

УДК 004.514+004.928

ПРОСТРАНСТВЕННАЯ ОРИЕНТАЦИЯ ОБЪЕКТОВ НА ОСНОВЕ ОБРАБОТКИ ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ ДЛЯ ГЕНЕРАЦИИ РАСКАДРОВОК

В. В. Кугуракова¹, Г. Ф. Сахибгареева², А. З. Нгуен³, А. М. Астафьев⁴

*¹⁻⁴Казанский (Приволжский) федеральный университет, Высшая школа
информационных технологий и интеллектуальных систем*

¹vlada.kugurakova@gmail.com, ²gulnara.sahibgareeva42@gmail.com,

³zung.nguyen.98@gmail.com, ⁴xiaomao472@gmail.com

Аннотация

Статья посвящена подходам в обработке текста на естественном языке для выяснения конкретного пространственного взаиморасположения объектов и трехмерной покадровой визуализации. Предложенные алгоритмы учитывают влияние явных ограничений извлеченных пространственных отношений и позволяют создавать вероятные трехмерные схемы расположения объектов на сцене. Такие пространственные схемы необходимы для уточнения воплощения замысла сценариста при проектировании видеоигр. Для создания удачных кадров использованы такие правила режиссуры, как учет плана, поворот камеры и другие композиционные нюансы.

Ключевые слова: преобразование текста в изображение, преобразование текста в сцену, преобразование текста в анимацию, понимание естественного языка, сценарий, видеоигра, сценарный прототип, положение камеры, режиссура.

ВВЕДЕНИЕ

Тема исследования связана с нашими попытками создать инструмент генерации раскадровок для сценарного прототипа видеоигры, которые являются черновым вариантом повествовательной составляющей, включая сценарий, внутриигровые тексты, инъекции предоставления возможностей для ведения диалога, повествовательные механики и другие инструменты передачи истории.

Архитектура разрабатываемого инструмента для генерации сценарного прототипа описана ранее в [1–3] и представлена на рис. 1.



Рисунок 1. Архитектура инструмента генерации сценарного прототипа.

Инструмент должен принимать на вход сценарий в виде текста на естественном языке. При этом мы считаем, что сценарий должен быть отформатирован по стандарту Голливуда [4] для лучшего распознавания сущностей.

Модуль редактирования должен давать возможность вносить изменения в сценарий, прототип и структуру сценария.

Модуль отображения должен реализовывать возможность просмотреть результат анализа документа: структура истории, сущности и статистика – количество персонажей, имена персонажей, реплики и т. д.

Генератор сценарного прототипа включает в себя три модуля, каждый из которых последовательно обрабатывает результат работы предыдущего:

1) модуль анализа текста должен распознавать необходимые сущности и связи между ними;

2) модуль анализа данных и построения структуры должен устанавливать целостную структуру сценария, связывать сущности между собой и строить структуру сценарного прототипа;

3) модуль визуализации должен подобрать необходимые трехмерные модели и распределить их по структуре сценарного прототипа; далее этот модуль должен отправить полученную визуализацию сценарного прототипа, а также визуализированную структуру сценария в Модуль воспроизведения.

Модуль воспроизведения должен позволять запускать получившийся сценарный прототип и диаграммы баланса (см. рис. 1).

В развитии идеи генерации сценарного прототипа [1] мы продвинулись на очередной этап – уточнение работы компонента визуализации. В данной работе мы представляем синтез прототипных решений для генерации трехмерной сцены для создания раскадровки. Для генерации используется текст, написанный на естественном языке.

Раскадровка – это промежуточный этап разработки генератора трехмерной анимации для интерактивного проекта. В данной работе рассматриваются такие нюансы генерации раскадровки, как постановка кадра и позиционирование предметов в пространстве.

Особое внимание посвящено пространственным предлогам, интерпретация которых помогает в генерации как общей трехмерной сцены, так и строения отдельного кадра.

Задачи пространственной ориентации объектов, создания трехмерной сцены и постановки кадра, решение которых мы предлагаем, включены в компонент визуализации.

СВЯЗАННЫЕ РАБОТЫ

Использование раскадровки в разработке игр сокращает время, деньги и другие ресурсы на создание компьютерной графики [2]. Более того, мы видим раскадровку как инструмент для тестирования повествовательной составляющей игры [1]. Компания Disney использует раскадровку как один из этапов разработки игр [5]. Функция генерации раскадровки по текстовому запросу есть в одной из версий Storyboarder, который мы использовали в качестве платформы для экспериментов [6].



(а) средний план; (б) крупный план; (в) очень крупный план

Рисунок 2. Примеры планов.

При генерации параметров для постановки кадра были выбраны следующие факты из практики кинорежиссуры ([7–11]): есть общий, средний и крупный планы, которые отличаются по количеству пространства, которое занимают в кадре ключевые объекты (см. рис. 2); есть 5 видов поворота камеры по оси X (с высоты птичьего полета, сверху вниз, на уровне глаз, снизу вверх, из-под ног) и 3 вида по оси Y (слева, спереди, справа), а также съемка со спины; расположение объектов регулируется правилами третей; камеры лучше двигать слева направо.

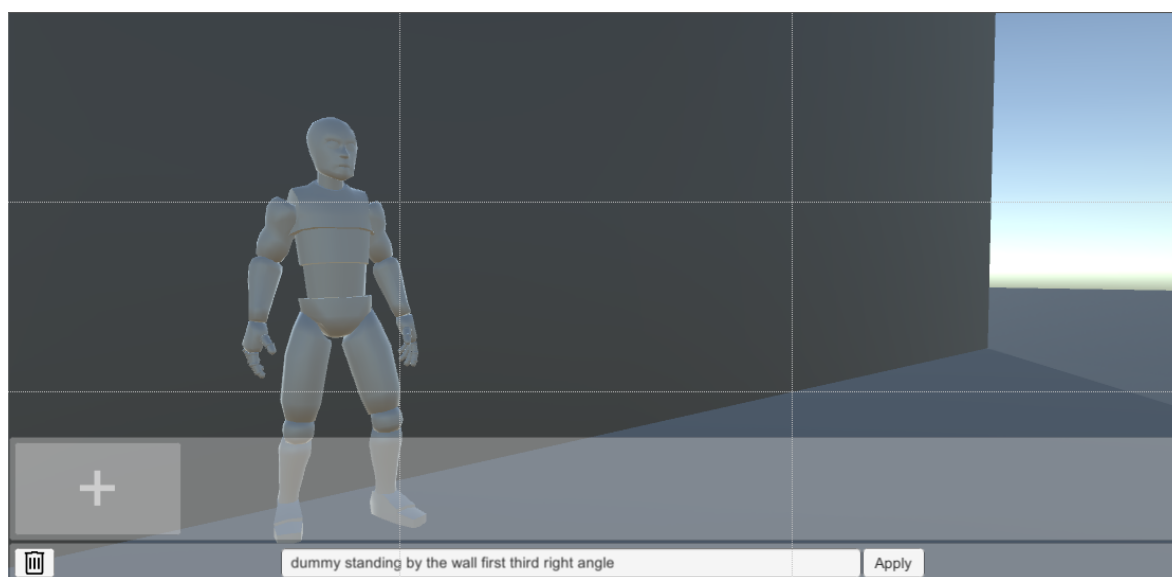


Рисунок 3. Изображение, разделенное по правилу третей.

Расположение объектов в соответствии с правилом третей – это принцип построения композиции, упрощающий правило золотого сечения. По этому принципу изображение следует рассматривать в виде девяти областей, которые разделены двумя линиями по вертикали и горизонтали, параллельными сторонам кадра, в определенных соотношениях. Важные части композиции обычно расположены на этих линиях или на их пересечениях, что акцентирует на них

внимание и преобладает над стандартным расположением объекта в центре кадра (см. рис. 2). Однако выбор тоже влияет на восприятие объектов в кадре. Направляющим фактором в данном случае является то, как люди воспринимают направление в кадре. Исходя из выводов исследования [12], считаем, что зрителям проще воспринимать движение слева направо, поэтому именно так и стоит располагать ключевые объекты. Однако бывают и менее тривиальные решения, где внимание зрителя сначала попадает в центр, затем – в правую, а затем – в левую части.

На разных этапах обработка текста происходит благодаря разным алгоритмам, это может быть система регулярных выражений, алгоритмы построения синтаксических и семантических сетей или обученные модели, но так или иначе важным является особенность цели интерпретации. В случае с генерацией раскадровки из постановки трехмерной сцены важными являются взаимное расположение объектов и камеры относительно друг друга, размеры локации и обстоятельства изменений в сцене.

В решении этой задачи помогает анализ пространственных предлогов, которые встречаются в английском тексте на естественном языке. Мы обратили внимание на следующие предлоги, которые помогают настроить сцену для статичных снимков: *by, in, to, with, nearby, between, front, behind, after, at, near, far, end, above, north, south, east, west, right, left, front*. Каждый предлог влияет на композицию объектов в кадре, а также на настройки камеры.

Естественный язык – простая и эффективная среда для описания визуальных идей и мысленных образов. Авторы [13], предвидя появление основанных на языке систем генерации трехмерных сцен, позволяющих обычным пользователям быстро создавать трехмерные сцены без необходимости изучать специальное программное обеспечение, еще в 2001 году представили собственный подход – *WordsEye* – систему для автоматического преобразования текста в репрезентативные 3D-сцены [14]. Интересно отметить и более поздние работы упомянутых авторов (см., например, [15; 16]), которые описывают детально процессы, необходимые для выяснения из текста относительного расположения объектов или уточнений месторасположения по описательным характеристикам.

Однако пространственное знание является важным аспектом мира и часто *не выражается явно* на естественном языке. WordsEye не решила проблему представления необъявленных фактов, которая является аналогом проблемы описания изображений (см., например, [17]). Эту проблему пытаются решать авторы [18] (кроме того, что извлекается представление пространственного знания из текста (слева, сверху и т. д.), изучается отображение между языком и геометрическими ограничениями, а также собирается статистика того, какие объекты встречаются в различных типах сцен, и рассматриваются их пространственные положения относительно друг друга.

НАША ИДЕЯ

Для получения удовлетворительного результата генерации мы разработали конвейер, который представлен в виде схемы (см. рис. 4).



Рисунок 4. Схема конвейера.

Обработка текста. Большая часть кода разработана на python; кроме этого, используется C#.

Для извлечения информации из текста проведен анализ следующих биб-

лиотек: nltk, gensim, CoreNLP, OpenNLP, spaCy. В разработке использована библиотека spaCy [19], т. к. она производит вычисления быстрее других инструментов, кроме того, из её достоинств стоит упомянуть функционал для визуализации результата разбора фраз.

Сущностями, которые необходимо разметить в сцене, считаются имена существительные. Позы, в которые необходимо разместить персонажей, определяются по глаголам, относящимся к именам существительным. Относительное размещение сущностей определяется пространственными предлогами.

Алгоритм обработки для извлечения информации об окружении и предметах в сцене, настройках камеры выглядит так:

1. Маркировка тегов предметов на сцене;
2. Синтаксический разбор предложения;
3. Выделение сущностей, их состояний и (при наличии) их поз;
4. Определение пространственных предлогов;
5. Выделение информации о размещении сущности в кадре и на сцене, а также параметров настройки камеры;
6. Проверка наличия сочетаний слов, которые были установлены правилами, например, определение главного и второстепенного объектов;
7. Составление таблицы из извлеченных блоков текста.

Таблица 1. Взаимоотношения между объектами

word	distance	rotation_side	rotation_up	bool_is_interna	main_object
near	0	-1	-1	-1	-1
far	2	-1	-1	-1	-1
end	2	-1	-1	-1	-1
north	-1	90	-1	-1	-1
south	-1	180	-1	-1	-1
east	-1	0	-1	-1	-1
west	-1	270	-1	-1	-1
right	-1	0	-1	-1	-1
left	-1	180	-1	-1	-1
front	-1	90	-1	-1	-1
behind	-1	270	-1	-1	-1
above	-1	-1	180	-1	-1
under	-1	-1	0	-1	-1

Пространственные предлоги. Чтобы определить влияние пространствен-

ных предлогов на сущности, необходимо построить дерево взаимоотношений между объектами, что имеет смысл сделать при помощи выбранного синтаксического анализатора spaCy.

В конечном итоге должна быть составлена таблица (см. Табл. 1) следующего вида: предлог, зависимый или управляемый объект, угол поворота камеры, расстояние.

Машинное обучение. Метод, предложенный в [18], позволяет формировать выводы на основании статистических данных. Так можно получить значения конкретной дистанции между объектами в сцене.

Сборка сцены и режиссура. Далее автоматически построенная трехмерная сцена, учитывающая как полученные на основе анализа пространственных отношений в тексте (предлоги), так и необъявленные факты, в обычной речи следующие из здравого смысла, но которые необходимо сформировать заранее, так и на учете конкретных дистанций между объектами, превращаются в кадры, которые оформлены по правилам режиссерской и операторской работы (см. [7–11]).

ОПИСАНИЕ МЕТОДА

Как было сказано выше, для сценарного прототипа большое значение имеет функциональная визуализация, которую необходимо получать автоматически с возможностью редактирования. Генерация раскадровок – необходимый этап для перехода к анимации. К построению кадров раскадровки мы подходим через создание трехмерной сцены и последующий режиссерский выбор фокуса и положения камеры. Опишем поэтапно создание трехмерной сцены и создание раскадровки.

Создание трехмерной сцены происходит по следующему алгоритму:

1. На вход подается блок текста на естественном языке;
2. После его обработки формируются данные о сцене, описанной в тексте.

а. Для обработки текста была выбрана библиотека spaCy, которая привлекает к себе внимание своей скоростью работы и обладает всеми нужными инструментами и готовыми моделями для обработки естественного текста, в особенности английского: стоит отметить, что главным инструментом этой библиотеки

для фокуса нашей работы, а именно, определения относительного расположения объектов, стала возможность составить из текста дерево зависимостей. Приятным дополнением `displaCy` является возможность рассмотреть, как был разобран текст, и увидеть взаимосвязь между словами в предложении (см. рис. 5). Поскольку синтаксическое дерево предоставляет еще и POS-тэги для слов, а также именованные связи, то основная идея заключается в использовании этих связей для составления правил, по которым будет выделяться нужная информация.

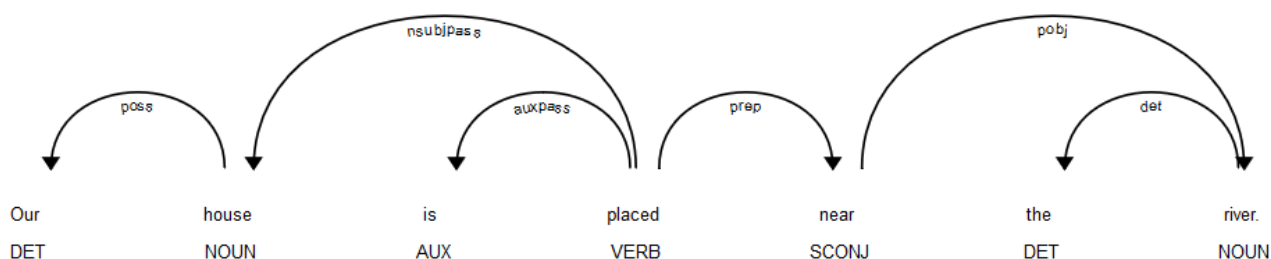


Рисунок 5. Утилита `displaCy` из библиотеки `sraCy` позволяет визуализировать взаимосвязи между словами в предложении.

b. В английском языке для обозначения позиции объекта относительно другого объекта используются пространственные предлоги. Рассмотрим пример, представленный рис. 6. Здесь родителем интересующего нас предлога «near» является глагол «to place», а существительные «house» и «river» являются потомками глагола и предлога соответственно. И в данном случае главным существительным будет «house», а второстепенным – «river». Таким образом можно создать правило, которое будет извлекать блоки текста, если в предложении появятся 2 существительных, глагол и предлог, связанные между собой определенным способом.

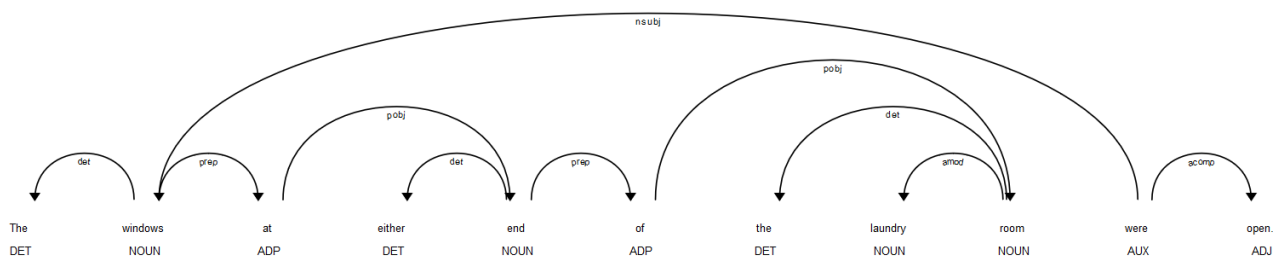


Рисунок 6. Пример взаимосвязей в предложении.

c. Следует отметить, что пространственные определения, такие как «near the river», можно ставить не только после глагола, но и после существительного. В

данном случае на самом деле просто опускается глагол «to be» для более красивой речи.

d. Если рассмотреть антоним предлогу «near», то можно заметить, что «far» всегда используется в связке с «from» (у «near» часто опускается «to»), образуя тем самым еще одно правило.

e. Иногда указанием на относительное месторасположение объектов служат не пространственный предлог, а существительное. Тогда эти два существительных связаны друг с другом сочетанием «at end of», состоящим из двух предлогов и существительного. Здесь существительное «end» играет ключевую роль, обозначая, что окно находится в конце комнаты, а два предлога не несут никакой смысловой нагрузки и стоят лишь в соответствии с правилами английского языка.

f. Немаловажно в данных правилах определить, какое существительное из двух будет главным существительным, а какое второстепенным. Если, например, в предложении «стол рядом со стулом» поменять местами существительные, то сути это не изменяет, однако если поменять в «стол справа от стула», то уже возникнет ошибка. Все изменяется в зависимости от того, относительно которого объекта происходит сравнение. В данном случае и в русском, и в английском языках главным объектом всегда будет существительное, находящееся в уточняющем дополнении.

g. Поскольку не все предлоги являются пространственными и не все существительные передают информацию о взаимном месторасположении объектов, то было принято решение использовать словарь. Благодаря нему, обрабатывая текст, инструмент будет исключать те блоки предложения, в которых не встречаются нужные слова. Также, поскольку в последствии потребуются расставить объекты на сцене, за каждым словом были закреплены некоторые числовые значения.

3. После создания дерева зависимостей библиотекой spaCy выделяются все существительные, которые есть в предложениях. Чтобы сопоставлять их с «объектами мира», каждый из них лемматизируется (то есть приводится в словарную форму). После этого проверяется наличие сочетаний слов, которые были установлены правилами, созданными ранее. Из извлеченных блоков текста со-

ставляется таблица с полями:

- a. Главный объект;
- b. Второстепенный объект;
- c. Взаимное месторасположение объектов;
- d. Числовые значения – удаленность одного объекта от другого и относительный угол.

4. Затем конкретизируются параметры о месторасположении объектов относительно друг друга.

a. Параметр Rotation содержит информацию о позиции объекта относительно другого объекта (справа, слева, сверху, снизу). В качестве численного параметра используются градусы (см. рис. 7). При размещении относительно не главного, а второстепенного объекта, значение пересчитывается ($Rotation - 180$).

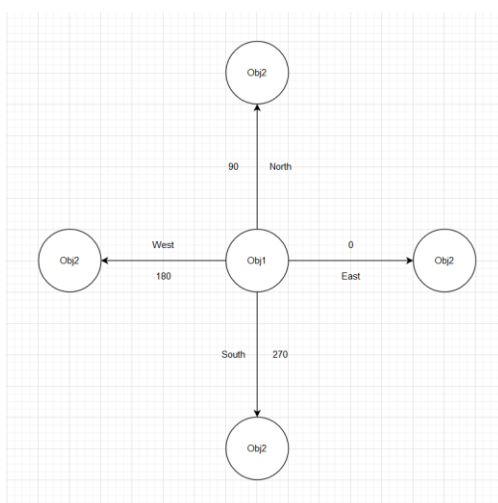


Рисунок 7. Градусы для обозначения позиции объекта относительно другого объекта.

5. Параметр Position (см. рис. 8) содержит информацию о расстоянии между объектами (от 0 до 2). Это обусловлено тем, что такие слова, как «близко» и «далеко», являются довольно расплывчатым определением расстояния между объектами. Человек воспринимает информацию, опираясь на свой накопленный опыт, поэтому для него «корабль около порта» воспринимается буквально, что корабль находится непосредственно у порта, а «человек у горы» – что человек находится либо у подножия горы, либо в двух трех километрах от нее.

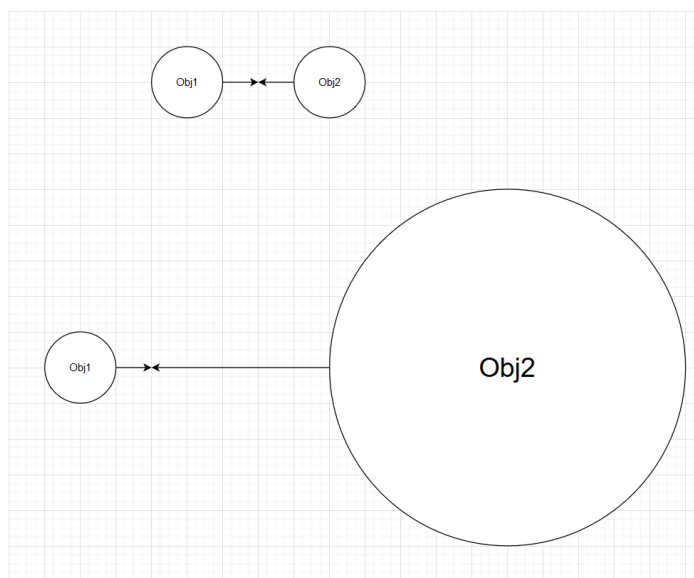


Рисунок 8. Расстояние между объектами задается радиусом.

6. На основании сформированных параметров создается карта местности с расположением объектов (см. [20]).

7. В работе [18] решают проблему уточнения числовых значений.

При генерации трехмерной сцены на основе текста на естественном языке важно не просто построить дерево зависимостей извлеченных сущностей, но и достроить описание сцены, чтобы на выходе получить адекватный результат. Когда на вход поступает запрос вроде «В комнате стол с тортом», на выходе должен получиться торт, который стоит на столе, а не под ним на полу. Такое уточнение дает модель, которую разработали в 2014 году [18], и она показала хорошие результаты (рис. 9).

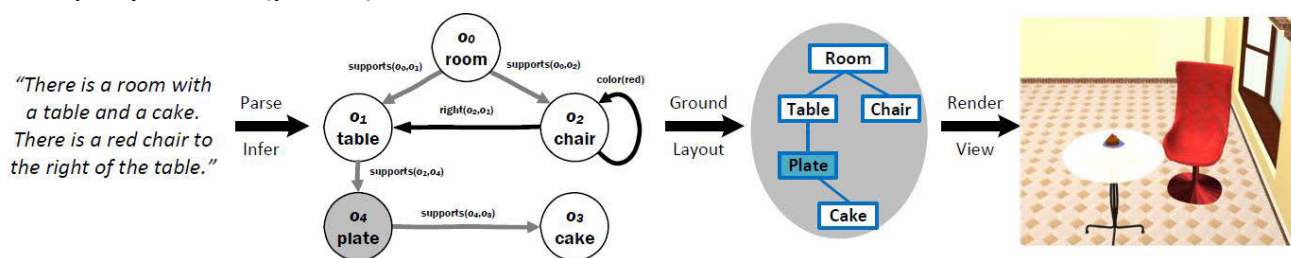


Рисунок 9. Конвейер генерации трехмерной сцены (по [18]).

Для обработки текстов в инструменте используется Stanford CoreNLP. В результате анализа текста происходит построение геометрического графа, в вершинах которого располагаются именованные трехмерных объектов, а ребрами являются пространственные отношения между ними. Пространственные отно-

шения определяются шаблонами. Чтобы получить их, в системе используются условная вероятность и правило Байеса. Полученные правила вносят изменения в граф, построенный ранее.

Кроме этого, модель активно обучается в процессе работы пользователя в редакторе полученной трехмерной сцены. Однако в [21] утверждают, что WordsEye превосходит разработку [18] с точки зрения понимания языка и объема репозитория трехмерных моделей.

Так как необходимо извлекать из текста неявную информацию, то безусловно нужно включить один из упомянутых в [18] подходов или иметь возможность выбора наилучшего решения из множества – как минимум, для уточнения вертикального позиционирования объектов, обрабатывая положения, формируемые соответствующими предлогами (ср. русские предлоги *с, над, под, на, при, у, около, за, перед, из-за, из-под* и т. п.).

Создание раскадровок

8. Во время разбора текста для получения общей трехмерной сцены с детальной проработкой расстановки объектов на ней параллельно происходит выделение тегов, как описывающих сцену, так и описывающих влияние на параметры постановки камеры. Важнейшей задачей являются обеспечение условий для постановки различных вариантов композиций и изменение состояний персонажей и объектов, присутствующих на собранной сцене.

Разработка возможностей, связанных с построением сцен и раскадровки для динамического действия, не является приоритетной, несмотря на некоторые сходства ее производства, значительно уступающих существующим различиям. Взаимодействие анимированных персонажей друг с другом относится к этим различиям, так как конкретная анимация персонажей за редким исключением задает композицию кадра. Однако в случаях, когда важно именно само взаимодействие персонажей, его проработка выходит за рамки задач раскадровки.

9. После синтаксического разбора предложений полученное формализованное описание включает конкретные предметы сцены, их состояния, пространственные отношения, а также информацию, необходимую для размещения объекта в кадре и на сцене.

10. По найденным тегам вызываются соответствующие методы, кото-

рые ищут группу объектов, относительно которой размещается субъект пространственного предлога. В разных предложениях разные приоритеты объекта и субъекта в кадре, взаимное расположение объектов. Помимо координат объектов, подлежащих изменениям на сцене, изменяется также их значение поворота.

11. Далее относительно готовой трехмерной сцены специальный компонент изменяет положение камеры в ней, исходя из заданных параметров. Параметры кадра соответствуют выведенным ранее типам съемки и различным углам поворота камеры. Их генерация происходит посредством компонента, который обрабатывает полученный результат разбора предложения, учитывая количество важных частей речи, конкретное влияние некоторых из них на тип кадра, количество основ предложения. Однако полученные таким образом параметры съемки уступают по приоритету заданным вручную, которые также собираются с помощью этого компонента из текстового описания.

12. Собранные таким образом параметры влияют на вычисление стартовой позиции камеры. Постановка камеры начинается со стартовой позиции, которая вычисляется, исходя из положения ключевых объектов таким образом, что все они находятся в поле зрения камеры. Далее это положение изменяется в зависимости от параметров. Исходя из полученных параметров, разными путями возможно дальнейшее редактирование положения и угла наклона камеры.



Рисунок 10. Интерфейс инструмента, выбран кадр для редактирования.

Когда постановка сцены и камеры завершена, становится доступна запись всех параметров данного кадра. Как только кадр устраивает пользователя, он

может приступить к созданию следующего, который будет начат с тех же объектов и типов съемки, что и предыдущий. Последовательность полученных кадров доступна для редактирования инструментом по нажатию на кадр. Также изменения возможны и путем прямого введения тегов для описания сцены (см. рис. 10).

Последовательно обрабатывая текстовое описание сцен, пользователь получает в конце работы последовательность раскадровки, представленную в виде графических растровых изображений, построенных на основе коллекции трехмерных моделей, с которыми ассоциированы извлекаемые теги-наименования объектов или персонажей (см. [22]).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Рассмотрим пример использования разработанного инструмента и сравнение результатов работы с использованными в фильме материалами. Для этого используем вариант словесного описания, происходящего на изображениях (см. рис. 11) на основе материалов киносериала «Мост» [23].

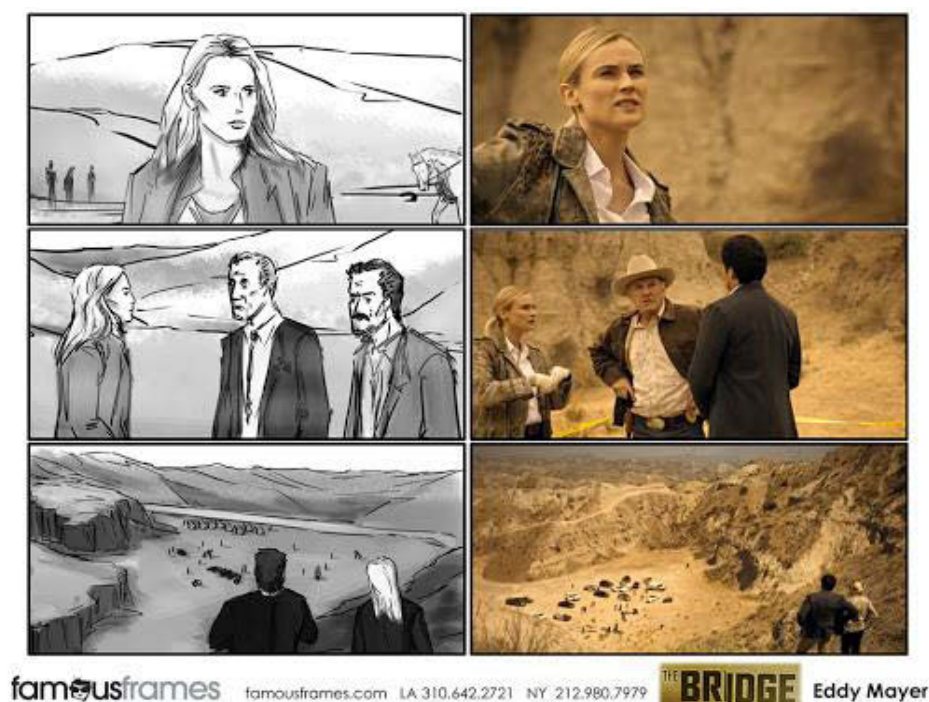


Рисунок 11. Сценарная раскадровка и итоговые кадры.

Первый в последовательности кадр удалось получить из описания “<character> is standing firmly by the cliff first third”. Таким образом, данные об

угле поворота камеры и типе съемки были получены, исходя из наличия в тексте наречий или дополнений, а также глагола.

Второй кадр был получен из “<character1> and <character2> are talking with <character3> by the cliff right angle”. По аналогии с первым, часть информации композиции кадра была получена из описания, но также была изменена с помощью добавления в текст конкретных тегирующих фраз, влияющих на композицию.



Рисунок 12. Полученная последовательность кадров.

Третий кадр был получен из описания: “<character1> and <character2> are staring at field backshot last third right angle”. Несмотря на значительные руко-творные изменения композиции, результат оказался лишь частично схож с оригиналом из-за различий в рельефе сцены. Таким образом, с применением описания и добавления некоторых тегов типа съемки получается последовательность кадров, в большей степени соответствующая рассмотренному примеру (рис. 12).

Также с помощью инструмента удалось воспроизвести стандартную диалоговую последовательность (см. рис. 13) – когда камера переключается с лица на лицо во время диалога.

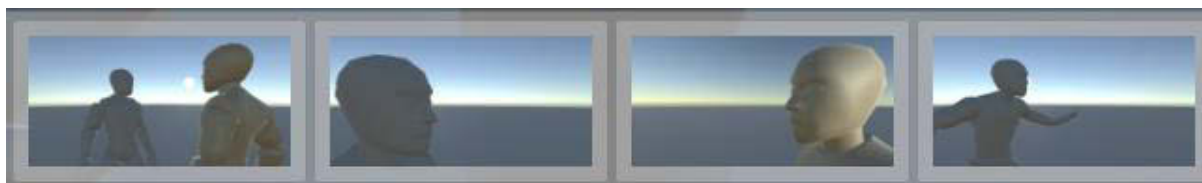


Рисунок 13. Диалоговая последовательность.

Исходя из приведенных выше примеров, можно сделать вывод о том, что выбранные подходы сборки сцены и конкретная реализация позволяют получить ценный результат при работе с прототипированием сценария в виде раскадровки. Наиболее полезен этот инструмент будет в работе именно со статичными компоновками.

ДАЛЬНЕЙШЕЕ РАЗВИТИЕ

Описываемый функционал инструмента генерации сценарного прототипа ещё имеет массу чёрных пятен, ещё только начаты исследования следующих функций:

- Создание реалистичного окружения – для этого необходимы распознавание элементов ландшафта (реки, озёра, дороги) и их размещение относительно точки зрения повествующего.
- Визуализация сюжетных линий и реализация вариативности сюжетов.
- Автоматическое проектирование сценария с разбивкой на отдельные сцены и кадры.
- Автоматическая примитивизация и автоматическое тегирование публичных трехмерных моделей для увеличения архива моделей, используемых при генерации раскадровок. Сложность такого подхода связана также с вынесением данных о модели из её бинарного представления, а информация о подобном декодировании не всегда доступна.
- Автоматический риггинг¹ модели персонажа для получения состояния по ключевым фразам, упоминающим некоторые характеристики: телосложение, пол, действие и т. п.

¹ Риггинг – расстановка вершин на меш (каркас) трехмерной модели для последующего создания анимации.

- Использование тегов для разметки эмоционального различия анимаций персонажей.
- Генерация светового решения сцены соответственно эмоциональному посылу действия и композиции кадра.

Конечно, важна и собственно реализация многокомпонентной архитектуры, которая подразумевает заменяемость функционала отдельных элементов инструмента генерации сценарного прототипа.

ЗАКЛЮЧЕНИЕ

Разработанный подход позволяет строить быстрые (относительно стандартного процесса создания раскадровок) визуализации на основе генерации трехмерной сцены, для чего были созданы правила, которые позволяют извлекать полезную информацию из текста и формировать итоговый кадр по тегам, описывающим положение камеры. Тестовая генерация раскадровки оказалась быстрее ручной сборки в 6 раз.

Таким образом, представлена основная часть инструмента для генерации сценарных и игровых прототипов [24] – как для компьютерных игр, так и для *serious games* или образовательных тренажеров (см., например, [25]).

СПИСОК ЛИТЕРАТУРЫ

1. Сахибгареева Г.Ф., Бедрин О.А., Кугуракова В.В. Разработка компонента генерации визуализации сценарного прототипа видеоигр // Научный сервис в сети Интернет: труды XXII Всероссийской научной конференции. М.: ИПМ им. М.В. Келдыша, 2020. С. 581–603.

2. Сахибгареева Г.Ф. Разработка решения для визуализации сценарного прототипа инструментами генерации раскадровок // Выпускная квалификационная работа магистратуры Высшей школы ИТИС. Казанский федеральный университет, Казань, 2020. 75 с.

URL: https://kpfu.ru/student_diplom/10.160.178.20_X87SBS2R8OQUD822Q1Y9MOGMYD_6QTSVCTR0JTXAC5Z54A_W6A_Sahibgareeva.pdf.

3. Сахибгареева Г.Ф. Разработка инструмента для создания игрового сценарного прототипа: выпускная квалификационная работа бакалавриата Высшей школы ИТИС. Казанский федеральный университет, Казань, 2018. 57 с.
URL: https://kpfu.ru/student_diplom/10.160.178.20_9311629_F_Sahibgareeva.pdf.
4. Хейг М. Голливудский стандарт: как написать сценарий для кино и ТВ, который купят / Майкл Хейг; Пер. с англ. М.: Альпина нон-фикшн, 2017. 388 с.
5. Lee N., Madej K. Disney Stories: Getting to Digital. Springer, 2012. 114 с.
6. Storyboarder: Wonder Unit. Режим доступа:
URL: <https://wonderunit.com/storyboarder/> (Дата обращения: 12.05.2020).
7. Thompson R. Grammar of the Shot. Focal Press, 1998. 219 p.
8. Hockrow R. Out of Order: Storytelling Techniques for Video and Cinema Editors. Peachpit Press, 2014. 240 p.
9. Dise J. Filmmaking 101: Camera Shot // Explora, 2017. Режим доступа:
URL: <https://www.bhphotovideo.com/explora/video/tips-and-solutions/filmmaking-101-camera-shot-types> (Дата обращения: 6.04.2020).
10. Gallea R., Ardizzone E., Pirrone R. Automatic aesthetic photo composition // International Conference on Image Analysis and Processing. Springer, Berlin, Heidelberg, 2013. P. 21–30.
11. Olivier P., Tsujii J. A computational view of the cognitive semantics of spatial prepositions // Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1994. P. 303–309.
12. Egizii M.L. et al. Which way did he go? Directionality of film character and camera movement and subsequent spectator interpretation // International Communication Association conference. Phoenix, AZ, 2012. Режим доступа:
URL: https://www.researchgate.net/publication/228448619_Which_Way_Did_He_Go_Directionality_of_Film_Character_and_Camera_Movement_and_Subsequent_Spectator_Interpretation (Дата обращения: 28.04.2020).
13. Coyne B., Sproat R. WordsEye: an automatic text-to-scene conversion system // Proceedings of the 28th annual Conference on Computer Graphics and Interactive Techniques, 2001. P. 487–496.

14. *Ulinski M., Coyne B., Hirschberg J.* Evaluating the WordsEye Text-to-Scene System: Imaginative and Realistic Sentences // Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 2018. P. 1493–1499.

15. *Coyne B., Sproat R., Hirschberg J.* Spatial relations in text-to-scene conversion // Computational Models of Spatial Language Interpretation, Workshop at Spatial Cognition, 2010. 8 p.

16. *Coyne B., Klapheke A., Rouhizadeh M., Sproat R., Bauer D.* Annotation tools and knowledge representation for a text-to-scene system // Proceedings of COLING 2012: Technical Papers. 2012. P. 679–694.

17. *Elliott D., Keller F.* Image description using visual dependency representations // Proceedings of Empirical Methods in Natural Language Processing (EMNLP). 2013. P. 1292–1302.

18. *Chang A.X., Savva M., Manning Ch.D.* Learning Spatial Knowledge for Text to 3D Scene Generation // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. P. 2028–2038.

19. spaCy: библиотека для NLP // Официальный сайт spaCy с документацией. Режим доступа: URL: <https://spacy.io> (Дата обращения: 10.06.2020).

20. *Нгуен А.З.* Генерация окружения на основе текстового описания: выпускная квалификационная работа бакалавриата Высшей школы ИТИС. Казанский федеральный университет, Казань, 2020. 45 с.

URL: https://kpfu.ru/student_diplom/10.160.178.20_OW9ARNOUY_98UU56A_D0JDQMEBFOF__66GZP8OQRFCLC4RY26N_F_Nguyen.pdf.

21. *Hassani K., Lee W.-S.* Visualizing Natural Language Descriptions // ACM Computing Surveys. 2016. V. 49(1). P. 1–34.

22. *Астафьев А.М.* Разработка инструмента для сборки сцен по тегам: выпускная квалификационная работа бакалавриата Высшей школы ИТИС. Казанский федеральный университет, Казань, 2020. 75 с.

URL: https://kpfu.ru/student_diplom/10.160.178.20_FPEBER9KDIZQVYJAE3VRTIFYWZB_CDDM972OPP2I28S0EEFABT_Astafev.pdf.

23. Episodes of “The Bridge” / Киносериял // BBC. 2020. Режим доступа: URL: <https://www.bbc.co.uk/programmes/b01gxlxj/episodes/guide> (Дата обращения: 24.04.2020).

24. Сахибгареева Г.Ф., Кугуракова В.В. Концепт инструмента автоматического создания сценарного прототипа компьютерной игры // Электронные библиотеки. 2018. Т. 21, №3-4. С. 235–249.

25. Антонов И.О., Зезегова К.В., Кугуракова В.В., Лазарев Е.Н., Хафизов М.Р. Программирование запахов для виртуального осмотра места происшествия // Электронные библиотеки. 2018. Т. 21, № 3-4. С. 301–313.

SPATIAL ORIENTATION OF OBJECTS BASED ON PROCESSING OF A NATURAL LANGUAGE TEXT FOR STORYBOARD GENERATION

V. V. Kugurakova¹, G. F. Sahibgareeva², A. Z. Nguyen³, A. M. Astafiev⁴

¹⁻³ *Kazan Federal University (KFU), Institute of Information Technology and Intelligent Systems (ITIS)*

¹vlada.kugurakova@gmail.com, ²gulnara.sahibgareeva42@gmail.com,

³zung.nguyen.98@gmail.com, ⁴xiaomao472@gmail.com

Abstract

The article is devoted to our approaches to processing text in natural language to clarify the specific spatial relationship of objects and three-dimensional frame-by-frame visualization. The proposed approach allows us to show how the explicit constraints of the extracted spatial relationships affect and makes it possible to create possible layouts of objects on the scene. Natural language interpretations for spatial knowledge can generate three-dimensional scenes, which in turn are necessary to translate the scriptwriter's intent into the design of video games. The work also takes into account the rules of directing to create successful shots. Among them, accounting for the plan, camera rotation, as well as compositional nuances.

Keywords: text-to-picture conversion, text-to-scene conversion, text-to-animation conversion, natural language understanding, video game, scenario, scenario prototype.

REFERENCES

1. *Sahibgareeva G.F., Kugurakova V.V., Bedrin O.A.* Razrabotka resheniya dlya vizualizacii scenarnogo prototipa instrumentami generacii raskadrovok // Vserossijskaya nauchnaya konferenciya "Nauchnyj servis v seti Internet". 2020. S. 581–603.
2. *Sahibgareeva G.F.* Razrabotka resheniya dlya vizualizacii scenarnogo prototipa instrumentami generacii raskadrovok // Vypusknaya kvalifikacionnaya rabota magistratury Vysšej shkoly ITIS. Kazanskij federal'nyj universitet, Kazan', 2020. 75 s.
URL: https://kpfu.ru/student_diplom/10.160.178.20_X87SBS2R8OQUD822Q1Y9MOGMYD_6QTSVCTROJTXAC5Z54A_W6A_Sahibgareeva.pdf.
3. *Sahibgareeva G.F.* Razrabotka instrumenta dlya sozdaniya igrovogo scenarnogo prototipa: vypusknaya kvalifikacionnaya rabota bakalavriata Vysšej shkoly ITIS. Kazanskij federal'nyj universitet, Kazan', 2018. 57 s. URL: https://kpfu.ru/student_diplom/10.160.178.20_9311629_F_Sahibgareeva.pdf.
4. *Heĭg M.* Gollivudskiĭ standart: Kak napisat' scenariĭ dlya kino i TV, kotoryĭ kupyat / Maĭkl Heĭg; Per. s angl. M.: Al'pina non-fikshn, 2017. 388 s.
5. *Lee N., Madej K.* Disney Stories: Getting to Digital. Springer. 2012. 114 p.
6. Storyboarder: Wonder Unit. URL: <https://wonderunit.com/storyboarder/> (Date of access: 12.05.2020).
7. *Thompson R.* Grammar of the Shot. Focal Press. 1998. 219 p.
8. *Hockrow R.* Out of Order: Storytelling Techniques for Video and Cinema Editors. Peachpit Press. 2014. 240 p.
9. *Dise J.* Filmmaking 101: Camera Shot // Explora. 2017. URL: <https://www.bhphotovideo.com/explora/video/tips-and-solutions/filmmaking-101-camera-shot-types> (Date of access: 6.04.2020).
10. *Gallea R., Ardizzone E., Pirrone R.* Automatic aesthetic photo composition // International Conference on Image Analysis and Processing. Springer, Berlin, Heidelberg. 2013. P. 21–30.
11. *Olivier P., Tsujii J.* A computational view of the cognitive semantics of spatial prepositions // Proceedings of the 32nd annual meeting on Association for

Computational Linguistics. Association for Computational Linguistics. 1994. P. 303–309.

12. *Egizii M.L. et al.* Which way did he go? Directionality of film character and camera movement and subsequent spectator interpretation // International Communication Association Conference. Phoenix, AZ, 2012.

URL: https://www.researchgate.net/publication/228448619_Which_Way_Did_He_Go_Directionality_of_Film_Character_and_Camera_Movement_and_Subsequent_Spectator_Interpretation (Date of access: 28.04.2020).

13. *Coyne B., Sproat R.* WordsEye: an automatic text-to-scene conversion system // Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001. P. 487–496.

14. *Ulinski M., Coyne B., Hirschberg J.* Evaluating the WordsEye Text-to-Scene System: Imaginative and Realistic Sentences // Proceedings of the Eleventh International Conference on Language Resources and Evaluation. 2018. P. 1493–1499.

15. *Coyne B., Sproat R., Hirschberg J.* Spatial relations in text-to-scene conversion // Computational Models of Spatial Language Interpretation, Workshop at Spatial Cognition. 2010. 8 p.

16. *Coyne B., Klapheke A., Rouhizadeh M., Sproat R., Bauer D.* Annotation tools and knowledge representation for a text-to-scene system // Proceedings of COLING 2012: Technical Papers. 2012. P. 679–694.

17. *Elliott D., Keller F.* Image description using visual dependency representations // Proceedings of Empirical Methods in Natural Language Processing (EMNLP). 2013. P. 1292–1302.

18. *Chang A. X., Savva M., Manning Ch.D.* Learning Spatial Knowledge for Text to 3D Scene Generation // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. P. 2028–2038.

19. spaCy: библиотека для NLP // Официальный сайт spaCy с документацией. URL: <https://spacy.io> (Date of access: 10.06.2020).

20. *Nguen A.Z.* Generaciya okruzeniya na osnove tekstovogo opisaniya: vypusknaya kvalifikacionnaya rabota bakalavriata Vysšej shkoly ITIS. Kazanskij federal'nyj universitet, Kazan', 2020. 45 s.

URL: https://kpfu.ru/student_diplom/10.160.178.20_OW9ARNOUY_98UU56A_DOJDQMEBFOF__66GZP8OQRFCLC4RY26N_F_Nguyen.pdf.

21. *Hassani K., Lee W.-S.* Visualizing Natural Language Descriptions // ACM Computing Surveys. 2016. V. 49(1). P. 1–34.

22. *Astaf'ev A.M.* Razrabotka instrumenta dlya sborki scen po tegam: vypusknaya kvalifikacionnaya rabota bakalavriata Vysšej shkoly ITIS. Kazanskij federal'nyj universitet, Kazan', 2020. 75 s.

URL: https://kpfu.ru/student_diplom/10.160.178.20_FPEBER9KDIZQVYJAE3VRTIFYWZB_CDDM972OPP2I28S0EEFABT_Astafev.pdf.

23. Episodes of “The Bridge” / movie series // BBC. 2020.

URL: <https://www.bbc.co.uk/programmes/b01gxlxj/episodes/guide> (Date of access: 24.04.2020).

24. *Sahibgareeva G.F., Kugurakova V.V.* Koncept instrumenta avtomaticheskogo sozdaniya scenarnogo prototipa komp'yuternoj igry // Elektronnye biblioteki. 2018. T. 21. No. 3-4. S. 235–249.

25. *Antonov I.O., Zezegova K.V., Kugurakova V.V., Lazarev E.N., Hafizov M.R.* Programmirovaniye zapahov dlya virtual'nogo osmotra mesta proisshestiya // Elektronnye biblioteki. 2018. T. 21, No. 3-4. S. 301–313.

СВЕДЕНИЯ ОБ АВТОРАХ



КУГУРАКОВА Влада Владимировна – к. т. н., доцент кафедры программной инженерии ИТИС КФУ, руководитель НИЛ разработки AR/VR приложений и компьютерных игр. Сфера научных интересов – генерация реалистичной визуализации трехмерных сцен и различные аспекты проектирования игр.

Vlada Vladimirovna KUGURAKOVA, PhD., Docent of ITIS KFU, Head of Laboratory «AR/VR applications and Gamedev». Research interests include different approaches in game design.

email: vlada.kugurakova@gmail.com.



САХИБГАРЕЕВА Гульнара Фаритовна – ассистент кафедры программной инженерии ИТИС КФУ. Сфера научных интересов – игровая сценаристика, нарративный дизайн, изучение вопроса эффективности создания сценарного прототипа и возможности автоматизировать данный процесс.

Gulnara Faritovna SAHIBGAREEVA – assistant of ITIS KFU. Sphere of research: game scripting, narrative design, studying the issue of the effectiveness of creating a scenario prototype and the ability to automate this process.

email: gulnara.sahibgareeva42@gmail.com



НГУЕН Ань Зунг – выпускник бакалавриата ИТИС КФУ. Сфера научных интересов – использование нейронных сетей в разработке игр.

An' Zung NGUYEN – Bachelor of the ITIS KFU. Research interests include neural networks for game development.

email: zung.nguyen.98@gmail.com



АСТАФЬЕВ Андрей Максимович – студент магистратуры ИТИС КФУ. Сфера интересов: разработка игр и AR/VR приложений на Unity.

Andrey Maksimovich ASTAFIEV – Masters student of ITIS KFU. Research interests include development in Unity.

Email: xiaomao472@gmail.com

Материал поступил в редакцию 8 ноября 2020 года