

УДК 681.3

РАЗРАБОТКА МОДУЛЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Мельников А. Ю., Кравченко К. В., Петренко А. А.

THE DEVELOPMENT OF DATA MINING MODULE AT DISTANCE LEARNING SYSTEM

Melnikov A. Y., Kravchenko K. V., Petrenko A. A.

Рассмотрена проблема интеллектуального анализа данных системы дистанционного обучения. Описано функционирование одной существующих систем дистанционного обучения. Спроектирована на языке моделирования UML подсистема, состоящая из двух частей – модуля анализа активности пользователей и модуля анализа результатов тестирования. Описана программная реализация системы на языке Perl.

Ключевые слова: интеллектуальный анализ данных, дистанционное обучение, информационная модель, унифицированный язык моделирования UML, веб-сайт.

Введение. В настоящее время все большее распространение получает Data Mining – интеллектуальный анализ данных – технология, которая предназначена для поиска в больших объемах данных неочевидных, объективных и полезных на практике закономерностей [1]. Очевидно, что модуль для такого анализа может стать весьма полезным и неотъемлемым элементом любой информационной системы. Решение таких задач интеллектуального анализа данных, как классификация, кластеризация, прогнозирование позволит сделать процесс принятия любого решения более обоснованным.

Дистанционное обучение – одна из форм организации образовательного процесса, основанная на использовании новейших телекоммуникационных и информационных технологий, а также на принципах индивидуального обучения и самоконтроля. В настоящее время создано множество систем дистанционного обучения. Как правило, основным элементом любой из них является веб-сайт, позволяющий разграничить доступ к образовательным ресурсам и организовать, с одной стороны, управление содержанием (со стороны преподавателя), а с другой – собственно учебный процесс.

Целью работы является разработка модуля интеллектуального анализа данных системы

дистанционного обучения и внедрение этого модуля в существующую систему.

Изложение основных материалов.

Рассматриваемая система дистанционного обучения [2] выполнена в виде веб-сайта на языке программирования Perl и соответствует большинству предъявляемым требованиям. Она обеспечивает работу пользователей трех уровней – администратора, преподавателя и студента, которым доступны функции:

- работа с личными данными – изменение пароля, электронного адреса;
- работа с предметами – список дисциплин, ресурсов и тестов;
- работа с преподавателями – фамилии, списки предметов;
- работа со списком групп – название, списки предметов;
- работа с группами – список студентов своей группы;
- работа с сообщениями – дистанционное общение (внутренняя почта).

В процессе своей деятельности система обеспечивает ведение следующих файлов данных:

- entry.dat – информация о посетителях системы (дата посещения сайта, ФИО, группа, уровень доступа, IP-адрес);
- messages.dat – информация о сообщениях, которыми обмениваются пользователи системы;
- testlist.dat – информация о перечне тестов;
- testresult.dat – информация о прохождении тестирования пользователями системы.

Накопленные системой данные могут быть подвергнуты интеллектуальному анализу.

Всех пользователей в зависимости от их активности необходимо разбить на группы, причем число групп может изменяться администратором. Таким образом, имеем задачу кластеризации, которую целесообразно решать итеративным методом k-средних [1]. К показателям активности

пользователей можно отнести, во-первых, сам факт входа в систему и вытекающие из него действия, во-вторых, пользование модулем обмена сообщениями и, в-третьих, прохождение тестирования. Анализ файлов данных привел к выводу 12-ти факторов, характеризующих активность пользователя:

1. сколько раз человек вообще был на сайте – число посещений
2. частота посещений в день
3. сколько дней прошло с последнего визита
4. сколько разных ip-адресов было для входа
5. сколько всего сообщений отправлено
6. сколько сообщений отправлено другим пользователям своего уровня
7. сколько сообщений отправлено другим пользователям уровня выше
8. сколько сообщений отправлено другим пользователям уровня ниже
9. число сообщений с приложениями
10. сколько всего сообщений получено
11. сколько всего проходил тестирований
12. среднее время, которое требовалось для ответа на один вопрос теста

Число сообщений, которые содержат приложения, берутся отдельно из-за того, что в приложениях содержатся решенные задачи, а этот показатель имеет отдельный вес для построения кластеров.

Еще одна задача, которая может быть здесь поставлена – прогнозирование. Исходя из имеющихся данных об активности пользователей, можно аппроксимировать эти данные для возможности получения прогноза будущей активности пользователей. В качестве системной единицы лучше взять не отдельного пользователя, а группу в целом. Данными будут сумма действий пользователей данной группы за сутки. К действиям можно отнести факт авторизации в системе (файл «entry.dat») и факт отправки какого-либо сообщения другому пользователю (файл «messages.dat»). Поскольку прохождение тестирования связано, в первую очередь, с учебным планом, а не с желанием пользователя, факт прохождения теста здесь не учитывается. Самый простой метод для решения задачи – метод наименьших квадратов.

Для проектирования модуля интеллектуального анализа данных используем нотацию языка UML [3-

4]. Диаграмма вариантов использования, отображающая работу администратора (предполагаем, что студенту данный раздел сайта будет недоступен), представлена на рис. 1. Структурно-логическая схема нашего модуля представлена в виде диаграммы классов на рис. 2. Файлы исходных данных изображены в виде классов условно – они представляют собой исключительно файлы данных и связаны с управляющим классом ассоциативно.

Реализация спроектированного модуля осуществлялась внедрением в уже имеющийся файл нового блока программного кода. В главное меню системы был добавлен новый пункт «Анализ» (рис. 3), который, в свою очередь, предоставляет ряд возможностей (рис. 4).

Результаты. Пользователь может выбрать следующее:

– «Посетители» – вывести общую информацию о посещениях системы с возможностью фильтрации;

– «Статистика посещений» – вывести информацию о посещениях каждого пользователя системы, по умолчанию сортировка осуществляется по полю «пользователь»;

– «Статистика сообщений» – вывести информацию о сообщениях, отправленных и принятым каждым активным пользователем системы;

– «Кластеризация пользователей» – объединить пользователей системы в несколько групп в зависимости от их активности;

– «Активность групп» – выявить закономерности в действиях групп и попытаться спрогнозировать этот фактор. Практическое применение данного раздела будет возможно после накопления достаточного объема статистических данных.

Еще одно направление работы – анализ результатов тестирования, который позволяет работать со списками успеваемости (включая показатели качества), с данными (кластерный анализ – разбиение студентов по группам успеваемости – лучшие, средние, худшие), с рейтингом по результатам тестирования

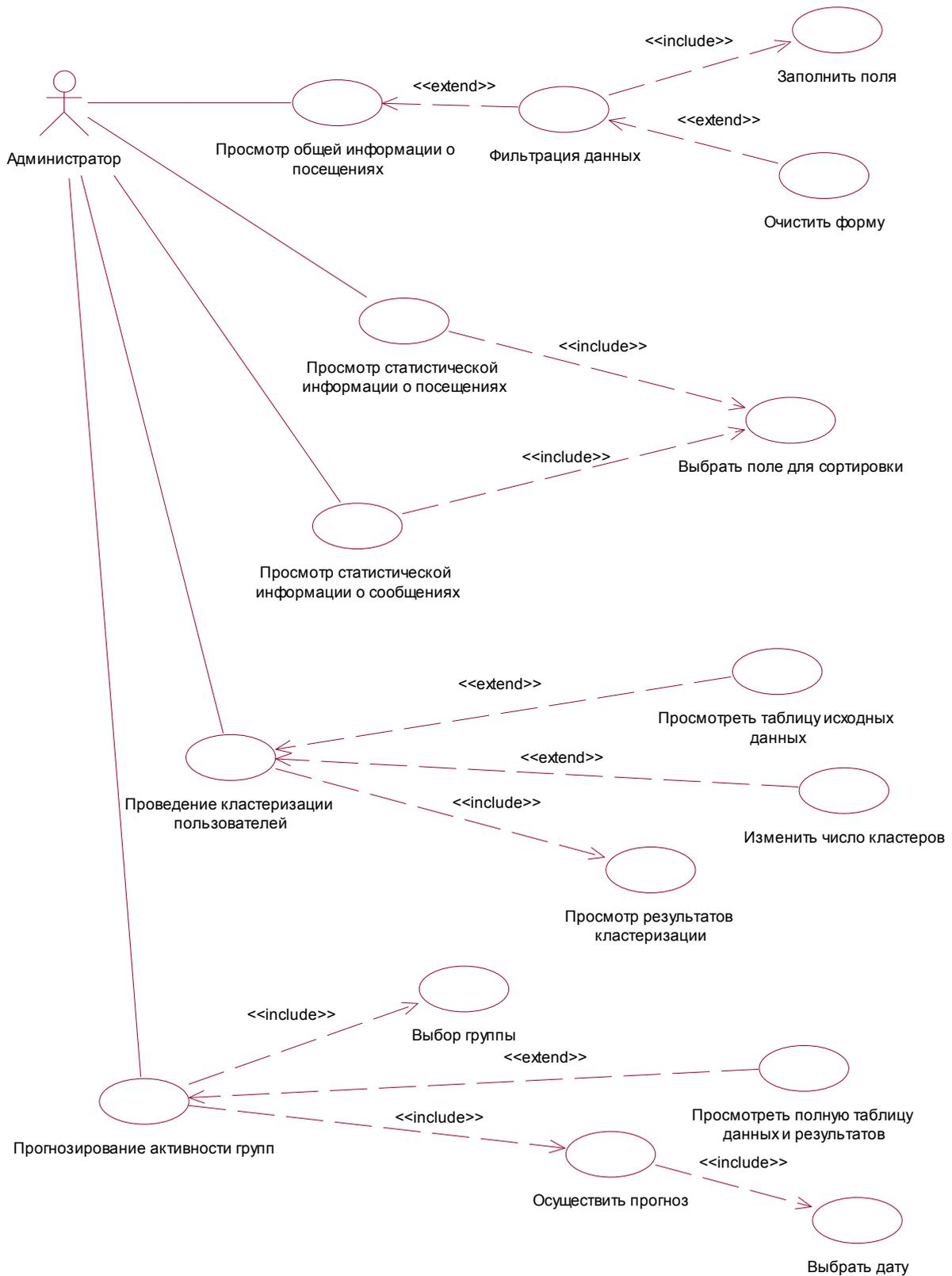


Рис. 1. Диаграмма вариантов использования

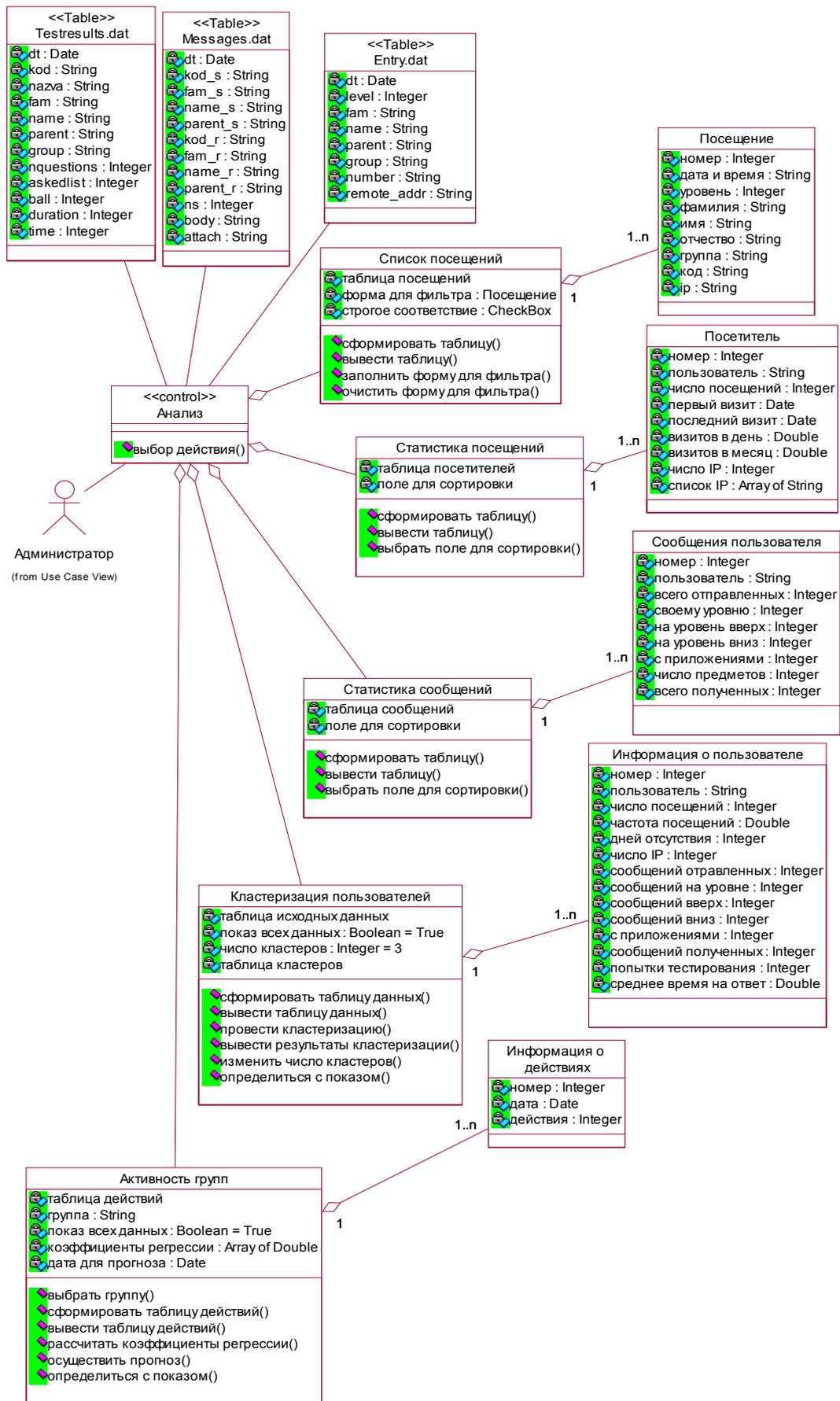


Рис. 2. Диаграмма классов

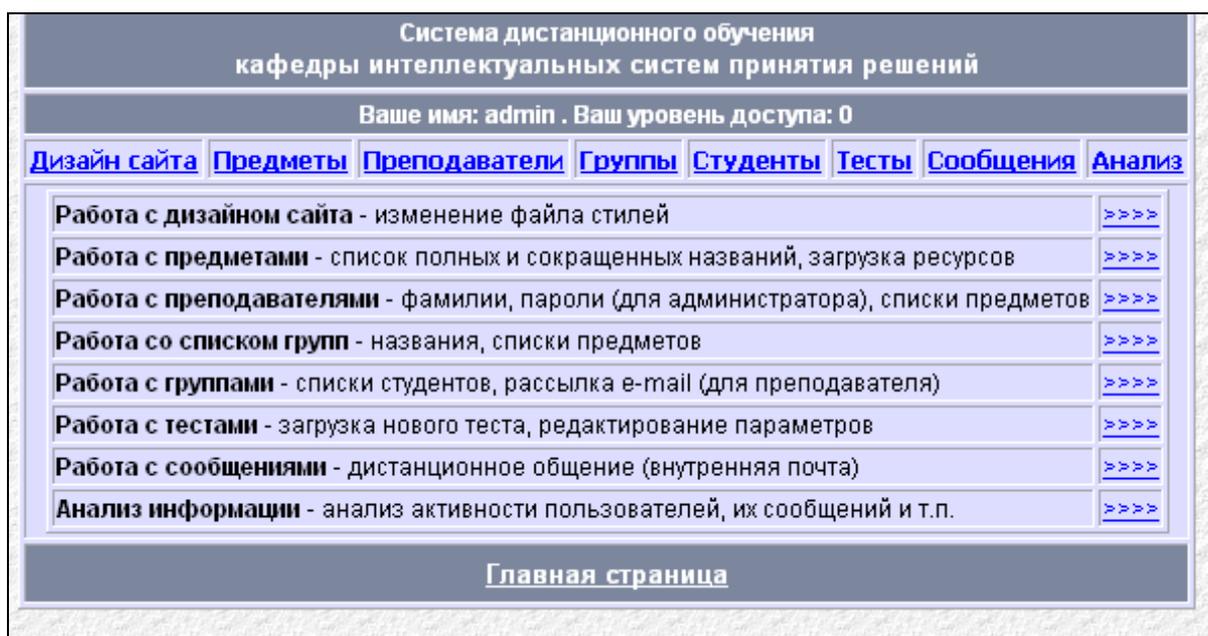


Рис. 3. Измененное главное меню системы

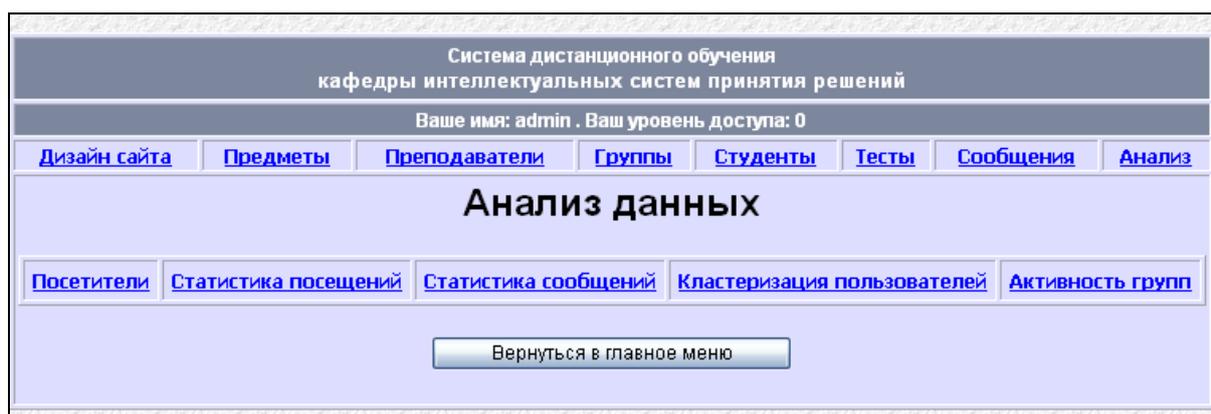


Рис. 4. Меню анализа данных

Выводы. Таким образом, был спроектирован и реализован модуль интеллектуального анализа данных для системы дистанционного обучения, позволяющий решать задачи кластеризации и прогнозирования при анализе активности пользователей и анализе результатов тестирования. Созданный модуль в настоящее время функционирует в системе дистанционного обучения кафедры интеллектуальных систем принятия решений ДГМА.

Л и т е р а т у р а

1. Чубукова И.А. Data Mining: Учебное пособие / И. А. Чубукова. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2006. – 382 с.
2. Мельников А.Ю. Информационная система для организации работы студентов заочной формы обучения с использованием интернет-технологий / Мельников А.Ю., Кушир Ю.В. // Восточно-

Европейский журнал передовых технологий. – Харьков, 2010. – № 3/4 (45). – С. 25 – 29.

3. Мельников А. Ю. Объектно-ориентированный анализ и проектирование информационных систем: учебное пособие / А. Ю. Мельников. – Изд. 2-е, перераб. и доп. – Краматорск: ДГМА, 2012. – 172 с.
4. Мельников А. Ю. Проектирование подсистемы интеллектуального анализа данных для системы дистанционного обучения / Мельников А. Ю., Кравченко К. В., Петренко А. А. // Качество образования: управление, сертификация, признание : сборник научных работ международной научно-методической конференции, 20–22 ноября 2013 года, г. Краматорск / под общ. ред. С. В. Ковалевского, д-ра техн. наук., проф. – Краматорск : ДГМА, 2013. – С. 99 – 101.

R e f e r e n c e s

1. Chubukova I. A. Data Mining: Textbook / I. A. Chubukova. – M.: Internet University of information technologies; BINOM. Knowledge Laboratory, 2006. - 382.

2. Melnikov A. Y. Information system for the organization of the students of correspondence courses using Internet technologies / Melnikov A.Y., Kushnir Y. V. // East European Journal of advanced technology. - Kharkov, 2010. - № 3/4 (45). - P. 25-29.
3. Melnikov A.Y. Object-oriented analysis and design of information systems: tutorial / A.Y. Melnikov. – Izd. 2-е, pererab. i dop. – Kramatorsk: DSEA, 2012. – 172 s.
4. Melnikov A. Y. Designing data mining subsystem for distance learning system / Melnikov A. Y., Kravchenko K. V., Petrenko A. A. // Quality of education: management, certification, recognition: the digest of scientific works of the international scientific conference, 20-22 November 2013, Kramatorsk / under total. ed. S.V. Kovalevsky, dr. tehn. sciences., prof. - Kramatorsk: DSEA, 2013. - P. 99 – 101.

Мельников О. Ю., Кравченко Х. В., Петренко А. О. Розробка модуля інтелектуального аналізу даних для системи дистанційного навчання

Розглянуто проблему інтелектуального аналізу даних системи дистанційного навчання. Спроектовано на мові моделювання UML підсистему, яка містить модуль аналізу активності користувачів і модуль аналізу результатів тестування. Наведено опис програмної реалізації системи.

Ключові слова: інтелектуальний аналіз даних, дистанційне навчання, інформаційна модель, уніфікована мова моделювання, UML, веб-сайт.

Melnikov A. Y., Kravchenko K. V., Petrenko A. A. The development of data mining module at distance learning system

Distance education as a form of organization of educational process based on the use of advanced telecommunications and information technology, as well as on the principles of individual learning and self-control is described. The problem of data mining of distance learning system is considered. A functioning of one of the existing systems of distance learning, which provides the user three levels – administrator, teacher and student, is described. The available functions are lists. The data files required for operation of the system are lists too. The 12 factors that characterize the activity of the user are identified and listed. The possibility of solving the problem of predicting future user activity is considered. The information model of new system on unified modeling language UML is described. This system contains the user activity analysis module and the testing results' analysis module. The program implementation of created model is described.

Keywords: data mining, distance learning, information model, Unified Modeling Language, web-site.

Мельников Александр Юрійович, к.т.н., доцент, доцент кафедри інтелектуальних систем прийняття рішень Донбаської державної машинобудівної академії (ДДМА), al_mel@mail.ru

Кравченко Христина Вікторівна, студентка спеціальності «Системи і методи прийняття рішень» Донбаської державної машинобудівної академії

Петренко Артем Олександрович, студент спеціальності «Системи і методи прийняття рішень» Донбаської державної машинобудівної академії

Рецензент: **Данич В.М.**, д.т.н., професор.

Стаття подана 03.03.14