



СПЕЦИАЛИЗАЦИЯ «АНАЛИЗ ДАННЫХ». ВИРТУАЛЬНОЕ ОБРАЗОВАТЕЛЬНОЕ ОКРУЖЕНИЕ С ПОДДЕРЖКОЙ СРЕДСТВ ОБРАЗОВАТЕЛЬНОЙ АНАЛИТИКИ

Мусабилов И.Л.¹, Сироткин А.В.¹

¹Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Аннотация

Рассмотрен опыт проектирования дополнительной специализации (майнора) по анализу данных, сочетающей аудиторную и онлайн-компоненты. Описана архитектура виртуального образовательного окружения для обучения анализу данных студентов социогуманитарных направлений с интегрированной подсистемой образовательной аналитики. В состав окружения входят интегрированная среда анализа данных и статистической графики, сервер аналитических приложений, система интерактивных упражнений, средства поддержки онлайн-коммуникации и мониторинга прогресса студента. Подсистема образовательной аналитики включает в себя средства анализа журналов работы в окружении, контекстные опросные инструменты с поддержкой генераторов имён и средства анализа онлайн-коммуникации студентов.

Ключевые слова: виртуальное образовательное окружение, анализ данных, образовательная аналитика, машинное обучение, социально-сетевой анализ.

Цитирование: Мусабилов И.Л., Сироткин А.В. Специализация «Анализ данных». Виртуальное образовательное окружение с поддержкой средств образовательной аналитики // Компьютерные инструменты в образовании, 2016. № 4. С. 32–42.

1. ВВЕДЕНИЕ

Изменение требований к образовательному процессу и бурное развитие информационных технологий привели к существенным изменениям образовательного ландшафта. Одним из таких изменений стало появление массовых открытых онлайн-курсов (MOOCs), делающих доступным образовательный контент ведущих университетов по всему миру и стимулирующих совершенствование инструментов виртуального обучения. По данным Национального центра образовательной статистики США по состоянию на 2013 год более 25% студентов бакалавриата принимали участие хотя бы в одном курсе с применением технологий дистанционного обучения [1]. Запущен проект «Национальная платформа открытого образования», ставящий целью включение онлайн-курсов в учебные планы вузов России.

Вместе с тем, стали очевидными и основные недостатки текущего поколения массовых онлайн-курсов, в том числе минимизация контакта с преподавателем, отсутствие детальной обратной связи, а также гигантский процент оттока студентов (по некоторым данным процент успешного окончания онлайн-курсов составляет 5–15% [2]).

Всё это обуславливает необходимость сбалансированного, свободного от излишнего технооптимизма подхода к разработке стратегии использования образовательных технологий, гибко сочетающей весь спектр доступных технологий традиционного, онлайн- и смешанного обучения и позволяющей избежать утраты сильных сторон конкретного вуза, программы, предмета с одновременным включением новых форматов, где это релевантно.

Поиск эффективного формата среди прочих факторов зависит от специфики курса. С одной стороны, базовые курсы, посвящённые введению в информатику и программирование или введению в анализ данных, являются очевидными кандидатами для перевода преимущественно в онлайн-формат за счет формирования большого количества типовых заданий, доступных студентам, и системы автопроверки самостоятельной работы студентов. Но для эффективной работы такой системы со всеми категориями студентов необходима также и система прогнозной образовательной аналитики, способная выделять потенциально проблемных студентов на ранних этапах. С другой стороны, контакт с инструкторами — преподавателем, тьюторами, учебными ассистентами — критически важен для таких курсов. Кроме того, учитывая важную роль связей дружбы и помощи в образовании студентов [4], одним из ключевых факторов успеха здесь является и поддержка, организационная и технологическая, группового со-обучения среди студентов как в асинхронной форме, характерной для виртуального обучения [3], так и в более традиционных формах.

В статье рассматривается опыт разработки виртуального образовательного окружения с интегрированной системой образовательной аналитики для обучения анализу данных студентов бакалавриата социогуманитарных и экономико-управленческих направлений в Санкт-Петербургском кампусе Высшей школы экономики. Начавшийся с введения факультативов по основам программирования и анализа данных для социогуманитарных направлений в 2013 году эксперимент по интеграции курсов, обеспечивающих развитие «вычислительного мышления» (computational thinking) [5, 6], привел к созданию дополнительной специализации («майнора») «Обработка и анализ данных», доступной студентам бакалавриата всех направлений и сочетающей аудиторный и онлайн-компоненты.

2. ОРГАНИЗАЦИОННЫЙ КОНТЕКСТ И МЕТОДИЧЕСКИЕ РЕШЕНИЯ

Первый набор студентов на специализацию «Обработка и анализ данных» произошел в 2015 году. В ее рамках студенты всех академических направлений на протяжении двух лет (на втором и третьем курсах) обучаются основам анализа данных, научного программирования в среде R и машинного обучения. Специализация состоит из четырех семестровых курсов, каждый из которых оценен в пять кредитов ECTS:

- Программирование для анализа данных и воспроизводимые исследования.
- Анализ данных и технологии работы с данными.
- Интеллектуальный анализ данных и основы машинного обучения.
- Приложения и практика анализа данных.

При этом, ввиду ограниченного количества аудиторных часов, слабой и неравномерной поддержки приобретаемых навыков за пределами специализации, курсы тесно интегрированы на уровне содержания, методических материалов, разбираемых примеров. Так, деревья, классификации и регрессии возникают ещё в первом семестре в контексте примеров "сегментации с учителем"(supervised segmentation) [7], иллюстрируемых на задачах "родных" для студента дисциплин, и в дальнейшем служат стартовой точкой курса по интеллектуальному анализу данных.

Особенности специализации обуславливают несколько вызовов как методических, так и технических. *Во-первых*, в её рамках обучаются студенты с различным уровнем подготовки: от полного отсутствия высшей математики на некоторых образовательных программах до двухлетнего курса (на момент начала обучения по специализации и параллельно с ним), включающего линейную алгебру, математическую статистику, основы теории вероятностей и анализа данных. Хотя невозможно полностью компенсировать эти различия, при разработке специализации были использованы некоторые методические приемы, направленные на их сглаживание. На ранних этапах, насколько возможно, математические выкладки заменяются на симуляции или подкрепляются иллюстративным материалом с сохранением возможности более глубокого погружения в материал. Кроме того, в специализации делается акцент на интуитивно понятных методах социально-сетевого и текстового анализа и подчёркивается на практических примерах их релевантность для разных задач из всех направлений, которые представляют студенты [8, 9]. Наконец, примеры начальной части курса включают большой объём предварительно написанного программного кода. Это, с одной стороны позволяет студентам, несмотря на ещё скромные знания, получать видимые результаты – отчеты, визуализации, интерактивные приложения. С другой стороны, такой подход сохраняет возможность для более «сильных» студентов глубже погружаться в тему за счет самостоятельного детального разбора кода примеров.

Во-вторых, студенты разных образовательных программ, как правило, не имеют практического опыта совместной работы, что может весьма затруднить их взаимодействие. Несмотря на то, что в некоторых случаях при формировании групп является важным сохранение существующих сетей дружбы и помощи [10], в рассматриваемом случае приоритет отдавался потенциальному формированию новых, в том числе междисциплинарных связей. Это привело к формированию групп первого года обучения, объединяющих студентов разных направлений, без учёта дружеских связей, принадлежности к одной учебной группе, совместного проживания в общежитии. Более того, проектные задания предполагают совместную работу без учета направлений студентов.

В-третьих, мотивации студентов, повлиявшие на их запись на специализацию, значительно различаются и, тем самым, формируют контрастные типы поведения и взаимодействия на данном курсе. Это привело к осознанию необходимости дополнительного анализа мотиваций студентов и рассмотрению возможности создания вариативного компонента специализации.

3. ТРЕБОВАНИЯ К ВИРТУАЛЬНОМУ ОКРУЖЕНИЮ И СИСТЕМЕ ОБРАЗОВАТЕЛЬНОЙ АНАЛИТИКИ

Описанный выше организационный контекст и акцент на социальной составляющей процесса обучения (в том числе на учете эффектов со-обучения [4]) обусловили необходимость совместного проектирования учебной программы и инструментальных

средств – виртуального образовательного окружения, нацеленного не только на поддержку процесса освоения студентами компетенций, связанных с анализом данных, но и на постоянное отслеживание прогресса в обучении и изменения отношения студентов к предмету. Были сформулированы несколько высокоуровневых требований к окружению. *Во-первых*, система образовательной аналитики должна являться неотъемлемым элементом виртуального окружения, а не низкоприоритетным дополнением. *Во-вторых*, помимо базовых метрик образовательного процесса, в ней необходимо учитывать социально-сетевые и социально-психологические метрики и модели на нескольких уровнях, описывающие не только индивидуальный прогресс студентов, но социально-сетевую динамику учебных групп специализации и проектных команд, а также паттерны онлайн-коммуникации студентов. *В-третьих*, различие в мотивациях студентов и динамика изменений мотиваций в процессе обучения должны быть включены в анализ вместе с сетевой динамикой паттернов коммуникации.

4. СОСТАВ ВИРТУАЛЬНОГО ОБРАЗОВАТЕЛЬНОГО ОКРУЖЕНИЯ

В качестве образовательного ядра специализации и, следовательно, технологического ядра виртуального образовательного окружения выступает язык программирования, среда анализа данных и статистической графики R [11], развиваемая международным сообществом разработчиков, существенную часть которого составляют специалисты по статистике, представляющие как академическое, так и профессиональное сообщества. О сильной профессиональной позиции R свидетельствует в том числе поддержка созданного для координации развития R консорциума со стороны таких компаний, как Microsoft, Google, HP, Oracle.

При построении окружения максимально задействовались проекты, относящиеся к свободному программному обеспечению и программному обеспечению с открытым исходным кодом, обладающие активным сообществом пользователей и, в некоторых случаях, коммерческой поддержкой.

Акцент на обучение студентов практике воспроизводимых исследований [14] и, в узком смысле, литературного программирования [13] поддержан на инструментальном уровне за счёт использования языка разметки RMarkdown [15], который, будучи довольно простым в изучении, скрывает от студента специфические особенности таких систем, как LaTeX, позволяя создавать отчёты в форматах PDF, Word-документов или HTML-страниц, с включением в отчёт структурированного текста, формул, результатов вычислений в виде таблиц и графиков. Помимо документов, подсистема работы с RMarkdown, интегрированная со средой RStudio позволяет создавать презентации в виде динамических HTML-страниц. При этом возможно частичное или полное обращение более подготовленных пользователей к использованию LaTeX и Beamer.

Еще одним базовым компонентом окружения становится swirl [16] — ПО для создания интерактивных упражнений на базе R. Возможность автоматизировать процесс знакомства с командами и методами в форме интерактивных упражнений особенно важна в условиях большого количества студентов и с учётом известных сложностей, испытываемых студентами младших курсов при знакомстве с (очень часто) первым языком программирования. Для swirl доступна начальная библиотека готовых курсов, соответствующих последовательности онлайн-специализации Data Science, разработанной Coursera и John Hopkins University, что даёт возможность прямого её использования для дополнительной практики или самостоятельного изучения. Однако более важной является воз-

возможность создания собственных интерактивных упражнений и курсов на основе этого материала. Для компьютерных практикумов на поток в 200 человек такая возможность особенно важна для поддержки формирования навыков дополнительными интерактивными упражнениями, а также в качестве дополнительного средства синхронизации компьютерных практикумов в параллельных группах.

Использование клиент-серверной версии интегрированной среды разработки RStudio позволило реализовать прозрачный для студента переход между работой в классе и вне класса в едином окружении, обеспечивая, совместно с другими компонентами окружения, следующие преимущества:

- отсутствие необходимости установки и дальнейшей поддержки локальной версии R на компьютерах студентов и преподавателей;
- использование централизованных учётных записей пользователей;
- наличие централизованного обмена файлами, что позволяет раздавать материалы для занятий (конспекты лекций, командные сценарии и базы данных) и собирать отчёты о выполненных заданиях;
- доступ к рабочей среде с любого компьютера (при наличии соединения с Интернет) и возможность публикации в открытом доступе презентаций и документов;
- возможность дистанционного входа инструктора в рабочее окружение студента при необходимости.

Использование клиент-серверной архитектуры на базе открытых стандартов также позволяет задействовать ряд сложных и современных технологий без установки, настройки и дальнейшей поддержки их на локальных компьютерах в компьютерных аудиториях или персональных компьютерах студентов. В качестве примера таких технологий можно назвать СУБД MySQL и MongoDB, сервер приложений DeployR, систему data workflow Drake, командные сценарии и программы на других языках программирования. Это позволяет решать три методически важных задачи:

- минимизацию входного барьера для знакомства студентов с новыми технологиями;
- поощрение самостоятельного углубленного знакомства студентов с технологиями в учебной, проектной и исследовательской деятельности;
- проблему расширяемости технологического стека в рамках новых курсов или проектно-исследовательской деятельности.

Важной частью окружения является и коммуникационная площадка для студентов и инструкторов, реализованная по принципу сайтов «вопрос-ответ», содержащих и функционал рейтинговой системы (рис. 1). Активность студентов подкрепляется учетом хороших вопросов и ответов в системе оценивания курсов.

Так как развитие почти всего использованного стороннего программного обеспечения идёт довольно быстрыми темпами, окружение спроектировано по принципу минимальной модификации используемых внешних компонент, который призван предотвращать потенциальные проблемы с обновлениями. Поэтому большая часть специфичных для окружения административных компонентов работает либо на уровне стандартных средств операционной системы, либо с помощью вариантов расширения, предусмотренных в стороннем ПО. Так, например, единая система авторизации использует стандартные механизмы операционной системы и средства HTTP-аутентификации на уровне проксирования запросов. Общая архитектурная схема окружения приведена на рис. 2.

3 votes	3 answers	37 views	error 83 while knitting	homework essay	written 1 day ago by valuganskiy • 80 • updated 1 day ago by ABC • 930
1 vote	1 answer	54 views	Не нитится в ворд	error homework essay	written 1 day ago by vvraksina • 20 • updated 1 day ago by sdtomasheva • 50
5 votes	6 answers	292 views	Tutorial: маленький лайфхак для тех, кто пишет эссе, а не наслаждается осенним солнцем	tutorial latex essay	written 11 days ago by avmikhaylov_1 • 130 • updated 1 day ago by mamaksimova • 10
26 votes	15 answers	396 views	Темы для эссе	hw latest essay topics	written 10 days ago by Rin • 200 • updated 2 days ago by sdtomasheva • 50
3 votes	2 answers	53 views	Fonts in Rmarkdown	rstudio R essay	written 2 days ago by eiefremova • 30 • updated 2 days ago by zrzr • 810
1 vote	1 answer	56 views	english vs russian	essay	written 2 days ago by purelistica • 520 • updated 2 days ago by ABC • 930
3 votes	1 answer	67 views	Formulas abuse in essay	hw why? essay	written 2 days ago by vikorshunov • 280 • updated 2 days ago by zrzr • 810
0 votes	1 answer	40 views	При попытке снитить pdf файл появляется error 43	rstudio devtools essay	written 2 days ago by liddle • 0
0 votes	1 answer	69 views	Ограничение по количеству знаков в эссе	essay	written 3 days ago by mobelyaev • 50 • updated 3 days ago by ABC • 930
5 votes	1 answer	134 views	Оценки за д/з и к/р	results homework	written 4 days ago by ashtikhonova • 40 • updated 3 days ago by zrzr • 810
0 votes	0 answers	74 views	Вопрос по оформлению	homework essay	written 4 days ago by mobelyaev • 50
1 vote	1 answer	72 views	Цитирование статьи не из сколара		

Recent Awards • All »

Recent Replies

- С: Как создавать библиографический список by purelistica • 520
должен быть отдельный .bib файл посмотри тут, как можно оформить ссылки в нем <http://qa.posoc.io...>
- С: Как создавать библиографический список by iasokolov_3 • 10
1. вместо [toc]{chapter}{References} надо вставлять что-нибудь свое? 2. Так нужно создать отдель...
- А: Как создавать библиографический список by valuganskiy • 80
Создай биб файл с ссылками] В начале добавь эту строчку после bibliography, указав название бибф...
- А: My own data frame by purelistica • 520
if you want to read data from the file this might be helpful <https://stat.ethz.ch/R-manual/R-dev...>
- А: Ссылки в документе в виде вопросов by ierkah • 180
У меня получилось! Если ты не сохраняла изменения в биб файле, то может быть, дело в этом. Провер...
- А: Ссылки в документе в виде вопросов by zkmishina • 40
А у вас оба файла в одной папке сохранены? У меня была такая же ошибка, когда bib и rmd в разных...
- С: Ссылки в документе в виде вопросов by ierkah • 180
Да, я так уже делала:(
- С: Ссылки в документе в виде вопросов by ABC • 930
Я бы сделал ссылку не кириллической, могут быть проблемы. Замените параметр201334 на

Рис. 1. Коммуникационная площадка виртуального окружения

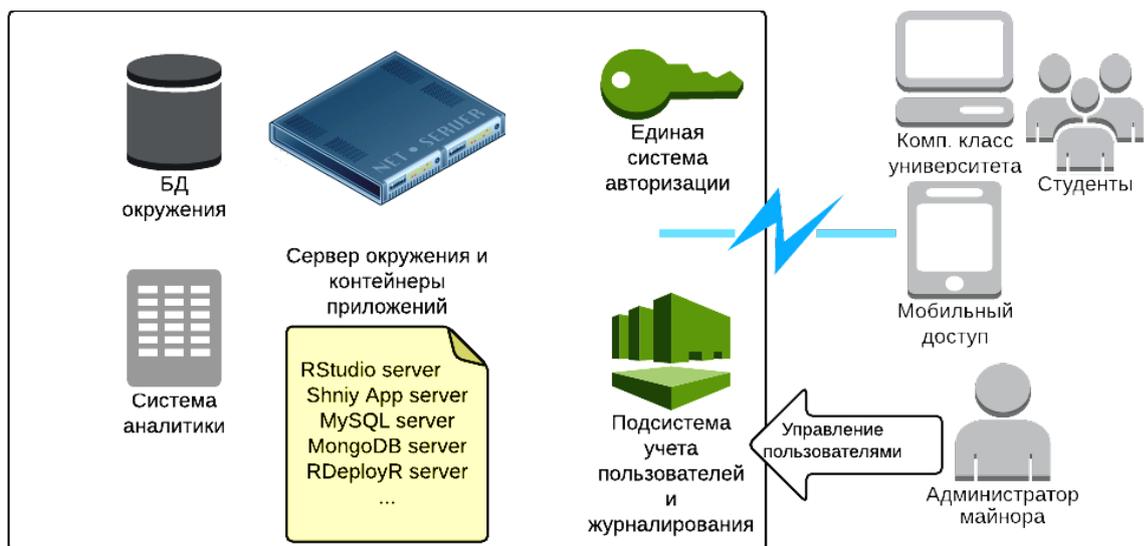


Рис. 2. Архитектура виртуального окружения

5. ОБРАЗОВАТЕЛЬНАЯ АНАЛИТИКА

Как уже было сказано раньше, важным компонентом построения эффективного виртуального образовательного окружения с целью перевода существенной части курса в онлайн является модуль поддержки образовательной аналитики. Поддержка образовательной аналитики постепенно становится стандартом de-facto для виртуальных образовательных окружений [16, 18], хотя существенная часть её современного инструментария специфична для массовых онлайн-курсов. Акцент на социально-сетевой составляющей обучения обусловил и специфику средств образовательной аналитики в рассматриваемом окружении.

Для измерения динамики мотиваций студентов используется подсистема контекстных опросов, позволяющих фиксировать изменения показателей групповой динамики и социально-психологических характеристик студентов в процессе обучения. Кроме того, для изучения влияния эффекта со-обучения на образовательные достижения в подсистему контекстных опросов включена возможность создания генераторов имён (name generators) [21], позволяющих студентам указывать конкретных студентов, с которыми они дружат или вступают в различные интеракции в процессе учебы. Пример опроса с генератором имён представлен на рис. 3.

1 Начало > 2 Текущие друзья > 3 Опрос часть 1 > 4 Опрос часть 2

← Назад **Дальше** →

Текущие друзья

Укажите ваше имя

Мусабилов Илья Леонидович

Выйти

Укажи, с кем из других студентов своего курса или учебных ассистентов ты дружишь

Начни вводить фамилию и выбери из выпадающего списка

Булыгин Денис x Воскресенский Вадим x Карелин Виктор x Поддубков Станислав x Харькина Дарья x гол|

Головина Анна Юрьевна

Головченко Анна Антоновна

Голубева Дарья Андреевна

Голубенкова Арсения Алексеевна

Голубенкова Арсения Алексеевна

Я выбрал(а) этот майнор, только чтобы учиться с друзьями

○ Не согласен(-сна) ○ ○ Полностью согласен(-сна)

Я выбрал(а) этот майнор, так как все остальные меня не привлекали

○ Не согласен(-сна) ○ ○ Полностью согласен(-сна)

Рис. 3. Опрос с генератором имён

Подсистема анализа онлайн-коммуникации студентов также играет важную роль. С точки зрения образовательной аналитики, коммуникация является важным фактором, влияющим на формирование внутри сети активного ядра агентов и возникновения эффекта Rich Club [17], приводящего к усилению проблем интеграции студентов, не вошедших в коммуникационное ядро. Выявление этого эффекта на ранних этапах должно позволить разработать меры, стимулирующие включение «выпадающих» из коммуникационной сети студентов. Предварительные наблюдения показывают, что этот эффект в рассматриваемом случае усиливается для студентов, мало представленных в когорте направлений. Кроме того, учитывая существенную разницу в подготовке обучающихся, можно предположить, что практика коммуникации на форуме может сыграть важную роль в процессе обмена опытом между студентами, обладающими разными уровнями в навыках программирования и знания математики. Иными словами, это дает возможность более опытным студентам делиться своим опытом, таким образом зарабатывая репутацию среди других студентов и помогая им повысить свои финальные оценки, и в то же время позволяет менее опытным студентам получать ответы на вопросы от одногруппников и быть вознагражденными, если вопросы хорошо сформулированы и могут быть полезными для всего сообщества. Этот же инструмент позволяет осуществлять и передачу знаний между когортами студентов, хотя некоторая степень редакторской обработки облегчает такого рода передачу

Важную роль в интеграции социально-психологической и социально-сетевой моделей с традиционными метриками образовательной аналитики, базирующимися на журналах событий, фиксирующих поведение студентов в образовательном окружении, играют аналитические социально-сетевые модели. Помимо классических моделей социально-сетевой динамики, таких как SIENA [19], разработанная T. Snijders, апробируются реляционные модели событий (relational event models) [20], которые успешно применяются в рамках социальной образовательной аналитики для анализа коэволюции потоков событий внутри массовых открытых онлайн курсов [22].

В рассматриваемом случае потоки состоят из трех частей:

- 1) поведенческих журналов и событий, происходящих в виртуальном окружении (время, потраченное на задание, количество ошибок, совершенных в процессе выполнения домашнего задания и т. д.);
- 2) событий, извлеченных из интеракций на форуме;
- 3) изменений в социальных и социально-психологических параметрах, измеренных при помощи контекстных опросов.

Такие потоки создают временные ряды, анализируя которые, можно выявлять как тренды и общие тенденции поведения, так и исключительные ситуации, выбивающиеся из общей схемы. Так можно выявлять студентов, которые тратят на конкретную тему существенно больше времени, чем мы от них ожидаем, а это, в свою очередь, может быть сигналом о необходимости точечной помощи студенту с конкретной тематикой или типом заданий.

Кроме анализа временных рядов, наличие текстов вопросов, а также наличие формализованного кода, написанного студентами для решения поставленных задач, позволяют применять методы текстовой аналитики, например тематической классификации и анализа тональностей.

6. ЗАКЛЮЧЕНИЕ

В статье описан подход к проектированию виртуального образовательного окружения для решения актуальной задачи обучения анализу данных студентов социогуманитарных и экономических специальностей. Подготовка в этой области повышает конкурентоспособность студентов в рамках их образовательных направлений, причём как на рынке труда, так и при продолжении обучения на более высокой ступени. Эффективное решение этой задачи невозможно без системного подхода к организации и проектированию содержимого обучения, инструментальной поддержки и получения обратной связи за счёт средств образовательной аналитики. Крайне важным является социальный аспект обучения, особенно в группах с существенной разницей в подготовке студентов, и проектирование образовательного процесса должно базироваться на детальном анализе социальной динамики студентов. Предложенная архитектура виртуального окружения позволяет детально фиксировать такого рода социальную динамику для последующего анализа.

Рассмотренный подход отражает опыт авторов в поиске эффективного сочетания традиционных и онлайн-компонент обучения, позволяющего не утратить экспертные знания конкретных преподавателей и контакт с аудиторией, с частичным переносом части функций, в том числе и накопления знаний между когортами студентов в онлайн-форму.

Список литературы

1. Enrollment in distance education course, by state: Fall 2012. Washington, DC: US Department of Education, National Center for Education Statistics, 2014.
2. Clow D. D. Clow. MOOCs and the funnel of participation // Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM, 2013. P. 185–189.
3. Hiltz S.R., Wellman B. Asynchronous Learning Networks As a Virtual Classroom // Commun. ACM, 1997. Vol. 40. № 9. P. 44–49.
4. Валеева Д.Р., Польшин О.В., Юдкевич М.М. Связи дружбы и помощи при обучении в университете // Вопросы образования, 2013. № 4. С. 70–84.
5. Wing J.M. Computational thinking // Communications of the ACM. 2006. Vol. 49. № 3. P. 33–35.
6. Хеннер Е.К. Вычислительное мышление // Образование и наука, 2016. № 2. С. 18–33.
7. Provost F., Fawcett T. Data Science for Business: What you need to know about data mining and data-analytic thinking. Sebastopol, Calif.: O'Reilly Media, 2013.
8. Easley D., Kleinberg J. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. New York: Cambridge University Press, 2010.
9. Cioffi-Revilla C. Introduction to Computational Social Science: Principles and Applications. London; New York: Springer, 2014.
10. Пронин А.С., Веретенник Е.В., Семенов А.В. Формирование учебных групп в университете с помощью анализа социальных сетей // Вопросы образования, 2014. № 3. С. 54–73.
11. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2015.
12. Gandrud C. Reproducible Research with R and R Studio. CRC Press, 2013.
13. Knuth D.E. Literate programming // The Computer Journal, 1984. Vol. 27. № 2. P. 97–111.
14. rmarkdown: Dynamic Documents for R / J.J. Allaire [et al.] R package version 0.8.1. 2015.
15. swirl: Learn R, in R. / N. Carchedi [et al.] R package version 2.2.21. 2015.
16. Siemens G., Gasevic D. Guest editorial-learning and knowledge analytics // Journal of Educational Technology & Society, 2012. Vol. 15. № 3. P. 1–2.
17. Vaquero L.M., Cebrian M. The rich club phenomenon in the classroom // Scientific reports, 2013. Vol. 3.

18. Ferguson R. Learning analytics: drivers, developments and challenges // International Journal of Technology Enhanced Learning, 2012. Vol. 4. Learning analytics. № 5–6. P. 304–317.
19. Huisman M., Snijders T.A. Statistical analysis of longitudinal network data with changing composition // Sociological methods & research, 2003. Vol. 32. № 2. P. 253–287.
20. Butts C.T. A relational event framework for social action // Sociological Methodology, 2008. Vol. 38. № 1. P. 155–200.
21. Robins G.L. Doing Social Network Research: Network-based Research Design for Social Scientists. SAGE Publications Ltd, 2015.
22. Vu D., Pattison P., Robins G. Relational event models for social learning in MOOCs // Social Networks, 2015. Vol. 43. P. 121–135.

Поступила в редакцию 11.11.2015, окончательный вариант — 09.06.2016.

Computer tools in education, 2016

№ 4: 32–42

<http://ipo.spb.ru/journal>

MINOR IN DATA SCIENCE. VIRTUAL LEARNING ENVIRONMENT WITH LEARNING ANALYTICS SUPPORT

Musabirov I.L.¹, Sirotkin A.V.¹

¹National Research University Higher School of Economics, Moscow, Russia

Abstract

This paper describes the design of a bootstrapped Virtual Learning Environment, developed for the undergraduate minor in Data Processing and Analysis. The environment is designed and deployed simultaneously with Learning Analytics subsystem to support students of diverse academic background studying in mixed groups. It includes integrated data analysis and statistical graphics environment, data applications server, interactive exercises system, tools for online communication and peer support, and progress tracking for students. Learning analytics subsystem includes analytical routines for behavioral logs, context-based surveys with a network component, and tools for analysis of online communication among the participants. The focus of Learning Analytics subsystem is on the social side of learning. Thus it includes advanced instruments for Social Network Analysis.

Keywords: *virtual learning environment, data analysis, learning analytics, machine learning, social network analysis.*

Citation: Musabirov, I. & Sirotkin, A., 2016. "Spetsializatsiya 'Analiz dannykh'. Virtual'noe obrazovatel'noe okruzhenie s podderzhkoi sredstv obrazovatel'noi analitiki" ["Minor in Data Science. Virtual Learning Environment with Learning Analytics Support."], *Computer tools in education*, no. 4, pp. 32–42.

Received 11.11.2015, the final version — 09.06.2016.

Ilya L. Musabirov, MA, MSc, Lecturer Department of Sociology Higher School of Economics, 190008, St. Petersburg, ul. Soyuza Pechatnikov, 16, NRU HSE, ilya@musabirov.info

Alexander V. Sirotkin, PhD, Associate Professor Department of Applied Mathematics Higher School of Economics, Senior Research Fellow SPIIRAS, 190008, St. Petersburg, ul. Soyuza Pechatnikov, 16, NRU HSE, avsirotkin@hse.ru



Наши авторы, 2016.

Our authors, 2016.

**Мусабилов Илья Леонидович,
преподаватель департамента социологии
НИУ ВШЭ,
190008, Санкт-Петербург, ул. Союза
Печатников, 16, НИУ ВШЭ,
ilya@musabirov.info**

**Сироткин Александр Владимирович,
кандидат физико-математических наук,
доцент департамента прикладной
математики и бизнес-информатики НИУ
ВШЭ, старший научный сотрудник
СПИИРАН,
avsirotkin@hse.ru**