

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и информационных технологий
Кафедра информатики

**РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ “ОЛИМПИАДА”
С МОДУЛЕМ АНАЛИЗА ДАННЫХ**
(выпускная квалификационная работа)

Выполнил:
студент 447М группы
Коломеец Р.Д.

Научный руководитель:
к. т. н., доцент
Рязанов М.А.

Работа защищена:

«__» _____ 2016 г.

Оценка: _____

Председатель ГЭК:

Ф.И.О

Допустить к защите:
врио зав. каф. информатики, к.ф.-м.н.,
Жариков А. В.

«__» _____ 2016 г.

Реферат

Тема выпускной квалификационной работы: «Разработка информационной системы “Олимпиада” с модулем анализа данных».

Цель работы – разработка информационной системы, позволяющей публиковать информацию о проводимых олимпиадах, результатах олимпиад, а также проводить последующий анализ данных для рекомендаций достижений организаторам и для рекомендаций олимпиад пользователям.

Объект исследования – организация олимпиад и методы анализа данных.

Предмет исследования – информационная система для организации олимпиад с модулем рекомендательной и рейтинговой систем.

В рамках работы был проведен анализ существующих информационных систем для организации олимпиад. Выявлены их достоинства и недостатки, ключевые функции и предложен вариант реализации собственной информационной системы, удовлетворяющей сформулированным требованиям.

Результатом работы является реализованная система “Олимпиада” с модулем анализа данных.

Система “Олимпиада” может быть использована как дополнительный модуль на любом образовательном портале, работающем на веб-приложении Moodle.

Объем работы 54 страницы, количество рисунков – 16, использованных источников литературы – 20, листингов – 2, таблиц – 1, приложений – 1.

Ключевые слова: информационная система, анализ данных, рейтинговая система, рекомендательная система, Moodle.

Оглавление

Введение.....	4
1. Анализ систем для организации олимпиад	6
1.1. Обзор существующих систем для организации и проведения олимпиад.....	6
1.2. Концепция системы для работы с олимпиадами.....	11
1.3. Выбор метода анализа данных	12
2. Разработка системы “Олимпиада”.....	18
2.1. Требования к системе и принцип работы.....	18
2.2. Разработка ядра системы “Олимпиада”	20
2.3. Разработка модуля анализа данных	33
3. Система “Олимпиада”. Тестирование полученных результатов ..	39
3.1. Описание разработанной системы “Олимпиада”	39
3.2. Тестирование рекомендательной и рейтинговой систем модуля анализа данных.....	47
Заключение	50
Библиографический список	51
Приложение	53

Введение

Актуальность. В портфолио любого школьника или студента значимую роль занимают дипломы, отражающие победы и участия в олимпиадах.

Ранее, основным источником информации о проводимых олимпиадах, для любого учащегося был его преподаватель. Но у современного школьника и студента появился незаменимый источник информации — интернет. К сожалению, даже такой источник данных не облегчает учащемуся сбор информации о проводимых в ближайшее время олимпиадах. Необходимо потратить много времени в поисках олимпиад нужного профиля. В связи с этим появляется потребность в создании ресурса, в котором будет собрана вся актуальная информация об олимпиадах, новостях и результатах. Дополнительно облегчить поиск олимпиад может рекомендательная система, “подсказывающая” пользователю олимпиады, которые могут быть ему интересны. Организаторам, в свою очередь, важно выявить способных участников. В этом им поможет рейтинговая система, составляющая общий рейтинг участников по итогам всех олимпиад.

Целью выпускной квалификационной работы является разработка информационной системы “Олимпиада”, состоящей из “ядра” в которое входит основной функционал системы, и дополнительных модулей анализа данных, состоящих из рекомендательной и рейтинговой систем.

Для достижения данной цели решаются следующие **задачи**:

1. Анализ существующих систем для работы с олимпиадами.
2. Выработка функциональных требований к информационной системе “Олимпиада”.
3. Разработка “ядра” системы и модулей анализа данных.
4. Тестирование информационной системы на соответствие функциональным требованиям.

Объект исследования. Организация олимпиад и методы анализа данных.

Предмет исследования. Информационная система для организации олимпиад с модулем рекомендательной и рейтинговой систем.

Практическая значимость. Система “Олимпиада” может быть использована как дополнительный модуль на любом образовательном портале, работающем на веб-приложении Moodle.

Апробация работы. Разработанная система проходит этап тестирования в рамках образовательного портала public.edu.asu.ru.

Структура и объем работы. Работа состоит из введения, трех глав, заключения, списка литературы и одного приложения. В работе содержится 16 рисунков, 2 листинга, 1 таблица. Общий объем работы составляет 54 страницы. Список литературы содержит 20 источников.

В первой главе производится анализ существующих систем для работы с олимпиадами, формируется концепция будущей системы. Также приводится обзор существующих методов построения рекомендательных и рейтинговых систем. Выбираются необходимые методы анализа.

Во второй главе рассказывается про основные моменты разработки: выработка требований к будущей системе, создание базы данных, реализация ролей пользователей, написание интерфейса системы, разработка рекомендательной и рейтинговой систем.

В третьей главе приводится описание реализованной системы “Олимпиада”. Производится тестирование модулей рейтинговой и рекомендательной систем.

1. Анализ систем для организации олимпиад

1.1. Обзор существующих систем для организации и проведения олимпиад

Предметная олимпиада — состязание учащихся учреждений среднего общего, высшего или профессионального образования, требующее от участников демонстрации знаний и навыков в области одной или нескольких изучаемых дисциплин [1].

Предметные олимпиады могут проводиться в различных форматах: в очной форме, заочной, онлайн олимпиада. Зачастую организаторы олимпиад создают интернет-ресурс или используют существующие, для информирования участников о различных событиях, связанных с олимпиадой, публикаций необходимых материалов для подготовки или проведения олимпиады. Организаторы небольших (региональных) олимпиад не всегда могут выделить средства из бюджета на создание такого интернет-ресурса. В большинстве таких случаев организаторы пользуются услугами сторонних ресурсов.

В свою очередь участникам сложно отслеживать информацию обо всех интересующих олимпиадах, а так же сложно найти информацию о предстоящих олимпиадах. Сбор информации с различных сайтов может занять много времени и не всегда удастся достичь нужного результата.

Интернет-ресурсы объединяющие информацию о различных олимпиадах очень удобны как для организаторов, так и для участников.

Существующие интернет-ресурсы с информацией об олимпиадах можно классифицировать следующим образом:

1. По типу ресурса:

- информационные ресурсы;
- системы проведения интернет-олимпиад.

2. По профилю:

- узконаправленные (профильные);
- многопрофильные.

3. По уровню участников:

- для школьников;
- для студентов.

Рассмотрим примеры существующих интернет-ресурсов.

Интернет-ресурс «Дом учителя» (см. рис. 1). Одним из направлений данного ресурса является информирование пользователей о проведении олимпиад города Барнаула [2]. Помимо информационной поддержки, ресурс не имеет иных возможностей для участника или организатора олимпиады.

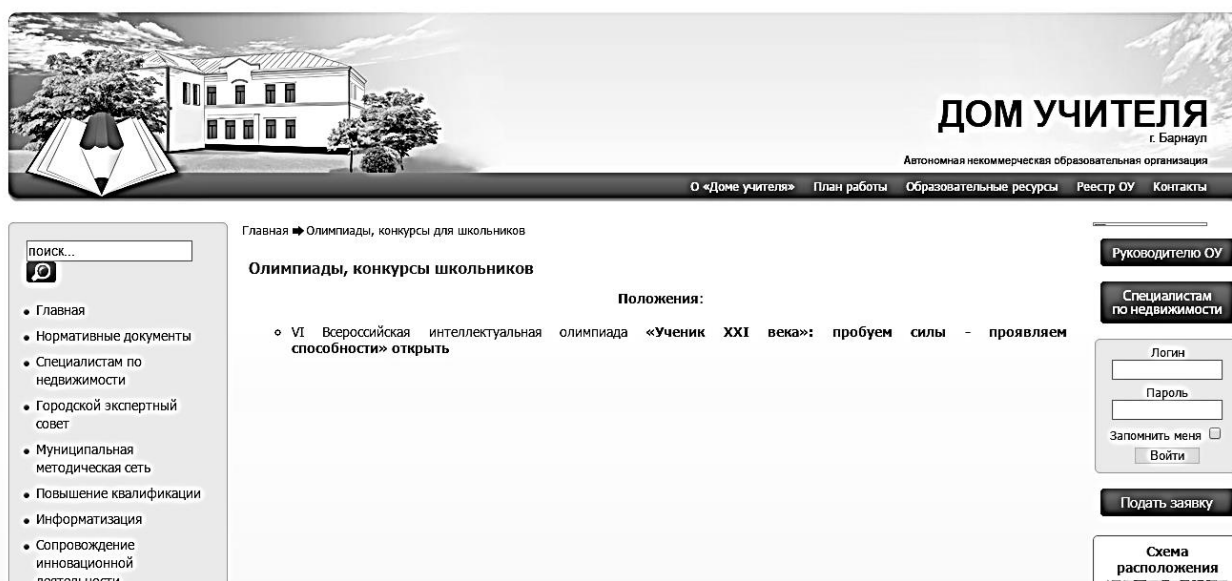


Рис. 1. Пример страницы интернет ресурса «Дом учителя»

Межрегиональная олимпиада школьников «Будущие исследователи — будущее науки» (см. рис. 2). Ресурс предназначен для организации межрегиональной олимпиады для школьников «Будущие исследователи — будущее науки» [3]. В рамках данного ресурса, пользователи могут следить за объявлениями, связанными с проведением олимпиады, ознакомиться с

нормативными документами, просмотреть календарь проведения олимпиад, узнать сроки, задания и результаты, получить материалы для подготовки.

Рис.2. Пример страницы ресурса «Будущие исследователи — будущее науки»

№	Название	Профиль	Предметы	Уровень
1	Всероссийская олимпиада школьников «Нанотехнологии - прорыв в будущее»	нанотехнологии	биология, математика, физика, химия	1
2	Всероссийский конкурс научных работ школьников «Юниор»	естественные науки инженерные науки	биология, экология физика, математика, информатика	3 3
3	Всероссийский турнир юных физиков	физика	физика	3
4	Всесибирская открытая олимпиада школьников	биология информатика математика	биология информатика математика	2 3 2

Рис.3. Пример страницы сайта Российского совета олимпиад школьников

Сайт Российского совета олимпиад школьников (см. рис. 3). Этот ресурс предназначен для организации всероссийских олимпиад [4]. Сайт предоставляет подробную информацию об олимпиадах утвержденных Российским советом олимпиад школьников. Здесь можно найти всю информацию об интересующей олимпиаде, задать вопросы, оценить прошедшую олимпиаду.

Для участника помимо получения информации об организации олимпиады больше нет никаких возможностей.

Организатору, для включения его олимпиады в утвержденный советом список олимпиад, необходимо пройти регистрацию на портале, отправить необходимый пакет документов в электронном виде и оригиналы по почте. В случае правильного оформления документов, совет рассмотрит представленную олимпиаду и вынесет решение о включении ее в список утвержденных олимпиад.



Рис. 4. Пример страницы информационного сайта Olimpiada.ru

Информационный сайт Olimpiada.ru (см. рис. 4) об олимпиадах и других мероприятиях для школьников. Здесь пользователь может найти объявления о предстоящих мероприятиях, материалы прошедших олимпиад, выездных школ и конференций (условия и решения задач, статистика, результаты) [5].



Рис. 5. Пример страницы центра «Эрудит»

Центр «Эрудит» (см. рис. 5) предлагает тематические, творческие и предметные конкурсы и олимпиады для дошкольников, учеников младшей, средней и старшей школы, студентов, воспитателей и учителей по различным направлениям и разной степени сложности. Задания разрабатываются ведущими педагогами школ и преподавателями вузов [6].

Интернет-ресурс «Эрудит» проводит заочные олимпиады. Участие в олимпиадах является платным.

Пользователи ресурса могут просмотреть рейтинги участников олимпиад. По данным, указанным на портале «Эрудит», рейтинг рассчитывается следующим образом: за каждый предметный или творческий конкурс участник получает определенное число баллов. Если участник занял 1-е место, то начисляется 10 баллов, если 2-е место - 9 баллов и т.д. Если участник не занял призового места, то начисляется 1 балл.

Такой подход к построению рейтинга не позволяет просмотреть лучших участников в рамках одного профиля, например, математики, физики или

химии. Также такой подход может не совсем верно оценивать любого из участников. Например, если первый участник занял 1-е место на одной олимпиаде из 10, и более не участвовал в конкурсах, а другой участник не выиграл ни в одной из олимпиад, но был участником 10-ти олимпиад, в таком случае оба этих участника займут одинаковое положение в рейтинге, что не совсем верно с точки зрения оценки качества знаний участников.

1.2. Концепция системы для работы с олимпиадами

Проанализировав некоторые из существующих систем, автор выявил в них ряд недостатков:

- ограниченные возможности для участников олимпиад;
- большинство ресурсов предназначено только для школьников;
- отсутствие возможности просмотра рейтинга участников по результатам нескольких олимпиад, в рамках одного профиля;
- отсутствие рекомендательных систем олимпиад для участников.

С учетом выше изложенного была составлена концепция разрабатываемой системы.

Ресурс должен позволять размещать олимпиады, как для школьников, так и для студентов. Это избавит абитуриента от поиска нового ресурса с олимпиадами для студентов и переноса своих достижений.

Организатором олимпиады могут быть как преподаватели школ, так и ВУЗов, пройдя регистрацию в системе в качестве организатора. Организаторы имеют возможность публикации новостей и важной информации, связанной с олимпиадой.

Для зарегистрированного в системе участника должна быть предусмотрена возможность подачи заявки на участие в выбранной олимпиаде и отслеживание статуса заявки. Результаты олимпиад должны храниться в системе, что позволит составить рейтинг участников.

Организатор должен иметь возможность зарегистрировать участника самостоятельно. В этом случае организатор передает участнику сгенерированный системой идентификатор для просмотра результатов олимпиады.

Результаты олимпиад должны быть доступны, как для зарегистрированных пользователей, так и без регистрации.

После авторизации, пользователю смогут рекомендоваться олимпиады, основанные на его интересах, указанных в анкете пользователя, и на основе его истории участия в олимпиадах.

Организаторы олимпиад должны иметь возможность сформировать рейтинги участников по выбранному профилю, например, по математике или биологии. При построении рейтинга, результатам олимпиад по выбранному профилю, должно отдаваться большее предпочтение, нежели результату по схожему профилю олимпиады.

1.3. Выбор метода анализа данных

Когда размер контента любого ресурса достигает объемов, которые человеку сложно обработать самостоятельно, не затратив на это много времени, наступает необходимость фильтрации контента самой системой.

Любая рекомендательная система основывается на анализе данных.

Анализ данных — область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов, извлечения знаний из экспериментальных (в широком смысле) данных; процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решений [7].

Рекомендательные системы — программы, которые пытаются предсказать, какие объекты (фильмы, музыка, книги, новости, веб-сайты) будут интересны пользователю, имея определенную информацию о его профиле [8].

Существуют три основные стратегии создания рекомендательных систем: фильтрация содержимого, коллаборативная фильтрация и контентная фильтрация.

При использовании метода фильтрации содержимого зачастую создаются профили пользователей и объектов с необходимой информацией:

- профили пользователей могут включать демографическую, географическую информацию или ответы на определённый набор вопросов;
- профили объектов могут включать названия олимпиад, имена организаторов, место проведения, научную область, предмет и т. п. – в зависимости от типа объекта [8].

Такой подход позволяет найти контент, наиболее точно подходящий пользователю, но такой подход не эффективен, например, для рекомендации новостей.

Коллаборативная фильтрация, совместная фильтрация — это один из методов построения рекомендаций в рекомендательных системах, использующий известные предпочтения (оценки) группы пользователей для прогнозирования неизвестных предпочтений другого пользователя. Его основное допущение состоит в следующем: те, кто одинаково оценивали какие-либо предметы в прошлом, склонны давать похожие оценки другим предметам и в будущем [9]. Такой подход более гибок, чем предыдущий и позволяет рекомендовать одновременно контент разной тематики. Основная проблема этого типа рекомендательных систем — «холодный старт»: отсутствие информации о только появившихся в системе пользователях или объектах [8].

Контентная фильтрация формирует рекомендацию на основе поведения пользователя. Например, этот подход может использовать ретроспективную информацию о просмотрах (в каких олимпиадах участвовал пользователь и характеристики этих олимпиад). Если какой-либо пользователь обычно

участвует в олимпиадах по программированию, то контентная фильтрация может использовать эту ретроспективную информацию для выявления подобного контента и предложения такого контента в качестве рекомендованного для этого пользователя [10].

Рассмотрев различные методы анализа данных, в качестве основного метода рекомендательной системы был выбран коллаборативный метод фильтрации.

Основным недостатком выбранного метода является ситуация «холодного старта», в которой окажется каждый пользователь в начале работы с системой.

Для решения этой проблемы был использован гибридный подход, который совместил в себе два вида фильтрации: коллаборативную и фильтрацию содержимого по анкетным данным. Благодаря такому подходу, система сможет рекомендовать пользователю олимпиады, как непосредственно его научных профилей, отмеченных в качестве его интересов, так и похожих.

Менее серьезной проблемой является нелогичное поведение пользователей. Например, пользователь 1 участвует в олимпиадах по математике, информатике и физике; пользователь 2 участвует в олимпиадах по математике, компьютерной графике и в олимпиаде по истории народов мира; пользователь 3 участвует в олимпиаде по математике, биологии и психологии. В таком случае пользователю 1 можно равнозначно рекомендовать олимпиады, в которых участвовали пользователи 2 и 3, так как у всех этих пользователей есть общий интерес – математика. Но такая рекомендация окажется не совсем верной.

Поведение многих пользователей можно описать какой-либо моделью, однако некоторые пользователи демонстрируют абсолютно нетипичное поведение. Наличие таких пользователей может приводить к искажению результатов рекомендательной системы и к снижению ее эффективности.

Целью алгоритма рекомендательной системы является получение

коэффициентов сходства между объектами анализа данных. Коэффициент сходства (мера сходства, индекс сходства) — безразмерный показатель, применяемый для количественного определения степени сходства исследуемых объектов [11]. Коэффициент сходства поможет избежать искажения результатов рекомендательной системы в связи с нетипичным поведением пользователей.

Другим направлением анализа данных является построение рейтинговых систем. В современном мире рейтинги стали достаточно популярными, благодаря возможности получения места (позиции) конкретного объекта, среди ряда схожих объектов. В результате построения рейтинга можно ранжировать участников системы олимпиад и рекомендовать их организаторам, например, для предложения продолжить обучение в их ВУЗе. В теории экономического анализа, для построения рейтингов применяются следующие экономико-математические методы, применимые и для построения рейтинга участников олимпиад:

1. Метод сумм — метод заключается в суммировании фактических значений показателей или их соотношения с плановым показателем (эталонным значением). В результате по итоговому значению показателей определяется место в ранжированном списке. Главным недостатком такого метода является некорректное сложение результатов различных по методу оценки олимпиад (масштабу оценочных баллов).
2. Метод расстояний — заключается в учете близости объектов к объекту-эталону. За эталон может быть принят объект с максимальными показателями. Расчет оценки расстояния проводится по формуле (1).

$$R_j = \sqrt{\sum_{i=1}^n k_i (1 - x_{ij}/y_i)^2}, \quad (1)$$

где i — номер показателя, x_i — значение i -го показателя для j -го объекта, y_i — эталонное значение i -го показателя, k — коэффициент значимости i -го показателя, R_j — комплексная оценка для j -го объекта. Полученные

результаты упорядочиваются по возрастанию, в результате чего лучшие объекты окажутся на первых позициях ранжированного списка, т.е. эти результаты ближе всех к эталонным. Данный метод учитывает значимость объекта, что является хорошим преимуществом перед другими методами. Недостатком метода является не наглядность результатов [12].

3. Метод сумм мест – метод заключается в предварительном ранжировании рассматриваемых данных по отдельным показателям и присвоению каждому объекту по каждому показателю нового параметра, определяющий его место среди других. Наилучшему объекту присваивается 1 место. Далее места по каждому показателю для объектов суммируются и ранжируются по возрастанию. Объект, имеющий наименьшую сумму мест, является наилучшим. Данный метод является самым распространенным. Недостаток метода заключается в ограничении количества данных по каждому показателю и отсутствии оценки значимости каждого из показателей. В каждом показателе, оцениваемое количество объектов должно быть одинаковым.
4. Метод суммы баллов – при формировании балльных оценок показателей задаются дискретные шкалы, которые определяют максимальное и минимальное количество баллов. Лучшим считается объект, получивший максимальное в сумме количество баллов. Все шкалы оценки объектов должны согласовываться между собой.
5. Метод многомерного сравнительного анализа – основывается на методе расстояний. В его основе лежит определение близости значения к эталону. Для оценки показателей можно учитывать как абсолютные значения, так и значение отклонения от эталона. Для получения оценки объекта необходимо получить отношение значения показателя к

эталонному значению (или максимальному баллу) и учесть коэффициент значимости показателя.

$$R_j = \sqrt{\sum k a_{ij}} = \sqrt{\sum k x_{ij} / x_{\text{этал}}}, \quad (2)$$

Лучшим является объект, получивший максимальную оценку [13].

В результате тестирования методов расчета рейтинговых оценок на экспериментальных данных был выбран метод расстояний. Метод расстояний показал более точные результаты относительно многомерного анализа и метода сумм баллов. Метод расстояний позволяет учитывать даже небольшие различия между участниками, что позволит корректно сформировать рейтинговый список.

Выводы

В настоящей главе, автором был проведен анализ существующих систем для работы с олимпиадами и методов анализа данных.

Были выявлены плюсы и минусы существующих систем, сформулирована концепция будущей системы.

В рамках рекомендательной системы был выбран гибридный метод рекомендаций, основанный на методе фильтрации содержимого по анкетным данным пользователя и методе коллаборативной фильтрации.

В рамках рейтинговой системы актуальным методом реализации оказался метод расстояний.

2. Разработка системы “Олимпиада”

2.1. Требования к системе и принцип работы

Перед началом разработки любого программного продукта необходимо сформулировать требования. Исходя из концепции системы, описанной в параграфе 1.2., автором были сформулированы следующие требования:

1. Все пользователи системы должны быть разграничены по ролям: администратор, организатор и участник.
2. У пользователя с ролью администратор должны быть возможности:
 - создавать, просматривать, редактировать и удалять любую новость в системе;
 - создавать, просматривать, редактировать и удалять любую олимпиаду в системе;
 - добавлять участников олимпиад;
 - принимать и отклонять заявки на участие в олимпиадах;
 - просматривать списки заявок на олимпиаду;
 - выгружать списки участников олимпиад в формате XLS или CSV;
 - загружать и просматривать результаты олимпиад;
 - назначать любого пользователя системы администратором или организатором.
3. У пользователя с ролью организатор должны быть возможности:
 - создавать, просматривать любые новости в системе, редактировать и удалять свои новости;
 - создавать, просматривать любые олимпиады в системе, редактировать и удалять свои олимпиады;
 - добавлять участников на олимпиаду;
 - принимать и отклонять заявки на участие в олимпиадах;
 - просматривать списки заявок на олимпиаду;

- выгружать списки участников олимпиад в формате XLS или CSV;
- загружать и просматривать результаты олимпиад;
- назначать любого пользователя системы организатором.

4. У пользователя с ролью участник должны быть возможности:

- просматривать новости в системе;
- просматривать олимпиады в системе;
- подавать заявку на олимпиаду по форме;
- просматривать статус заявки;
- просматривать свои результаты.

Стоит отметить, что система должна допускать одновременное назначение пользователю всех трех ролей: администратор, организатор и участник.

Кроме этого у пользователей с ролью организатор и участник должны быть личные страницы, на которых присутствует дополнительная информация из их профиля, а также отображаются модули анализа данных.

Модули анализа данных могут подключаться отдельно для каждой роли пользователей и отображаться на личной странице пользователя.

Для пользователя с ролью организатор модуль анализа данных отображает рейтинговые листы участников. Рейтинговый лист представляет собой список участников, набравших наибольший рейтинг по итогам прошедших олимпиад необходимого профиля. Списки в дальнейшем могут быть использованы организатором, например, для приглашения участников на обучение.

Для пользователя с ролью участник модуль анализа данных отображает рекомендованные ему олимпиады. Олимпиады рекомендуются двумя различными способами. Так как в начале работы системы еще нет достаточной базы пользователей и олимпиад, невозможно будет составить рекомендации, основанные на поведении пользователей. Поэтому в начале работы системы,

рекомендации строятся на основе методов фильтрации содержимого по анкетным данным пользователя. При появлении достаточной истории поведения пользователей, рекомендации можно выстраивать на основе коллаборативной фильтрации. Система обучается на имеющейся у нее информации, основываясь на олимпиадах и участниках олимпиад. После обучения системы, список олимпиад, рекомендованных пользователю, будет зависеть не только от интересов, указанных в профиле пользователя, но и от “похожести” олимпиад.

2.2. Разработка ядра системы “Олимпиада”

Архитектура системы

За основу системы “Олимпиада” была взята система управления курсами Moodle [14], также известная как система управления обучением или виртуальная обучающая среда. Она представляет собой свободное (распространяющееся по лицензии GNU GPL [15]) веб-приложение, представляющее возможность создавать сайты для онлайн обучения.

Ядро системы “Олимпиада” представляет собой блок для Moodle, веб-страницы для пользователя и файлы библиотеки. Все эти файлы могут быть дополнительно установлены в любой момент. Файлы должны быть скопированы в директорию Moodle, в соответствии с архитектурой актуальной версии системы.

Для пользователя системы Moodle все выглядит как подключаемый блок, который можно добавить или скрыть на любой странице. Наполнение блока различается в зависимости от прав пользователя.

Блок — это стандартный контейнер для отображения виджетов на страницах Moodle. Страница Moodle обычно содержит блоки справа и слева. В режиме редактирования блоки можно добавлять, скрывать, удалять перемещать по странице [16].

Блок отображается как для зарегистрированного пользователя, так и для незарегистрированного.

Создание базы данных

Для хранения информации, необходимой для корректной работы системы необходимо было спроектировать и создать базу данных.

Автором был составлен перечень необходимых требований:

- просмотр любой, опубликованной ранее новости;
- просмотр любой, созданной ранее олимпиады;
- просмотр дополнительной информации о пользователе;
- просмотр информации о поданных заявках;
- просмотр результатов олимпиад;
- просмотр участников олимпиад;
- хранение данных, необходимых для рекомендательной и рейтинговых систем.

Исходя из этих требований, появилась необходимость в создании следующих сущностей:

- сущность новость, в которой хранится информация о новости: ее заголовок, текст, автор;
- сущность олимпиада, в которой хранится информация об олимпиаде: ее название, профиль, описание, место проведения, дата проведения, даты подачи заявок, уровень олимпиады (студент или школьник), организатор олимпиады;
- сущность участник, в которой хранится информация об участнике: имя, фамилия, отчество, адрес электронной почты, уровень участника (курс/класс), место учебы, ФИО руководителя, статус участника, его идентификационный номер;

- сущность результаты, в которой хранится информация о результатах олимпиады: место в олимпиаде, набранные баллы, максимально допустимый балл, призовое место;
- сущность пользователь, в которой хранится информация о пользователе системы: место учебы, уровень (курс/класс), руководитель, профиль пользователя;
- сущность роль, в которой хранится информация о роли пользователя системы: имя роли;
- сущность роли пользователей, в которой хранится информация, связывающая идентификаторы роли и пользователя;
- сущность профиль пользователя, в которой хранится информация о профиле: имя профиля и идентификатор научной области;
- сущность интересы пользователя, в которой хранится информация, связывающая идентификаторы пользователей и профилей;
- сущность научная область, в которой хранится информация о научной области профиля: имя научной области.

Листинг 1. Создание сущности Пользователь в СУБД MySQL

```
create table if not exists mdl_olmp_user(
id bigint not null auto_increment unique,
id_sys_user bigint not null,
grade text, student boolean,
school boolean, place_of_study text,
mentor text, profile_id bigint,
foreign key fk15(profile_id) references mdl_olmp_profile(id),
foreign key fk1(id_sys_user) references mdl_user(id)
on delete restrict on update cascade
);
```

Кроме этих сущностей в базе данных также были созданы сущности, необходимые для реализации модулей анализа данных.

Модель базы данных представлена на рис. 6.

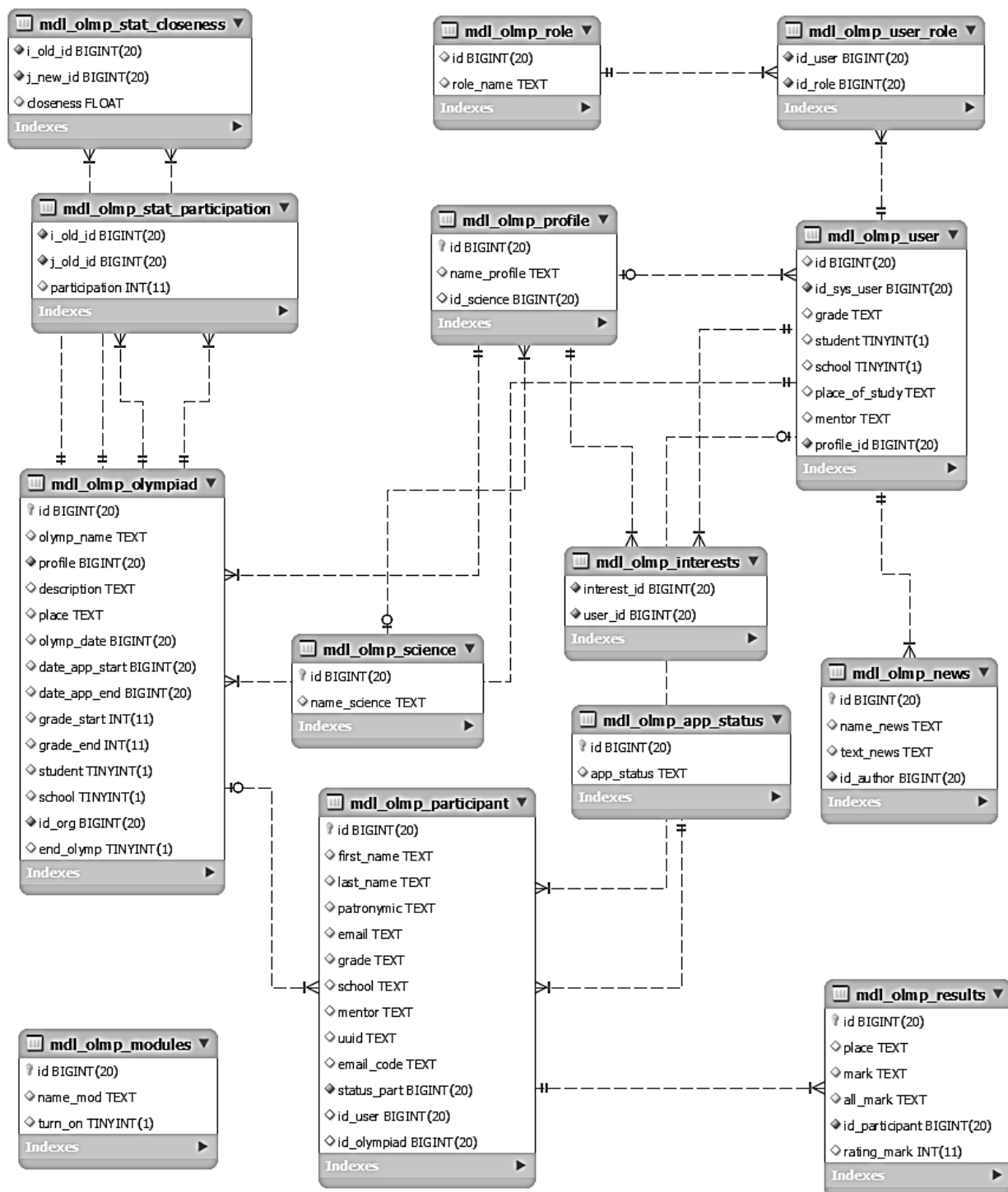


Рис. 6. Модель базы данных системы “Олимпиада”

Создание пользователей и ролей

В системе Moodle представлена хорошая и проработанная система аутентификации и авторизации. Существуют всевозможные плагины для различных способов аутентификации в системе, такие как CAS, IMAP, LDAP, NNTP и другие [17]. В связи с этим при разработке системы решено было воспользоваться стандартными средствами.

Но так как в системе существуют свои собственные роли администратор, организатор и участник, необходимо было каким-то образом предоставить пользователю различный функционал без изменения стандартных методов и системных таблиц.

Для этого была создана дополнительная таблица OLMP_USER, которая дополняет системную таблицу USER и соединяется с ней связью один-к-одному. В ней находится дополнительная информация о пользователе, которая необходима для корректной работы системы, такая как: профиль пользователя, его интересы и так далее. Так же именно с этой таблицей происходит связь всех остальных таблиц.

С таблицей OLMP_USER связана таблица OLMP_ROLE, в которой заполнены роли пользователей системы “Олимпиада”: администратор, организатор и участник.

В связи с разработкой собственных ролей пользователей, доступ к страницам системы было решено осуществлять на их же основе. Вся логика веб-страниц была “обернута” в специальные методы проверки ролей. Если пользователю с его ролью разрешено посещать необходимую страницу, то на этой странице будет отображен интересующий его контент. В ином случае пользователю выводится информационное сообщение о запрете доступа.

Сценарии использования

По требованиям, представленным в параграфе 2.1, были составлены сценарии использования, представляющие основные пути поведения пользователя в системе. Сценарий использования, вариант использования, прецедент или же пользовательский сценарий (англ. Use Case) — в разработке программного обеспечения и системном проектировании это описание поведения системы, которым она отвечает на внешние запросы. Другими словами, сценарий использования описывает, «кто» и «что» может сделать с рассматриваемой системой [18].

UC - CreateOlymp: Создание олимпиады

1. Пользователь с ролью организатор или администратор нажимает на ссылку создать олимпиаду.
2. Система отображает форму на странице с полями, необходимыми для заполнения.
3. На форме пользователь выбирает профиль олимпиады из выпадающего списка и заполняет все оставшиеся поля.
4. Система сохраняет олимпиаду.
5. Система предлагает пользователю создать новость об опубликованной олимпиаде. Если пользователь не собирается создавать новость, то сценарий заканчивается, иначе переход на шаг 6.
6. Пользователь нажимает на ссылку “Создать новость”.
7. Система отображает форму для создания новости с полями, необходимыми для заполнения. Поля автоматически заполняются по шаблону, при этом, подставляются актуальные сведения о созданной только что олимпиаде.
8. Пользователь нажимает сохранить.
9. Система сохраняет новость.

UC - CreateApply: Подача заявки на олимпиаду

1. Пользователь с ролью участник нажимает на ссылку подать заявку на олимпиаду.
2. Система отображает форму на странице с полями, необходимыми для заполнения.
3. На форме пользователь выбирает олимпиаду из выпадающего списка. Список формируется таким образом, что в нем присутствуют только актуальные олимпиады, то есть те, у которых не закончился срок подачи заявки на текущий момент. Олимпиады, упорядочиваются по дате создания. Пользователь заполняет все оставшиеся поля.
4. Система формирует и выводит на экран уникальный идентификатор пользователя (UUID), который пользователь должен сохранить и использовать в дальнейшем для проверки статуса заявки и результатов.
5. Система сохраняет участника олимпиады со статусом “не рассмотрена”.

UC - ConsiderationApply: Рассмотрение заявки на олимпиаду

1. Пользователь с ролью организатор заходит на страницу с олимпиадами.
2. Пользователь выбирает нужную ему олимпиаду.
3. Система отображает олимпиаду с информацией о ней и кнопками, среди которых есть кнопка “список заявок”.
4. Пользователь нажимает кнопку “список заявок”.
5. Система отображает список участников олимпиады со статусом заявки “не рассмотрена”.
6. Пользователь нажимает на необходимого ему участника.

7. Система отображает всю информацию об участнике, а также отображает кнопки “Принять заявку” и “Отклонить заявку”.
8. Если пользователь нажимает на кнопку “Принять заявку”, то система меняет статус заявки на “одобрена” – пользователю выводится соответствующее сообщение. Если пользователь нажимает на кнопку “Отклонить заявку”, то система меняет статус заявки на “отклонена” – пользователю выводится соответствующее сообщение.

UC - CheckResult: Проверка результатов заявки и олимпиады

1. Пользователь с ролью участник нажимает на ссылку “Проверка результатов”.
2. Система отображает на странице форму с полем “Уникальный идентификатор”.
3. Пользователь вводит свой идентификатор, выданный ему в сценарии UC – CreateApply на шаге 4.
4. Дальнейшее поведение системы зависит от введенного идентификатора:
 - если система не нашла такой идентификатор среди поданных заявок, то выводится сообщение “заявка не найдена”;
 - если система нашла идентификатор и заявка отклонена, то выводится сообщение “заявка отклонена”;
 - если система нашла идентификатор, заявка одобрена, но олимпиада еще не прошла, то выводится сообщение “заявка одобрена”;
 - если система нашла идентификатор, заявка одобрена, олимпиада прошла, но результатов в БД еще нет, выводится сообщение “олимпиада прошла, ожидайте результатов”;

— если система нашла идентификатор, заявка одобрена, олимпиада прошла, и результаты олимпиады уже опубликованы, то выводится сообщение с описанием результатов участника.

**UC - AddParticipant: Добавление участника
организатором/администратором вручную**

1. Пользователь с ролью организатор нажимает на ссылку “Добавить участника”
2. Система отображает форму на странице с полями, необходимыми для заполнения.
3. На форме пользователь выбирает олимпиаду из выпадающего списка. Список формируется таким образом, что в нем присутствуют только актуальные олимпиады, то есть те, у которых не закончился срок подачи заявки на текущий момент. Олимпиады, упорядочиваются по дате создания. Пользователь заполняет все оставшиеся поля.
4. Система формирует и выводит на экран уникальный идентификатор пользователя (UUID), который организатор может сообщить участнику олимпиады для проверки статуса заявки и результатов.
5. Система сохраняет участника олимпиады со статусом “одобрена”.

**UC – CreateOrganizer/Administrator: Назначение пользователя
системы Moodle организатором/администратором**

1. Пользователь с ролью администратор или организатор нажимает на ссылку “Создать организатора”/“Создать администратора”.
2. Система отображает форму на странице с полями: “логин пользователя” и “оставить в роли участника?”. Поле логин должно быть заполнено актуальным логином пользователя системы Moodle. Поле “оставить в роли участника?” имеет тип checkbox и может быть заполнено значением “да” или “нет”. Если поле имеет значение “да”,

то пользователь, после назначения ему новой роли, остается, также, в роли “участник”, если имеет значение “нет”, то роли “участник” он лишается.

3. Пользователь нажимает кнопку “Назначить”.
4. Если введенный логин найден в системе, то система назначает пользователю роль “организатор”/“администратор”, иначе выводится сообщение “пользователь не найден”.

UC – UploadResults: Загрузка результатов

1. Пользователь с ролью организатор или администратор заходит на страницу с интересующей его олимпиадой.
2. Пользователь нажимает на кнопку “Список участников”.
3. Система отображает данные всех участников олимпиады с возможностью загрузки списка в формате CSV или XLS.
4. Пользователь нажимает кнопку “Скачать”, предварительно выбрав из выпадающего списка формат “Таблица Excel”.
5. В скачанной таблице, кроме столбцов, отображенных на странице списка участников, также добавляются столбцы – “Всего баллов”, “Кол-во набранных баллов” и “Призовое место”. Эти столбцы организатору необходимо заполнить результатами олимпиады по каждому участнику.
6. После заполнения результатов пользователь возвращается на страницу олимпиады и нажимает кнопку “Загрузить результаты”.
7. Система отображает форму на странице с полем: “Файл”.
8. Пользователь нажимает кнопку “Выберите файл” и загружает файл с заполненными результатами. Затем пользователь нажимает кнопку “Загрузить данные”.

9. Система автоматически сопоставляет результаты олимпиады с пользователями системы Moodle и добавляет результаты в базу данных.

Также в системе есть и другие сценарии пользователя, но они имеют или более простое поведение или частично дублируют уже описанные сценарии, поэтому автором было принято решение не приводить их здесь.

Загрузка результатов олимпиады

При работе с любой системой организации олимпиад самое продолжительное время у пользователя, безусловно, уходит на добавление результатов олимпиады, поэтому автором была поставлена задача минимизировать это время. Необходимо было также учесть, чтобы заполнение результатов можно было производить по частям или несколькими организаторами сразу. Выбранный метод должен быть доступным для пользователя, удобным и интуитивно понятным. В связи с этим при заполнении результатов было решено воспользоваться стандартными электронными таблицами, такими как Microsoft Excel или OpenOffice Calc. Алгоритм загрузки результатов был приведен в сценарии использования UC – UploadResults.

Для решения поставленной задачи автор воспользовался открытой библиотекой RHPExcel. Библиотека RHPExcel состоит из набора классов для языка программирования PHP, которые позволяют писать и читать из различных форматов электронных таблиц, таких как Excel (.xls) BIFF, Excel 2007 (OfficeOpenXML) XLSX, CSV, Libre / OpenOffice Calc (.ods) , Gnumeric, PDF, HTML. Этот проект построен вокруг стандарта Microsoft OpenXML и PHP. Библиотека обеспечивает возможность установки мета-данных таблиц (автор, название, описание и т.д.), работу с несколькими листами, различными шрифтами и стилями шрифтов, границами ячеек, заливкой, градиентами,

добавлением изображений в таблицу, расчетными формулами, преобразованиями между типами файлов и многим другим [19].

Функция загрузки результатов представлена в листинге 2.

Листинг 2. Функция загрузки результатов олимпиады

```
public function uploadResult($destination_directory, $filename){

    $idString = "Идентификатор";
    $allMarkString = 'Всего баллов';
    $markString = 'Кол-во набранных баллов';
    $prizeString = 'Призовое место';
    $dangerMessage = "Во время загрузки данных произошла ошибка.
    Пожалуйста, повторите попытку позже или обратитесь к
    администратору системы.";

    $PHPExcel_file =
    PHPExcel_IOFactory::load($destination_directory.$filename);

    $worksheet = $PHPExcel_file->getSheet(0);

    $rows_count = $worksheet->getHighestRow();

    $columns_count =
    PHPExcel_Cell::columnIndexFromString($worksheet->
    >getHighestColumn());    $arrayColumnNames = array();

    for ($column = 0; $column < $columns_count; $column++) {
        $arrayColumnNames[$worksheet->
    >getCellByColumnAndRow($column, 1)->getCalculatedValue()] =
    $column;
    }
```

Листинг 2 (продолжение). Функция загрузки результатов олимпиады

```
for ($row = 2; $row <= $rows_count; $row++) {
    $uuid = $worksheet-
>getCellByColumnAndRow($arrayColumnNames[$idString], $row)-
>getCalculatedValue();
    $allMark = $worksheet-
>getCellByColumnAndRow($arrayColumnNames[$allMarkString], $row)-
>getCalculatedValue();
    $mark = $worksheet-
>getCellByColumnAndRow($arrayColumnNames[$markString], $row)-
>getCalculatedValue();
    $prize = $worksheet-
>getCellByColumnAndRow($arrayColumnNames[$prizeString], $row)-
>getCalculatedValue();
    global $DB;
    $participant = $this->getParticipantByUUID($uuid);
    $result = new stdClass();
    if(!is_null($prize)){
        $result->place = $prize;
    }
    $result->mark = $mark;
    $result->all_mark = $allMark;
    $result->id_participant = $participant->id;
    try{
        $this->insertResult($result);
    } catch(Exception $ex){
        echo $this->show_message("danger", $dangerMessage);
    }
}
}
```

2.3. Разработка модуля анализа данных

Немаловажную роль в информационной системе “Олимпиада” играют модули анализа данных. Они размещаются на личной странице пользователя. Всего в системе предусмотрено два модуля: рейтинговые листы для организаторов и рекомендации по олимпиадам для участников.

Как уже было сказано в параграфе 2.1. настоящей главы, рейтинговые листы для организаторов представляют собой список участников, занявших лидирующее место в рейтинге, по итогам прошедших олимпиад необходимого профиля. Рекомендации по олимпиадам для участников, в свою очередь, формируются двумя способами: на основе методов фильтрации содержимого по анкетным данным пользователя и на основе коллаборативной фильтрации.

Прогноз составляется индивидуально для каждого пользователя, хотя используемая информация собрана от многих участников.

Рейтинговая система.

Расчет рейтинга пользователей становится возможным после появления первых результатов олимпиад в системе. Методы формирования данных для рейтинговой системы вызываются каждый раз, когда пользователь публикует результаты олимпиады. В момент загрузки результатов, вызывается функция, которая частично просчитывает комплексную оценку для каждого участника и сохраняет ее в базу данных.

Расчет комплексной оценки проводится по формуле (3).

$$R_j = \sqrt{\sum_{i=1}^n k_i (1 - x_{ij}/y_i)^2}, \quad (3)$$

где i – номер олимпиады, x_i – количество полученных баллов пользователя j в олимпиаде i , y_i – максимально возможное количество баллов олимпиады i , k_i – коэффициент значимости олимпиады i .

Рейтинг строится динамически в момент генерации web-страницы и является уникальным для каждого организатора с определенным профилем.

При построении рейтинга система берет частично посчитанную комплексную оценку за каждую олимпиаду, в которой участвовал пользователь и умножает ее на соответствующий коэффициент k :

- если профиль олимпиады совпадает с профилем организатора, то $k = 0,0000001$;
- если научная область олимпиады близка к научной области организатора, то $k = 0,0001$;
- иначе, $k = 0,0007$.

После этого, по формуле, все значения складываются, и из результата берется корень.

Принадлежность профиля к научной области олимпиады регулируется соответствующими записями в базе данных. Коэффициенты k были вычислены опытным путем и в дальнейшем при необходимости могут быть изменены.

Рекомендательная система. Фильтрация содержимого по анкетным данным пользователя

У каждого участника есть личная страница. На ней размещена подробная информация о пользователе: класс/курс, место учебы, руководитель, профиль, а также его интересы. Пользователь может указать эту информацию в любой момент, после того, как первый раз зайдет на свою личную страницу. В том случае, если пользователь указал свой профиль и интересы – у него начинает работать фильтрация содержимого по анкетным данным, то есть система рекомендует ему олимпиады с тем же профилем, что и у него или, если хотя бы один из его интересов совпадает с профилем олимпиады. Кроме этого рекомендации основаны на организаторе олимпиады. Если у организатора олимпиады указан профиль, совпадающий с профилем участника или одним из его интересов, такая олимпиада так же будет рекомендована пользователю.

Рекомендательная система. Коллаборативная фильтрация

У каждой олимпиады есть свои свойства: профиль и организатор. Кроме информации, описывающей олимпиаду, известны статистические данные – количество пользователей, которые участвовали в разных олимпиадах.

На момент обучения системы существует множество олимпиад, которые уже были проведены и по которым есть статистика участия, и множество олимпиад, регистрация на участие в которых еще не закончена, и которые подлежат рекомендации.

Для пользователя, который участвовал хотя бы в одной олимпиаде, необходимо предложить новые олимпиады для участия (если таковые найдутся).

За основу алгоритма рекомендательной системы был взят алгоритм, занявший первое место на Международном соревновании “ECML/PKDD Discovery Challenge 2011” [20].

Пусть информацию об олимпиаде можно записать в виде n -мерного вектора $f = (f_1, f_2, \dots, f_n)$. Если n – число профилей всех олимпиад, то бинарный вектор f описывает профиль конкретной олимпиады: единицы помечают номера соответствующих профилей (такой вектор называется “характеристическим вектором профиля олимпиады”). Аналогично можно описать и организатора олимпиады.

Для каждого вида информации можно оценить, насколько похожи олимпиады. Например, для i -й олимпиады, представленной вектором $f(i) = (f_1(i), f_2(i), \dots, f_n(i))$ и j -й олимпиады, представленной вектором $f(j) = (f_1(j), f_2(j), \dots, f_n(j))$ их близость можно оценить с помощью измененной косинусной меры:

$$\langle f(i), f(j) \rangle = \frac{f_1(i)f_1(j) + \dots + f_n(i)f_n(j)}{\sqrt{f_1(i)^2 + \dots + f_n(i)^2 + \varepsilon} \sqrt{f_1(j)^2 + \dots + f_n(j)^2 + \varepsilon}} \quad (4)$$

Алгоритм: для заданной олимпиады, в которой участвовал пользователь перебрать все новые олимпиады и вычислить близость к каждой, просуммировав для всех представленных видов информации (профиль, организатор). Вместе с близостью к рассматриваемой олимпиаде необходимо учитывать близость к похожим олимпиадам с точки зрения поведения пользователей. Например, если прошедшая олимпиада относиться к профилю физика, а среди новых нет олимпиад из этого профиля, то можно порекомендовать из “похожего” профиля, например из математики.

Пусть множество прошедших олимпиад индексируется номерами из I , пусть $f(i)$ – вектор описания i -й олимпиады, пусть m'_{ij} – оценка близости i -й и j -й олимпиады с точки зрения поведения пользователя. Тогда пусть

$$f'(i) = \sum_{j \in I} \left(m'_{ij} \frac{f(j)}{\sqrt{f_1(j)^2 + \dots + f_n(j)^2 + \varepsilon}} \right) \quad (5)$$

и близость к новой i -й олимпиаде вычисляется с помощью суммы $\langle f'(i), f(t) \rangle$ по всем видам информации.

Пусть L – число олимпиад, m_{ij} – число пользователей, которые приняли участие в i -й и j -й олимпиаде при $i \neq j$. Тогда

$$m'_{ij} = \frac{m_{ij}}{\sum_{t=1}^L m_{it}} \quad (6)$$

Ниже перечислены виды информации, использованные для оценки близости. Для каждого вида указывается вектор $\gamma = (\langle f'(i), f(j_1) \rangle, \dots, \langle f'(i), f(j_r) \rangle)$, где $J = \{j_1, \dots, j_r\}$ – множество номеров новых олимпиад.

1. Близость профилей – γ_{prof} . $f(j)$ – характеристический вектор профилей олимпиад, т.е. бинарный вектор, в котором i -й элемент равен единице, если профиль олимпиады принадлежит t -у профилю.
2. Близость организаторов – γ_{org} . $f(j)$ – характеристический вектор организаторов олимпиад, т.е. бинарный вектор, в котором i -й элемент

равен единице, если t -й организатор является организатором олимпиады.

Решением алгоритма является значение $\gamma = \gamma_{prof} + \gamma_{org}$.

Таким образом, получается, что у каждой прошедшей олимпиады присутствует значение близости к каждой новой олимпиаде.

Обучение системы можно проводить двумя способами: “на лету” и отложенным запуском. Запуск “на лету”, означает, что при добавлении новой олимпиады, системе необходимо будет просчитать близость этой олимпиады к каждой прошедшей олимпиаде, а при завершении олимпиады, считать ее прошедшей и, соответственно, считать ее близость до каждой новой олимпиады. В этом способе есть существенный минус: при добавлении олимпиады система моментально начинает считать близость, и в случае, если организатор ошибся (была создана не та олимпиада или профиль олимпиады был указан неверно), необходимо будет остановить процесс и сделать пересчет значений близости. Фактически придется следить за любым изменением информации об олимпиаде, что впоследствии увеличивает нагрузку на систему в целом. Поэтому автором был выбран второй способ обучения – отложенный запуск.

Обучение отложенным запуском означает, что администратор системы Moodle по своему усмотрению в определенный момент времени или через заданный временной интервал запускает обучение. После этого составляется список новых олимпиад и список олимпиад, ставших за это время прошедшими. Затем определяется близость каждой новой олимпиады к каждой прошедшей олимпиаде и каждой олимпиады, ставшей прошедшей ко всем актуальным (у которых есть возможность подачи заявки) олимпиадам системы.

После того, как в системе посчитаны близости между прошедшими и новыми олимпиадами, последние необходимо рекомендовать пользователю. Они отображены на личной странице по ссылке “Рекомендованные

олимпиады”. Список рекомендованных олимпиад формируется по следующему алгоритму:

1. Составляется список олимпиад, в которых принимал участие пользователь.
2. Новые олимпиады фильтруются по уровню олимпиады (для школьника/для студента) и по классу/курсу олимпиады.
3. Выбираются близости между олимпиадами, посещенными пользователем, и новыми (отфильтрованными).
4. Близости упорядочиваются по убыванию.
5. Из полученного списка выбираются первые n (устанавливается в системе) записей.

В том случае, если данных в системе достаточно и у пользователя заполнены все необходимые поля – работают оба типа рекомендаций. В первую очередь олимпиады рекомендуются по анкетным данным пользователя. Если рекомендованных олимпиад недостаточно, то в рекомендованный список добавляются олимпиады, выбранные методом коллаборативной фильтрации.

Выводы

В настоящей главе, автором была поставлена задача рассказать про основные разработанные механизмы системы “Олимпиада”.

В рамках текущей главы был описан принцип работы системы. Автором было выделено ядро информационной системы: БД, реализация прав пользователей, пользовательские сценарии и веб-страницы, основанные на них. Рассказано про архитектуру системы.

Также были описаны дополнительные модули системы – модули анализа данных, и основные алгоритмы, которые были использованы в разработке рекомендательной системы.

3. Система “Олимпиада”. Тестирование полученных результатов

3.1. Описание разработанной системы “Олимпиада”

Разработанная система имеет различные интерфейсы для участника, организатора, администратора и анонимного пользователя. Пользователь, который заходит в систему видит подключаемый блок, который можно добавить или скрыть на любой странице. Наполнение блоков различается в зависимости от прав пользователя. На рис. 7 представлены блоки (слева направо, сверху вниз) незарегистрированного пользователя, пользователя с ролью участник, пользователя с ролью администратор и с ролью организатор.

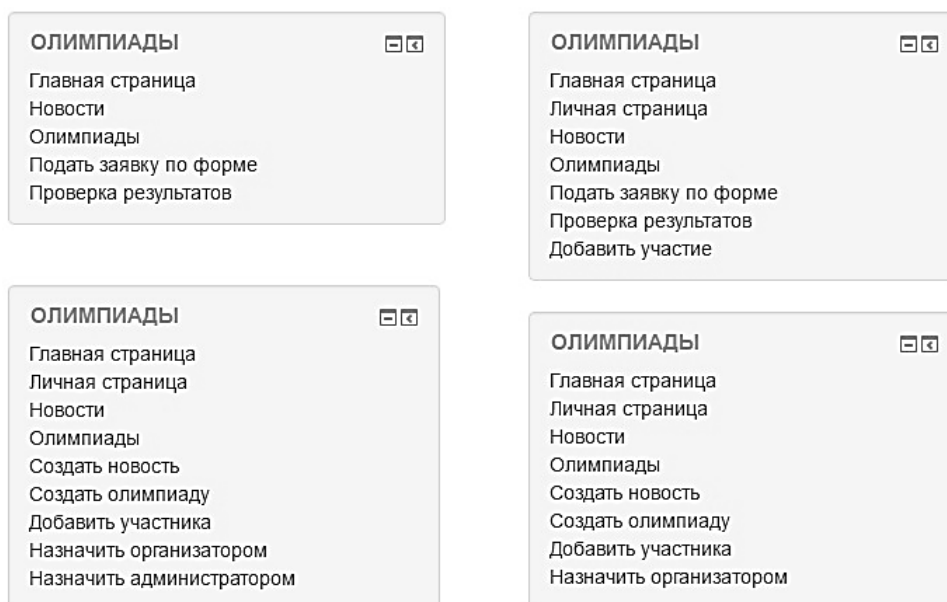


Рис. 7. Пример блоков для различных ролей пользователя

Первой ссылкой в блоке является “Главная страница”. При нажатии на нее пользователь попадает на главную страницу системы, на которой отображены последние новости (рис. 8). При нажатии на ссылку “Олимпиады” пользователю будут показаны последние добавленные в систему олимпиады.

ИС "Олимпиада"

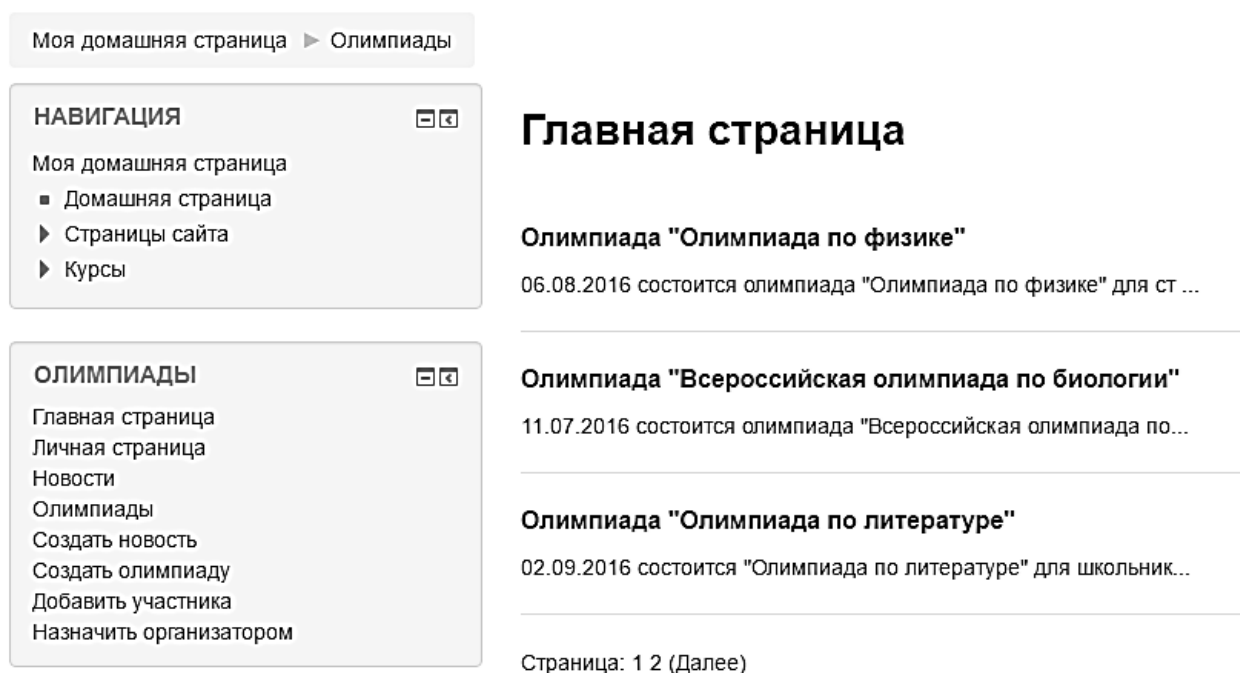


Рис. 8. Главная страница системы "Олимпиада"

Чтобы начать работу в системе – организатор олимпиады – кафедра, преподаватель или сотрудник кафедры, создает событие олимпиады (рис. 9). После создания олимпиады организатору предлагается опубликовать новость об олимпиаде. Новость создается по заранее назначенному шаблону, но может быть отредактирована. Это сделано для того, чтобы пользователю не приходилось вручную снова вводить все параметры олимпиады. После этого созданная новость и олимпиада публикуются и их можно просмотреть.

Пользователь с ролью организатор может посмотреть по каждой олимпиаде список заявок, список участников, результаты олимпиады, а также загрузить результаты, удалить или отредактировать олимпиаду (рис. 10). Стоит отметить, что редактирование или удаление олимпиады организатором возможно, только если он сам создал эту олимпиаду.

ИС "Олимпиада"

Моя домашняя страница ▶ Олимпиады ▶ Создание олимпиады

НАВИГАЦИЯ

Моя домашняя страница

- Домашняя страница
- Страницы сайта
- Курсы

ОЛИМПИАДЫ

Главная страница
Личная страница
Новости
Олимпиады
Создать новость
Создать олимпиаду
Добавить участника
Назначить организатором

Создание олимпиады

Название олимпиады

Олимпиада по физике

Профиль олимпиады

физика

Описание

Рис. 9. Создание олимпиады

ИС "Олимпиада"

Моя домашняя страница ▶ Олимпиады ▶ Олимпиада

НАВИГАЦИЯ

Моя домашняя страница

- Домашняя страница
- Страницы сайта
- Курсы

ОЛИМПИАДЫ

Главная страница
Личная страница
Новости
Олимпиады
Создать новость
Создать олимпиаду
Добавить участника
Назначить организатором

Олимпиада

Олимпиада по физике

Профиль олимпиады: физика

Место проведения: АлтГУ, корпус "К", ауд. 310

Дата проведения: 20.05.2016

Прием заявок: 20.04.2016 - 10.05.2016

Уровень участника: для студентов с 3 по 4 курс

Описание:

Редактировать

Удалить

Список заявок

Список участников

Результаты олимпиады

Загрузить результаты

Рис. 10. Просмотр информации об олимпиаде

Участники, в свою очередь, могут подать заявку на олимпиаду по форме. После принятия заявки выдается уникальный идентификатор (рис. 11). По данному идентификатору участник в дальнейшем сможет узнать свои результаты. Для этого ему необходимо перейти по ссылке “Проверка результатов” и ввести в поле идентификатор. В зависимости от корректности идентификатора, статуса заявки или наличия результатов ему будет показано соответствующее сообщение. После размещения результатов олимпиады в системе, пользователь также сможет увидеть их на странице самой олимпиады. Все олимпиады, в которых когда-либо участвовал пользователь, доступны ему на личной странице.

Для пользователя с ролью организатор просмотр результатов возможен на странице олимпиады. Для этого необходимо нажать кнопку “Результаты олимпиады”.

ИС "Олимпиада"

Моя домашняя страница ▶ Олимпиады ▶ Подать заявку

НАВИГАЦИЯ

Моя домашняя страница

- Домашняя страница
- Страницы сайта
- Курсы

ОЛИМПИАДЫ

Главная страница
Личная страница
Новости
Олимпиады
Подать заявку по форме
Проверка результатов
Добавить участие

Подать заявку

Заявка принята. Ваш номер для проверки заявки: 5403b650-7722-4c58-9267-8db588087eb7

Рис. 11. Пример успешной подачи заявки на олимпиаду

Также организатор может добавить участника вручную, если у него нет возможности подать заявку на олимпиаду онлайн. Идентификаторы, в дальнейшем, могут быть распечатаны и выданы участникам, зарегистрированным оффлайн, на самой олимпиаде.

После подачи заявки участником, организатор должен утвердить их или отклонить (рис. 12).

ИС "Олимпиада"

Моя домашняя страница ► Олимпиады ► Заявка

НАВИГАЦИЯ

Моя домашняя страница

- Домашняя страница
- Страницы сайта
- Курсы

ОЛИМПИАДЫ

Главная страница
Личная страница
Новости
Олимпиады
Создать новость
Создать олимпиаду
Добавить участника
Назначить организатором

Заявка

ФИО: Петров Петр Иванович

Место учебы: АлтГУ

Курс/Класс: 4

Руководитель: Иванов А.А.

Адрес электронной почты: petrov@ya.ru

Статус заявки: Не рассмотрена

Утвердить заявкуОтклонить заявку

Рис. 12. Пример рассмотрения заявки организатором

Все списки (список участников, список результатов, рейтинговый список) в системе “Олимпиада” являются интерактивными: можно задать сортировку для каждого столбца, как по убыванию, так и по возрастанию, а также настроить видимость. Организатор может просмотреть и выгрузить любой список в формате CSV или XLS (рис. 13).

Список результатов олимпиады

Скачать табличные данные как CSV-файл (разделитель - запятая) Скачать

Фамилия	Имя	Отчество	Адрес электронной почты	Курс/ Класс	Место учебы	Руководитель	Идентификатор	Всего баллов	Кол-во баллов	Призовое место
Петров	Петр	Иванович	petrov@ya.ru	4	АлпГУ	Иванов И.А.	afc60389-63df-49ec-97b5-16d169ea5883	50	48	1
Сидоров	Илья	Игоревич	fedorenko@ya.ru	3	АлпГУ	Федоренко М.А.	cf8bf6f2-1b52-4985-a894-67920bc9240a	50	45	2
Глинина	Екатерина	Андреевна	glinina@ya.ru	3	АлпГУ	Обухов Л.В.	9cb0e181-7f1b-4e33-b2f3-d8fe6b8c0118	50	38	
Богданов	Федор	Русланович	hasanov@ya.ru	3	АлпГУ	Хазанов В.Р.	9bbe2c28-009a-4870-937e-69b7a2c2b5d2	50	41	3

Рис. 13. Просмотр списка результатов олимпиады

Добавление результатов олимпиады осуществляется методом загрузки документа в систему с заполненными результатами. После первой загрузки результатов, олимпиада считается завершенной, и ее результаты начинают учитываться в рейтинговой системе.

В случае, если пользователь не раз участвовал в олимпиадах, но не был зарегистрирован в системе, а после решил все-таки зарегистрироваться, он может добавить идентификаторы всех олимпиад, в которых он участвовал, и они автоматически привяжутся к его профилю (рис. 14). В дальнейшем, это позволит ему видеть все олимпиады, в которых он участвовал на одной странице, и получать рекомендации по новым олимпиадам по схожей тематике.

Для пользователей с ролью организатор или администратор в системе предусмотрена возможность назначения других пользователей, соответственно, организатором или администратором.

ИС "Олимпиада"

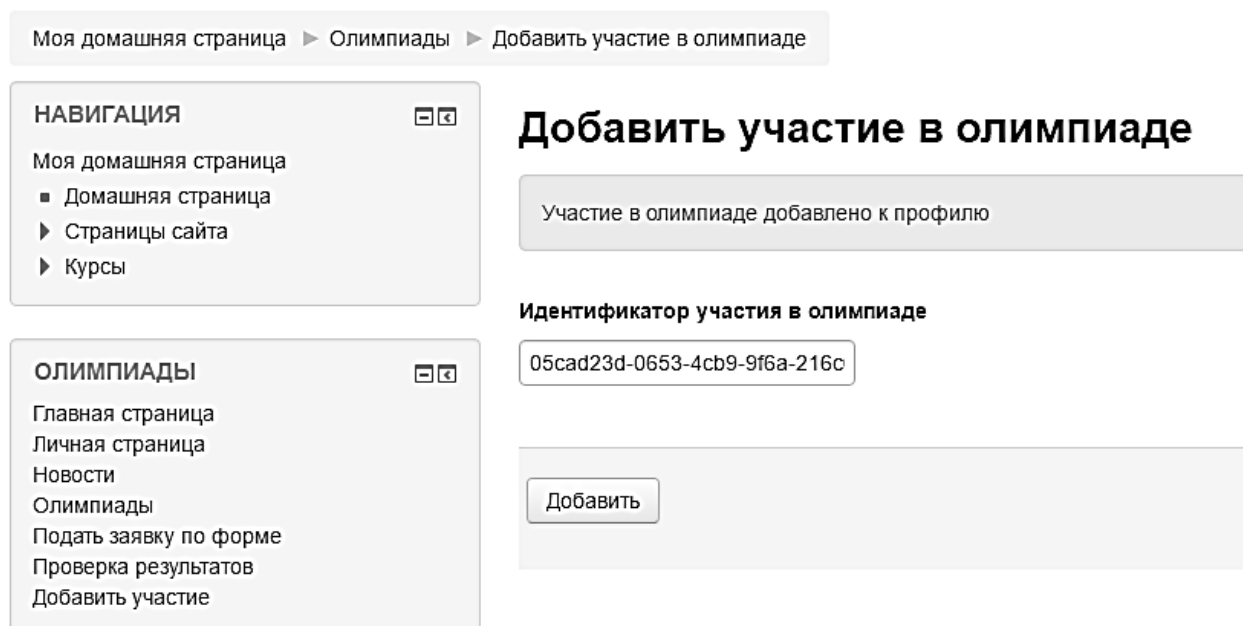


Рис. 14. Пример добавления участия в старой олимпиаде

Важной особенностью системы является личная страница. На ней отображается дополнительная информация о пользователе, списки олимпиад и модули рекомендательной и рейтинговой систем. Наполнение личной страницы зависит от роли пользователя. Дополнительную информацию о себе пользователь указывает сам и в любой момент может ее отредактировать.

Для пользователя с ролью организатор или администратор на личной странице располагается ссылка на рейтинговый список пользователей системы, а также ссылка на предстоящие и собственные олимпиады (рис. 15). Собственные олимпиады – это олимпиады, автором которых является текущий пользователь.

Для пользователя с ролью участник на личной странице располагается ссылка на рекомендованные олимпиады, а также ссылка на предстоящие и пройденные олимпиады (рис. 16).

ИС "Олимпиада"

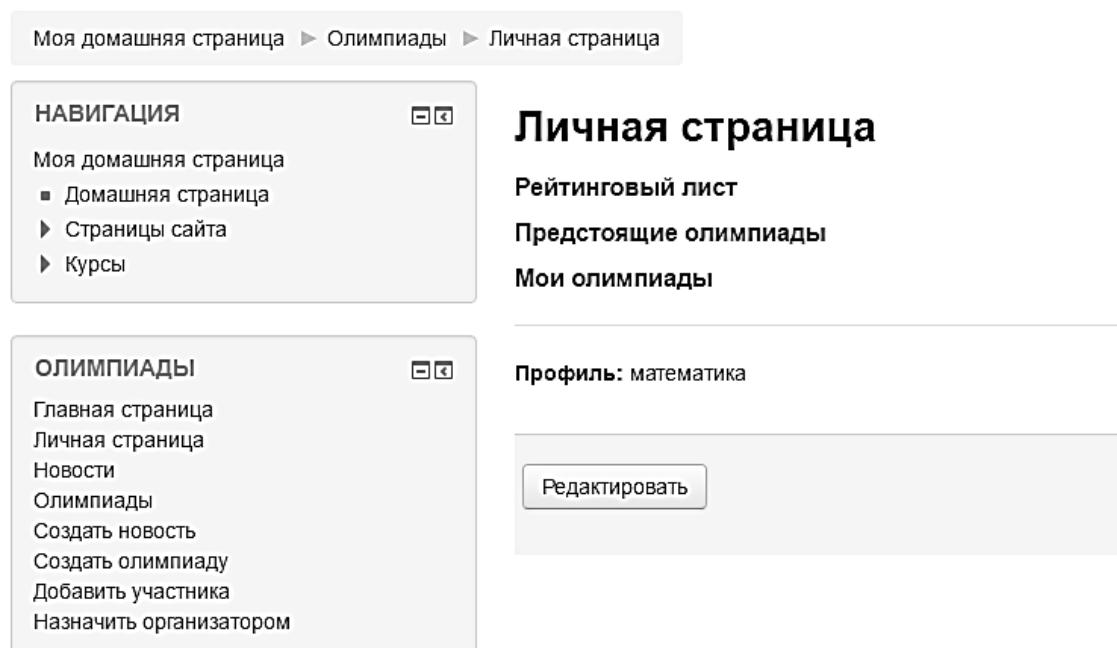


Рис. 15. Пример личной страницы организатора или администратора

ИС "Олимпиада"

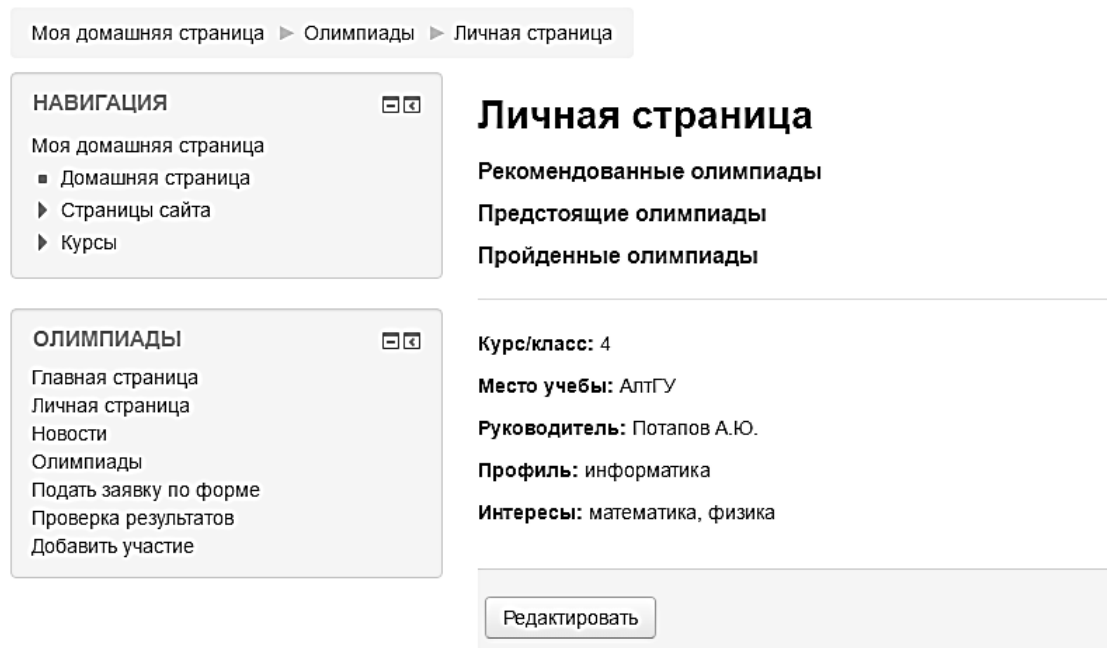


Рис. 16. Пример личной страницы участника

3.2. Тестирование рекомендательной и рейтинговой систем модуля анализа данных

Тестирование рекомендательной системы

Для оценки качества выдачи рекомендательной системы авторами оригинального алгоритма использовалась формула:

$$\frac{1}{z} \frac{\{r_1, \dots, r_{\min(S, R, z)}\} \cap \{s_1, \dots, s_{\min(S, R, z)}\}}{\min(S, R, z)}, \quad (7)$$

где $\{r_1, \dots, r_{\min(S, R, z)}\} \in R$ – множество олимпиад, являющихся результатом рекомендательной системы, $\{s_1, \dots, s_{\min(S, R, z)}\} \in S$ – эталонная выборка олимпиад, z – размер выборки.

Для тестирования были подготовлены данные, содержащие 50 прошедших и 10 новых олимпиад разного профиля. Были сгенерированы 50 пользователей системы. В каждой прошедшей олимпиаде было заведено от 10 до 30 участников. Также была составлена эталонная выборка из 10 пользователей системы для рекомендации новых олимпиад, по которой проверялось качество работы алгоритма.

Таблица 1. Данные для оценки качества алгоритма

Номер пользователя	R	S	Основное направление
1	1, 2, 12	1, 2, 3, 12	математика
2	5, 4, 11	5, 4	биология
3	10, 2, 9	10	лингвистика
4	9, 1, 2	9	история
5	7, 6, 11	7, 6	психология
6	14, 3, 2	14, 3, 2, 12	робототехника
7	13, 1, 2	13, 1, 2	криптография
8	12, 3, 1	12, 3, 1, 14	механика
9	3, 12, 1	3, 12, 1, 14, 2	физика
10	6, 7, 11	6, 7	социология

Пусть результатом рекомендательной системы является множество олимпиад R . Множество S является эталонной выборкой олимпиад. Олимпиады в каждой из выборок множества R упорядочены по значению близости. Выборка из множества S составлена автором самостоятельно, на основе знаний об основном профиле пользователя из списка новых олимпиад. Показатель $z = 3$. При оценке качества работы рекомендательной системы по формуле (7) автор заметил, что все рекомендации выдавались со 100% совпадением. Полученные результаты говорят о том, что система однозначно точно рекомендует олимпиады по основному профилю пользователя, но не объясняют, учитывается ли при этом история участия пользователей. При дальнейшем анализе рекомендованных олимпиад было выявлено, что все олимпиады, которые не совпадали по основному профилю, были либо из близких по тематике к основному профилю, либо созданы организатором с таким же профилем. На основании полученных результатов, можно говорить, что алгоритм рекомендательной системы работоспособен и выдает требуемые результаты.

Тестирование рейтинговой системы

Тестирование рейтинговой системы проводилось на созданной тестовой выборке. Было сгенерировано 14 олимпиад различных профилей. Олимпиады имели различные максимально возможные баллы для оценки результатов участников. Было создано 50 участников. Полученные баллы за олимпиады и факт участия были заполнены произвольно. Баллы некоторых участников были откорректированы автором вручную, для анализа показательных ситуаций. Например, когда участник участвовал только в олимпиадах своего профиля или когда он занимал только призовые места.

На полученной тестовой таблице, автору удалось откорректировать величину влияющего на результат коэффициента.

Лучшим участником рейтинга является пользователь, итоговое значение которого расположено ближе к эталонному значению. Из этих соображений и выбирались значения коэффициентов значимости. Ярким примером для оценки подобранного коэффициента были показательные ситуации, описанные выше. Опытным путем были подобраны следующие коэффициенты: коэффициент для олимпиад с однозначно совпадающим профилем рейтинга имеет наименьшее значение 0,0000001, олимпиадам близкого профиля установлено значение 0,0001, значения олимпиад остальных профилей будут умножены на коэффициент равный 0,0007. Два последних коэффициента не отличаются разрядностью, для исключения негативного влияния значений олимпиад отдаленного профиля на результаты олимпиад профиля, совпадающего с профилем, строящегося рейтинга.

Для исключения большого влияния количества пройденных пользователем олимпиад на итоговую оценку, было принято решение усреднить результат метода расстояний, поделив его на количество олимпиад, в которых текущий пользователь принимал участие и получил какой-либо балл. За счет подобранного коэффициента, при делении результата на количество олимпиад качество оценки олимпиады с основным профилем не изменилось.

Полученные результаты на тестовой выборке показали, что выбранный метод позволяет достичь поставленной цели и удовлетворяет всем условиям.

Выводы

В настоящей главе были представлены результаты проделанной работы, приведены примеры функционирования и выполнено тестирование полученной системы. Созданная система удовлетворяет требованиям, поставленным автором во второй главе.

Заключение

В данной выпускной квалификационной работе ставилась цель разработать информационную систему для организации олимпиад. Основными возможностями такой системы являются:

1. Публикация информации о проводимых олимпиадах.
2. Хранение результатов олимпиад и удобный доступ к ним, как для участника системы, так и для организатора.
3. Составление рейтинга участников для организаторов.
4. Рекомендации олимпиад участникам.

В рамках работы, автором был проведен анализ существующих информационных систем для работы с олимпиадами, выработана концепция будущей системы, определены требования и описан принцип работы системы. Были разработаны модули анализа данных и “ядро” информационной системы.

По итогам работы была получена работоспособная система, удовлетворяющая всем поставленным требованиям, и проведено ее тестирование.

По данным результатам можно сделать вывод, что поставленные задачи были выполнены и цель работы достигнута.

Библиографический список

1. Предметная олимпиада — Википедия - [Электронный ресурс].
-URL: https://ru.wikipedia.org/wiki/Предметная_олимпиада
2. Автономная некоммерческая образовательная организация «Дом учителя» - [Электронный ресурс].
-URL: <http://domu-barnaul.ru/>
3. Университет Лобачевского - [Электронный ресурс].
-URL: <http://www.unn.ru>
4. Российский совет олимпиад школьников - [Электронный ресурс].
-URL: <http://www.rsr-olymp.ru/>
5. Олимпиады для Школьников - [Электронный ресурс].
-URL: <http://info.olimpiada.ru>
6. Конкурсы и олимпиады школьников, студентов и педагогов - Эрудит - [Электронный ресурс].
-URL: <http://erudyt.ru>
7. Анализ данных — Википедия - [Электронный ресурс].
-URL: https://ru.wikipedia.org/wiki/Анализ_данных
8. Рекомендательная система - Википедия - [Электронный ресурс].
-URL: https://ru.wikipedia.org/wiki/Рекомендательная_система
9. Коллаборативная фильтрация — Википедия - [Электронный ресурс].
-URL: https://ru.wikipedia.org/wiki/Коллаборативная_фильтрация
10. Рекомендательные системы: Часть 1. Введение в подходы и алгоритмы - [Электронный ресурс].
-URL: <https://www.ibm.com/developerworks/ru/library/os-recommender1>
11. Коэффициент сходства — Википедия - [Электронный ресурс].
-URL: https://ru.wikipedia.org/wiki/Коэффициент_сходства
12. Тема 9 Способы и методы детерминированной комплексной оценки результатов хозяйственной деятельности - [Электронный ресурс]. -

URL:<http://edu.dvgups.ru/METDOC/EKMEN/BU/TEA/METHOD/PRAKTIKUM/TUMALI/2.htm>

13. Свиридова Нина Владимировна, Шопырин Александр Владимирович
Анализ методик построения рейтингов инновационной деятельности и активности // Известия ВУЗов. Поволжский регион. Общественные науки. 2013. №4 (28) - [Электронный ресурс].

URL: <http://cyberleninka.ru/article/n/analiz-metodik-postroeniya-reytingov-innovatsionnoy-deyatelnosti-i-aktivnosti>

14. Moodle - Википедия- [Электронный ресурс].

-URL: <https://ru.wikipedia.org/wiki/Moodle>

15. GNU_General_Public_License — Википедия - [Электронный ресурс].

-URL: https://ru.wikipedia.org/wiki/GNU_General_Public_License

16. Блок — MoodleDocs - [Электронный ресурс].

-URL: <https://docs.moodle.org/archive/ru/Блок>

17. Authentication - MoodleDocs - [Электронный ресурс].

-URL: <https://docs.moodle.org/30/en/Authentication>

18. Сценарий использования — Википедия - [Электронный ресурс].

-URL: https://ru.wikipedia.org/wiki/Сценарий_использования

19. PHPExcel - Home - [Электронный ресурс].

-URL: <http://phpexcel.codeplex.com/>

20. Дьяконов А. Г. Алгоритмы для рекомендательной системы: технология lenkor // Бизнес-информатика. 2012. №1 (19). - [Электронный ресурс].

-URL:<http://cyberleninka.ru/article/n/algoritmy-dlya-rekomendatelnoy-sistemy-tehnologiya-lenkor>

Исходный код информационной системы “Олимпиада”

Выпускная квалификационная работа выполнена мной совершенно самостоятельно. Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

« ____ » _____ Г.

(подпись выпускника)

(Ф.И.О.)