

ПРИМЕНЕНИЕ РУЧНОГО РЕДАКТОРА И ПОЛУАВТОМАТИЧЕСКОГО ПАРСЕРА ИСХОДНЫХ ТЕКСТОВ ДЛЯ ФОРМИРОВАНИЯ КОРПУСА ИГРОВЫХ СЦЕНАРИЕВ

Н. Р. Нурлыгаянов^[0000-0002-7969-0221]

Казанский (Приволжский) федеральный университет, г. Казань, Россия

nikita25@mail.ru

Аннотация

Предложен практический подход к формированию специализированного корпуса игровых сценариев на основе сочетания ручной разметки и полуавтоматического парсинга исходных текстов. Метод объединяет графический редактор для экспертной разметки, модульный парсер, выполняющий первичную структуризацию исходных текстов на основе формальных правил и эвристик, а также визуализацию связей между сценами. Такое сочетание обеспечивает итеративный цикл экспертной валидации, необходимый для получения корректного и воспроизводимого корпуса.

Ключевые слова: NLP, разработка видеоигр, парсинг, валидация, визуализация, игровые сценарии, корпус текстов.

ВВЕДЕНИЕ

Создание сюжета и диалогов – это один из самых трудоемких этапов разработки видеоигр. В нелинейных нарративах каждое ветвлениекратно увеличивает нагрузку на команду, а изменения на поздних стадиях влекут непропорционально высокие издержки на озвучивание и локализацию [1, 2]. Наиболее остро эти проблемы стоят в AAA¹ сюжетно-ориентированных играх с их максимальными требованиями к качеству [3].

¹ AAA (произносится как «трипл-эй») — неофициальный термин, используемый в игровой индустрии для обозначения игр с самыми высокими бюджетами разработки и маркетинга. Происхождение термина не имеет однозначного объяснения, но предполагается, что он возник по аналогии с системой оценки облигаций, где AAA обозначает наивысший рейтинг. В контексте видеоигр AAA указывает на наивысшее качество, масштаб и ожидаемый коммерческий успех проекта. AAA-игра характеризуется крупными инвестициями, франшизами, включающими несколько

Современные игроки ожидают нелинейных нарративных структур, значительной вариативности последствий совершаемых действий, персонализированных и контекстно-зависимых диалогов [4–7].

Целью работы является создание воспроизводимого процесса и инструментария для преобразования разнородных исходных сценариев в стандартизированный корпус, пригодный для аналитики, обучения и интеграции с игровыми системами. Предложенный подход сочетает полуавтоматический парсинг (формальные правила и эвристики для первичной структуризации) с экспертной ручной разметкой в интерактивном редакторе. Итеративная валидация с применением графового анализа позволяет выявлять и исправлять логические несоответствия (циклы, дедлоки), обеспечивая высокое качество итогового корпуса при контролируемых затратах ручного труда.

В работе представлены концептуальная модель данных корпуса, архитектура разработанных приложений (ручной редактор и полуавтоматический парсер с веб-интерфейсом), методы визуализации и верификации диалоговых структур, а также рассмотрены ограничения предложенного подхода и направления его дальнейшего развития.

ИНДУСТРИАЛЬНАЯ ВОСТРЕБОВАННОСТЬ

Крупные студии инвестируют в решения на базе NLP и генеративного искусственного интеллекта (ИИ) для автоматизации создания текстов и облегчения работы сценаристов (например, платформы Microsoft × Inworld AI) [8–11].

Трансформеры применимы для генерации NPC²-реплик, но требуют адаптации под игровую лексику [12]. Учет контекста и предшествующих событий способен повысить качество сценариев и сократить ручную постобработку, однако существуют ограничения (сложности с долгосрочным контекстом), а для надежной интеграции таких подходов в производственный цикл необходимы организационные меры (глоссарии, отлаженные процессы локализации) [12, 13].

названий, множеством персонажей и длинными сюжетными линиями, а также большой базой поклонников.

² NPC (Non-playable character, неигровой персонажей) — персонаж компьютерной игры, управляемый программой, а не игроком. Выполняет нарративные (сюжетные, диалоговые), и/или процедуральные функции (торговля, выдача заданий), выступая элементом игрового мира и средством коммуникации с ним.

АНАЛИЗ СУЩЕСТВУЮЩИХ КОРПУСОВ ТЕКСТОВ ВИДЕОИГР

Общие текстовые корпуса (Google Ngram, Wikipedia Dumps) не отражают специфику игровых сценариев: ситуативные реплики, игровой сленг и авторскую терминологию. Для анализа нарративов требуются специализированные корпуса, учитывающие мультимодальность³, связь с игровой логикой и динамический контент (квесты, диалоги) [14–17].

Ранее было предложено создание отдельного корпуса для каждой игры [18]. Примером альтернативного подхода служит проект Video Game Dialogue Corpus, агрегирующий данные из RPG для создания унифицированного лингвистического ресурса [19] (рис. 1).

Специализированный игровой корпус — это целенаправленно составленное и аннотированное собрание текстов из конкретных игр или серий. Он служит референсной базой для анализа нарративных решений, жанровой лексики и принципов построения диалогов [18]. Кроме того, такие корпуса используют для обучения моделей генерации естественного языка (NLG), обеспечивая дообучение алгоритмов для автоматизированного прототипирования диалогов и текстовых описаний.

Таким образом, специализированные корпуса способствуют оптимизации процесса разработки сценариев, предоставляя структурированные данные для последующего анализа и автоматизации.

Корпусы используют для обучения языковых моделей с целью генерации правдоподобного игрового контента [20]. Большие языковые модели (Large Language Models, LLM) позволяют осуществлять синтез фоновых текстов, описаний предметов и сценарных ветвей на основе соответствующих обучающих выборок из специализированных корпусов. Это создает потенциал для оперативного наполнения игровых сред дополнительным контентом с применением методов машинного обучения (Machine Learning, ML). Ключевым ограничивающим фактором в данном случае является качество исходного корпуса: чем выше его адекватность и репрезентативность, тем больше результаты синтеза контента соответствуют ожидаемым характеристикам.

³ Под мультимодальностью игрового текста понимается его свойство существовать не изолированно, а в неразрывной связи с анимацией, звуковым рядом и действиями игрока, где значение реплики или события рождается именно из их взаимодействия.

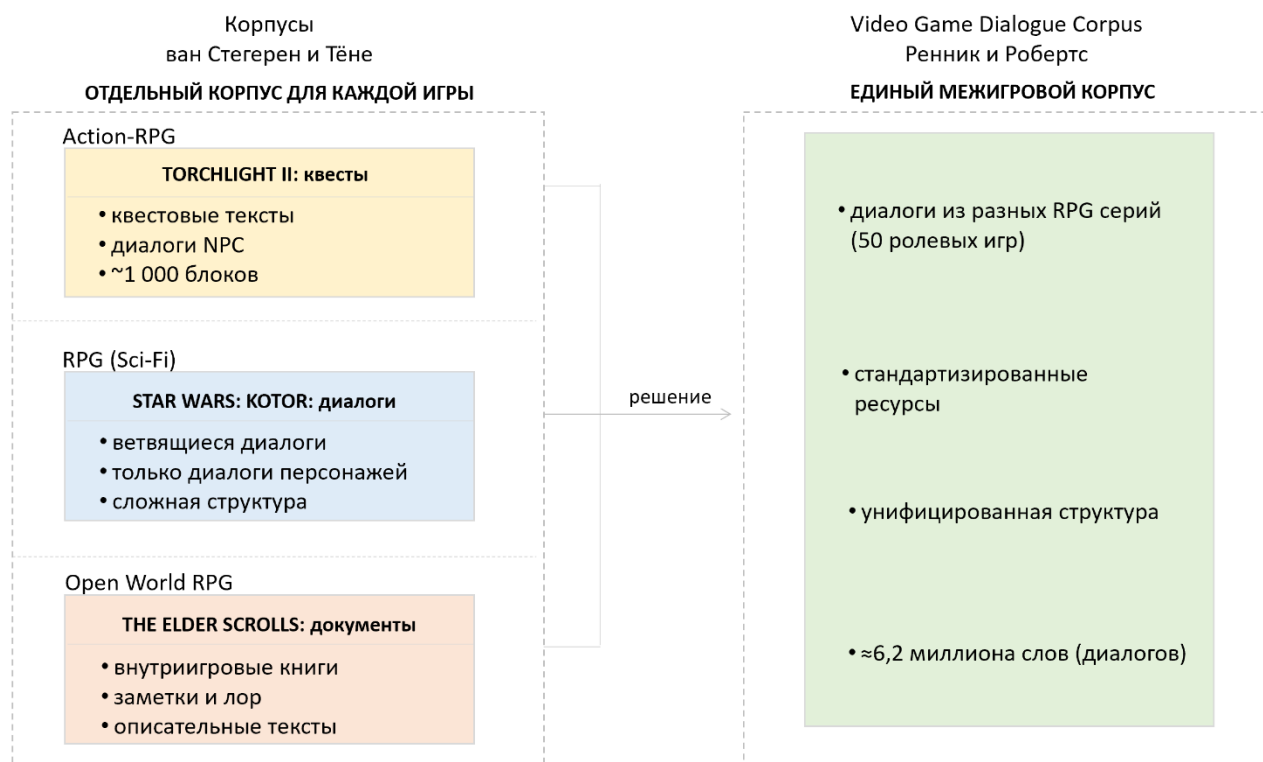


Рис. 1. Концепция унификации: от множества разнородных корпусов к единой структуре данных.

Специализированный корпус служит нормативным референс-ресурсом для сценаристов, обеспечивая стилистическую консистентность и жанровую специфику. В студийном производстве он сокращает адаптацию новых авторов и используется для контроля качества, включая верификацию терминологии. Генеративные модели, обученные на корпусе, способны создавать нарративные прототипы, инициируя авторскую разработку [21, 22].

Корпусы также применяют в локализации и персонализации контента: на их основе генерируют адаптивные ответы NPC и сценарии, учитывающие характеристики персонажей и контекст взаимодействия с игроком [4, 23–29]. Таким образом, специализированные игровые корпуса выступают эталоном для лингвистического обеспечения локализации, адаптивной генерации нарратива и оптимизации разработки.

Игровые тексты имеют сложную структуру: ветвящиеся диалоговые деревья, метки событий, индикаторы состояний, эмоциональную окраску и пара-

метры озвучки. Корпус должен объединять нарративные, описательные и интерфейсные данные, а также учитывать высокую концентрацию неологизмов и вымышленных имен [30].

Основная методологическая проблема заключается в ограниченности и неоднородности доступных корпусных данных, она связана со сложностью сбора, разработкой под конкретные задачи и отсутствием единых стандартов хранения и аннотации. Это затрудняет сравнительный анализ и вынуждает исследователей каждый раз собирать данные заново [18]. Для преодоления этих проблем Ван Стегерен и Тене предложили критерии качества корпусов, которые учтены в настоящей работе. Формирование корпусов игровых текстов опирается на два типа источников [18]: прямое извлечение из файлов игры (репрезентативно, сохраняет контекст, но технически сложно и юридически рискованно) и фанатские ресурсы (доступны в переносимых форматах, но уступают в полноте контекста и требуют верификации).

Типовая процедура извлечения текстов из видеоигр (рис. 2) требует обязательного документирования каждого этапа (используемые утилиты, параметры, ограничения) для обеспечения воспроизводимости результата. В целях повышения научной ценности итоговый корпус должен включать для каждой текстовой единицы комплекс метаданных, фиксирующих ее внутриигровой контекст.

Для структурирования извлеченных знаний применяются онтологии — формальные модели предметной области, описывающие концепты, их свойства и взаимосвязи [31]. Процесс разработки онтологии итеративен и включает пять этапов [32] (рис. 3).

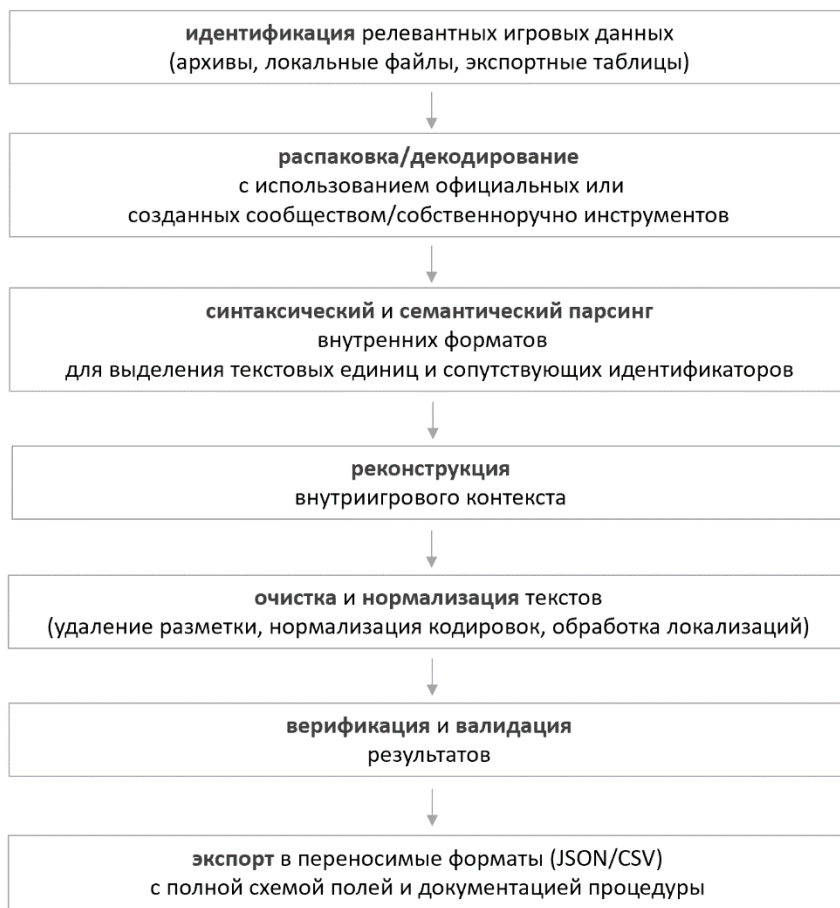


Рис. 2. Типовой процесс извлечения текстовых данных из видеоигр.



Рис. 3. Этапы создания онтологии.

Альтернативой онтологическому моделированию служит ручная разметка данных, однако этот подход, как отмечено в [33], является ресурсоемким, трудномасштабируемым и зависит от экспертов. Краудсорсинг⁴ снижает затраты, но порождает проблему вариативного качества аннотаций, особенно в сложных предметных областях с контекстными зависимостями. В связи с этим для построения корпусов игровых сценариев мы предлагаем отказаться от стратегии полной ручной аннотации в пользу воспроизводимого программного конвейера, который комбинирует формальные правила, эвристики и итеративную валидацию. Такой подход, демонстрирующий эффективность в ряде исследований [34–36], позволяет достичь оптимального баланса между качеством данных и трудозатратами на их подготовку.

ПРЕДЛАГАЕМЫЙ ПОДХОД

Для преодоления существующего ограничения возможно использование предложенной ранее модели [37], в которой сценарий трактуется как упорядоченный набор диалогов. Ранее в модели использовался один, основной файл в формате JSON, содержащий структурированные данные, представляющие диалоговую систему. Каждая запись в нем описывает отдельную реплику.

В настоящей работе модель расширена вторым, вспомогательным файлом (рис. 4) в формате JSON, предназначенным для хранения метаданных персонажей. Его структура включает поле `speaker` (идентификатор персонажа, согласованный с основным файлом) и поле `aliases`, представляющее собой список псевдонимов, используемых для данного персонажа в сценарии. Разделение на два файла обеспечивает компактность основного хранилища и гибкость при загрузке метаданных.

```
"26": [  
  "Oracle",  
  "The Witch"  
],
```

Рис. 4. Пример записи из дополнительного файла.

⁴ Краудсорсинг – привлечение широкого круга неспециалистов через онлайн-платформы для выполнения задач по сбору или разметке данных.

Предложенная модель данных открывает возможности для различных практических применений, часть из которых будет рассмотрена далее. Преобразование исходных сценариев в JSON можно осуществлять вручную, но предлагается использовать специально разработанные инструменты, обеспечивающие частичную или полную автоматизацию. Их использование позволит построить интерактивную визуализацию для многоаспектного анализа.

После загрузки данных в веб-браузере пользователь получает доступ к графовым представлениям. Основным является общий граф, где вершины соответствуют диалогам (наборам реплик с общим `id_dialogue`), а ребра – связям между диалогами. Направленные связи формируются на основе анализа полей `requires` и `opens`. Если хотя бы одна реплика диалога *A* ссылается хотя бы на одну реплику диалога *B* в поле `requires`, между вершинами устанавливается отношение зависимости. Если реплика в диалоге *A* содержит реплику диалога *B* в поле `opens`, создается связь, означающая активацию целевого диалога. Эти типы отношений визуально дифференцируются. Изолированные вершины (не имеющие связей) маркируются отдельным цветом, сигнализируя о возможных структурных проблемах.

Система автоматически детектирует циклы в графе, выделяя потенциально проблемные последовательности вида диалог *A* → ... → диалог *B*. Такие конструкции, указывающие на возможное сюжетное заикливание, визуально маркируются. Их наличие свидетельствует о нарушении прогрессии состояния в сценарии и представляет особую проблему для нелинейных нарративов с множеством ветвлений, где ручная верификация затруднена. Инструмент не только визуализирует циклические зависимости, но и делает доступными стратегии коррекции: модификацию полей `requires/opens`, введение промежуточных состояний или реструктуризацию нарративной ветви. Аналогичным образом при работе с графом отдельного диалога (вершины — реплики) проводится детекция локальных циклов: алгоритм проверяет, не образуют ли переходы внутри диалога ссылки на уже пройденные вершины, блокирующие развитие сцены. Данный подход обеспечивает двухуровневый контроль целостности сценария – на уровне общей сюжетной структуры и уровне отдельных диалоговых последовательностей.

Помимо циклических зависимостей модель обеспечивает формальную верификацию сценариев на наличие критических тупиковых ситуаций – дедлоков (deadlocks). В контексте нарративного дизайна дедлок определяется как взаимная блокировка, при которой прогресс в диалоге *A* обусловлен выполнением условия, активируемого диалогом *B*, в то время как для доступа к диалогу *B* необходимо условие из диалога *A*. Эта взаимообусловленность создает логический тупик, блокирующий дальнейшее развитие сюжета.

Детекция дедлоков осуществляется через анализ графа условных зависимостей, построенного на основе атрибутов *requires* и *opens* для каждого диалога. Простейшим примером является конфигурация:

- диалог *A*: {"requires": ["flag_X"], "opens": ["flag_Y"]};
- диалог *B*: {"requires": ["flag_Y"], "opens": ["flag_X"]},

где возникает классическая взаимная блокировка: активация *flag_X*, требуемого в диалоге *A*, возможна только в диалоге *B*, который, в свою очередь, требует предварительной активации *flag_Y*, устанавливаемого исключительно в диалоге *A*.

Для автоматического обнаружения взаимных блокировок (дедлоков) граф анализируется на наличие компонент сильной связности⁵ (применяется алгоритм Тарьяна [38]), которые визуально маркируются в интерфейсе как проблемные зоны.

Предложенная модель формирует методологическую основу для создания стандартизированного корпуса игровых сценариев, обеспечивающего интеграцию разнородных нарративных данных в единую структурированную базу. Формирование корпуса осуществляется в несколько этапов.

Начальный этап включает сбор исходных текстовых сценариев из игр различных жанров с обязательным соблюдением установленных критериев репрезентативности и качества.

⁵ Компонента сильной связности — это максимальное подмножество вершин ориентированного графа, в котором для любой пары вершин *U* и *V* существует ориентированный путь из *U* в *V* и из *V* в *U* (т. е. вершины взаимно достижимы). Каждая вершина графа принадлежит ровно одной такой компоненте.

Этап конвертации заключается в автоматизированном парсинге исходных данных с помощью специализированного инструментария. Данные трансформируются в соответствии с унифицированной JSON-схемой (основной и вспомогательный файлы). Применение алгоритмов полуавтоматического парсинга позволяет выполнить первичную структуризацию диалогов и сформировать унифицированное представление сценария. Идентификация диалоговых ветвлений, условий активации и спикеров в неоднозначных случаях предполагает последующую экспертную корректировку в интерактивном интерфейсе.

Этап валидации и семантической аннотации выполняется посредством интерактивной системы. Используя общий граф и графы отдельных диалогов, исследователь может выявлять и корректировать структурные аномалии (например, изолированные диалоги или некорректные условные связи) и семантические несоответствия, включая резкие изменения тональности в эмоциональных арках персонажей. Веб-интерфейс системы предоставляет возможность прямого редактирования с последующим экспортом исправленных данных.

На этапе агрегации JSON-файлы, прошедшие валидацию, объединяются в единый репозиторий. Практическая ценность созданного корпуса заключается в его применении для обучения предметно-ориентированных NLP-моделей (например, тонкой настройки GPT для генерации игровых диалогов), проведения сравнительного анализа жанровых нарративных паттернов и интеграции с системами процедурной генерации контента (PCG) для автоматизированного прототипирования игр.

ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ

Ручной редактор

Ручной редактор [39] реализован на C# с использованием WPF и предназначен для ручной обработки сценариев в соответствии с предложенной моделью. Инструмент позволяет формировать оба используемых JSON-файла и постепенно заполнять их, не требуя одновременного открытия внешних редакторов.

Приложение предназначено для двух основных сценариев применения: целенаправленной ручной аннотации для получения эталонных основной и вспомогательной разметок (полезно для настройки алгоритмов парсинга или по-

следующей визуализации) и создания эталонных наборов данных в случаях, когда автоматические алгоритмы (например, используемые в *полуавтоматическом парсере исходных текстов*) дают неудовлетворительный результат, что может потребовать дообучения.

Полуавтоматический парсер исходных текстов

Полуавтоматический парсер исходных текстов [40] реализован на Python и JavaScript с использованием библиотеки Pyvis (визуализация на базе VisJS) [41] и предназначен для автоматизации полного цикла обработки текстовых игровых сценариев: от парсинга исходного файла до генерации структурированных JSON-данных и их интерактивной веб-визуализации.

Приложение разделено на два компонента: модуль парсинга и генерации JSON (Python) и веб-интерфейс визуализации и правки данных (HTML/JS, VisJS через Pyvis).

Пользователю доступны три предобученные модели:

- 1) DistilBERT [42] — облегченная версия BERT, обеспечивающая быстрый вывод при умеренной точности; целесообразна для использования при ограниченных вычислительных ресурсах;
- 2) Twitter-XLM-Roberta [43] — XLM-Roberta-base, дообученная на корпусе твитов; эффективна для анализа неформального языка, сленга и эмоционально окрашенных текстов;
- 3) BERT Multilingual [44] — модель, предобученная на 104 языках; подходит для единообразного анализа в мультязычных корпусах без развертывания отдельных моделей.

Пользователь может пропустить шаг парсинга в интерфейсе и загрузить уже готовые JSON-файлы (например, созданные с помощью *ручного редактора*).

Общий граф визуализирует диалоги как узлы. На каждом узле отображаются id_dialogue и число реплик.

Узлы кодируются по наличию связей; изолированные узлы и циклы помечаются для ручной проверки. Система автоматически детектирует циклы и дед-

локи, подсвечивая вовлеченные узлы и ребра. Пользователю доступны получение пояснений о проблемных элементах (например, список записей, образующих цикл) и способы разрешения (например, ручное связывание).

Для изучения и редактирования доступны графы конкретных диалогов.

Объединённый граф сформирован по выбранным спикерам: диалоги, в которых участвуют все выбранные персонажи, объединяются в общий представительный граф.

Инструмент позволяет построить динамику тональности для визуализации эмоциональных арок. Для каждого диалога рассчитывается среднее значение тональности по всем репликам спикера.

При запуске веб-приложение автоматически уведомляет о найденных циклах (если они обнаружены) и позволяет перейти к проблемным участкам для их устранения.

Таким образом, *полуавтоматический парсер исходных текстов* обеспечивает полный цикл работы со сценарием: генерацию JSON по предложенной структуре, интерактивный анализ, визуальную диагностику структурных проблем и экспорт скорректированных данных. Приложение может использоваться как совместно с *ручным редактором* для подготовки корпусных данных, так и автономно для формирования и коррекции разметки, применяемой в дальнейших исследованиях и процедурах автоматической генерации прототипов игр.

НАПРАВЛЕНИЯ ДОРАБОТКИ РАЗРАБОТАННЫХ ПРИЛОЖЕНИЙ

Разработанный инструментарий обеспечивает структурированный анализ игровых нарративов, позволяя выявлять ошибки в эмоциональной динамике, оценивать баланс ролей персонажей и готовить данные для систем диалогового ИИ. В целях дальнейшего повышения универсальности, надежности и масштабируемости системы выделен ряд перспективных направлений для доработки.

Экспериментальные результаты и ограничения

Анализ эмоциональной окраски. Использование моделей-трансформеров позволило получить гибкое и контекстуально-зависимое представление о динамике эмоций. Возможность переключения между моделями в *полуавтомати-*

ческом парсере исходных текстов позволяет выявлять расхождения в интерпретациях, строить динамику тональности на графах диалогов и фиксировать эмоциональные пики сюжета.

Эксперименты показали, что ни одна из универсальных моделей не демонстрирует стабильно корректных результатов «из коробки» для всего многообразия игровых текстов. Это обусловлено следующими факторами:

- 1) уникальная лексика: вымышленные имена, топонимы, термины вселенной (например, «драконорождённый», «каджит» в случае *The Elder Scrolls V: Skyrim*) отсутствуют в стандартных обучающих корпусах, что приводит к некорректной токенизации и потере смысла;
- 2) сложный контекст: реплики NPC часто зависят от внутриигровых событий, выбора игрока или скрытых параметров, что выходит за рамки семантики отдельного предложения;
- 3) жанровая вариативность: стилистика диалогов в фэнтези-RPG, научной фантастике или постапокалипсисе кардинально различается, что снижает эффективность единой модели;
- 4) структурная сложность: ветвящиеся диалоги с условиями требуют понимания не только текста, но и логики переходов между репликами.

Как следствие, универсальные модели часто фрагментируют игровую лексику, неверно интерпретируют внутриигровые роли (например, антагонист или союзник) и демонстрируют низкую точность в задачах классификации.

Для преодоления этих ограничений необходима адаптация моделей. В перспективе можно рассмотреть два подхода: дообучение модели на корпусе конкретной игры, что позволит ей усвоить специфическую лексику и нарративные паттерны, и тонкая настройка, адаптирующая лишь часть параметров под сеттинг без потери общих лингвистических знаний.

Ограничения локальных LLM для автоматизации. Для гарантии целостности и воспроизводимости корпуса первичное преобразование сырого текста в JSON-структуру выполняется детерминированными парсерами. Это исключает неконтролируемые вариации, свойственные LLM. Сами языковые модели задействуются на поздних стадиях — при анализе тональности. Кроме того, в будущем

возможно их применение для задач, где допустима и даже желательна креативность: написание вариантов реплик, обогащение текста комментариями, адаптация под стиль персонажа.

Сравнение результатов применения программного комплекса для игр с линейным и нелинейным сценарием. Для валидации предложенного подхода и демонстрации возможностей разработанного инструментария было проведено сравнение двух сценариев, радикально различающихся по своей структуре: линейный сценарий коммерческой игры *Assassin's Creed: Bloodlines* (рис. 5) и нелинейный, ветвящийся сценарий авторской игры *Empty Life* (рис. 6).



Рис. 5. Фрагмент общего графа *Assassin's Creed: Bloodlines*.

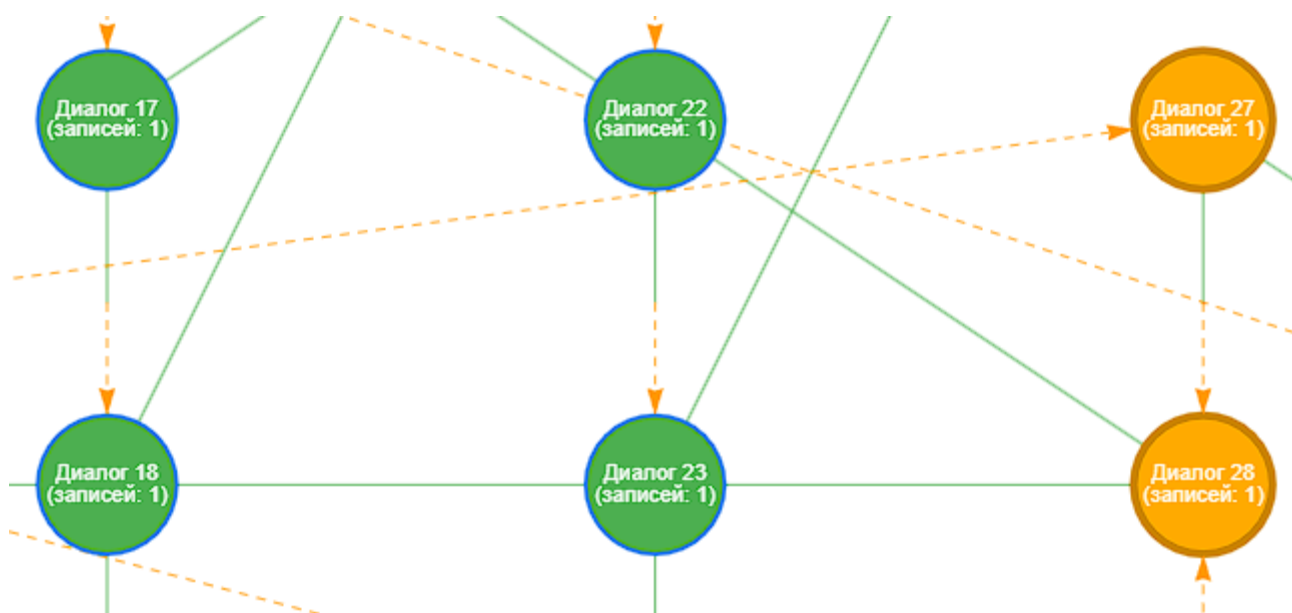


Рис. 6. Фрагмент общего графа *Empty Life*. Оранжевые узлы относятся к одному циклу.

На рис. 5–6 приведены узлы общих графов, где каждый узел представляет собой диалог. Число записей в диалоге — это число реплик.

Первый сценарий был преобразован в унифицированный JSON-формат с использованием одного из алгоритмов, встроенных в *полуавтоматический парсер исходных текстов*. Второй сценарий сформирован при помощи связи *ручного редактора* и *полуавтоматического парсера исходных текстов*.

На основе полученных данных построены графы диалоговых структур и рассчитан ряд количественных метрик, позволяющих объективно оценить сложность нарратива.

Метрики анализа. Для сравнения были использованы следующие характеристики:

- общее количество диалогов (уникальных id_dialogue) и реплик (записей);
- средняя длина диалога (реплик на диалог);
- количество узлов ветвления (диалогов, у которых поле opens содержит более одного идентификатора);
- количество точек схождения (диалогов, на которые ссылаются более одного другого диалога в поле requires);
- наличие циклов;

- плотность графа (отношение числа ребер-связей к числу узлов-диалогов).

Как видно из результатов количественного анализа (табл. 1), сценарий *Assassin's Creed: Bloodlines* представляет собой классическую линейную последовательность. Все 26 диалогов выстроены в цепочку, где каждый последующий диалог открывается только после завершения предыдущего. Отсутствие узлов ветвления и точек схождения подтверждает, что у игрока нет возможности влиять на развитие сюжета. Граф сценария является простым путем, а единственный терминальный узел соответствует финальной сцене игры. Плотность графа близка к минимальной, что характерно для строго последовательных нарративов. Инструментарий в данном случае выполняет прежде всего функцию структурирования и визуализации, позволяя быстро убедиться в отсутствии логических ошибок (циклов, «мертвых» веток).

Табл. 1. Сравнительные метрики сценариев *Assassin's Creed: Bloodlines* и *Empty Life*.

Метрика	Assassin's Creed: Bloodlines	Empty Life
Общее количество диалогов	26	34
Общее количество реплик	443	34
Средняя длина диалога	17	1
Узлы ветвления	0	31 (91 % от всех диалогов)
Точки схождения	0	15 (44 % от всех диалогов)
Наличие циклов	нет	да (обнаружен 1 цикл)
Терминальные узлы (концовки)	1 (финальная реплика)	11 (32 % от всех диалогов)
Плотность графа	0,04 (линейная цепь)	0,21 (высокая связность)

Сценарий *Empty Life* демонстрирует принципиально иную, нелинейную структуру. При сопоставимом общем количестве диалогов (34 против 26) средняя длина диалога составляет всего одну реплику, что указывает на высокую степень декомпозиции нарратива: каждое игровое состояние представлено отдельным диалогом-вершиной. Более 90 % диалогов являются узлами ветвления, предоставляя игроку выбор из двух или более вариантов продолжения. Наличие 15 точек схождения (диалогов, к которым можно прийти различными путями) свидетельствует о конвергенции сюжетных линий. Относительно высокая плотность графа (0/21) отражает более сложную сеть взаимосвязей между состояниями.

Наиболее важным результатом применения инструментария к *Empty Life* стала автоматическая детекция цикла в графе зависимостей. Такая структура означает, что игрок теоретически может бесконечно перемещаться по данному кольцу, не достигая концовки. В контексте авторского замысла это может быть как преднамеренным художественным приемом (симуляция «замкнутого круга» в жизни персонажа), так и ошибкой проектирования. Инструмент не только визуально маркирует подобные циклы, но и предоставляет разработчику полный список вовлеченных вершин для дальнейших анализа и коррекции.

Выполненное сравнение наглядно демонстрирует, что предложенный программный комплекс позволяет не просто визуализировать сценарий, но и проводить его объективный количественный анализ. Метрики, извлекаемые из унифицированного JSON-представления, дают четкую характеристику нарративной сложности: линейный сценарий описан минимальными значениями ветвлений и плотности, тогда как нелинейный характеризуется высокими показателями связности, наличием множества концовок и потенциально проблемных циклов.

Особую ценность представляет возможность автоматического обнаружения циклов и дедлоков. В линейном сценарии такие аномалии отсутствуют, и валидация сводится к проверке последовательности. В нелинейном же сценарии ручное выявление циклов крайне затруднительно, и автоматизированная диагностика становится незаменимым инструментом контроля качества, позволяя сценаристам и гейм-дизайнерам целенаправленно анализировать подозрительные участки нарратива.

Таким образом, разработанный инструментарий (рис. 7) успешно решает задачу анализа сценариев любого уровня сложности, предоставляя как наглядную визуализацию, так и строгие количественные метрики для оценки структурной целостности и выявления логических ошибок.

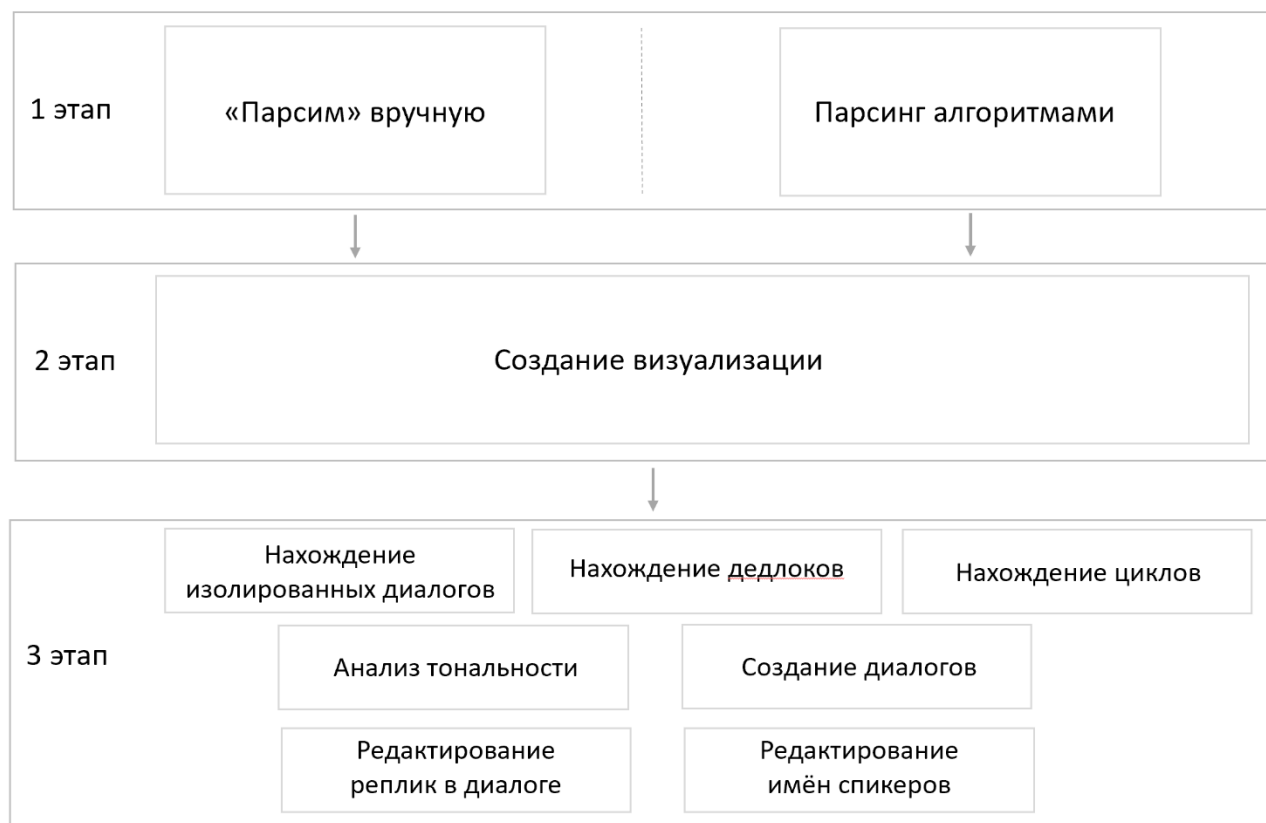


Рис. 7. Предлагаемая последовательность применения разработанного программного комплекса

Перспективные направления развития

На основе проведенного анализа были сформулированы следующие ключевые векторы развития инструментария:

- 1) повышение точности и устойчивости парсинга. Внедрение гибридного подхода, комбинирующего предобученные ML-модели для распознавания шаблонов и набор детерминированных правил для обработки нестандартных или поврежденных форматов данных;

- 2) расширение набора аналитических метрик;
- 3) анализ баланса и развития ролей. Разработка методов для количественной оценки значимости и экранного времени персонажей, оптимизации второстепенных ролей и трекинга развития нарративных арок ключевых героев;
- 4) сравнительный кросс-жанровый анализ. Применение инструментария к массивам сценариев из разных игр и жанров для выявления устойчивых паттернов (например, интересно сравнить сложность диалоговых деревьев в RPG с линейностью в экшн-играх) и оценки исторической эволюции приемов нарративного дизайна;
- 5) интеграция с системами разработки. Создание плагинов для экспорта структурированных данных и аналитических отчетов напрямую в популярные игровые движки (Unreal Engine, Unity).

Реализация указанных направлений позволит трансформировать *полуавтоматический парсер исходных текстов* в комплексную платформу для полного цикла работы с игровым нарративом – от анализа и верификации существующих сценариев до проектирования и интеграции новых, что существенно расширит его приложение как в академической сфере, так и в индустрии игр.

ЗАКЛЮЧЕНИЕ

Предложен комплексный программно-методический подход к формированию стандартизированного корпуса игровых сценариев, основанный на интеграции ручного редактора и полуавтоматического парсера исходных текстов. Разработанный инструментарий реализует концептуальную модель данных, впервые описанную в работе [40]. Данная модель позволяет формализовать нарративные конструкции, обеспечивая структурную целостность и логическую непротиворечивость игровых текстов.

Разработанный программный комплекс образует единый рабочий конвейер, поддерживающий полный цикл обработки сценариев: от первичного парсинга и разметки до визуализации, верификации и экспорта данных. Его применение позволяет эффективно сочетать точность экспертной ручной правки с производительностью автоматизированного анализа, что существенно снижает трудозатраты на подготовку корпусных данных.

Ключевыми практическими результатами работы являются:

- стандартизация представления игровых сценариев;
- автоматизированная валидация структурной целостности;
- контекстуальный анализ тональности и динамики диалогов.

Разработанный подход обладает значительным потенциалом для применения как в академических исследованиях (вычислительная лингвистика, анализ нарративов, кросс-жанровые сравнения), так и в индустрии разработки игр – для контроля качества сценариев, обучения диалоговых ИИ-систем, локализации и процедурной генерации контента.

Перспективные направления развития инструментария включают:

- повышение точности парсинга за счет гибридных ML-правил;
- интеграцию с игровыми движками (Unreal Engine, Unity);
- расширение аналитических метрик (сюжетная сложность, UX-метрики);
- поддержку динамических и персонализированных сценариев.

ЛУДОГРАФИЯ

Bethesda Game Studios. (2025) [2011.] *The Elder Scrolls V: Skyrim*. [видеоигра][загрузка][Windows, Xbox 360, PlayStation 3, Xbox One, PlayStation 4, Switch, Xbox Series X|S, PlayStation 5, Switch 2]. Bethesda Softworks.

Griptonite Games. (2009). *Assassin's Creed: Bloodlines*. [видеоигра][UMD, загрузка][PSP]. Ubisoft.

СПИСОК ЛИТЕРАТУРЫ

1. Engström H. Game Development Research. Skövde: University of Skövde, 2020, 242 p.
2. Oliver G. Renegade or Paragon?: Categorizing Narrative Choice in Video Game Storylines // Dialogue: The Interdisciplinary Journal of Popular Culture and Pedagogy. 2020. Vol. 7, Iss. 1. No. 9. P. 1-13.
3. Goh E., Al-Tabbaa O., Khan Z. Unravelling the complexity of the Video Game Industry: An integrative framework and future research directions // Telematics and Informatics Reports. 2023. Vol. 12. No. 100100. P. 1–18.

4. *Ashby T., Webb B. K., Knapp G., Searle J., Fulda N.* Personalized Quest and Dialogue Generation in Role-Playing Games: A Knowledge Graph- and Language Model-based Approach // Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). 2023. P. 1–20.
5. *Burgess J., Jones C.* Exploring Player Understandings of Historical Accuracy and Historical Authenticity in Video Games // Games and Culture: A Journal of Interactive Media. 2022. Vol. 17. Iss. 5. P. 816–835.
6. *Erb V., Lee S., Doh Y.Y.* Player-Character Relationship and Game Satisfaction in Narrative Game: Focus on Player Experience of Character Switch in The Last of Us Part II // Frontiers in Psychology. 2021. Vol. 12. No. 709926. P. 1–11.
7. *Zhang X., Huang Q., Shi Z., Zhang K.* How Cultural Elements Shape Game Evaluations: The Role of Cultural Authenticity and Perceived Effort // Computers in Human Behavior. 2024. Vol. 162. No. 108452.
8. *Inworld AI.* Xbox partners with Inworld AI to build generative AI tools for game development. URL: <https://inworld.ai/blog/xbox-partners-with-inworld-ai-to-build-generative-ai-tools-for-game-development>
9. Microsoft. Project Explora, powered by Inworld AI & Microsoft. URL: <https://developer.microsoft.com/en-us/games/products/project-explora/>.
10. *Klose T. Inworld AI.* Will Bring Generative AI to Microsoft Xbox Game Development. URL: <https://voicebot.ai/2023/11/13/inworld-ai-will-bring-generative-ai-to-microsoft-xbox-game-development/>
11. Microsoft. Xbox and Inworld AI partner to empower game creators with the potential of Generative AI. URL: <https://developer.microsoft.com/en-us/games/articles/2023/11/xbox-and-inworld-ai-partnership-announcement/>
12. *Huang J.* Generating Dynamic and Lifelike NPC Dialogs in Role-Playing Games Using Large Language Models. Lappeenranta: LUT University, 2024. 78 p.
13. IGDA Localization SIG. Best Practices for Game Localization. URL: <https://igda-website.s3.us-east-2.amazonaws.com/wp-content/uploads/2021/04/09142137/Best-Practices-for-Game-Localization-v22.pdf>
14. *Deterding S.* Contextual Autonomy Support in Video Game Play: A Grounded Theory // Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems New York: ACM, 2016. P. 3931–3943.

15. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice* Cambridge: Cambridge University Press, 2011. 295 p.
16. Small S., Stromer-Galley J., Strzalkowski T. Multi-modal annotation of quest games in Second Life // *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*. 2011. P. 171–179.
17. Summerville A., Snodgrass S., Guzdial M., Holmgård C., Hoover A.K., Isaksen A., Nealen A., Togelius J. Procedural Content Generation via Machine Learning (PCGML) // *IEEE Transactions on Games*. 2018. Vol. 10, No. 3. P. 257–270.
18. van Stegeren J., Theune M. Fantastic strings and where to find them: The quest for high-quality video game text corpora // *12th Intelligent Narrative Technologies Workshop (INT 2020) CEUR*. 2021. P. 1–8.
19. Rennick S., Roberts S. The video game dialogue corpus // *Corpora*. 2024. Vol. 19, No. 1. P. 93–106.
20. van Stegeren J., Myśliwiec J. Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation // *Proceedings of the 16th International Conference on the Foundations of Digital Games*. 2021. P. 1–8.
21. Kybartas B., Bidarra R. A survey on story generation techniques for authoring computational narratives // *IEEE Transactions on Computational Intelligence and AI in Games*. 2016. Vol. 9, No. 3. P. 239–253.
22. Riedl M.O., Young R.M. Narrative planning: Balancing plot and character // *Journal of Artificial Intelligence Research*. 2010. Vol. 39. P. 217–268.
23. Ginel M.R., Theroine S. Machine Translation and Gender Biases in Video Game Localisation: A Corpus-Based Analysis // *Journal of Data Mining & Digital Humanities*. 2022. P. 1–10.
24. Hansen D., Houlmont P.Y. A Snapshot into the Possibility of Video Game Machine Translation // *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track)*. 2022. Vol. 3. P. 257–269.
25. How Developers Localize Game Characters' Names, Voices, and More: Facts & Finer Points. URL: <https://igda.org/resources-archive/how-developers-localize-game-characters-names-voices-and-more-facts-finer-points/>

26. *Nathan D.* Sound and Unsound Practices in Documentary Linguistics: Towards an Epistemology for Audio and Video Corpora // *Language Documentation and Description*. 2010. Vol. 7. P. 262–288.
27. *Pirrone M., D’Ulizia A.* The Localization of Software and Video Games: Current State and Future Perspectives // *Information*. 2024. Vol. 15. No. 10.
28. *Raja R., Vats A.* Parallel Corpora for Machine Translation in Low-Resource Indic Languages: A Comprehensive Review // *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*. 2025. P. 129–143.
29. *Liu X., Xie Z., Jiang S.* Personalized Non-Player Characters: A Framework for Character-Consistent Dialogue Generation // *AI*. 2025. Vol. 6. No. 5.
30. *Bergsma T., van Stegeren J., Theune M.* Creating a Sentiment Lexicon with Game-Specific Words for Analyzing NPC Dialogue in The Elder Scrolls V: Skyrim // *Workshop on Games and Natural Language Processing*. 2020. P. 1–9.
31. *Gruber T.R.* A Translation Approach to Portable Ontology Specifications // *Knowledge Acquisition*. 1993. Vol. 5, No. 2. P. 199–220.
32. *Noy N.F. et al.* Ontology Development 101: A Guide to Creating Your First Ontology. 2001. P. 1–28.
33. *Snow R. et al.* Cheap and Fast – But Is It Good? Evaluating Non-Expert Annotations for Natural Language Tasks // *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 2008. P. 254–263.
34. *Mintz M. et al.* Distant Supervision for Relation Extraction without Labeled Data // *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing*. 2009. P. 1003–1011.
35. *Ratner A.J. et al.* Data Programming: Creating Large Training Sets, Quickly // *Advances in Neural Information Processing Systems*. 2016. Vol. 29. P. 1–9.
36. *Anastassiou P. et al.* Seed-tts: A family of high-quality versatile speech generation models // *arXiv*. 2024. No. arXiv:2406.02430. P. 1–18.
37. *Nurlygaianov N.R., Kugurakova V.V.* A New Approach to Creating a Corpus of Video Game Texts // *Russian Digital Libraries Journal*. 2024. Vol. 27, No. 4. P. 578–597.

38. Tarjan R.E. Depth-First Search and Linear Graph Algorithms // SIAM Journal on Computing. 1972. Vol. 1, No. 2. P. 146–160.

39. Нурлыгаянов Н.Р., Кузуркова В.В., Шубин А.В. Интерактивный редактор для представления сценариев компьютерных игр в виде JSON с данными о диалогах, их зависимостях и персонажах: свидетельство о государственной регистрации программы для ЭВМ RU № 2025684890 от 18.09.2025 / правообладатель: федеральное государственное автономное образовательное учреждение высшего образования «Казанский (Приволжский) федеральный университет». Заявка № 2025683882 от 20.08.2025.

40. Нурлыгаянов Н.Р., Кузуркова В.В., Шубин А.В. Программный комплекс для визуализации, интерактивного анализа и редактирования диалоговых структур в игровых сценариях: свидетельство о государственной регистрации программы для ЭВМ RU № 2025685255 от 22.09.2025 / правообладатель: федеральное государственное автономное образовательное учреждение высшего образования «Казанский (Приволжский) федеральный университет». Заявка № 2025684052 от 25.08.2025.

41. Perrone G., Unpingco J., Lu H. Network Visualizations with Pyvis and VisJS // arXiv. 2020. No. arXiv:2006.04951. P. 1–10.

42. Sanh V. et al. DistilBERT: A Distilled Version of BERT // arXiv. 2019. No. arXiv:1910.01108. P. 1–5.

43. Barbieri F., Anke L.E., Camacho-Collados J. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond // Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2021. P. 258–266.

44. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019. P. 4171–4186.

USING A MANUAL EDITOR AND A SEMI-AUTOMATIC PARSER OF SOURCE TEXTS TO BUILD A CORPUS OF GAME SCRIPTS

N. R. Nurlygaianov^[0000-0002-7969-0221]

Kazan (Volga region) Federal University, Kazan, Russia

nikita25@mail.ru

Abstract

A practical approach is proposed for building a specialized corpus of game scripts based on a combination of manual annotation and semi-automatic parsing of the source texts. The method integrates a graphical editor for expert annotation, a modular parser that performs initial structuring of the source texts using formal rules and heuristics, and a visualization of links between scenes. This combination provides an iterative cycle of expert validation necessary to obtain a correct and reproducible corpus.

Keywords: NLP, video game development, parsing, validation, visualization, game scripts, text corpus.

LUDOGRAPHY

Bethesda Game Studios. (2025) [2011.] *The Elder Scrolls V: Skyrim*. [video game][download][Windows, Xbox 360, PlayStation 3, Xbox One, PlayStation 4, Switch, Xbox Series X|S, PlayStation 5, Switch 2]. Bethesda Softworks.

Griptonite Games. (2009). *Assassin's Creed: Bloodlines*. [video game][UMD, download][PSP]. Ubisoft.

REFERENCES

1. Engström H. Game Development Research. Skövde: University of Skövde, 2020, 242 p.
2. Oliver G. Renegade or Paragon?: Categorizing Narrative Choice in Video Game Storylines // Dialogue: The Interdisciplinary Journal of Popular Culture and Pedagogy. 2020. Vol. 7, Iss. 1. No. 9. P. 1-13.
3. Goh E., Al-Tabbaa O., Khan Z. Unravelling the complexity of the Video Game Industry: An integrative framework and future research directions // Telematics and Informatics Reports. 2023. Vol. 12. No. 100100. P. 1–18.

4. Ashby T., Webb B. K., Knapp G., Searle J., Fulda N. Personalized Quest and Dialogue Generation in Role-Playing Games: A Knowledge Graph- and Language Model-based Approach // Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). 2023. P. 1–20.
5. Burgess J., Jones C. Exploring Player Understandings of Historical Accuracy and Historical Authenticity in Video Games // Games and Culture: A Journal of Interactive Media. 2022. Vol. 17. Iss. 5. P. 816–835.
6. Erb V., Lee S., Doh Y.Y. Player-Character Relationship and Game Satisfaction in Narrative Game: Focus on Player Experience of Character Switch in The Last of Us Part II // Frontiers in Psychology. 2021. Vol. 12. No. 709926. P. 1–11.
7. Zhang X., Huang Q., Shi Z., Zhang K. How Cultural Elements Shape Game Evaluations: The Role of Cultural Authenticity and Perceived Effort // Computers in Human Behavior. 2024. Vol. 162. No. 108452.
8. Inworld AI. Xbox partners with Inworld AI to build generative AI tools for game development. URL: <https://inworld.ai/blog/xbox-partners-with-inworld-ai-to-build-generative-ai-tools-for-game-development>
9. Microsoft. Project Explora, powered by Inworld AI & Microsoft. URL: <https://developer.microsoft.com/en-us/games/products/project-explora/>.
10. Klose T. Inworld AI. Will Bring Generative AI to Microsoft Xbox Game Development. URL: <https://voicebot.ai/2023/11/13/inworld-ai-will-bring-generative-ai-to-microsoft-xbox-game-development/>
11. Microsoft. Xbox and Inworld AI partner to empower game creators with the potential of Generative AI. URL: <https://developer.microsoft.com/en-us/games/articles/2023/11/xbox-and-inworld-ai-partnership-announcement/>
12. Huang J. Generating Dynamic and Lifelike NPC Dialogs in Role-Playing Games Using Large Language Models. Lappeenranta: LUT University, 2024. 78 p.
13. IGDA Localization SIG. Best Practices for Game Localization. URL: <https://igda-website.s3.us-east-2.amazonaws.com/wp-content/uploads/2021/04/09142137/Best-Practices-for-Game-Localization-v22.pdf>
14. Deterding S. Contextual Autonomy Support in Video Game Play: A Grounded Theory // Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems New York: ACM, 2016. P. 3931–3943.

15. *McEnery T., Hardie A.* Corpus Linguistics: Method, Theory and Practice Cambridge: Cambridge University Press, 2011. 295 p.
16. *Small S., Stromer-Galley J., Strzalkowski T.* Multi-modal annotation of quest games in Second Life // Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. 2011. P. 171–179.
17. *Summerville A., Snodgrass S., Guzdial M., Holmgård C., Hoover A.K., Isaksen A., Nealen A., Togelius J.* Procedural Content Generation via Machine Learning (PCGML) // IEEE Transactions on Games. 2018. Vol. 10, No. 3. P. 257–270.
18. *van Stegeren J., Theune M.* Fantastic strings and where to find them: The quest for high-quality video game text corpora // 12th Intelligent Narrative Technologies Workshop (INT 2020) CEUR. 2021. P. 1–8.
19. *Rennick S., Roberts S.* The video game dialogue corpus // Corpora. 2024. Vol. 19, No. 1. P. 93–106.
20. *van Stegeren J., Myśliwiec J.* Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation // Proceedings of the 16th International Conference on the Foundations of Digital Games. 2021. P. 1–8.
21. *Kybartas B., Bidarra R.* A survey on story generation techniques for authoring computational narratives // IEEE Transactions on Computational Intelligence and AI in Games. 2016. Vol. 9, No. 3. P. 239–253.
22. *Riedl M.O., Young R.M.* Narrative planning: Balancing plot and character // Journal of Artificial Intelligence Research. 2010. Vol. 39. P. 217–268.
23. *Ginel M.R., Theroine S.* Machine Translation and Gender Biases in Video Game Localisation: A Corpus-Based Analysis // Journal of Data Mining & Digital Humanities. 2022. P. 1–10.
24. *Hansen D., Houlmont P.Y.* A Snapshot into the Possibility of Video Game Machine Translation // Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track). 2022. Vol. 3. P. 257–269.
25. How Developers Localize Game Characters' Names, Voices, and More: Facts & Finer Points. URL: <https://igda.org/resources-archive/how-developers-localize-game-characters-names-voices-and-more-facts-finer-points/>

26. *Nathan D.* Sound and Unsound Practices in Documentary Linguistics: Towards an Epistemology for Audio and Video Corpora // *Language Documentation and Description*. 2010. Vol. 7. P. 262–288.
27. *Pirrone M., D’Ulizia A.* The Localization of Software and Video Games: Current State and Future Perspectives // *Information*. 2024. Vol. 15. No. 10.
28. *Raja R., Vats A.* Parallel Corpora for Machine Translation in Low-Resource Indic Languages: A Comprehensive Review // *Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)*. 2025. P. 129–143.
29. *Liu X., Xie Z., Jiang S.* Personalized Non-Player Characters: A Framework for Character-Consistent Dialogue Generation // *AI*. 2025. Vol. 6. No. 5.
30. *Bergsma T., van Stegeren J., Theune M.* Creating a Sentiment Lexicon with Game-Specific Words for Analyzing NPC Dialogue in The Elder Scrolls V: Skyrim // *Workshop on Games and Natural Language Processing*. 2020. P. 1–9.
31. *Gruber T.R.* A Translation Approach to Portable Ontology Specifications // *Knowledge Acquisition*. 1993. Vol. 5, No. 2. P. 199–220.
32. *Noy N.F. et al.* *Ontology Development 101: A Guide to Creating Your First Ontology*. 2001. P. 1–28.
33. *Snow R. et al.* Cheap and Fast – But Is It Good? Evaluating Non-Expert Annotations for Natural Language Tasks // *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 2008. P. 254–263.
34. *Mintz M. et al.* Distant Supervision for Relation Extraction without Labeled Data // *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing*. 2009. P. 1003–1011.
35. *Ratner A.J. et al.* Data Programming: Creating Large Training Sets, Quickly // *Advances in Neural Information Processing Systems*. 2016. Vol. 29. P. 1–9.
36. *Anastassiou P. et al.* Seed-tts: A family of high-quality versatile speech generation models // *arXiv*. 2024. No. arXiv:2406.02430. P. 1–18.
37. *Nurlygaianov N.R., Kugurakova V.V.* A New Approach to Creating a Corpus of Video Game Texts // *Russian Digital Libraries Journal*. 2024. Vol. 27, No. 4. P. 578–597.

38. Tarjan R.E. Depth-First Search and Linear Graph Algorithms // SIAM Journal on Computing. 1972. Vol. 1, No. 2. P. 146–160.

39. Nurlygaianov N.R., Kugurakova V.V., Shubin A.V. Interaktivnyi redaktor dlya predstavleniya stsensariy komp'iuternykh igr v vide JSON s dannymi o dialogakh, ikh zavisimostyakh i personazhakh: svidetel'stvo o gosudarstvennoi registratsii programmy dlya EVM RU No. 2025684890 ot 18.09.2025 / pravoobladatel': federal'noe gosudarstvennoe avtonomnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya "Kazanskii (Privolzhskii) federal'nyi universitet". Zayavka No. 2025683882 ot 20.08.2025.

40. Nurlygaianov N.R., Kugurakova V.V., Shubin A.V. Programmnyi kompleks dlya vizualizatsii, interaktivnogo analiza i redaktirovaniya dialogovykh struktur v igrovykh stsensariyakh: svidetel'stvo o gosudarstvennoi registratsii programmy dlya EVM RU No. 2025685255 ot 22.09.2025 / pravoobladatel': federal'noe gosudarstvennoe avtonomnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya "Kazanskii (Privolzhskii) federal'nyi universitet". Zayavka No. 2025684052 ot 25.08.2025.

41. Perrone G., Unpingco J., Lu H. Network Visualizations with Pyvis and VisJS // arXiv. 2020. No. arXiv:2006.04951. P. 1–10.

42. Sanh V. et al. DistilBERT: A Distilled Version of BERT // arXiv. 2019. No. arXiv:1910.01108. P. 1–5.

43. Barbieri F., Anke L.E., Camacho-Collados J. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond // Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2021. P. 258–266.

44. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019. P. 4171–4186.

СВЕДЕНИЯ ОБ АВТОРЕ



НУРЛЫГАЯНОВ Никита Рамильевич — ассистент кафедры индустрии разработки видеоигр ИТИС КФУ. Сфера научных интересов — автоматизация процесса разработки компьютерных игр.

Nikita Ramilevich NURLYGAIANOV — assistant at the Department of Video Game Development Industry of ITIS KFU. Research interest is automation of the computer game development process.

email: nikita25@mail.ru

ORCID: 0000-0002-7969-0221

Материал поступил в редакцию 10 мая 2026 года