Alexander Andreevich FERENETS – senior lecturer of Software Engineering of Institute of Information Technologies and Intelligent Systems KFU.

Email: ist.kazan@gmail.com

ORCID: 0000-0002-7859-9901

Материал поступил в редакцию 20 июля 2024 года

УДК 004.85

ЧИСЛОВЫЕ ИНСТРУМЕНТЫ И ВИРТУАЛЬНЫЕ АССИСТЕНТЫ ДЛЯ ПОДДЕРЖКИ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

В. В. Наумова¹ [0000-0002-3001-1638], А. А. Загумёнов² [0000-0002-0501-5362],

В. С. Ерёменко³ [0000-0002-5250-5743]

^{1,3}ФГБУН Государственный геологический музей им. В.И. Вернадского РАН,
Москва;

²ФГБУН Институт автоматизи и процессов управления ДВО РАН,
Владивосток;

¹naumova_new@mail.ru, ²trueepikvic@gmail.com, ³vitaer@gmail.com

Аннотация

Описаны основные этапы применения информационных технологий для научных исследований в геологии. Рассмотрен ряд цифровых технологий ближайшего будущего, которые в настоящее время имеют перспективы применения в геологии. Первые результаты работ в *Государственном геологическом музее (ГГМ) им. В.И. Вернадского РАН* показали перспективность применения очередных шагов развития ИТ-технологий при создании информационно-вычислительных систем для поддержки геологических исследований.

Ключевые слова: *искусственный интеллект, числовые инструменты, виртуальные ассистенты, интеллектуальные агенты, поддержка научных исследований в геологии.*

ВВЕДЕНИЕ

Математические методы для анализа геологических данных используются достаточно давно. Современное состояние математических методов в геологии отражено в десятках монографий и сотнях публикаций, из которых следует особо отметить работы Д.А. Родионова, В.Н. Бондаренко, А.Б. Каждана, Н.Н. Боровко, Р.И. Дубова. Среди зарубежных авторов отметим тех, чьи работы переведены на русский язык: Ч. Крамбейн, Д. Лоули, А. Максвелл, Ж. Матерон, Р.Л. Миллер, Д.Г. Криге, Дж.С. Девис.

Среди программного обеспечения для анализа числовых данных, используемого геологами-исследователями, можно отметить такие пакеты прикладных программ для ПК, как Statistica, BMDP, StatSoft. Для анализа пространственных данных – ArcGis, CorelDraw, Adobe, Surfer и др.

В связи с тем, что в последние годы для решения научных геологических задач вычисления становятся возможными в интернете, геологи используют облачное хранение данных, стандартные и специализированные программные платформы. В этот период отмечено широкое распространение территориально распределенных информационно-вычислительных систем, создаваемых геологическими службами, университетами, научными организациями, библиотеками и естественнонаучными музеями.

Актуальной задачей становится обеспечение открытого доступа к совместимым цифровым данным и программным системам. В качестве примеров можно привести следующие информационно-вычислительные системы:

Digital Earth Australia (<http://www.ga.gov.au/dea/home>). Digital Earth Australia (DEA) – это реализация правительством Австралии платформы анализа с открытым исходным кодом, разработанной в рамках инициативы Open Data Cube (ODC). Программа DEA предоставляет код, документацию, руководства по использованию, учебные пособия и поддержку для международных пользователей Open Data Cube.

U.S. Geoscience Information Network (USGIN) (<http://usgin.org>). Основная цель USGIN: облегчить открытый доступ к совместимым цифровым данным и программному обеспечению в науках о Земле. USGIN-стандарты, протоколы и задачи – наследие National Geothermal Data System (NGDS), системы совместного использования данных, которая обеспечивает доступ к информации о геотермических ресурсах.

Ausgin – Австралийская информационная сеть в области наук о Земле (<http://www.geoscience.gov.au>). В ней широко используются веб-сервисы – главным образом, веб-картографические сервисы (WMS), но также Web Feature Services (WFS) и Web Coverage Services (WCS).

OpenGeoscience BGS. Британская геологическая служба имеет широкий спектр наборов данных и постоянно расширяет доступ к ним к ним, публикуя

большое количество данных на портале OpenGeoscience BGS (<http://www.bgs.ac.uk/opengeoscience>). OpenGeoscience – бесплатный сервис, где можно просматривать карты, загружать данные, сканировать фотографии и другую информацию.

Современное геологическое программное обеспечение с открытым исходным кодом [1]:

QGIS — это надежная и удобная в использовании географическая информационная система (ГИС), пользующаяся популярностью у геологов по всему миру. Обширная библиотека плагинов делает QGIS невероятно универсальным для широкого спектра геологических приложений.

GemPy — инструмент 3D-геологического моделирования с использованием Python Power. Для геологов, знакомых с Python, GemPy предлагает мощный способ создания сложных 3D-геологических моделей. Подход к неявному моделированию и вероятностные особенности делают его идеальным для решения сложных геологических сценариев.

Разработанная при поддержке Geoscience Australia, Loop представляет собой комплексную платформу для построения, визуализации и анализа 3D геологических и геофизических моделей.

Acquire GIM Suite — программное обеспечение для управления геологическими данными, разработанное для эффективного сбора, управления геологическими данными и доступа к ним.

Datamine Fusion объединяет все данные по разведке и добыче в безопасную и доступную базу данных; поддерживает комплексное управление данными по геологии, геохимии и геотехнике.

Plexer — это платформа управления геологическими данными, предназначенная для централизации и управления всеми данными, относящимися к геологоразведке и добыче полезных ископаемых. Основное внимание уделяется доступу к данным в режиме реального времени с удобным интерфейсом.

Leapfrog Geo — это программное обеспечение для 3D-геологического моделирования, разработанное Seequent. Оно предназначено для предоставления геологам интуитивно понятных и мощных инструментов для построения и интерпретации сложных геологических моделей.

Для анализа спутниковых данных одним из лидеров среди облачных сервисов является An Open Source Geospatial Data Management & Analysis Platform (<https://www.opendatacube.org>). Open Data Cube (ODC) — это некоммерческий проект с открытым исходным кодом, который был создан для обеспечения доступа, управления и анализа больших объёмов данных ГИС по мониторингу Земли. Он представляет собой общий аналитический фреймворк, содержащий наборы структурированных данных и инструментов, которые позволяют проводить анализ больших коллекций пространственных данных. ODC был разработан для анализа пространственных данных на больших временных промежутках, однако его можно использовать на любых наборах пространственных данных. Данные могут представлять собой модели высот, геофизические сетки, интерполированные поверхности и т. д. Ключевой особенностью ODC является возможность сохранения каждого уникального элемента набора пространственных данных, в отличие от многих других методов работы с большими коллекциями пространственных данных.

Следующий значительный этап в развитии информационных технологий — применение методов и технологий искусственного интеллекта.

Далее рассмотрим ряд перспективных цифровых технологий, способных радикально изменить развитие сразу многих областей, которые в настоящее время уже применяются или имеют перспективы применяться в геологических исследованиях.

ЦИФРОВЫЕ ИНСТРУМЕНТЫ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

Цифровые инструменты — это программное обеспечение, приложения и устройства, используемые для работы с цифровыми данными. Они позволяют пользователям создавать, редактировать, хранить, обрабатывать и передавать информацию в цифровом формате.

Уже сегодня для широкого применения доступны цифровые инструменты, которые ещё несколько десятилетий назад казались недостижимым будущим.

Появилось множество сервисов и чат-ботов на основе ChatGPT и других нейросетей, которые помогают в самых разных сферах и задачах. Чаще всего

поручают ChatGPT как **персональному помощнику**: поиск и обобщение информации из интернета; проверку кода; роль персонального репетитора английского; интерпретацию и пересказ фильмов, книг и сериалов; составление индивидуальных подборок по интересам; разработку программы домашних тренировок; мониторинг вакансий, авиабилетов и т. д.; генерацию текстов, ответов на электронные письма, поздравлений; разъяснение сложных или обширных тем.

ИНТЕЛЛЕКТУАЛЬНЫЕ АГЕНТЫ В ГЕОЛОГИИ

Агенты искусственного интеллекта (ИИ) – это программное обеспечение, которое умеет взаимодействовать с окружающей средой, собирать данные и на их основе самостоятельно определять и выполнять задачи, позволяющие добиться заранее определенных целей.

Примеры некоторых областей применения агентов на основе искусственного интеллекта в геологии:

- *Анализ данных.* Методы машинного обучения позволяют эффективно обрабатывать сейсмические данные, данные о составе пород, спутниковые снимки и выявлять скрытые закономерности.
- *Моделирование геологических процессов.* Например, прогнозирование землетрясений, моделирование распространения волн цунами и анализ движения тектонических плит.
- *Предсказание катастроф.* ИИ позволяет улучшить точность предсказаний природных катастроф, анализируя исторические данные и выявляя признаки, предшествующие событиям.
- *Интерпретация изображений.* Автоматическое распознавание и классификация изображений, полученных с помощью спутников и дронов, ускоряет процесс картирования и анализа геологических объектов.
- *Автоматизация картирования.* С помощью алгоритмов машинного обучения можно анализировать спутниковые снимки и данные лидара для определения характеристик рельефа, типов почв и других геологических особенностей.

- Минералогия. ИИ применяется для классификации минералов и прогнозирования их местоположения. Машинное обучение помогает анализировать данные о составе пород и их физических свойствах для определения вероятных месторождений полезных ископаемых.

Также ИИ может быть использован для создания трёхмерных моделей геологических структур и их визуализации. Это позволяет более точно определять распределение полезных ископаемых, оценивать риски геологических процессов и принимать обоснованные решения при разработке месторождений.

ВИРТУАЛЬНЫЕ АССИСТЕНТЫ ГЕОЛОГА НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Виртуальные ассистенты на основе ИИ – это программные агенты, которые могут выполнять задачи (или сервисы) для пользователя на основе информации, введенной пользователем, а также информации, полученной из различных интернет-ресурсов.

Примеры некоторых современных цифровых инструментов и виртуальных ассистентов для научных исследований в геологии [2]:

1. *Цифровой геолог*. ПО, разработанное специалистами Университета МИСИС, помогает в поиске чёрных и цветных металлов, фосфатных руд и других полезных ископаемых. Алгоритм использует данные, полученные в ходе разведки и эксплуатации месторождения.

2. *Сервис для распознавания и классификации горных пород по фотографиям*. Его создали исследователи Сколтеха. Сервис позволяет увеличить объём данных и в 20 раз ускорить анализ.

3. *Проект «Цифровой керн»*. Это цифровая лаборатория исследований керна материала. Полученные данные позволяют выполнять достоверную оценку запасов месторождения, снижая различные риски и повышая доходность проектов.

4. *Инструмент на основе технологий компьютерного зрения*. Он может анализировать и интерпретировать огромное количество микроскопических фотоснимков срезов пласта. Компьютер сам находит и выделяет нужные

сегменты на изображении породы, отмечая все важные показатели и свойства.

ПРОЕКТ ГГМ РАН В ОБЛАСТИ РАЗРАБОТКИ ЧИСЛОВЫХ ИНСТРУМЕНТОВ, ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ И ВИРТУАЛЬНЫХ АССИСТЕНТОВ ДЛЯ ПОДДЕРЖКИ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

На основе развиваемой Информационно-аналитической среды для поддержки научных исследований в геологии (<https://geologyscience.ru>) [3], в ГГМ РАН, осуществляется разработка числовых инструментов, интеллектуальных агентов и виртуальных ассистентов, которые могут быть использованы как для этой Среды, так и для других систем.

Основная цель данного Проекта заключается в организации единой точки доступа к геологическим данным на территорию России и системам их обработки и анализа с использованием возможностей сервисов поиска данных в территориально распределенных разнородных источниках, а также с использованием цифровых инструментов в виде различных сервисов анализа разнородной информации [4].

Список цифровых инструментов (сервисов) для <https://geologyscience.ru>:

Сервисы поиска информации: текстовый, географической, музейный.

Коллекция сервисов текстового поиска:

- по международной библиотеке DataCite;
- по Порталу открытых данных ГГМ РАН;
- по Государственным геологическим картам (ВСЕГЕИ);
- по Цифровому репозиторию научных статей по геологии ГГМ РАН;
- по Энциклопедии «Геология России» ГГМ РАН.

Сервисы анализа таблиц количественных данных [5], а также доступ к интерактивному облачному сервису Excel (Microsoft) и сервису анализа петролого-геохимических данных (ИФЗ РАН):

Сервисы визуализации количественных данных.

Сервисы с использованием искусственного интеллекта:

- Сервисы нейросетевого анализа геологических текстов;
- Чат-бот по месторождениям РФ [6];

- Чат-бот создания программного кода.

Вышеперечисленные сервисы являются самостоятельными веб-сервисами, имеющими возможности по их интеграции в различные вычислительные среды и информационные системы [7].

На основе созданного кабинета пользователя <https://geologyscience.ru> разрабатывается цифровой ассистент геолога-исследователя [8] (рис. 1).

Цифровой ассистент на основе ИИ — это система автоматизации взаимодействия с пользователем в диалоговом формате. Сервис осуществляет текстовые и голосовые консультации, обработку заявок и поддержку по направлениям деятельности Портала.

Основная цель цифрового ассистента для пользователя Портала <https://geologyscience.ru> – сопровождение и поддержка научных исследований в геологии с использованием данных и сервисов Информационно-аналитической среды ГГМ РАН в интерактивном режиме на естественном языке.

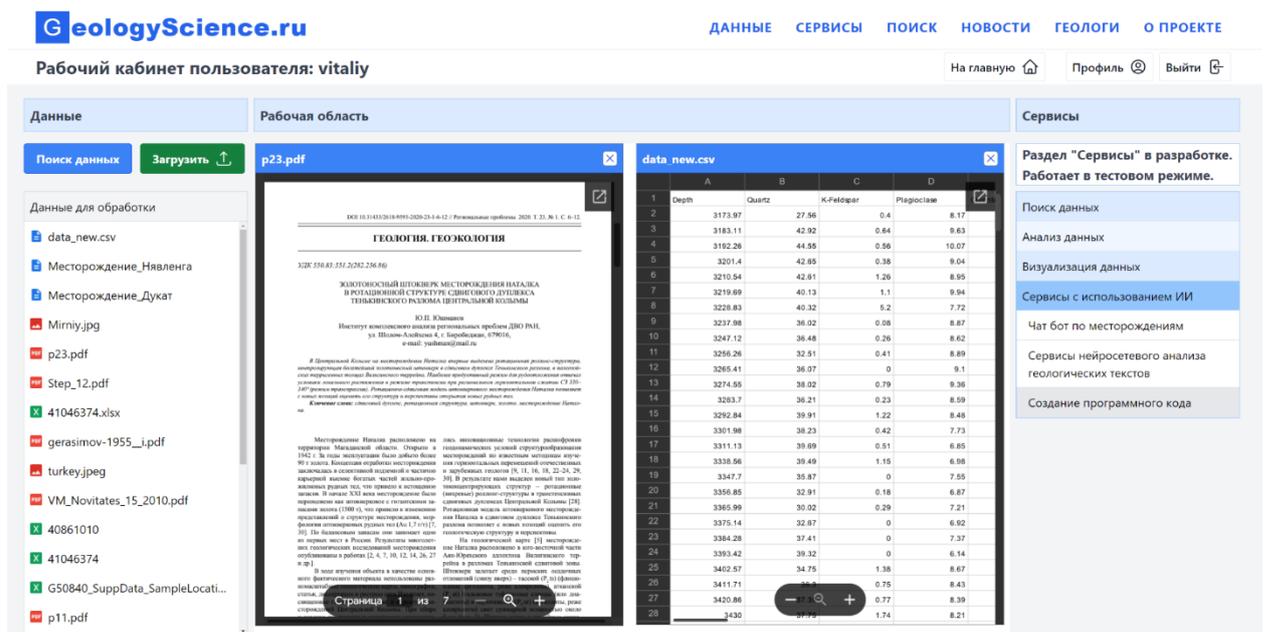


Рис. 1. Виртуальный кабинет геолога-исследователя

Базовой основой цифрового ассистента геолога-исследователя на Портале <https://geologyscience.ru> является разрабатываемый чат-бот по постановке геологической задачи, определению необходимой информации и выбора сервисов для ее решения.

Основные функции виртуального ассистента на основе больших языковых моделей:

1. Обработка сообщений;
2. Постановка задачи;
3. Формулировка заданий агентам;
4. Обработка результата;
5. Генерация ответа.

В свою очередь агенты:

- Имитируют деятельность человека;
- Взаимодействуют со средой (другим человеком или агентами);
- Имеют восприятие – способность обрабатывать данные разных модальностей;
- Имеют память и доступ к базе данных и знаний;
- Могут совершать набор действий (взаимодействие с внешним миром, памятью, базой знаний, обработка данных).
- Могут быть выстроены в сеть или цепочку агентов

ЗАКЛЮЧЕНИЕ

- Рассмотрен ряд цифровых технологий ближайшего будущего, которые в настоящее время имеют перспективы применения в геологии.
- Первые результаты работ показали перспективность применения очередных шагов развития ИТ-технологий при создании информационно-вычислительных систем для поддержки геологических исследований.
- Показана возможность интеграции сервисов на основе ИИ с существующими данными и классическими сервисами с программным или пользовательским интерфейсами без необходимости их существенной доработки.

В настоящее время осуществляется разработка первой версии открытой программной платформы интеграции данных, сервисов и тематических агентов на основе искусственного интеллекта – основы для цифрового ассистента геолога-исследователя.

Благодарности

Работа выполняется в рамках Государственной темы «Цифровая платформа интеграции и анализа геологических и музейных данных», тема № 1021061009468-8-1.5.1.

СПИСОК ЛИТЕРАТУРЫ

1. Обзор CorePlan. URL: <https://www.coreplan.io/blog/top-12-geological-databases-for-mineral-exploration-in-2024>

2. Разведка, добыча и прогнозирование: как цифровые решения помогают геологам.

URL: <https://sber.pro/digital/publication/razvedka-dobicha-i-prognozirovanie-kak-tsifrovie-resheniya-pomogayut-geologam/?ysclid=m35mgi851607776006>

3. *Наумова В.В., Еременко В.С., Загуменнов А.А., Патук М.И.* Научный портал Geologyscience.ru: текущее состояние и перспективы развития // Геоинформатика. 2023. № 3. С. 33–43.

4. *Наумова В.В., Патук М.И., Ерёменко А.С., Загумённов А.А., Ерёменко В.С.* Цифровые технологии будущего — современные решения в науках о Земле // Геоинформатика. 2024. № 3. С. 53–63.

5. *Загуменнов А.А., Наумова В.В., Еременко В.С.* Облачный сервис многомерной обработки количественных данных для решения геологических задач // Вестник НГУ. Серия: Информационные технологии. 2021. Т. 19, №3. С. 40–49.

6. *Загуменнов А.А., Наумова В.В.* Сервисы на основе методов искусственного интеллекта для поддержки научных исследований в геологии // Материалы Международной конференции «Марчуковские научные чтения 2024» (МНЧ-2024), 7–11 октября 2024 г., Академгородок, Новосибирск, Россия. С. 125.

7. *Aleksey A. Zagumennov, Vera V. Naumova.* Development of the computing node for processing satellite imagery and spatial data for earth sciences // Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021). Novosibirsk, Russia, August 24–27, 2021. P. 272–279.

8. Ерёмченко В.С., Наумова В.В., Загумёнов А.А. Разработка виртуальных пространств с интеллектуальным адаптивным интерфейсом для цифровых геологических систем // Информационные и математические технологии в науке и управлении. 2024. № 1 (33). С. 146–153.

DIGITAL TOOLS AND VIRTUAL ASSISTANTS TO SUPPORT SCIENTIFIC RESEARCH IN GEOLOGY

V.V. Naumova¹ [0000-0002-3001-1638], A.A. Zagumennov² [0000-0002-0501-5362],

V.S. Eremenko³ [0000-0002-5250-5743]

^{1,3}Vernadsky State Geological Museum of Russian Academy of Sciences, Moscow;

²Institute of Automation and Control Processes FEB RAS, Vladivostok;

¹naumova_new@mail.ru, ²truepikvic@gmail.com, ³vitaer@gmail.com

Abstract

The article describes the main stages of the application of information technologies for scientific research in geology. A number of digital technologies of the near future, which currently have prospects for application in geology, are considered. The first results of the work at the GGM RAS showed the prospects of applying the next steps in the development of OT technologies in the creation of information and computing systems to support geological research.

Keywords: *artificial intelligence, numerical tools, virtual assistants, intelligent agents, support for scientific research in geology.*

REFERENCES

1. CorePlan Review.

URL: <https://www.coreplan.io/blog/top-12-geological-databases-for-mineral-exploration-in-2024>

2. Razvedka, добыча i prognozirovanie: kak cifrovye resheniya pomagajut geologam.

URL: <https://sber.pro/digital/publication/razvedka-dobicha-i-prognozirovanie-kak-tsifrovie-resheniya-pomogayut-geologam/?ysclid=m35mgi851607776006>

3. *Naumova V.V., Eremenko V.S., Zagumennov A.A., Patuk M.I.* Nauchnyj portal Geologyscience.ru: tekushhee sostojanie i perspektivy razvitija // Geoinformatika. 2023. № 3. S. 33–43.

4. *Naumova V.V., Patuk M.I., Eremenko A.S., Zagumennov A.A., Eremenko V.S.* Cifrovye tehnologii budushhego — sovremennye reshenija v naukah o Zemle // Geoinformatika. 2024. № 3. S. 53–63.

5. *Zagumennov A.A., Naumova V.V., Eremenko V.S.* Oblachnyj servis mnogomernoj obrabotki kolichestvennyh dannyh dlja reshenija geologicheskikh zadach // Vestnik NGU. Serija: Informacionnye tehnologii. 2021. T. 19, №3. C. 40–49.

6. *Zagumennov A.A., Naumova V.V.* Servisy na osnove metodov iskusstvennogo intellekta dlja podderzhki nauchnyh issledovanij v geologii // Materialy Mezhdunarodnoj konferencii «Marchukovskie nauchnye chtenija 2024» (MNCh-2024), 7–11 oktjabrja 2024 g, Akademgorodok, Novosibirsk, Russia. S. 125.

7. *Aleksey A. Zagumennov, Vera V. Naumova.* Development of the computing node for processing satellite imagery and spatial data for earth sciences // Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021). Novosibirsk, Russia, August 24–27, 2021. P. 272–279.

8. *Eremenko V.S., Naumova V.V., Zagumennov A.A.* Razrabotka virtual'nyh prostranstv s intellektual'nym adaptivnym interfejsom dlja cifrovyh geologicheskikh sistem // Informacionnye i matematicheskie tehnologii v nauke i upravlenii. 2024. № 1 (33). 146–153.

СВЕДЕНИЯ ОБ АВТОРАХ



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

Email: naumova_new@mail.ru,

ORCID: 0000-0002-3001-1638

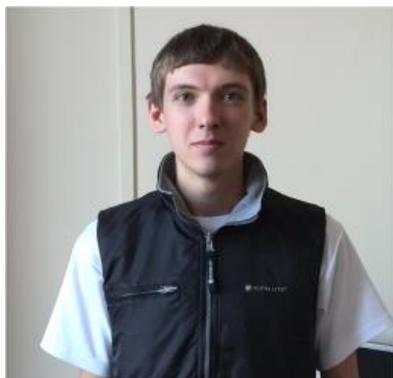


ЗАГУМЁННОВ Алексей Андреевич – программист (внештатный сотрудник), Государственный Геологический музей им. В.И. Вернадского РАН, Москва; младший научный сотрудник, Институт автоматизации и процессов управления ДВО РАН, Владивосток.

Aleksey A. ZAGUMENNOV – contract programmer of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow; Junior researcher of Institute of automatic and control processes of FEB RAS, Vladivostok.

Email: truepikvic@gmail.com,

ORCID: 0000-0002-0501-5362



ЕРЁМЕНКО Виталий Сергеевич – научный сотрудник, Государственный геологический музей им. В.И. Вернадского РАН, Москва.

Vitaliy S. EREMENKO –researcher of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

Email: vitaer@gmail.com,

ORCID: 0000-0002-5250-5743

Материал поступил в редакцию 25 сентября 2024 года

УДК 004.85

ИСПОЛЬЗОВАНИЕ СЕМАНТИЧЕСКОГО ПОИСКА ДЛЯ ВЫБОРА И РАНЖИРОВАНИЯ НАУЧНЫХ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ

М. И. Патук¹ [0000-0003-3036-2275], В. В. Наумова² [0000-0002-3001-1638]

ФГБУН Государственный геологический музей им. В.И. Вернадского РАН,
Москва;

¹patuk@mail.ru, ²naumova_new@mail.ru

Аннотация

Агрегирование научной информации играет важную роль для комплексного анализа геологических объектов. В настоящей работе мы рассматриваем потенциал и возможности семантического поиска для выбора тематически близких геологических публикаций. Проанализированы различные языковые модели в контексте нахождения сходства и различия между текстами при описании месторождений полезных ископаемых. Показано значительное улучшение результатов поиска после дополнительной тренировки языковых моделей. Представлены два веб-сервиса, основанных на методе расчета семантической близости текстов с количественной оценкой меры близости.

Ключевые слова: искусственный интеллект, машинное обучение, обработка естественного языка, семантический поиск, геология.

ВВЕДЕНИЕ

Научные геологические публикации состоят, как правило, из трех типов информации: текста, числовых данных и пространственных данных. Основой для анализа служат, в основном, числовые данные. Реже анализу подвергаются пространственные данные, например, карты, схемы, разрезы, спутниковые изображения и др. Сам текст публикаций содержит анализ, выполненный автором. Но геологические объекты, которые описаны в публикациях, являются комплексными образованиями. И каждая отдельная статья отражает лишь некоторые свойства описываемого объекта.

Таким образом, необходим числовой инструмент для выбора близких по изучаемому объекту публикаций. Такой инструмент уже существует – это семантический поиск, при котором способ и технология поиска информации основаны на использовании контекстного (смыслового) значения запрашиваемых фраз, вместо словарных значений отдельных слов или выражений при поисковом запросе [1]. Этот подход активно развивается в последнее десятилетие в связи с бурным развитием методов обработки естественного языка (NLP) [2]. Одна из задач обработки естественного языка – семантическая схожесть текстов (Semantic Text Similarity, STS) [3]. Это направление достаточно хорошо освещено в литературе. Основным импульс развитию семантического поиска дал поиск информации в интернете, а именно, такие поисковые системы, как Google и Yandex.

Ключевой вехой на этом пути стал выпуск большой языковой модели BERT [4]. BERT – это языковая модель, основанная на архитектуре трансформер [5] и предназначенная для предобучения языковых представлений с целью их последующего применения для решения широкого спектра задач обработки естественного языка [6]. В настоящее время только на портале HuggingFace насчитывается 15 различных BERT моделей [7], а с учетом клонов количество BERT-подобных моделей насчитывает несколько сотен.

Анализ литературных источников

До наступления эры больших языковых моделей семантическая близость текстов определялась, в основном, на основе словарей и онтологий, формальных описаний терминов предметной области и отношений между ними. Хороший обзор этих методов приведен в работе [8].

Следующим этапом в семантическом поиске можно считать применение таких методов, как TF-IDF [9], BM25 [10] и Word2Vec [11]. От лексического подхода предыдущих методов данные методы отличает использование векторных представлений текстов. Хотя векторное представление слов в этих методах является статичным и не зависит от контекста, они до сих пор активно применяются для поиска информации [12]. Отсутствие зависимости от контекста

в методе Word2Vec может компенсироваться построением расширенного запроса с применением синонимов [13]. Мейнстримом в данной области является использование больших языковых моделей типа BERT, дополненных различными методиками выбора и обработки текста [14, 15], хотя некоторые авторы считают, что поиск с помощью больших языковых моделей следует дополнить использованием ключевых слов [16].

Методика работы

Анализ литературы по семантическому поиску информации с использованием больших языковых моделей типа BERT позволяет принять следующую схему работы:

1. На портале HuggingFace выбираем подходящую языковую модель.
2. Создаем обучающий набор данных (датасет).
3. Дообучаем (fine-tuning) модель на созданном наборе данных.
4. Производим оценку полученного результата.
5. Применяем выбранную языковую модель для поиска информации.

Метод поиска, близкий к описанному выше, используется в электронной библиотеке Elibrary [17]. Насколько можно судить, он основан на языковой модели sci-rus-tiny [18], для обучения которой были использованы данные этой библиотеки [19]. Такой поиск присутствует на основной странице в виде пункта «Нейропоиск» и на странице конкретной публикации в виде пункта «Найти близкие по тематике публикации». Проверка работы этого поиска на запросах геологической тематики показала, что он хорошо находит основную тематику запроса, но представляет слишком широкие результаты в рамках этой тематики. Несколько обобщенные результаты поиска становятся понятны, если обратить внимание на структуру обучающего набора данных. Все естественные науки составляют в нем 11%, а геология, естественно, еще меньше.

Модельными геологическими объектами в нашей работе были выбраны месторождения твердых полезных ископаемых. Подбирались статьи, относящиеся к их описанию, с сайтов интернет-библиотек Elibrary [17] и CyberLeninka

[20], на русском и английском языках. Обучающий набор данных был создан на основе пар наименований статей и пар абстрактов с указанием степени схожести в интервале от 0 (совсем не похожи) до 5 (идентичные).

Дообучение (fine-tuning) моделей выполнялось по следующей методике [21]. Необходимость дообучения языковых моделей при их применении в конкретных предметных областях неоднократно демонстрировалось в литературе [22]. Данное обстоятельство связано с тем, что первоначальное обучение моделей происходит на больших наборах данных (от сотен тысяч до сотен миллионов записей), имеющих в интернете. Эти данные, как правило, взяты из Википедии и социальных сетей. Как следствие, информации из требуемой нам предметной области там или очень мало, или нет вообще. Есть редкие примеры, когда модель обучается на корпусе научных статей [19]. Но и в этом случае авторы стремятся охватить наиболее широкий круг областей знаний, и в результате конкретная предметная область оказывается «сжатой» в узкую полосу представлений. Для того чтобы языковая модель наиболее подробно отражала нюансы требуемой предметной области, необходимо выполнить тонкую настройку ее многочисленных параметров. Эта задача решается путем дополнительного обучения модели на наборе соответствующих примеров.

В процессе обучения языковых моделей части слов, отдельные слова, предложения или целые фрагменты обучающего текста преобразуются в числовые многомерные векторы, так называемые эмбединги (embedding) или векторные представления. При таком обучении векторное представление конкретного слова оказывается зависимым от стоящих рядом с ним слов, т. е. от контекста. К полученным векторным представлениям можно применять методы математической обработки, для измерения меры их близости. Анализ литературы показал, что в большинстве случаев для определения степени близости векторов используется косинусное сходство [23] (угол между векторами, значение которого находится в интервале $[0, 1]$): 1 означает, что вектора совпадают (максимальное сходство), 0 – вектора перпендикулярны (максимальное различие).

Выбор языковых моделей выполнялся по следующим критериям: поддерживаемые языки – русский и английский, тренировка на задаче – семантическая схожесть текстов (Sentence Similarity). Размер модели не более ~ 2 Гб. Модели большего размера проблематично тренировать на персональном компьютере или облачных платформах типа Yandex Cloud.

В процессе производства языковых моделей создаются, как правило, 3 версии модели: маленькая (small), основная (base) и большая (large). Но иногда, для специальных целей, ограничиваются только одной версией. Мы выбрали 4 модели разного размера: 2 маленьких, 1 основную (среднюю) и 1 большую. Первая – sci-rus-tiny – маленькая модель, обученная на текстах научных статей с сайта Elibrary.ru [18]. Вторая – rubert-tiny2 – маленькая модель, обученная на общеупотребительной лексике [24]. Третья – multilingual-e5-base – модель среднего размера, мультиязычная – 94 языка, обученная на данных из Википедии и новостных сайтов [25]. Четвертая – E5-large-en-ru – большая языковая модель [26]. Это очищенная модель – multilingual-e5-large [27] – в которой оставлены только русский и английский языки. За счет этого удалось значительно уменьшить размер модели и сократить количество параметров.

Для оценки качества моделей в литературе используются бенчмарки (benchmark) [28]. Поскольку на интересующей нас задаче – семантического сходства текстов описаний месторождений твердых полезных ископаемых – таких бенчмарков нет, то нами был создан собственный бенчмарк. Были выбраны фрагменты текстов описания золоторудных месторождений на русском и английском языках. Фрагменты описаний железорудных месторождений и фрагменты описаний месторождений меди. При использовании этого бенчмарка мы ожидали, что метрика сходства между текстами описания месторождений золота (не зависимо от языка фрагмента) будет большой – в интервале 0.5–1.0, а метрика сходства между текстами описания месторождений золота и месторождений железа и меди гораздо меньше – в интервале 0.0–0.5.

Результат проверки показал, что для исходных выбранных моделей наши ожидания не оправдались. Для удобства визуализации мы объединили результаты в 4 пары – золото–золото, золото–gold, золото–железо, золото–медь

– с усреднением результатов по каждой паре. Все 4 пары консолидированных примеров дали практически одинаковые результаты на нашем бенчмарке – в районе 0.7–0.9. Такой результат еще раз подтверждает тезис о необходимости дообучения моделей на примерах из исследуемой предметной области (в нашем случае – это Геология – Месторождения полезных ископаемых).

Далее мы выполнили дообучение выбранных моделей на нашем наборе данных, варьируя величину обучающего датасета, и повторно выполнили проверку на нашем бенчмарке. По результатам измерений выделился явный лидер – модель E5-large-en-ru. Зависимость 4 пар консолидированных примеров от величины обучающего датасета приведена на Рис. 1. Видно, что метрики разных типов месторождений значительно разделяются. Полученные результаты полностью отвечают нашим ожиданиям. Указанная модель была выбрана для дальнейшей работы.

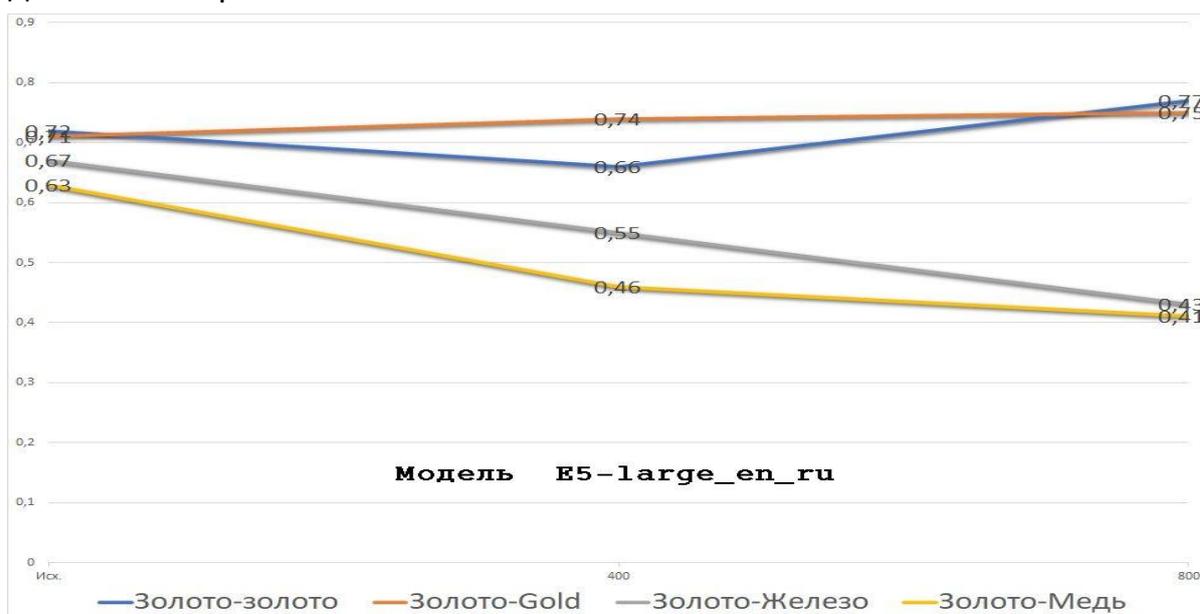


Рис. 1. Зависимость качества модели от величины обучающего датасета.

РЕЗУЛЬТАТЫ

Цель первого этапа нашей работы – выбрать семантически близкие публикации из Репозитория ГГМ РАН (<https://repository.geologyscience.ru/>) [29] для дальнейшего анализа. Для уменьшения временных затрат на поиск была создана база данных на PostgreSQL. В таблицу помещены наименования ста-

тей, их ссылки на архив публикаций и заранее рассчитанные векторные представления наименований статей и абстрактов. Таким образом, поиск статей по искомому запросу преобразуется в вычисление косинусной близости векторного представления запроса с векторными представлениями статей. Максимальные значения косинусной близости будут отражать искомые публикации.

Поиск семантически близких публикаций

в архиве публикаций repository.geologyscience.ru с тематикой "Науки о Земле"
на основе дополнительно тренированной нейросетевой языковой модели [d0rj/e5-large-en-ru](https://arxiv.org/abs/2308.12750)
[Патук М.И. ГГМ РАН](#)

Строка запроса	Кол-во результатов
<input type="text"/>	5 ▾
<input type="button" value="Найти"/> <input type="button" value="↶"/>	
Строка запроса → <i>результат</i>	
<i>золоторудные месторождения в черносланцевых толщах</i>	
Наименование статьи	Сходство
Информативность геофизических методов при поисках золотого оруденения в черносланцевых толщах	0.795
Золото-сульфидные месторождения в углеродисто-терригенных толщах	0.765
Петролого-геохимические свидетельства геолого-генетической однородности гидротермальных месторождений золота, образованных в черносланцевом и несланцевом субстрате	0.733
Геолого-геохимическая модель нового нетрадиционного золото-платиноидного оруденения в черносланцевых горизонтах офиолитовых поясов	0.724
Самородное золото в рудах и россыпях глухаринского узла, магаданская область	0.723

© 2024 Государственный геологический музей им. В.И. Вернадского РАН

Рис. 2. Веб-сервис поиска публикаций

Описанная технология была реализована в виде веб-сервиса и доступна по адресу <https://service.geologyscience.ru/> [30]. После введенного запроса и указания количества публикаций для отображения производится расчет, и результаты выводятся на экран (Рис. 2). Клик по статье в результатах открывает страницу со статьей в архиве публикаций. Чем подробнее будет составлена строка запроса, тем более точным будет результат. Объединенные текстовые данные выбранных статей будут использованы для дальнейшего анализа.

Определение близости двух текстов геологической направленности

на основе дополнительно тренированной нейросетевой языковой модели [d0rj/e5-large-en-ru](#)
определяется косинусная близость двух текстов (1.0 - тексты максимально близки, 0.0 - тексты не совпадают)
[Патук М.И. ГГМ РАН](#)

Наименование 1	Наименование 2
Абстракт 1	Абстракт 2

Введите наименования статей и абстракты в каждое из окон. Для расчета используются первые 500 слов абстракта. Допускаются русский и английский языки.

© 2024 Государственный геологический музей им. В.И. Вернадского РАН

Рис. 3. Веб-сервис определения близости текстов

На основе данной модели можно решать еще одну актуальную задачу – экспресс-анализ схожести двух публикаций на основе расчета метрики сходства фрагментов их текстов, на русском или английском языках. Для этого реализован еще один веб-сервис – Определение близости двух текстов геологической направленности (Рис. 3). Реализована такая же технология работы. Выполняется расчет векторных представлений фрагментов текста и выводится результат их косинусной близости (Рис. 4).

ВЫВОДЫ

Проведенная нами работа показала, что современные методы искусственного интеллекта с успехом могут применяться для анализа текстов научных статей. Возможности семантического поиска значительно расширяют возможности поиска информации. Полученные языковые модели могут с успехом применяться в смежных областях анализа – определения схожести текстов. У приведенных методов есть свои естественные ограничения. Они хорошо работают только в той предметной области, на которой была выполнена дополнительная тренировка моделей.

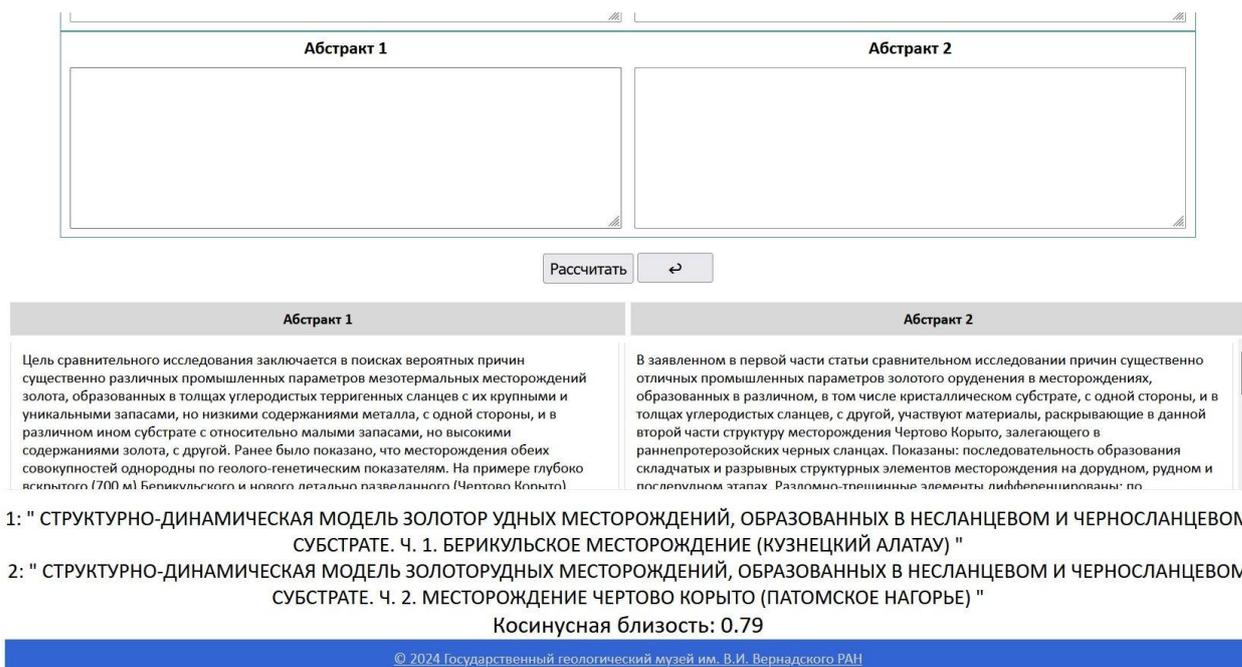


Рис. 4. Веб-сервис определения близости текстов. Результат расчета.

Для улучшения качества работы нашей языковой модели мы дополнили систему логированием запросов, чтобы выбрать направление дальнейшего совершенствования работы системы.

Работы выполняются в рамках Государственного задания ГГМ РАН по теме № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музейных данных».

СПИСОК ЛИТЕРАТУРЫ

1. Семантический поиск.
URL: https://ru.wikipedia.org/wiki/Семантический_поиск (дата обращения 10.09.2024)
2. Ваш путеводитель по миру NLP (обработке естественного языка),
URL: <https://habr.com/ru/companies/otus/articles/705482/> (дата обращения 10.09.2024)
3. Semantic similarity. URL: https://en.wikipedia.org/wiki/Semantic_similarity (дата обращения 10.09.2024)
4. Devlin J., Chang M.W., Lee K., Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding.
URL: arxiv.org/abs/1810.04805 (2018)

5. Объясняем простым языком, что такое трансформеры.
URL: <https://habr.com/ru/companies/mws/articles/770202/> (дата обращения 10.09.2024)
 6. BERT (языковая модель).
URL: [https://neerc.ifmo.ru/wiki/index.php?title=BERT_\(языковая_модель\)](https://neerc.ifmo.ru/wiki/index.php?title=BERT_(языковая_модель)) (дата обращения 10.09.2024)
 7. BERT community. URL: <https://huggingface.co/google-bert> (дата обращения 10.09.2024)
 8. *Akila D., Jayakumar C.* Semantic Similarity – A Review of Approaches and Metrics // International Journal of Applied Engineering Research. 2014. Vol. 9, No. 24. P. 27581–27600.
 9. TF-IDF. URL: <https://ru.wikipedia.org/wiki/TF-IDF> (дата обращения 10.09.2024)
 10. Окапи BM25. URL: https://ru.wikipedia.org/wiki/Окапи_BM25 (дата обращения 10.11.2024)
 11. Word2vec. URL: <https://ru.wikipedia.org/wiki/Word2vec> (дата обращения 10.09.2024)
 12. *Краснов Ф.В., Смазневич И.С., Баскакова Е.Н.* Проблема потери решений в задаче поиска схожих документов: Применение терминологии при построении векторной модели корпуса // Бизнес-информатика. 2021. Т. 15. № 2. С. 60–74. <https://doi.org/10.17323/2587-814X.2021.2.60.74>
 13. *Атаева О.М., Серебряков В.А., Тучкова Н.П.* Модель поиска схожих документов в семантической библиотеке // Научный сервис в сети Интернет: труды XXIII Всероссийской научной конференции (20–23 сентября 2021 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2021. С. 54–64.
<https://doi.org/10.20948/abrau-2021-24>
 14. *Kanerva J., Kitti H., Chang L.-H., Vahtola T., Creutz M., Ginter F.* Semantic search as extractive paraphrase span detection // Lang Resources and Evaluation. 2024. <https://doi.org/10.1007/s10579-023-09715-7>
 15. *Denghui Yang, Dengyun Zhu, Hailong Gai, Fucheng Wan.* Semantic Similarity Calculating based on BERT // Journal of Electrical Systems. 2024. Vol. 20, No. 2. P. 73–79.
-

16. *Kuang M. et al.* Multi-task Learning Based Keywords Weighted Siamese Model for Semantic Retrieval // In: Kashima H., Ide T., Peng W/C. (Eds.) *Advances in Knowledge Discovery and Data Mining. PAKDD 2023. Lecture Notes in Computer Science.* 2023. Vol. 13937. Springer, Cham.

https://doi.org/10.1007/978-3-031-33380-4_7

17. Elibrary. URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения 10.11.2024)

18. Sci-rus-tiny. URL: <https://huggingface.co/mlsa-iai-msu-lab/sci-rus-tiny> (дата обращения 10.09.2024)

19. ruSciBench — бенчмарк для оценки эмбедингов научных текстов. URL: <https://habr.com/ru/articles/781032/> (дата обращения 10.09.2024).

20. Научная электронная библиотека «КиберЛенинка». URL: <https://cyberleninka.ru/> (дата обращения 10.09.2024)

21. Fine-tuning BERT for Semantic Textual Similarity with Transformers in Python.

URL: <https://thepythoncode.com/article/finetune-bert-for-semantic-textual-similarity-in-python> (дата обращения 10.09.2024)

22. *Lawley C.J.M., Raimondo S., Chen T., Brin L., Zakharov A., Kur D., Hui J., Newton G., Burgoyne S.L., Marquis G.* Geoscience language models and their intrinsic evaluation // *Applied Computing and Geosciences.* 2022. Vol. 14, 100084. P. 1–10. <https://doi.org/10.1016/j.acags.2022.100084>

23. Cosine similarity. URL: https://en.wikipedia.org/wiki/Cosine_similarity (дата обращения 10.09.2024)

24. Rubert-tiny2. URL: <https://huggingface.co/cointegrated/rubert-tiny2> (дата обращения 10.09.2024)

25. Multilingual-e5-base. URL: <https://huggingface.co/intfloat/multilingual-e5-base> (дата обращения 10.09.2024)

26. E5-large-en-ru. URL: <https://huggingface.co/d0rj/e5-large-en-ru> (дата обращения 10.09.2024)

27. Multilingual-e5-large.

URL: <https://huggingface.co/intfloat/multilingual-e5-large> (дата обращения 10.09.2024)

28. Тест производительности.

URL: https://ru.wikipedia.org/wiki/Тест_производительности (дата обращения 10.09.2024)

29. Патук М.И., Наумова В.В., Ерёменко В.С. Цифровой репозиторий "geologyscience.ru": открытый доступ к научным публикациям по геологии России // Электронные библиотеки. 2020. Т. 23, № 6. С. 1324–1338.

<https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

30. Geologyscience.ru – Виртуальный ассистент – Сервисы с использованием ИИ – Сервисы нейросетевого анализа геологических текстов. URL: <https://service.geologyscience.ru/> (дата обращения 10.09.2024)

USING SEMANTIC SEARCH TO SELECT AND RANK GEOLOGICAL PUBLICATIONS

Mikhail I. Patuk¹ [0000-0003-3036-2275], Vera V. Naumova² [0000-0002-3001-1638]

State Geological Museum named after Vladimir Vernadsky of RAS, Moscow

¹patuk@mail.ru; ²Naumova_new@mail.ru

Abstract

The aggregation of scientific information is essential for a comprehensive analysis of geological objects. This paper explores the potential and possibilities of semantic search to select thematically similar publications in the geological domain. Various language models are examined in the context of identifying similarities and differences in texts describing mineral deposits. After additional training of language models, a significant improvement in search results is demonstrated. Two web services are presented, based on a method for calculating the semantic similarity between texts and providing a quantitative assessment of their similarity.

Keywords: artificial intelligence, machine learning, natural language processing, semantic search, geology.

REFERENCES

1. Semantic search. URL: https://en.wikipedia.org/wiki/Semantic_search (date of access 10.09.2024)
2. Your guide to the world of NLP (Natural Language Processing). URL: <https://habr.com/ru/companies/otus/articles/705482/> (date of access 10.09.2024)
3. Semantic similarity. URL: https://en.wikipedia.org/wiki/Semantic_similarity (date of access 10.09.2024)
4. *Devlin J., Chang M.W., Lee K., Toutanova K.* Bert: pre-training of deep bidirectional transformers for language understanding. arxiv.org/abs/1810.04805 (2018).
5. We explain in simple terms what transformers are. URL: <https://habr.com/ru/companies/mws/articles/770202/> (date of access 10.09.2024)
6. BERT (language model). URL: [https://neerc.ifmo.ru/wiki/index.php?title=BERT_\(языковая_модель\)](https://neerc.ifmo.ru/wiki/index.php?title=BERT_(языковая_модель)) (date of access 10.11.2024)
7. BERT community. URL: <https://huggingface.co/google-bert> (date of access 10.09.2024)
8. *AkilaD., Jayakumar C.* Semantic Similarity- A Review of Approaches and Metrics // International Journal of Applied Engineering Research. 2014. Vol. 9, No. 24. P. 27581–27600.
9. TF-IDF. URL: <https://en.wikipedia.org/wiki/Tf-idf> (date of access 10.09.2024)
10. Okapi BM25. URL: https://en.wikipedia.org/wiki/Okapi_BM25 (date of access 10.09.2024)
11. Word2vec. URL: <https://en.wikipedia.org/wiki/Word2vec> (date of access 10.09.2024).

12. *Krasnov F., Smaznevich I., Baskakova E.* The problem of loss of solutions in the task of searching similar documents: Applying terminology in the construction of a corpus vector model // *Business Informatics*. 2021. Vol. 15. 60–74. <https://doi.org/10.17323/2587-814X.2021.2.60.74>.

13. *Ataeva O.M., Serebryakov V.A., Tuchkova N.P.* Search model for similar documents in the semantic library // *Scientific service in Internet: Proceedings of the XXIII All-Russian Scientific Conference (20–23 September 2021)*. M.: Keldysh Institute of Applied Mathematics, 2021. P. 54–64 (in Russian). <https://doi.org/10.20948/abrau-2021-24>

14. *Kanerva J., Kitti H., Chang L.-H., Vahtola T., Creutz M., Ginter F.* Semantic search as extractive paraphrase span detection // *Lang Resources and Evaluation*. 2024. <https://doi.org/10.1007/s10579-023-09715-7>

15. *Denghui Yang, Dengyun Zhu, Hailong Gai, Fucheng Wan.* Semantic Similarity Calculating based on BERT // *Journal of Electrical Systems*. 2024. Vol. 20, No. 2. P. 73–79.

16. *Kuang M. et al.* Multi-task Learning Based Keywords Weighted Siamese Model for Semantic Retrieval // In: *Kashima H., Ide T., Peng W/C. (Eds.) Advances in Knowledge Discovery and Data Mining. PAKDD 2023. Lecture Notes in Computer Science*. 2023. Vol. 13937. Springer, Cham. https://doi.org/10.1007/978-3-031-33380-4_7

17. Elibrary. URL: <https://www.elibrary.ru/defaultx.asp> (date of access 10.09.2024).

18. Sci-rus-tiny. URL: <https://huggingface.co/mlsa-iai-msu-lab/sci-rus-tiny> (date of access 10.09.2024).

19. ruSciBench — A benchmark for evaluating the quality of embeddings for scientific texts. URL: <https://habr.com/ru/articles/781032/> (date of access 10.09.2024)

20. Scientific electronic Library «CyberLeninka». <https://cyberleninka.ru/> (date of access 10.09.2024).

21. Fine-tuning BERT for Semantic Textual Similarity with Transformers in Python.

URL: <https://thepythoncode.com/article/finetune-bert-for-semantic-textual-similarity-in-python> (date of access 10.09.2024).

22. Lawley C.J.M., Raimondo S., Chen T., Brin L., Zakharov A., Kur D., Hui J., Newton G., Burgoyne S.L., Marquis G. Geoscience language models and their intrinsic evaluation // *Applied Computing and Geosciences*. 2022. Vol. 14. 100084. P. 1–10. <https://doi.org/10.1016/j.acags.2022.100084>

23. Cosine similarity. URL: https://en.wikipedia.org/wiki/Cosine_similarity (date of access 10.09.2024).

24. Rubert-tiny2. URL: <https://huggingface.co/cointegrated/rubert-tiny2c> (date of access 10.09.2024)

25. Multilingual-e5-base.

URL: <https://huggingface.co/intfloat/multilingual-e5-base> (date of access 10.09.2024).

26. E5-large-en-ru. URL: <https://huggingface.co/d0rj/e5-large-en-ru> (date of access 10.09.2024).

27. Multilingual-e5-large.

URL: <https://huggingface.co/intfloat/multilingual-e5-large> (date of access 10.09.2024).

28. Benchmark (computing).

URL: [https://en.wikipedia.org/wiki/Benchmark_\(computing\)](https://en.wikipedia.org/wiki/Benchmark_(computing)) (date of access 10.09.2024).

29. Patuk M.I., Naumova V.V., Eryomenko V.S. Digital repository "geology-science.ru": open access to scientific publications on russian geology // *Russian Digital Library Journal*. 2020. Vol. 23, no. 6. P. 1324–1338 (in Russian). <https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

30. Geologyscience.ru – Virtual Assistant – Services using AI – Neural network analysis services for geological texts, <https://service.geologyscience.ru/> (date of access 10.09.2024).

СВЕДЕНИЯ ОБ АВТОРАХ



ПАТУК Михаил Иванович – к. г.-м. н., н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail I. PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

Email: patuk@mail.ru

ORCID: 0000-0003-3036-2275



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

Email: naumova_new@mail.ru

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 24 сентября 2024 года

УДК 004.89; 005.

ПРОЕКТИРОВАНИЕ ИНСТРУМЕНТАРИЯ ДЛЯ СОЗДАНИЯ ИГРОВОГО ПРОЦЕССА ЧЕРЕЗ СИСТЕМАТИЗАЦИЮ ИГРОВЫХ МЕХАНИК

А. В. Шубин¹ [0000-0002-6203-3268], В. В. Кугуракова² [0000-0002-1552-4910]

^{1, 2}Казанский федеральный университет, Институт информационных технологий и интеллектуальных систем.

¹shubin.aleksey.kpfu@gmail.com, ²vlada.kugurakova@gmail.com

Аннотация

Представлен новый подход к разработке инструмента, направленного на упрощение рабочего процесса игрового дизайнера. Выработаны требования, разработан сценарий работы и уточнены основные параметры для разрабатываемого инструмента. Основная задача инструмента заключается в ускорении и облегчении выбора подходящих игровых механик без необходимости тратить ценное время на длительный анализ других видеоигровых проектов.

Чтобы обеспечить более эффективную работу геймдизайнеров при подборе игровых механик, был проведён анализ разнообразных подходов к классификации игровых механик. В процессе исследования были рассмотрены различные методы классификации игровых механик, их разбор и анализ показали, какие классификации в большей степени подходят для декомпозиции игровой механики. Результаты исследования позволили выявить ключевые аспекты игровой механики, которые будут служить фундаментом для разработки инструмента.

Настоящее исследование представляет собой важный этап в создании инструмента, который, будучи внедренным, позволит оптимизировать процесс геймдизайна и ускорить разработку видеоигр.

Ключевые слова: игровой дизайн, классификация, игровые механики, автоматизация, видеоигры.

ВВЕДЕНИЕ

Разработка видеоигр – сложный и комплексный процесс, который требует взаимодействия специалистов различных областей. Наличие такой мультидисциплинарной команды разработки позволяет отдельным членам команды иметь лишь навыки относительно своей области действий, без необходимости в изучении смежных задач.

При этом непосредственная разработка видеоигр (программирование в игровом движке, создание 3D-моделей и анимаций, создание звукового сопровождения и внутриигровых изображений) подразумевает трудоёмкую, ручную работу с использованием инструментов, которые позволяют ускорить, сделать процесс более удобным. В свою очередь, процесс создания игрового дизайна подразумевает большое количество аналитической работы и творческого вдохновения для поиска и создания наиболее удачных решений, подходящих к конкретно разрабатываемой игре.

Современный геймдизайнер имеет довольно широкий спектр задач и ответственностей во время создания видеоигр:

- разработка общей концепции и механик будущей видеоигры,
- проектировка игрового мира и основ игрового уровня,
- прототипирование и сбор обратной связи от пользователей,
- сотрудничество с другими членами команды и руководство их работы.

При этом ключевой задачей геймдизайнера остаются выбор и разработка игровых механик, включенных в видеоигру. Игровая механика – репрезентация игровых правил, представленная игроку в виде чёрного ящика. Именно игровые механики поддерживают интерактивную составляющую любой видеоигры. Воздействия игрока через игровые механики ведут к изменению видеоигровой системы, которая даёт обратную связь о воздействии обратно игроку (рис. 1). Система динамически регулирует доступ к игровым механикам, постепенно открывая новые по мере прогресса игрока и выполнения определенных условий.

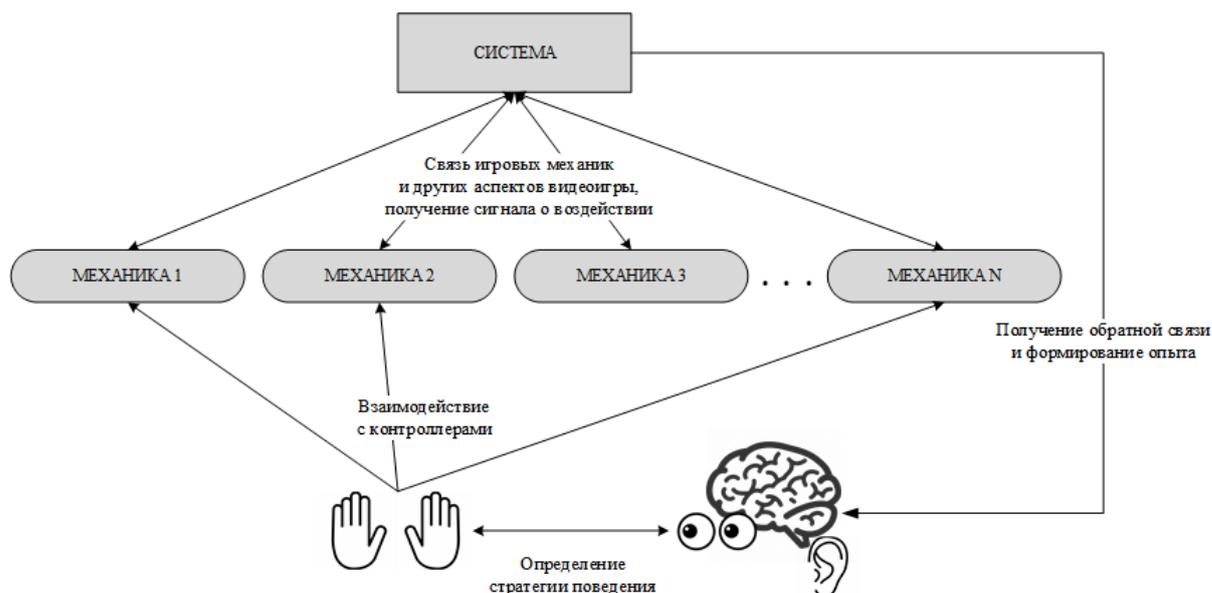


Рис. 1. Схема взаимодействия игровой системы с игроком через игровые механики.

Одна из проблем современного геймдизайна заключается в ограниченной доступности инструментария для ускорения процесса разработки. Многие существующие решения (такие как Game Maker¹, RPGMaker², Machinations³, Twine⁴) либо предъявляют высокие требования к техническим компетенциям геймдизайнеров, либо ограничивают свободу реализации игровых проектов, что сужает спектр эффективно применяемых инструментов [1].

Зачастую геймдизайнеры вынуждены самостоятельно изучать новые технологии и референсные проекты, что может занять много времени и усилий. В ином случае разработчик ограничивается лишь классическими приёмами старых настольных игр, как, например, создание бумажных или других физических прототипов из доступных материалов.

Итак, можно судить о потребности в инструменте, который мог бы облегчить работу геймдизайнера по выбору наиболее оптимальных вариантов игровой механики и был бы удобен в использовании.

¹ <https://gamemaker.io/en>

² <https://www.rpgmakerweb.com>

³ <http://machinations.io>

⁴ <https://twinery.org>

ТАКСОНОМИИ ИГРОВЫХ МЕХАНИК

Термин «игровая механика» можно встретить уже в середине XX века [2], он до сих пор используется среди причастных к соответствующей индустрии людей: игроков, разработчиков, учёных и специалистов по геймификации. Наиболее краткое и точное определение данного термина приведено в [3]: «*игровая механика – “кирпичики” игрового процесса*».

Для лучшего понимания структуры игровой механики эффективно разбить ее на составные элементы. В работе [4] выделены 4 основных элемента игровой механики: принцип, изменение, обратную связь и цикличность.

- Принцип – набор действий и их влияние на игровое пространство.
- Изменение – это смена состояния объекта на игровой сцене.
- Обратная связь – это способ передачи игроку об изменениях через органы человеческого восприятия (зрение, слух, осязание, обоняние и т. д.).
- Цикличность – одна из важнейших составляющих игровой механики. Встреченная лишь раз механика будет восприниматься игроком как бесполезная информация, что приведёт к непониманию, ощущению лишнего.

Среди публикаций разработчиков видеоигр можно найти значительное количество исследований и работ, посвящённых классификации игровых механик [5–7]. Это неслучайно, так как игровые механики представляют собой фундаментальные строительные блоки игрового опыта [8]. Классификация игровых механик позволяет разбираться в том, какие элементы составляют основу и влияют на процесс игры и как они взаимодействуют между собой. Также некоторые классификации помогают понять зависимости механик и соответствующие аспекты игрового опыта, будь то стратегическое мышление, взаимодействие с другими игроками или создание эмоциональных переживаний у игрока. Это сильно отличает современного разработчика видеоигр на фоне предыдущего поколения – современные геймдизайнеры становятся всё более заинтересованными в понимании структуры и взаимосвязей между игровыми механиками, поскольку дает им возможность создавать игры более осознанно и целенаправленно.

Каждая созданная классификация обычно рождается с определенными целями и задачами, которые могут варьироваться в зависимости от контекста

и нужд. Эти цели влияют на выбор параметров классификации, способы группировки механик и глубину анализа, что, в итоге, формирует специфику и фокус конкретной классификации игровых механик.

Но создание универсальной классификации игровых механик на данный момент является задачей, не решённой до конца. Р. Зубек [9] также подтверждает факт, что за всё время существования игровой индустрии, со второй половины XX века, геймдизайнеры так и не смогли прийти к общей таксономии различных типов игровых механик. Дж. Шелл в [3] отметил, что игровые механики довольно сложно классифицировать из-за их «таинственности», подразумевая сложные связи и неочевидные отношения между объектом и механикой, а также взаимодействие механик друг с другом.

Простой пример: стена – игровой объект, который не является игровой механикой. При этом игрок не может проходить через стену, что является ограничением или правилом игры, что порождает игровую механику. Другой пример: игрок может управлять перемещением персонажа (бегать) и прыгать; это две механики, которые тесно связаны друг с другом, потому что игрок может прыгать на бегу, а также управлять полётом во время прыжка. Если отнять у игрока способность к перемещению, появится другая механика, где игрок может только прыгать, но не управляет перемещением, следовательно, потребуются дополнительные изменения, чтобы реализовать новую механику. Например, в мобильной видеоигре *Flappy Bird*⁵ игрок может поднимать персонажа выше при помощи нажатий на экран устройства, при этом перемещение персонажа автоматизировано и не может контролироваться игроком. К тому же, в некоторых играх (например, *Papers, Please* или *Five Nights At Freddy's*) может отсутствовать перемещение в каком-либо виде. Следовательно, при отсутствии данной механики можно судить о наличии правила, которое запрещает перемещения, либо возможность перемещения несущественна и не является ключевым фактором игры, следовательно, не влияет на игровой процесс.

⁵ Список упомянутых видеоигр можно найти в разделе «Лудография».

Классификация игровых механик G-Model

G-Model [10] представляет собой классификацию 23 механик, распределяя их на 4 слоя: обратная связь, нарратив, экономика и социализация. Эти слои соответствуют аспектам игрового процесса, привлекательным для различных групп игроков. Каждая механика может относиться к одному или нескольким слоям, затрагивая разные аспекты видеоигры.

В контексте G-Model игровая механика определяется как «инструмент модуляции поведения игрока». Следовательно, наиболее ключевым для данной классификации является влияние игровых механик на обучение игрока, делающее его более увлекательным и запоминающимся. Данная методика предназначена для интеграции игровых механик в обучающие и рабочие процессы и делает их более увлекательными и запоминающимися.

Структура G-Model напоминает классификацию Р. Бартла [11], разделяющую игроков на 4 категории по их предпочтениям в видеоиграх. Это позволяет подобрать набор механик, подходящий как отдельным категориям, так и всем игрокам одновременно.

Классификация Шелла

Дж. Шелл [3] предложил классификацию в попытке затронуть все механики, которые так или иначе могут быть реализованы в любой видеоигре. Эта классификация стремится разделить игровые механики на группы с целью проведения более детального анализа их характеристик и взаимосвязей.

Данная классификация насчитывает семь типов игровых механик:

- Пространство – тип области, внутри которой происходит действие игры.
- Время – игровая условность, которая позволяет контролировать темп игры: сделать его протяжённым или более коротким, в зависимости от потребностей.
- Объекты – «существительные» любой игры, которые в свою очередь имеют свойства. Под свойством понимается любая информация, характеризующая объект и его поведение.

- Действия – «глаголы» игры, которые описывают все доступные игроку манипуляции над игрой.
- Правила – функция видеоигр, которая обуславливает ограничения и наличие остальных механик [12] и определяют цели для игрока.
- Навык – категория, затрагивающая общее понимание и способности игрока к игре. Эта категория механик зависят от игрока, а не от игры.
- Шанс – вид механик, создающих неопределённость даже в понятной системе.

Таким образом, Шелл делит игровые механики на типы для более глубокого анализа игровых механик. Понимание аспектов, влияющих на игровую механику, позволяет создателям игр более глубоко подходить к реализации игровой концепции и игрового процесса.

Классификация Зубека

Классификация Зубека [9] обращает внимание на тесную связь между механиками и жанром игры. Это означает, что игровые механики могут относиться лишь к определенным жанрам, что способствует формированию уникального игрового опыта для игроков. В отличие от более детальной классификации Шелла, классификация Зубека не была направлена на сложную декомпозицию игровых механик, а лишь объединяет механики по общим свойствам:

- механики контроля – вариант репрезентации игрока в видеоигре и набор действий, которые игрок может совершать для взаимодействия с игровым миром,
- механики прогресса – отображение прогресса игрока и предоставление обратной связи,
- механики неопределённости – случайность, скрытая информация и другие источники неопределённости, влияющие на игровой процесс,
- механики управления ресурсами – возможность распоряжаться игровыми ресурсами, такими как здоровье, валюта, оружие и другие, а также механики, связанные с данными ресурсами,

- механики специфичных жанров – определённый набор действий, который отражает принадлежность специфичному жанру (шутер, гонка, стратегия и др.).

Такое деление в классификации может быть особенно полезно для начинающих разработчиков, которые недавно начали изучать игровой дизайн и не знают о зависимостях между игровыми объектами, подсистемами и игровыми механиками.

Классификация «Атомарные механики»

Особого внимания заслуживает классификация [13], разработанная С. Гиммельрейхом. Данная таксономия позволяет разложить сложные игровые взаимодействия на базовые составляющие, открывая новые перспективы для понимания и проектирования игрового процесса. Игровая механика может быть разложена на «атомы» (рис. 2) в каждой из трёх категорий:

- физические механики – правила и возможности, изменяющие состояние игрового пространства,
- ментальные механики – правила и возможности, использующие или изменяющие когнитивную модель игрока,
- экономические механики – правила и возможности для управления игровыми ресурсами.

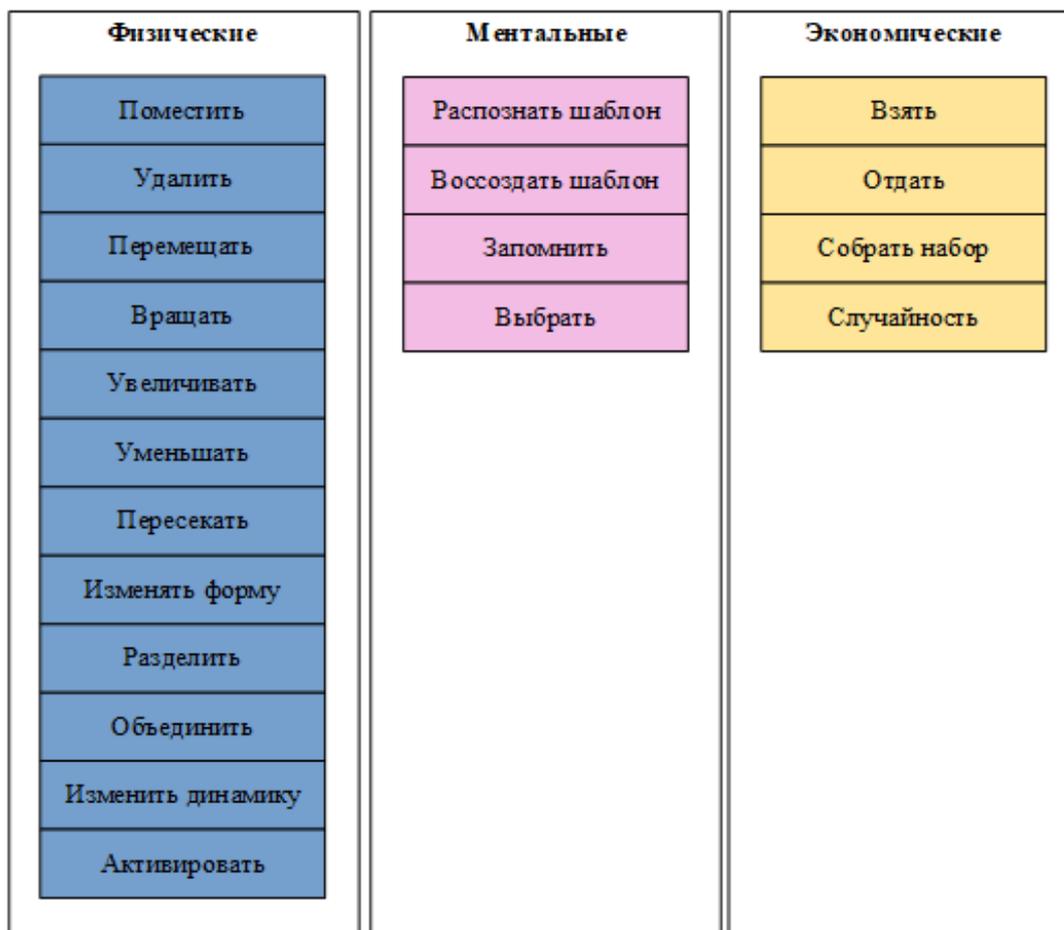


Рис. 2. Атомарные механики, разделённые по категориям

Практическое применение данной классификации реализуется в виде колоды карт. Каждая карта в этой колоде представляет отдельную атомарную механику. Затем случайный выбор карт из колоды позволяет стимулировать творческий процесс при создании игровых механик. Комбинируя вытянутые карты, разработчики могут экспериментировать с различными сочетаниями базовых элементов, придумывая новые и оригинальные игровые механики. Этот метод не только облегчает генерацию идей, но и способствует более глубокому пониманию структуры и взаимосвязей между различными аспектами игрового процесса.

ДЕТАЛИЗАЦИЯ СТРУКТУРЫ ИГРОВОЙ МЕХАНИКИ

Любая из существующих классификаций имеет свои уникальные преимущества и может быть полезна в различных контекстах разработки игр. Разработчики могут выбирать ту, которая лучше соответствует их конкретным потребностям и целям в создании игры. При этом в рамках разработки данного инструмента ни одна из описанных ранее классификаций не может в полной мере покрыть необходимые потребности.

Классификация G-Model [10] предоставляет готовый и удобный набор игровых механик, но их набор ограничен не только числом механик, но и способом применения, так как, в первую очередь, данная классификация ориентирована на геймификацию процессов, а не на описание всех игровых механик видеоигр.

Классификация Шелла [3] детально описывает элементы пространства видеоигры и свойства игровых механик, но при этом не имеет достаточной детализации, чтобы описать игровую механику данными категориями. Кроме этого, эта классификация не имеет чёткой иерархии для определения взаимосвязей между элементами игровой механики.

В классификации Зубека [9] поднят важный вопрос о том, что игровые механики делятся на отдельные системы, которые могут содержать в себе набор механик, который во многом зависит от жанра и фокуса видеоигры. При этом в классификации отсутствует какая-либо детализация.

Классификация «Атомарные механики» [13] дают возможность её простого практического применения – предоставленный набор атомарных механик и их интерпретаций открывает большое число комбинаций для дальнейшего этапа генерации и обдумывания идей. Недостатком этой классификации является тот факт, что генерация новых идей напрямую зависит от объёма креативности дизайнера и ограниченного спектра полученного им опыта. Другой проблемой является тот факт, что полноты этой классификации недостаточно, чтобы детализировано описать любую игровую механику.

Следовательно, необходима разработка расширенной классификации игровых механик. Такая классификация должна максимально полно описы-

вать структуру игровой механики, а сами механики должны сохранять основные четыре элемента [4]: принцип, изменение, обратная связь и цикличность. Таким образом, новая классификация (далее MGMT – сокр. от англ. Multifaced Game Mechanics Taxonomy) разовьет названную структуру и будет строиться на большей детализации, сохраняя основные положительные черты других классификаций (рис. 3).

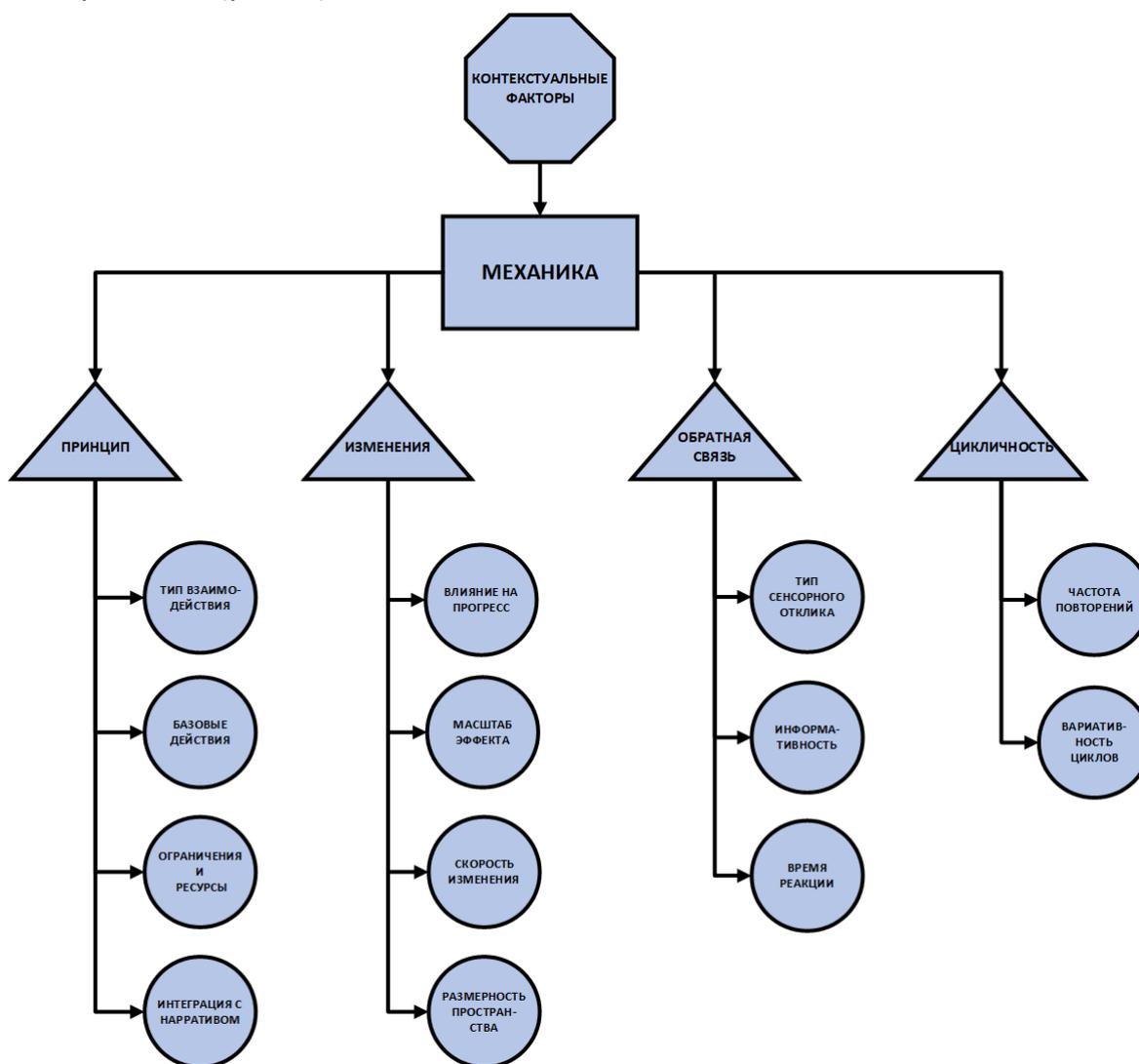


Рис. 3. Расширенное описание свойств игровой механики

Под контекстуальными факторами (такими, как жанр игры, целевая аудитория, сеттинг и т. д.) понимаются характеристики, ключевые для всей игры. Как упоминалось в классификации Зубека [9], именно контекстуальные факторы во многом и определяют конечный набор игровых механик и зачастую появляются в самом начале разработки.

Каждая игровая механика основывается на четырех ключевых свойствах, каждое из которых, в свою очередь, характеризуется набором дополнительных атрибутов. Так, например, свойство «принцип» игровой механики может быть описано следующими атрибутами:

- **Тип взаимодействия**, который включает в себя масштаб взаимодействия (индивидуальное или с группой объектов), время взаимодействия (однократное или продолжительное) и др.
- **Базовые действия** – набор ключевых операций, которые могут выполнять различные игровые объекты в рамках игровой механики. В данном случае можно использовать физические атомарные механики (рис. 2).
- **Ограничения и ресурсы** – свойства механики, ограничивающие её использование или позволяющие получить новые возможности.
- **Интеграция с сюжетом** описывает степень важности механики для сюжета и нарратива, а также тип подачи информации (явный или скрытый).

Из-за внушительного количества параметров, которые требуется держать во внимании при выборе игровой механики, использование предложенной многогранной классификации будет проблематичным для человека. В связи с этим применение автоматизированных инструментов, которые позволят быстрее предлагать подходящие варианты, будет незаменимым при использовании подобной детализированной классификации.

Процесс анализа и классификации игровых проектов можно существенно ускорить, применив системы интеллектуального анализа текстов, аналогичные тем, что используются, например, в юриспруденции [14]. Автоматизированный инструмент для выявления общих значимых сущностей позволит эффективно категоризировать игровые проекты, выделяя общие черты и механики, характеризующие как игровой процесс, так и нарративную составляющую.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ИНСТРУМЕНТА ДЛЯ ГЕНЕРАЦИИ ИГРОВЫХ МЕХАНИК

Анализ различных подходов, предлагаемых опытными геймдизайнерами по классификации игровых механик, позволяет сделать вывод о целесообразности разработки инструмента-ассистента для генерации игровых механик.

В [15] была предпринята попытка описания похожего инструмента для прототипирования видеоигр на основе сценария. В [16] представлен функционал генерации игровой сцены, наполненной различными игровыми объектами. В [17] был изложен функционал для балансирования видеоигр. Результаты этих публикаций показали, что разработка инструментов для автоматического создания игровых прототипов имеет смысл и может являться ценной практикой для проверки гипотез и выбора направления разработки игры.

В текущих реалиях набирает обороты объём использования больших языковых моделей (Large Language Models, LLM) и таких инструментов, как ChatGPT или Claude. Их применение для поиска информации, персонализированного обучения, оформления текстов, автоматизации рутинных задач и поиска новых идей является эффективным, что обуславливается активным развитием этих инструментов и быстрым расширением аудитории [18]. Учитывая новизну подхода, можно надеяться, что применение этого инструментария в автоматизации и поддержке работы станет перспективным.

Несмотря на наличие у LLM обширных данных из различных областей, включая знания о видеоиграх и геймдизайне, включение новых знаний в обученные языковые модели [19] и использование структурированной базы знаний может улучшить процедурную генерацию контента в задачах, требующих специфических знаний [20]. Следовательно, предобучение LLM датасетами, размеченными на основе классификации MGMT, позволит решить основную задачу – повышение качества сгенерированных ответов, при этом перекладывая манипуляции с низкоуровневой детализированной структурой игровой механики на LLM.

Функционал проектируемого инструмента-ассистента генерации игровых механик заключается в автоматическом подборе различных аспектов видеоигры на основе ключевой идеи, вводимой пользователем. Данный инструмент позволит предлагать игровые механики и связанные с ними элементы сюжета и нарратива. Схема работы (см. рис. 4) может быть представлена в следующем виде:

- Пользователь вводит ключевую идею игры, которая может включать в себя часть сюжета и возможности игрока.
- Автоматически подбираются параметры, определяющие мир и механики видеоигры (сеттинг, жанр игры, сюжетный конфликт и т. д.).
- Пользователь получает возможность редактировать параметры работы инструмента для получения нового результата.
- На основе выбранных параметров предлагается выбор игровых механик, которые могли бы соответствовать данной игре.

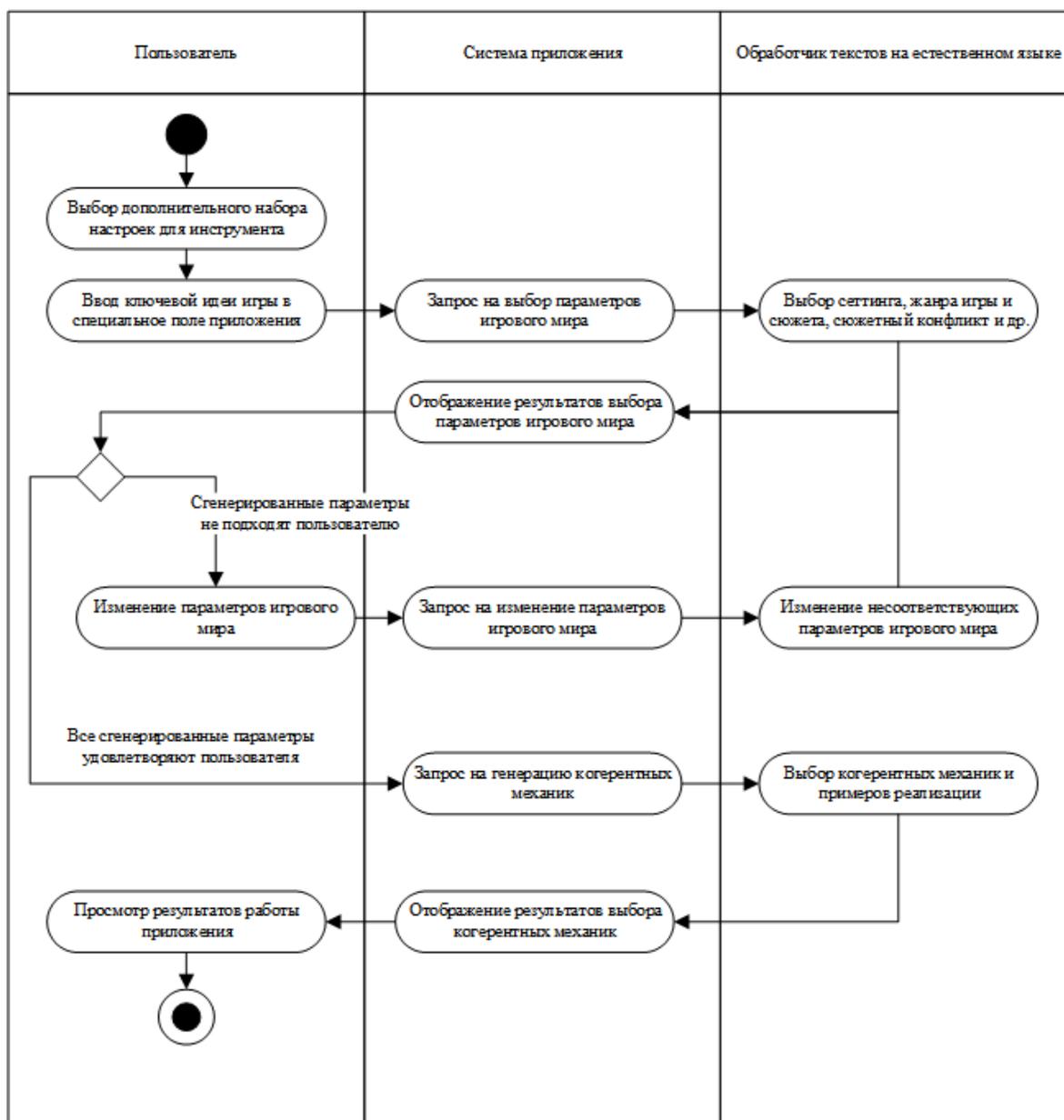


Рис. 4. Диаграмма деятельности работы инструмента-ассистента

Пользовательский интерфейс должен включать поля ввода, в которых геймдизайнеры могут задавать параметры игры, а также редактировать варианты, предложенные искусственной нейронной сетью.

Быстрая генерация вариантов для последующей творческой переработки геймдизайнером – основная цель такого инструмента, переход от ручного перебора к автоматизации будет способствовать более эффективной и творческой разработке игр.

ЗАКЛЮЧЕНИЕ

Процесс работы геймдизайнера практически никак не изменился со времён зарождения этой индустрии, но она уже шагнула далеко вперёд, а комплексность и разнообразие видеоигр заметно возросли.

В ходе настоящего исследования был проведён анализ различных классификаций игровых механик, которые могли бы определить наиболее важные свойства всех игровых механик, стать основой для разрабатываемого инструмента. Результаты показали, что классификации Шелла и Зубека имеют плохое практическое применение в декомпозиции игровых механик на отдельные составляющие и не подходят для поиска и выбора механик. G-Model предусматривает наличие в классификации конкретного набора механик, чего будет недостаточно для описания всех игровых механик. В связи с этим, пересматривая подходы более ранних классификаций, нами была предложена расширенная классификация игровых механик MGMT, в том числе, включающая в себя основную часть «Атомарных механик» и решающая проблему её недостаточной детализации.

С использованием MGMT была предложена концептуальная модель инструмента-ассистента для генерации игровых механик. Разрабатываемый инструмент позволит оптимизировать процесс разработки концепции видеоигры за счёт автоматизации процесса подбора подходящих игровых решений. Инструмент будет способствовать гармоничному сочетанию игровых механик, сократит время разработки, сделает возможными эксперименты и инновации в геймдизайне для улучшения игрового опыта.

Благодарности

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

ЛУДОГРАФИЯ

Dong Nguyen (2013). *Flappy Bird* [Mobile] [iOS, Android], Gears.

Lucas Pope (2013). *Papers, Please* [Puzzle] [Windows, Linux iOS, Android, OS X, PlayStation Vita], 3909 LLC.

Scott Cawthon. Five Nights at Freddy's [Point-and-click] [Windows, Android, iOS, Nintendo Switch, PlayStation 4, Xbox One, Windows Phone, PlayStation 5], ScottGames.

СПИСОК ЛИТЕРАТУРЫ

1. *Gajewski S., el Mawas N., Heutte J.* A Systematic Literature Review of Game Design Tools // in International Conference on Computer Supported Education, CSEDU, 2022. Vol. 2.
2. *Greenlaw P.S., Herron L.W., Rawdon R.H.* Business simulation in industrial and university education // Englewood Cliffs. United States. Prentice-Hall, 1962. 376 p.
3. *Schell J.* The Art of Game Design: A Book of Lenses. Burlington. United States. Morgan Kaufmann Publishers, 2008. 520 p.
4. *Гиммельрейх С.Э.* Механика, геймплей, динамика — как устроен игровой процесс в играх // netology.ru. 2020.
URL: <https://netology.ru/blog/09-2020-igrovoy-process>.
5. *Aarseth E., Smedstad S.M., Sunnana L.A* Multidimensional typology of games // In 2003 DiGRA International Conference: Level Up, November, 2003.
6. *Elverdam C., Aarseth E.* Game Classification and Game Design: Construction Through Critical Analysis // Games and Culture. 2007. Vol. 2, No. 1. P. 3–22. <https://doi.org/10.1177/1555412006286892>
7. *Proulx J.N., Romero M., Arnab S.* Learning Mechanics and Game Mechanics Under the Perspective of Self-Determination Theory to Foster Motivation in Digital Game Based Learning // Simulation and Gaming. 2017. Vol. 48, Issue 1, P. 81–97. <https://doi.org/10.1177/1046878116674399>
8. *Fabricatore C.* Gameplay and game mechanics design: a key to quality in videogames // In OECD-CERI Expert Meeting on Videogames and Education, October, 2007. URL: <http://dx.doi.org/10.13140/RG.2.1.1125.4167>
9. *Zubek R.* Elements of Game Design. MIT Press. 2020. 240 p.
10. Gmodel. URL: <https://gl.sberlabs.com/en/gmodel/index>.
11. *Bartle R.A.* Designing virtual worlds. Berkeley. United States. New Riders. 2004. 937 p.

12. *Parlett D.* The Oxford History of Board Games, Oxford University Press, 1999. 386 p.

13. Игровые механики: Часть II // *gdcuffs.com*. 2018.
URL: https://gdcuffs.com/game_mechanics_deconstruct_2/

14. *Алексеев А.А., Катасёв А.С., Зуев Д.С., Тутубалина Е.В., Хасьянов А.Ф.* Интеллектуальная информационная система поддержки принятия судебных решений в сфере экономического правосудия // Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2018. С. 17–27.
URL: <http://keldysh.ru/abrau/2018/theses/60.pdf>
<https://doi.org/10.20948/abrau-2018-60>

15. *Сахибгареева Г.Ф., Кугуракова В.В.* Редактор интерактивной структуры для инструмента генерации сценарных прототипов // Электронные библиотеки. 2022. Т. 24, №.6. С. 1184–1202.
<https://doi.org/10.26907/1562-5419-2021-24-6-1184-1202>

16. *Козар Б.А., Кугуракова В.В., Сахибгареева Г.Ф.* Структуризация сущностей естественного текста с использованием нейронных сетей для генерации трехмерных сцен // Программные продукты и системы. 2022. №3. С. 329–339. <http://doi.org/10.15827/0236-235X.139.329-339>

17. *Сахибгареева Г.Ф., Кугуракова В.В., Большаков Э.С.* Инструменты балансирования игр // Электронные библиотеки. 2023. Т. 26, № 2. С. 225–251.
<https://doi.org/10.26907/1562-5419-2023-26-2-225-251>

18. *Prillaman M.* Is ChatGPT making scientists hyper-productive? The highs and lows of using AI // *Nature*. 2024. Vol. 627.
<https://doi.org/10.1038/d41586-024-00592-w>

19. *Wang R. et al.* K-ADAPTER: Infusing Knowledge into Pre-Trained Models with Adapters // in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. <https://doi.org/10.18653/v1/2021.findings-acl.121>

20. *Lewis P. et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks // in Advances in Neural Information Processing Systems, 2020, December. URL: <https://arxiv.org/pdf/2005.11401>.

DESIGNING A TOOL FOR CREATING GAMEPLAY THROUGH THE SYSTEMATIZATION OF GAME MECHANICS

A. V. Shubin¹ [0000-0002-6203-3268], V. V. Kugurakova² [0000-0002-1552-4910]

^{1, 2}*Institute of Information Technologies and Intelligent Systems, Kazan Federal University, 35 Kremlyovskaya st., Kazan, 420008*

¹shubin.aleksey.kpfu@gmail.com, ²vlada.kugurakova@gmail.com

Abstract

A new approach to the development of a tool aimed at simplifying the workflow of a game designer is presented. The requirements are elaborated, the work scenario is developed and the main parameters for the developed tool are specified. The main objective of the tool is to speed up and facilitate the selection of proper game mechanics without the need to spend valuable time on lengthy analysis of other videogame projects.

To provide more effective work of game designers in the selection of game mechanics, we analyzed a variety of approaches to the classification of game mechanics. In the process of the research various methods of classification of game mechanics were considered, the analysis revealed which classifications are more suitable for decomposition of game mechanics. The results of the research allowed us to identify key aspects of game mechanics, which will serve as a foundation for the development of the tool.

This research represents an important step in creating a tool that will optimize the game design process and increase the speed of videogame development.

Keywords: *game design, classification, game mechanics, automatization, videogame.*

LUDOGRAPHY

Dong Nguyen (2013). *Flappy Bird* [Mobile] [iOS, Android], Gears.

Lucas Pope (2013). *Papers, Please* [Puzzle] [Windows, Linux iOS, Android, OS X, PlayStation Vita], 3909 LLC.

Scott Cawthon. Five Nights at Freddy's [Point-and-click] [Windows, Android, iOS, Nintendo Switch, PlayStation 4, Xbox One, Windows Phone, PlayStation 5], ScottGames.

REFERENCES

1. *Gajewski S., el Mawas N., Heutte J.* A Systematic Literature Review of Game Design Tools // in International Conference on Computer Supported Education, CSEDU, 2022. Vol. 2.
2. *Greenlaw P.S., Herron L.W., Rawdon R.H.* Business simulation in industrial and university education // Englewood Cliffs. United States. Prentice-Hall, 1962. 376 p.
3. *Schell J.* The Art of Game Design: A Book of Lenses. Burlington. United States. Morgan Kaufmann Publishers, 2008. 520 p.
4. *Himmelreich S.* Механика, геймплей, динамика – как устроены игровые процессы в играх. 2020. URL: <https://netology.ru/blog/09-2020-igrovoy-process> (In Russian)
5. *Aarseth E., Smedstad S.M., Sunnana L.A.* Multidimensional typology of games // In 2003 DiGRA International Conference: Level Up, November, 2003.
6. *Elverdam C., Aarseth E.* Game Classification and Game Design: Construction Through Critical Analysis // Games and Culture. 2007. Vol. 2, No. 1. P. 3–22. <https://doi.org/10.1177/1555412006286892>
7. *Proulx J.N., Romero M., Arnab S.* Learning Mechanics and Game Mechanics Under the Perspective of Self-Determination Theory to Foster Motivation in Digital Game Based Learning // Simulation and Gaming. 2017. Vol. 48, Issue 1, P. 81–97. <https://doi.org/10.1177/1046878116674399>
8. *Fabricatore C.* Gameplay and game mechanics design: a key to quality in videogames // In OECD-CERI Expert Meeting on Videogames and Education, October, 2007. <http://dx.doi.org/10.13140/RG.2.1.1125.4167>
9. *Zubek R.* Elements of Game Design. MIT Press. 2020. 240 p.
10. Gmodel. URL: <https://gl.sberlabs.com/en/gmodel/index>.
11. *Bartle R.A.* Designing virtual worlds. Berkeley. United States. New Riders. 2004. 937 p.

12. *Parlett D.* The Oxford History of Board Games, Oxford University Press, 1999. 386 p.

13. *Igrovye mehaniki: Chast II.* 2018. Available at: https://gdcuffs.com/game_mechanics_deconstruct_2/ (In Russian)

14. *Alekseev A.A., Katasev A.S., Khassianov A.F., Tutubalina E.V., Zuev D.S.* Intellectual information decision support system in the field of economic justice // In Scientific service in the Internet: Proceedings of the XX All-Russian Scientific Conference, 2018, September. P. 17–27.

<http://doi.org/10.20948/abrau-2018-60> (In Russian)

15. *Sahibgareeva G.F., Kugurakova V.V.* Interactive Structure Editor for Scenario Prototyping Tool // Russian Digital Libraries Journal, 2021. Vol. 24, No. 6. P. 1184–1202.

<https://doi.org/10.26907/1562-5419-2021-24-6-1184-1202> (In Russian)

16. *Kozar B.A., Kugurakova V.V., Sahibgareeva G.F.* Structuring natural text entities using neural networks for generating 3D-scenes // Software and Systems, 2022, No. 3. P. 329–339. <http://dx.doi.org/10.15827/0236-235X.139.329-339> (In Russian)

17. *Sahibgareeva G.F., Kugurakova V.V., Bolshakov E.S.* Game Balance Tools // Russian Digital Libraries Journal. 2023. Vol. 26, No. 2. P. 225–251.

<https://doi.org/10.26907/1562-5419-2023-26-2-225-251> (In Russian)

18. *Prillaman M.* Is ChatGPT making scientists hyper-productive? The highs and lows of using AI // Nature. 2024. Vol. 627.

<https://doi.org/10.1038/d41586-024-00592-w>

19. *Wang R. et al.* K-ADAPTER: Infusing Knowledge into Pre-Trained Models with Adapters // in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. <https://doi.org/10.18653/v1/2021.findings-acl.121>

20. *Lewis P. et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks // in Advances in Neural Information Processing Systems, 2020, December. URL: <https://arxiv.org/pdf/2005.11401>.

СВЕДЕНИЯ ОБ АВТОРАХ



ШУБИН Алексей Витальевич – ассистент кафедры программной инженерии Института ИТИС КФУ.

Aleksey Vitalevich SHUBIN –assistant at the Department of Software Engineering of the Institute of ITIS KFU.

Email: shubin.aleksey.kpfu@gmail.com

ORCID: 0000-0002-6203-3268



КУГУРАКОВА Влада Владимировна – кандидат технических наук, и. о. зав. кафедрой индустрии разработки видеоигр Института ИТИС КФУ, руководитель НИЛ Digital Media Lab.

Vlada Vladimirovna KUGURAKOVA, Ph.D. of Engineering Sciences, Head of the Video Game Development Industry Department of ITIS KFU, Head of Laboratory «Digital Media Lab».

Email: vlada.kugurakova@gmail.com

ORCID: 0000-0002-1552-4910

Материал поступил в редакцию 24 сентября 2024 года