

1.6. ВЕБ-ОРИЕНТИРОВАННАЯ РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ИНТЕРНЕТ-БРОНИРОВАНИЯ ОТЕЛЕЙ

Ясницкий М.В., Васяева Т.А., Мартыненко Т.В.

Донецкий национальный технический университет, г. Донецк, ДНР

В статье рассмотрены этапы построения веб-ориентированной системы интернет-бронирования отелей. Выполнен обзор существующих систем и выделены их основные достоинства и недостатки. Разработана рекомендательная система интернет-бронирования отелей.

Введение

Все больше и больше людей, отправляясь в путешествие, отдают предпочтение бронированию жилья и билетов через интернет. Первые компьютерные системы резервирования CRS (Computer Reservation System), появились в середине 60-х гг. XX в. и позволили ускорить этот процесс. На сегодняшний день большая часть систем бронирования работает в режиме реального времени. Следующим этапом развития компьютерных систем бронирования стало объединение в глобальную систему бронирования различных туристских услуг GDS (Global Distribution Systems) [5]. Теперь сервисы, ранее доступные в рамках одной компании (бронь или покупка авиа-, ж/д или автобусных билетов, заказ гостиницы, и т.п.), реализованы в одной системе, но доступ к GDS имеют лишь агенты туристических компаний. ADS (Alternative Distribution System) разработанные в противовес GDS, предоставляют услуги по бронированию услуг частным клиентам. Заходя на любой из порталов ADS, клиент может самостоятельно выбрать нужную ему гостиницу и тип номера на определенные даты, забронировать размещение в режиме реального времени и получить подтверждение о бронировании на свой электронный ящик. Подобные системы бронирования чаще всего взимают плату, в виде процента от стоимости услуги.

Постановка проблемы

В последние годы количество систем, позволяющих забронировать жилье в режиме реального времени, непрерывно растет, и пользователю становится все труднее и труднее делать выбор путем простого просмотра контента. Вследствие чего создаются инструменты, позволяющие помочь в поиске, предлагая более референтную информацию для него. Такие программные средства получили название рекомендательные системы [2].

В работе рассматривается разработка веб-ориентированной системы, которая будет предлагать пользователю наиболее предпочтительные варианты временного жилья; обеспечивать возможность выбрать, забронировать и оплатить жилье на сайте системы.

Основные задачи и функции системы:

- учет и регистрация клиентов;
- учет и регистрация пользователей;
- учет и регистрация гостиниц и комнат;
- поиск подходящего жилья;
- формирование заказа;
- подтверждение оплаты;
- учет заказов.

Анализ аналогичных подсистем

На данный момент существует множество систем, предоставляющих услуги по бронированию жилищного фонда. При изучении аналогов разрабатываемой системы были выделены следующие 3 системы: Airbnb.ru, Ostrovok.ru, Hotels.com. Сравнительный анализ существующих систем с разрабатываемой системой приведен в табл. 1.

Таблица 1 – Анализ аналогичных веб-ориентированных систем

Функции /Система	Airbnb.ru	Ostrovok.ru	Hotels.com	Разрабатываемая веб-ориентированная система
Бронирование номера	+	+	+	+
Учет ранее использованных услуг клиентом	-	-	-	+
Учет клиентов	+	+	+	+
Учет гостиниц	+	+	+	+
Системы оплаты	Apple Pay, Google Pay, PayPal, карты	Apple Pay, Google Pay, PayPal, карты	Карты, PayPal, Яндекс-Моней	Карты, Web Money, PayPal, ЯндексМоней
Учет геолокации	-	-	-	+
Сдача в аренду жилищного фонда	+	-	-	+
Рекомендательная система	-	-	-	+

К тому же важной особенностью нашей системы является наличие рекомендаций в соответствии с предпочтениями пользователя.

Математическая постановка задачи к формированию рекомендаций

Пусть:

$U = \{u_1, u_2, \dots, u_n\}$ – множество пользователей,

$P = \{p_1, p_2, \dots, p_m\}$ – множество объектов временного жилья,

R – матрица рейтингов размера $n * m$, где на месте $r_{i,j}$ стоит оценка, если пользователь u_i оценил объект p_j , и пусто – в противном случае;

N – желаемое число рекомендаций, которые мы хотим получить от системы.

Требуется для пользователя u_i найти N -мерный вектор рекомендаций $(p_{i_1}, p_{i_2}, \dots, p_{i_N})$, где p_{i_k} – объекты временного жилья, которые не оценены этим пользователем, но есть в матрице рейтингов R ; $k \in N$.

То есть наша задача – предсказать отсутствующие оценки $r_{i,j}$ в матрице рейтингов и для каждого пользователя u_i на основе спрогнозированных оценок сформировать список из N -объектов, которые наиболее точно соответствуют предпочтениям этого пользователя.

Принципы построения рекомендательных систем

1. *Построение рекомендательных систем на основе фильтрации содержимого.* Для формирования рекомендации на основе фильтрации содержимого используется информация, которая хранится в системе: данные профилей пользователей, включающие индивидуальную информацию, например, возраст, социальный статус, род деятельности, место проживания, а также характеристики, отражающие заинтересованность пользователя в определенных объектах; данные профилей объектов заинтересованности, которые включают все интересующие параметры [3].

Фильтрация на основе содержимого предполагает, что круг интересов пользователя постоянен в течение времени. Задачей фильтрации на основе содержимого является вычисление множества объектов, наиболее близких интересам текущего пользователя. Базовым способом определения сходства считается извлечение ключевых слов из контекста, содержащихся у объекта интереса, или из метаданных, которыми проаннотирован объект.

Главными достоинствами данного подхода является возможность рекомендаций для новых пользователей; также в рекомендации могут попадать новые, еще не оцененные объекты. Можно выделить и более высокую скорость разработки. К недостаткам относят, как правило, низкую точность.

2. *Рекомендательные системы, основанные на знаниях.* Рекомендации, основанные на знаниях, зачастую не зависят от описания объектов с помощью метаданных или их оценки, а основываются на более глубоких правилах для выявления объектов интереса. В качестве знаний выступает информация об объектах интереса [4], например, множество правил. На основе выбранных предпочтений пользователя правила описывают, какие именно объекты должны быть рекомендованы.

Главным достоинством данного метода является возможность исключения рекомендаций объектов, уже не актуальных для текущего пользователя, недостатком же является высокая сложность построения систем и сбора данных для формирования знаний.

3. *Построение рекомендательных систем на основе коллаборативной фильтрации.* Коллаборативная фильтрация [1] – технология прогнозирования предпочтений пользователя, использующая известные предпочтения (оценки) других пользователей. Основное допущение состоит в следующем: те, кто одинаково оценивал какие-либо объекты в прошлом, склонны давать похожие оценки другим объектам и в будущем. Такой подход основан исключительно на поведении данного пользователя, с учетом поведения других пользователей со сходными характеристиками.

Основные подходы к коллаборативной фильтрации основаны на сходстве пользователей (user-based) и сходстве элементов (item-based). Целью обоих подходов является выделение похожих объектов и объединение их в группы на основе матриц оценок.

User-based [6] фильтрация находит k ближайших пользователей, чьи оценки схожи с текущим пользователем, и использует их оценки для прогнозирования предпочтений текущего пользователя. Главным достоинством данного подхода является высокая точность. Недостатком данного подхода является высокий порог входа. Это означает, что, не имея данных об интересах пользователя, рекомендации подобрать практически невозможно.

Отличием item-based [6] от user-based фильтрации является не использование поведения пользовательских оценок, а использование сходства между оценками моделей объектов. Как правило, два элемента, имеющие одинаковые оценки пользователей, похожи, а это значит, что пользователи должны иметь аналогичные предпочтения для подобранных объектов. Достоинством данного подхода является возможность вычисления степени близости в отложенном режиме, так как рейтинг объектов доступен до формирования рекомендаций.

4. *Гибридные подходы.* Рекомендательные системы, объединяющие вышеперечисленные подходы, называются гибридными. Они сочетают сильные стороны данных подходов для создания методов, которые могут работать эффективнее в узкопрофильных системах.

Гибридный подход объединяет возможности базовых подходов, благодаря чему позволяет добиться более высокой точности. Широко распространен гибридный подход при построении классификаторов. При построении гибридных рекомендательных систем используют следующие стратегии [7]:

- взвешенная стратегия. Спрогнозированная оценка для объекта рассчитывается как средне-взвешенное арифметическое оценок, спрогнозированных различными алгоритмами;
- стратегия переключения. Перед построением рекомендации вычисляется значение некоторого критерия, на основании которого принимается решение о выборе алгоритма для построения рекомендаций. Таким критерием может выступать сравнение количества действий пользователя с неким выбранным заранее порогом, начиная с которого к пользователю можно применять модель коллаборативной фильтрации;
- смешанная стратегия. Основана на том, что прогнозы отдельных рекомендаций отображаются в одном интегрированном результате;
- каскадная стратегия является итеративным методом построения рекомендательных систем. Первый алгоритм играет роль грубого фильтра, а все последующие алгоритмы корректируют оценки.

Основные подходы к построению рекомендательных систем сравнивались по следующим параметрам (табл. 2):

- адаптивность – подразумевает автоматический учет оценок пользователя при формировании последующих рекомендаций;
- быстрое развертывание – отсутствие необходимости детальной информации о свойствах объекта;
- интуитивность – возможность отображения релевантного объекта, даже без инициирования поиска пользователем.
- прозрачность – возможность обоснования результата работы рекомендательного алгоритма;
- холодный старт – подразумевает необходимость предоставления начальных оценок до начала работы алгоритма.

Таблица 2 – Сравнение характеристик основных рекомендательных подходов.

Подход	Коллаборативная фильтрация	Фильтрация на основе содержимого	Фильтрация на основе знаний
Адаптивность	Да	Да	Нет
Быстрое развертывание	Да	Да	Нет
Интуитивность	Да	Нет	Нет
Прозрачность	Нет	Нет	Да
Холодный старт	Да	Да	Нет

Разработка алгоритма формирования рекомендаций на основе коллаборативной фильтрации

Любой из подходов к коллаборативной фильтрации (user-based и item-based) предполагает определение близости объектов. Поэтому на начальном этапе необходимо рассчитать расстояние между пользователями или объектами. Для расчета расстояния использовалась метрика Евклидова расстояния. Данная метрика является наиболее интерпретируемой мерой близости объектов [8]. Для алгоритма user-based сходство между двумя пользователями x и y определяется следующим образом:

$$w(x, y) = \sqrt{\sum_i^n (x_{u,i} - y_{u,i})^2},$$

где $x_{u,i}$ – оценка объекта текущим пользователем, $y_{u,i}$ – оценка объекта другим пользователем, n – множество объектов.

После расчёта расстояния между пользователями производится выбор ближайших соседей относительно текущего пользователя. После выбора N лучших соседей формируется прогнозируемая оценка неоцененных объектов текущего пользователя по следующей формуле:

$$\bar{r}_{u,j} = \frac{\sum_{i \in N} (w_{x,y} * r_{y,j})}{\sum_{i \in N} (|w_{x,y}|)},$$

где $\bar{r}_{u,j}$ – предсказанный рейтинг текущего пользователя и для объекта j , $r_{y,j}$ – рейтинг пользователя y для объекта j , $w_{x,y}$ – сходство пользователей x и y .

Затем сформированный массив оценок сортируется от большего значения к меньшему и выводится в качестве рекомендаций пользователю.

При формировании рекомендаций алгоритмом item-based, для расчета расстояния использовалась формула:

$$w(s, p) = \sqrt{\sum_i^m (x_{u,i} - y_{u,i})^2},$$

где $x_{u,i}$ и $y_{u,i}$ – оценки объекта пользователями, m – множество пользователей, оценивших объект.

И тогда прогнозируемая оценка определяется следующим образом:

$$\bar{r}_{u,j} = \frac{\sum_{i \in N} (w_{s,p} * r_{y,j})}{\sum_{i \in N} (|w_{s,p}|)},$$

где $\bar{r}_{u,j}$ – вероятность понравившегося объекта для пользователя и относительно объекта j , $r_{y,j}$ – рейтинг пользователя u для объекта j , $w_{s,p}$ – сходство объектов x и y .

Также хорошо зарекомендовал себя алгоритм формирования рекомендаций ALS [9]. Алгоритм является итеративным, он разлагает матрицу предпочтений на произведение матриц пользователей и объектов. Алгоритм принимает набор обучающих данных и дополнительные параметры для создания модели:

- rank — количество рекомендаций;
- maxIter — количество итераций;
- alpha — эмпирический коэффициент, влияющий на расчет степени доверия;
- regParam — коэффициент регуляризации, позволяющий управлять переобучением модели.

Разработка системы

Use Case (рис. 1) описывает сценарий взаимодействия участников. В нашей системе один тип пользователей может регистрировать свои гостиницы или другую жилую площадь. Владелец жилья заполняет необходимые данные, такие как: название, адрес, тип и фотографии гостиницы, а также ее комнат, возможное для размещения количество человек, посуточная стоимость, предоставляемые дополнительные услуги. Другой тип пользователя – это клиенты (желающие снять жилье). Клиенты задают параметры поиска, выбирают, бронируют или сразу заказывают жилье и услуги. Зарегистрированные клиенты после завершения заказа имеют возможность оставить отзыв.

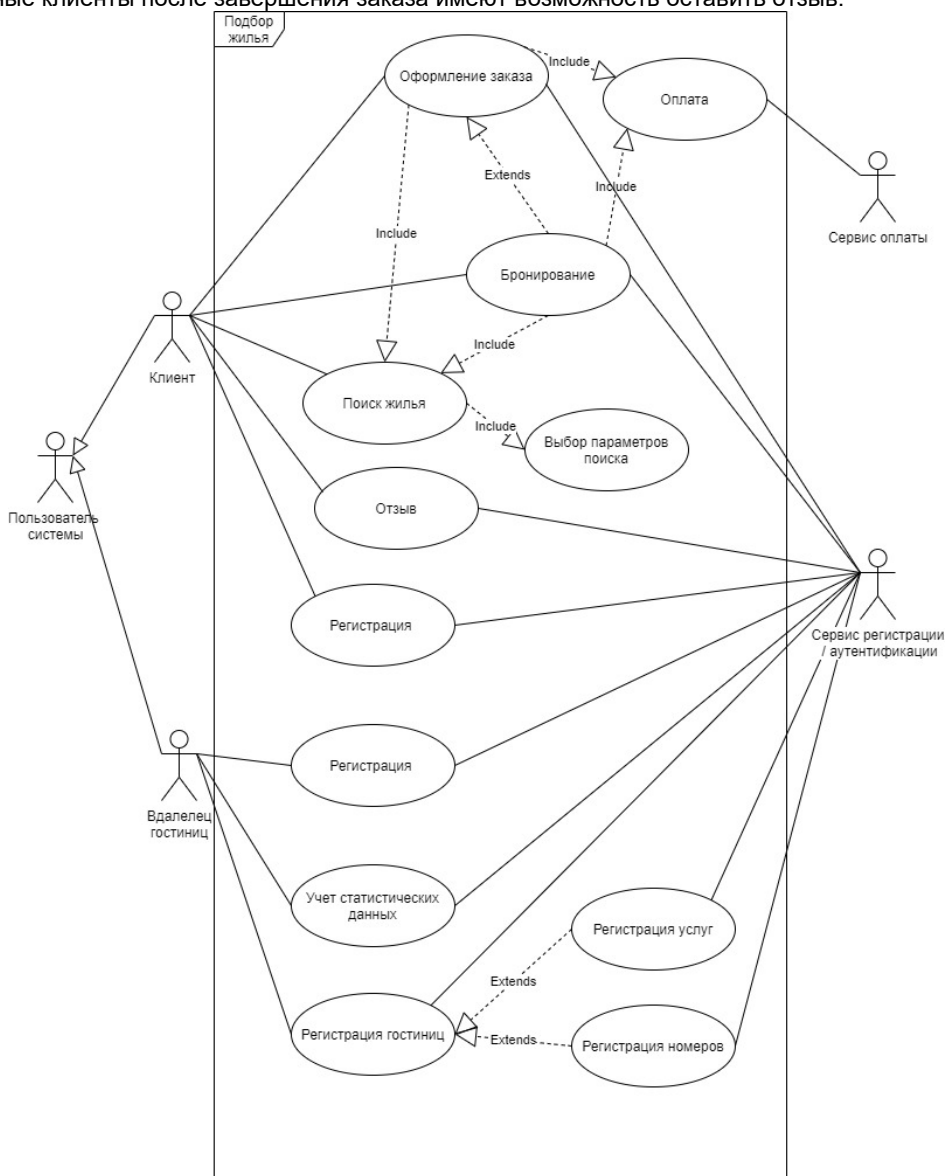


Рисунок 1. Диаграмма вариантов использования

Разработанная физическая модель данных (рис. 2) содержит подробную информацию о каждом атрибуте, а также описание всех таблиц. Перечень сущностей физической модели данных и их описание приведены в табл. 3.

Таблица 3 – Сущности и атрибуты физической модели данных.

Сущность	Описание	Атрибуты
Users	Контактная информация о пользователях	Код, ник-нейм, имя, фамилия, отчество, e-mail, пароль, статус
Feedback	Данные, оставленные пользователем о качестве предоставляемых услуг	Код, отзыв, рейтинг, дата, код заказа, код гостиницы
Order	Данные о заказах, совершенных покупателями	Код, дата заказа, дата начала, дата конца, код пользователя, код гостиницы
Payment	Данные о статусе оплаты заказа пользователем	Код, статус, код заказа
Favorite	Данные об избранных гостиницах пользователем	Код, код пользователя, код гостиницы
Like	Данные о понравившихся пользователю гостиницах	Код, код пользователя, код гостиницы
Hotels	Данные о гостиницах	Код, название, описание, звёзды, номер телефона, номер дома, индекс, статус, код типа гостиницы, код улицы
Hotel_image	Данные фотографий отеля	Код, код гостиницы, расположение фотографии
Index_hotel_service	Связывающая таблица отелей и услуг	Код, код гостиницы, код услуги
Service_hotel	Информация об услугах	Код, название услуги
Country	Информация о странах	Код, название
City	Информация о городах	Код, название, код страны
District	Информация о районах	Код, название, код города
Street	Информация об улицах	Код, название, код района
Type_hotel	Информация об виде гостиниц	Код, название типа
Room	Информация о номерах гостиницы	Код, количество человек, стоимость, площадь, описание, код типа номера, код отеля
Index_room_service	Связывающая таблица номеров и услуг	Код, код номера, код услуги
Service_room	Информация об услугах номера гостиницы	Код, название услуги
Type_room	Информация о типе номера гостиницы	Код, название типа
Room_image	Данные фотографий номера гостиницы	Код, код номера, расположение (фотографии)

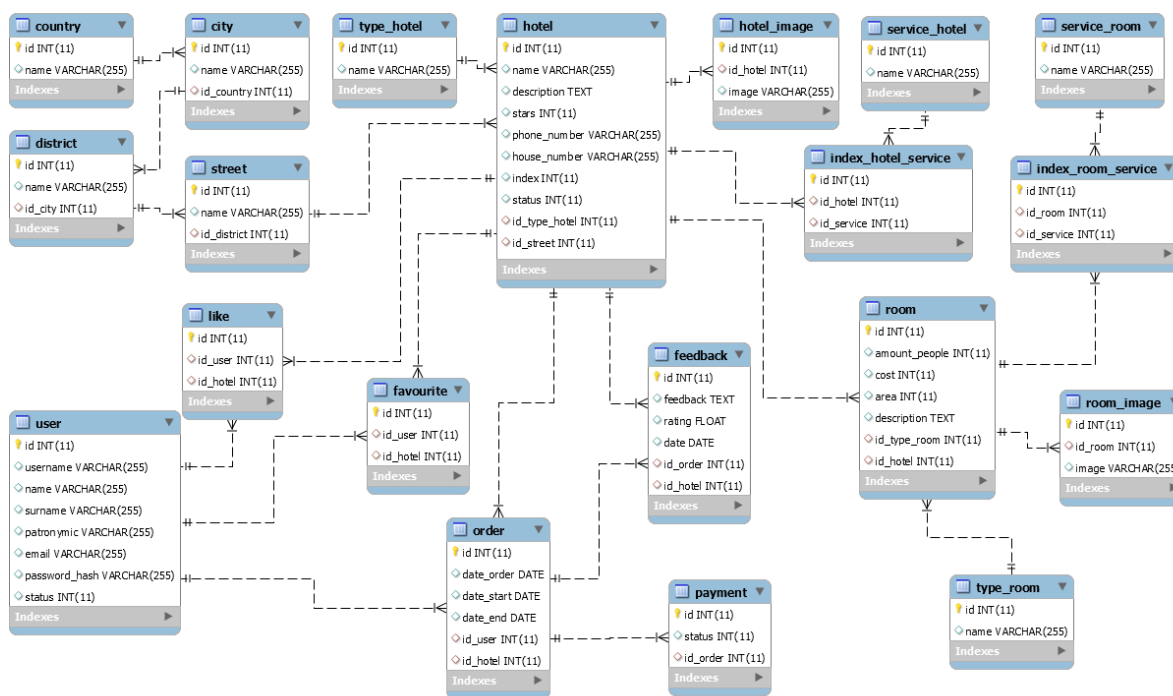


Рисунок 2 – Физическая модель данных

Диаграмма классов разработанной системы представлена на рис. 3

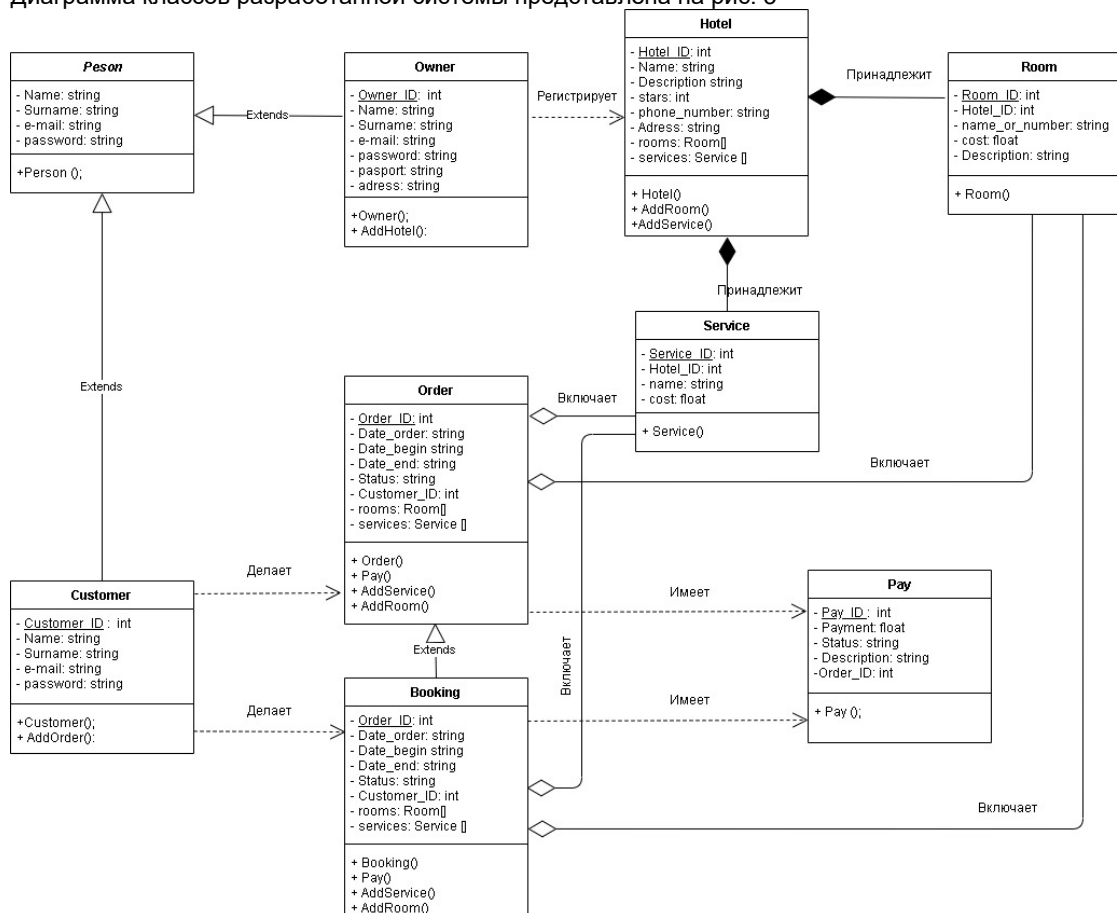


Рисунок 3 – Диаграмма классов

Для хранения данных используется СУБД MySQL. Данная СУБД обладает всеми необходимыми функциями для реализации поставленной задачи. В качестве Web-сервера был выбран Apache. Apache позволяет подключать различные модули, добавляющие в него новые возможности, а также он обладает надежностью, гибкими настройками и безопасностью.

Для реализации поставленной задачи выбран язык PHP совместно с фреймворком Yii2.

Схема развертывания системы представлена на рис. 4

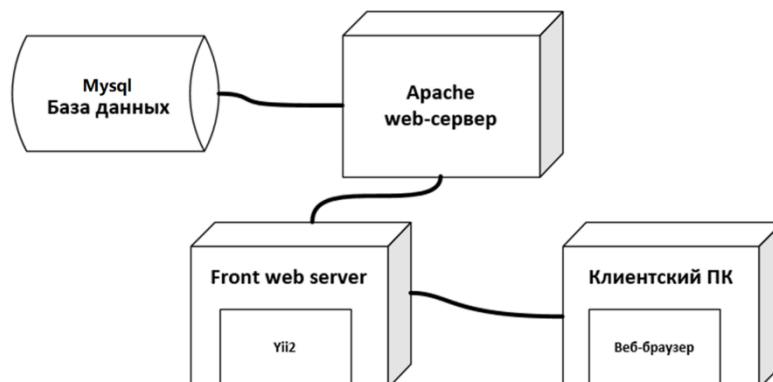


Рисунок 4 – Схема развертывания системы

Метрики точности и экспериментальные исследования

Все метрики качества можно условно разделить на три категории:

- Prediction Accuracy — оценивают точность предсказываемого рейтинга;
- Decision support — оценивают релевантность рекомендаций;
- Rank Accuracy метрики — оценивают качество ранжирования выдаваемых рекомендаций.

Нет единой рекомендуемой метрики на все случаи жизни, и каждый, кто занимается тестированием рекомендательной системы, подбирает её под свои цели.

Для оценки точности предсказываемого рейтинга будем использовать среднюю абсолютную ошибку (MAE) и среднеквадратическую ошибку (RMSE).

MAE является достаточно популярной и широко используемой метрикой [10]. Она показывает отклонения рекомендаций от реальных оценок пользователей.

$$MAE = \frac{\sum_i^n |p_{u,i} - a_{u,i}|}{N},$$

где $p_{u,i}$ – прогнозируемая оценка для пользователя u и объекту i ; $a_{u,i}$ – фактическая оценка; N – общее количество оценок по набору объектов.

RMSE, также как и MAE, измеряет среднюю величину ошибки рекомендаций. Ключевая причина использования данной метрики состоит в том, что Среднеквадратическая ошибка (RMSE) стала уже стандартным способом измерения ошибки модели при прогнозировании количественных данных, хотя популярность к ней пришла совсем недавно, а именно – после использования в конкурсе Netflix Prize [11]. Рассчитывается RMSE следующим образом:

$$RMSE = \sqrt{\frac{\sum_i^n (p_{u,i} - a_{u,i})^2}{N}},$$

где $p_{u,i}$ – прогнозируемая оценка для пользователя u и по объекту i ; $a_{u,i}$ – фактическая оценка; N – общее количество оценок по набору объектов.

Для оценки релевантности рекомендаций и качества ранжирования выдаваемых рекомендаций будем использовать метрику – Precision at K , которая является наиболее часто используемой метрикой качества ранжирования. Когда алгоритм ранжирования выдает оценки релевантности каждого объекта, отобрав первые из них с наибольшим рейтингом, можно посчитать долю релевантных объектов [12]. Формула расчета приведена ниже.

$$p@K = \frac{\sum_{k=1}^K (p^{-1}(k))}{K},$$

где $p^{-1}(k)$ – элемент p , оказавшийся на k -ой позиции

Заключение

В статье рассмотрены все этапы построения веб-ориентированной системы для подбора временного жилья. Приведены и описаны подходы к реализации рекомендательных систем. Выделены основные достоинства и недостатки существующих решений. Разработана веб-ориентированная рекомендательная система интернет-бронирования отелей на основе коллаборативной фильтрации.

Литература

1. Федоренко В.И., Киреев В.С./Анализ подходов к построению гибридных рекомендательных систем в задаче рекомендации фильмов / Международный научно-технический журнал «ТЕОРИЯ. ПРАКТИКА. ИННОВАЦИИ» [Электронный ресурс] // <https://docplayer.ru/60855201-Mezhdunarodnyu-nauchno-tehnicheskij-zhurnal-teoriya-praktika-innovacii-iyun-2017-avtomatika-vychislitel'naya-tehnika.html>
2. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems – An Introduction. Cambridge University Press, 2010. 360. P
3. Pazzani M., Billsus D. Learning and revising user profiles: The identification of interesting web sites // Machine Learning – Special issue on multistrategy learning, 1997. Vol. 27, Issue 3. – С. 313–331.
4. Rajendra Arvind Akerkar, Priti Srinivas Sajja, Knowledge-Based Systems. Jones and Bartlett Publishers – С. 18-36.
5. Global Distribution System. [Электронный ресурс] // <http://www.businessdictionary.com/definition/Global-Distribution-System-GDS.html/>
6. Carlos Pinela. Recommender Systems – User-Based and Item-Based Collaborative Filtering [Электронный ресурс] // <https://medium.com/@cfpinela/recommender-systems-user-based-and-item-based-collaborative-filtering-5d5f375a127f>
7. Burke R. Hybrid web recommender systems // The adaptive web. – 2007. – С. 377-408.
8. EuclideanMetric [Электронный ресурс] // <https://mathworld.wolfram.com/EuclideanMetric.html>
9. Collaborative Filtering [Электронный ресурс] // <https://spark.apache.org/docs/2.2.0/ml-collaborative-filtering.html>
10. Absolute Error & Mean Absolute Error (MAE) [Электронный ресурс] // <https://www.statisticshowto.com/absolute-error/>
11. Netflix Prize [Электронный ресурс] // https://en.wikipedia.org/wiki/Netflix_Prize
12. Rajdeep Dua, Manpreet Singh Ghotra, Nick Pentreath. Machine Learning with Spark – Second Edition. Packt Publishing, 2017. 532. Pс.

References in Cyrillics

1. Fedorenko V.I., Kireev V.S. Analiz podxodov k postroeniyu gibridny`x rekomendatel`ny`x sistem v zadache rekomendacii fil`mov // Mezhdunarodny`j nauchno-tehnicheskij zhurnal «TEORIYA. PRAKTIKA. INNOVACII» [E`lektronny`j resurs] // <https://docplayer.ru/60855201-Mezhdunarodnyu-nauchno-tehnicheskij-zhurnal-teoriya-praktika-innovacii-iyun-2017-avtomatika-vychislitel'naya-tehnika.html>

Ключевые слова

коллаборативная фильтрация, User-Based, Item-Based, рекомендательная система, веб-ориентированная система, физическая модель данных, схема развертывания системы.

Ясницкий Максим Валерьевич, студент гр. ИСм-19, направление подготовки 09.04.02 Информационные системы и технологии, факультет компьютерных наук и технологий, Донецкий национальный технический университет, max.yasnitsky@gmail.com

Васяева Татьяна Александровна, к.т.н., доц., доцент кафедры автоматизированных систем управления, факультет компьютерных наук и технологий, Донецкий национальный технический университет, vasyaeva@gmail.com.

Мартыненко Татьяна Владимировна, к.т.н., доц., доцент кафедры автоматизированных систем управления, факультет компьютерных наук и технологий, Донецкий национальный технический университет, tatyana.v.martynenko@gmail.com

Yasnitsky M.V., Vasyaeva T.A., Martynenko T.V, Web-based Hotel Online Representation recomended Sistem

Keywords

Collaborative Filtering, User-Based, Item-Based, Recommender Systems, Web-Based System, Physical Data Model, System Deployment Diagram.

DOI: 10.34706/DE-2021-02-06

JEL classification: С 45 – Нейронные сети и относящиеся к ним темы; Z 33 – Маркетинг и финансы

Abstract

The article describes the stages of building a web-based online hotel reservation system. A review of existing systems is carried out and their main advantages and disadvantages are highlighted. A recommendation system for hotel online reservation has been developed.