

ПРОБЛЕМЫ ВЫСШЕГО ОБРАЗОВАНИЯ

УДК 004.416.6:004.051:004.055

Методика автоматизированной обработки зачетно-экзаменационных ведомостей в полнофункциональной ИУС «Деканат»

В.К.Григорьев, А.В.Грушин

**Московский государственный институт радиотехники, электроники и автоматики
(технический университет)**

Рассмотрены проблемы компьютерной поддержки управления учебным процессом на примере авторской полнофункциональной информационно-управляющей системы «Деканат» МИРЭА. Выделен процесс выпуска и приема зачетно-экзаменационных ведомостей. Предлагается модель структуры ведомости и рассматривается методика распознавания оценок на основе штрихкодовой идентификации ведомостей. Приведены результаты экспериментального исследования методики распознавания.

Анализ результатов информатизации вузов показывает, что используемые системы недостаточно охватывают учебную деятельность вузов [1]. Необходимость учета особенностей вуза при автоматизации управления учебным процессом породила ряд разработок в этой области [2, 3].

Одним из важнейших узлов управления учебным процессом является деканат, который должен отслеживать информационный поток «Студенты» и на его основе управлять переводом, отчислением студентов, начислением стипендии путем выпуска приказов различных типов, а также проводить мониторинг успеваемости путем выпуска и учета зачетно-экзаменационных ведомостей и индивидуальных направлений на экзамен (зачетно-экзаменационных листов). Поддержку всех информационных и управляющих процессов деканата осуществляет полнофункциональная информационно-управляющая система (ИУС) «Деканат» МИРЭА [4].

Полнофункциональная ИУС «Деканат» содержит пять логических слоев: *база данных, интерфейсный слой взаимодействия между базой данных и системой, функциональные модули*, которые через *слой преобразования на основе разграничения полномочий* организуют *автоматизированные рабочие места пользователей*. Система имеет двухуровневую клиент-серверную архитектуру «толстый клиент/сервер» (рис.1) и реализована инструментальными средствами Delphi7 и Interbase7.1. Система зарегистрирована в фонде алгоритмов и программ [5]. Наиболее критичным периодом работы деканата с точки зрения появления большого объема новой информации, принятия на ее основе управленческих решений является сессия (табл.1). В ИУС «Деканат» предложены и реализованы решения для компьютерной поддержки технологических процессов, составляющих сессию.



Рис. 1. Структурная схема ИУС «Деканат»

Таблица 1

Основные информационные потоки деканата

Название потока	Объем потока	Период
Приказ о зачислении	300 человек	Перед началом учебного года (10 дней)
Мониторинг успеваемости	$N_{оценок} = N_{студентов} \cdot (N_{экзаменов} + N_{зачетов} + N_{перезэкзаменовок})$. Около 1700 человек, 5 экзаменов, 10 зачетов, 0,4 перезэкзаменовки (более 27500 оценок)	Сессия (три недели)
Перевод с курса на курс	1700 человек	По окончании летней сессии (10 дней)
Отчисление в связи с окончанием	250 человек · (диплом + приложение) = 500 печатных документов	После приказа об отчислении (10 дней)

Компоненты ИУС «Деканат», поддерживающие технологические процессы сессии. Выделим основные процессы, составляющие зачетно-экзаменационную сессию:

- выпуск большого числа зачетно-экзаменационных ведомостей для студентов, допущенных к сдаче экзаменов, выпуск ведомостей для перезэкзаменовок;
- регистрация оценок, проставленных в зачетно-экзаменационных ведомостях;
- прием студентов в деканате, анализ актуальной информации о студенте;
- выдача индивидуальных зачетно-экзаменационных листов для сдачи и пересдачи зачетов и экзаменов;
- регистрация оценок, проставленных в зачетно-экзаменационных листах.

В ИУС «Деканат» для поддержки процессов выпуска ведомостей и регистрации оценок с точки зрения их автоматизации используются следующие механизмы.

Модель зачетно-экзаменационной ведомости. Построим инфологическую модель зачетно-экзаменационной ведомости:

$$Mo = (Header, SL, RL, ES, EN), \tag{1}$$

где Header – заголовок зачетно-экзаменационной ведомости; SL – список студентов группы; RL – список оценок и дат их проставления (рукописный текст); ES – список подписей преподавателя (рукописный текст); EN – ФИО преподавателя (рукописный текст).

Из (1) видно, что все информационные поля заполняются преподавателем в виде рукописного текста и для выполнения автоматического ввода информации необходимо осуществлять его распознавание. Данное решение крайне сложно в реализации и обладает низкой надежностью. Для решения задачи распознавания оценок в зачетно-экзаменационной ведомости наиболее целесообразно введение дополнительных полей для простой фиксации оценки и маркировка ведомости штрих-кодом для создания ее эталона и установления соответствия между эталоном и выпускаемой ведомостью [6]. С этой целью в модель были введены поля для машинной обработки. Была построена новая инфологическая модель ведомости (рис.2), которая представляет собой расширение (1):

$$Md = (\text{Header}, \text{BC}, \text{SL}, \text{RL}, \text{RAL}, \text{ES}, \text{EN}, \text{KMD}, \text{KMM}), \quad (2)$$

где **BC** – штрих-код; **RAL** – поле оценок для машинной обработки; **KMD** – день проведения контрольного мероприятия (КМ), поле для машинной обработки; **KMM** – месяц проведения КМ, поле для машинной обработки.

Опорные метки:

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ИНСТИТУТ РАДИОТЕХНИКИ,
ЭЛЕКТРОНИКИ И АВТОМАТИКИ (ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ)

Header ЗАЧЕТНО-ЭКЗАМЕНАЦИОННАЯ ВЕДОМОСТЬ

по дисциплине Математический анализ (Экзамен)

факкультет ВМС группа ВБ-1-06

2007/2008 уч.год 3 семестр Экзаменатор

BC **EN**

№	Ф.И.О. студента	Шифр	Дата	Оценка	для машинной обработки						Подпись экзаменатора
					н.я.	зач	неуд	удо	хор	отл	
1	Валиев Т.С.	061001	19.10.07	хорошо					✓		С.С.
2	Дормаков Е.Н.	061003	19.10.07	отлично						✓	Е.Н.
3	Доронкина А.В.	061004	19.10.07	отлично						✓	А.В.

SL **RL** Экзаменатор **RAL** **ES**

Дата зачета / экзамена:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
																			×												

Декан

1	2	3	4	5	6	7	8	9	10	11	12
											×

KMM **KMD**

Опорные метки:

Рис.2. Модифицированная форма ведомости с обозначением полей модели Md

Структура новых элементов модели (2) содержит следующие элементы.

Штрихкод – предназначен для автоматической идентификации зачетно-экзаменационных ведомостей при приеме. Наносится на ведомости при печати.

Поле оценок для машинной обработки – представляет собой матрицу размерности [SL,6]. Каждый элемент может принимать одно и только одно значение из набора: не-явка, зачет, незачет/неуд, удовлетворительно, хорошо, отлично.

Дата проведения контрольного мероприятия (день и месяц проведения КМ) – представляет собой два вектора размерности [31] и [12] соответственно. Заполненным должно быть одно и только одно значение в каждом из векторов.

На основе (2) была модифицирована структура базы данных (БД) системы по работе с зачетно-экзаменационными ведомостями (рис.3). В БД были добавлены две таблицы: «Активные ведомости» и «Список студентов ведомости». В первой таблице сохраняется ссылка на проводимое КМ (ID контрольного мероприятия, где содержится

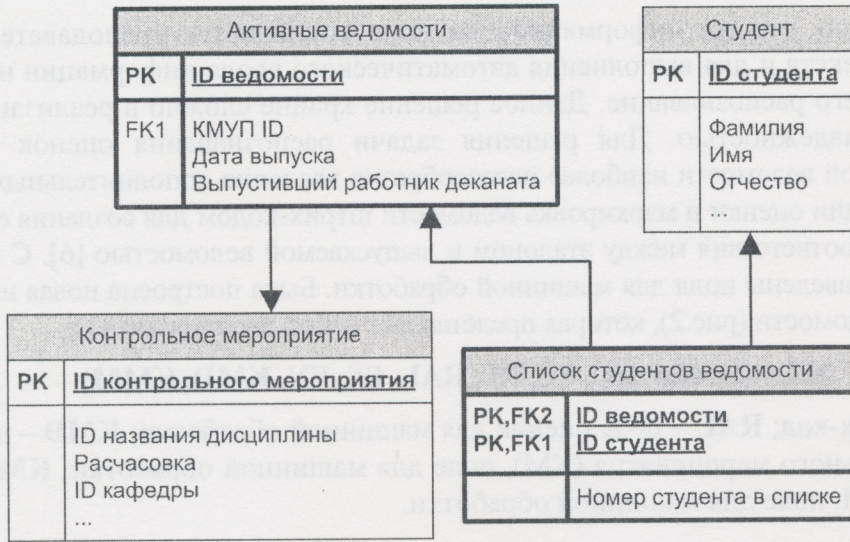


Рис.3. Фрагмент БД по работе с ведомостями

ссылка на номер семестра). Ключ «ID ведомости» из таблицы «Активные ведомости» заносится в штрихкод выпускаемой ведомости и используется для ее идентификации при сканировании. Таблица «Список студентов ведомости» содержит ссылки на студентов ведомости, включая порядковый номер каждого студента в списке. Данные из таблицы «Список студентов ведомости» используются для установления соответствия между распознанной оценкой ведомости и конкретным студентом группы.

В системе «Деканат» работа с зачетно-экзаменационными ведомостями разделена на этапы в следующей последовательности: 1) выпуск ведомостей; 2) заполнение ведомостей преподавателями; 3) сканирование ведомостей, распознавание оценок; 4) контроль и сохранение результатов. При выпуске ведомости происходит печать и сохранение информации о ней в БД (рис.4, А3). В описанные выше таблицы БД заносятся данные о КМ и списочном составе, а на лист ведомости впечатываются поля Header, BC и SL.

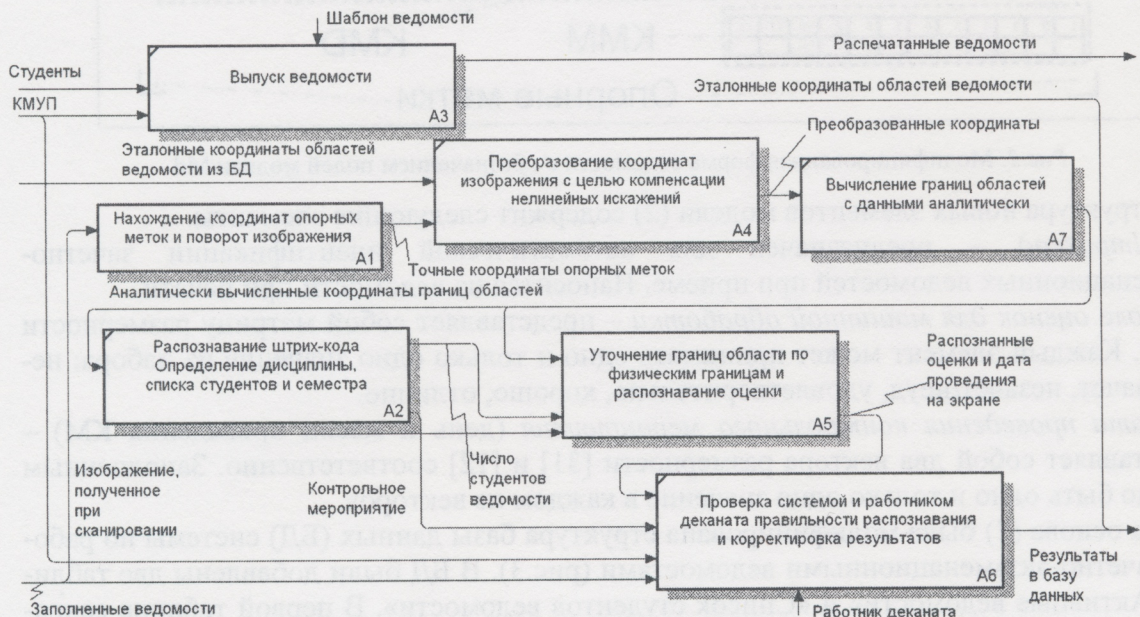


Рис.4. Функциональная модель ИУС «Деканат» по работе с ведомостями

При заполнении ведомостей преподаватель помимо заполнения стандартных полей RL, ES, EN заполняет также поля для машинной обработки RAL, KMD и KMM (см. рис.2).

Ввод оценок из зачетно-экзаменационных ведомостей в БД системы «Деканат» производится сканированием и последующим автоматическим распознаванием (см. рис.4). По окончании процесса распознавания на экране появляется вся информация по ведомости: список студентов группы, оценки, дата проведения КМ (рис.5).

					н.я.	зач.	неуд.	удо.	хор.	отл.
1	Горушкин А.В.	071002								✓
2	Гуреев В.М.	071004								
3	Гусаров Г.А.	071326							✓	
4	Денисов С.В.	071005							✓	
5	Дударевич А.Г.	071006							✓	
6	Львов И.Э.	071237								
7	Метелкин С.А.	071010							✓	
8	Минаев В.С.	071011								✓
9	Осокин Д.В.	071012								✓
10	Семичев И.Д.	071015								
11	Терентьев Ф.А.	071238							✓	
12	Чистяков П.С.	071017								

а

б

Рис. 5. Печатный вариант ведомости (а), интерфейс приема ведомостей (б)

Распознавание зачетно-экзаменационных ведомостей. Алгоритм распознавания ИУС «Деканат» МИРЭА построен на основе экстенционального метода «голосования», который сводится к математическому расчету координат информационных областей с последующей коррекцией по реальным границам областей. Для определения априорных координат информационных полей используется принцип расчета по четырем опорным меткам, расположенным по углам листа (см. рис.2).

Распознавание зачетно-экзаменационной ведомости начинается с определения координат опорных меток (рис.4, А1). Для надежного поиска опорных меток применен «рекурсивный алгоритм заполнения с затравкой» и последующий сравнительный анализ найденных областей с критериями эталонной опорной метки. Принцип его работы состоит в следующем: на основе фрагмента изображения, где производится поиск опорной метки, строится двумерный массив элементов типа boolean. Черная точка отображается как «Истина», белая – как «Ложь». Поиск проводится построчно, начиная сверху, слева направо. Ищем первый попавшийся элемент «Истина». Как только этот элемент найден, включается алгоритм заполнения, который рекурсивным образом «за-

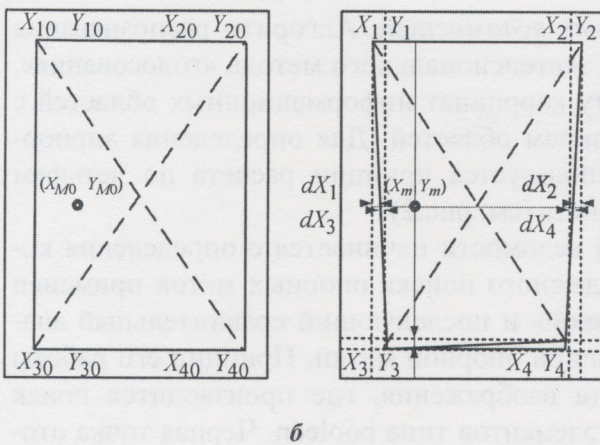
ливает» найденную и все прилегающие к ней элементы «Истина» значением «Ложь». Во время заполнения вычисляются следующие значения найденной связной области: длина, ширина, X и Y описывающего ее прямоугольника, а также вес данной области (число элементов «Истина»). После окончания работы алгоритма поиска получаем массив связных областей, из которого по заданному критерию выбираем координаты искомой опорной метки. Как показали результаты экспериментальных исследований, параметры описывающего прямоугольника и вес дают возможность устойчивой идентификации опорных меток и отделение их от других надписей ведомости. Преимущество алгоритма поиска, в частности, состоит в том, что посторонние записи на ведомости не оказывают влияния на точность нахождения положения опорных меток. Отметим, что эти записи не должны «наезжать» на опорные метки, т.е. не должны иметь с ними общих и примыкающих точек.

После нахождения опорных меток выполняется процедура определения центра изображения и производится поворот относительно него на средний угол, вычисленный по координатам меток по формуле

$$\varphi = \frac{(a+b+c+d)}{4},$$

где a, b, c, d – углы между прямыми, проведенными через опорные метки и соответствующими осями (рис.6,а).

Далее выполняется корректировка линейных искажений изображения ведомости, которые появляются как при печати, так и во время сканирования.



Пусть X_1, X_2, X_3, X_4 – найденная координата X опорной метки соответствующего номера; Y_1, Y_2, Y_3, Y_4 – найденная координата Y опорной метки соответствующего номера; (X_{m0}, Y_{m0}) – координата искомой точки на эталоне; dX_L, dX_R – значение коэффициентов деформации по левой и правой стороне изображения; dY_T, dY_B – значение коэффициентов деформации по верхней и нижней стороне изображения.

Определяем коэффициенты деформации ведомости:

$$\begin{aligned} dX_L &= ((X_3 - X_1) / (Y_1 - Y_3)) \cdot Y_{m0}, \\ dX_R &= ((X_4 - X_2) / (Y_4 - Y_2)) \cdot Y_{m0}, \\ dY_T &= ((Y_2 - Y_1) / (X_2 - X_1)) \cdot X_{m0}, \\ dY_B &= ((Y_4 - Y_3) / (X_4 - X_3)) \cdot X_{m0}. \end{aligned} \quad (3)$$

Исходя из (3) получаем общие коэффициенты

$$\begin{aligned} dx &= dX_L \cdot (X_2 - X_{m0}) + dX_R \cdot (X_{m0} - X_1), \\ dy &= dY_T \cdot (Y_4 - Y_{m0}) + dY_B \cdot (Y_{m0} - Y_2). \end{aligned} \quad (4)$$

Отсюда получаем формулу преобразования координат эталонной ведомости в координаты информационных полей отсканированного изображения (рис.6,б):

Рис.6. Линейные преобразования отсканированного изображения ведомости: а – поворот изображения; б – компенсация линейных искажений

$$X_m = X_{m0} + dx, Y_m = Y_{m0} + dy \quad (5)$$

Можно рассматривать искажения по вертикали и горизонтали (5) независимо, так как погрешность их влияния друг на друга имеет более высокий порядок малости.

Далее распознается штрихкод (см. рис.4, A2). Считывание штрихкода в стандарте Code128C происходит из общего отсканированного изображения ведомости. Наличие в штрихкоде контрольной свертки делает идентификацию с помощью штрихкодов очень удобной и надежной (99,97%). Распознавание штрихкода реализуется в соответствии с описанием стандарта [7] в несколько итераций:

1) для считывания штрихкода проводится сканирующая прямая через область штрихкода;

2) в случае успешного распознавания штрихкода значение возвращается;

3) в случае возникновения ошибки при распознавании проводится сканирующая прямая через соседнее место штрихкода со смещением на несколько точек;

4) при повторении ошибки (п. 3) 25 раз (определяется высотой штрихкода на ведомости) возвращается сообщение о невозможности чтения штрихкода.

При исследовании работы алгоритма распознавания оценок в зачетно-экзаменационных ведомостях в ходе экспериментов выяснилось, что существуют такие нелинейные искажения бланка зачетно-экзаменационной ведомости при печати, при которых вычисленные аналитически области сильно смещены относительно реальных, несмотря на правильно определенные координаты опорных меток. Это происходит, например, при дефекте в протяжном устройстве принтера, когда в одной части изображение сжато, а в другой растянуто. Поэтому было добавлено уточнение вычисленных априорных координат по изображениям границ областей путем голосования (см. рис.4, A5). Принцип его заключается в следующем: вместе с аналитически рассчитанной областью задается область погрешности, внутри которой путем анализа изображения проводится поиск реальной границы области. Если граница найдена в заданной области, уточняется ее значение, если нет – координата остается рассчитанной аналитически. На рис.7 показана ситуация, где в четырех случаях из пяти проводится уточнение по реальным границам информационной области оценок и в одном случае из двух координат выбирается аналитически рассчитанная (4-й пик на графике), так как уточненная координата выходит за пределы погрешности вычисленной координаты. Происходит это из-за присутствия посторонних записей в области полей для машинной обработки, которые мешают правильной коррекции.

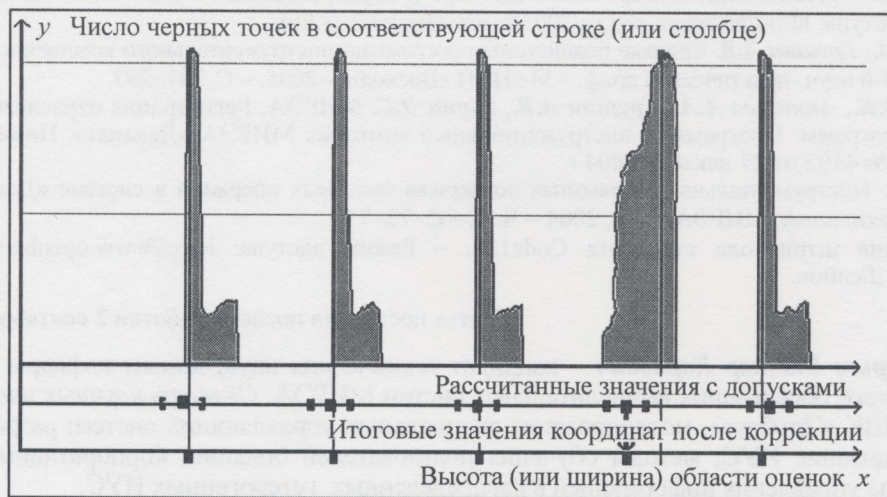


Рис.7. Уточнение аналитически рассчитанных границ информационных областей ведомости для компенсации нелинейных искажений

После того как все информационные области найдены, проводится расчет мощности множества черных точек в каждой из них и сравнение с критериями эталонной ведомости. Порог чувствительности установлен на уровне 5% от общего числа точек области. Заполнение на меньшую величину воспринимается как шум и игнорируется. После установления заполненных областей проводятся сравнения с критериями, установленными в модели (2), «сборка» даты по дню и месяцу и результаты выводятся на экран пользователя.

Результаты работы системы распознавания оценок в зачетно-экзаменационных ведомостях за летнюю сессию 2006/2007 учебного года представлены в табл.2.

Таблица 2

Результаты работы подсистемы распознавания оценок

Деканат	Зачетно-экзаменационные ведомости			
	Всего принято	Общее число ошибок	Неправильное заполнение ведомостей	Ошибки распознавания
ВМС	1401	98 (7%)	87 (6,2%)	11 (0,8%)
Кибернетика	776	50 (6,4%)	44 (5,64%)	6 (0,76%)
РТС	2206	190 (8,5%)	169 (7,55%)	21 (0,95%)

Анализ ошибок, возникающих при приеме ведомостей, показал, что наибольшее количество ошибок появляется вследствие неправильного заполнения ведомостей (незаполненные поля, заполнение более одной позиции в строке, посторонние записи в областях для машинной обработки). Среди ошибок распознавания наиболее частой ошибкой является определение менее четырех опорных меток вследствие неровного положения ведомости в сканере таким образом, что часть ее оказывается вне области сканирования.

Литература

1. *Голосов А., Полотнюк И.* Реформа образования и информатизация в вузах // Открытые системы. – № 1. – 2007. – Режим доступа: www.osp.ru/os/2007/01/3999187/p2.html.
2. *Стасышин В.М.* Информатизация учебного процесса и система «Деканат» // ТУ ИНФОРМ. – 2005. – № 2 (139). – Режим доступа: <http://inform.nstu.ru/index.php?type=inform&statat=228>.
3. *Горелик Е.П., Петров А.О., Рубцов Е.Б.* Система АСУ «Деканат» для автоматизации управления учебным процессом. – Московский авиационный институт (Государственный технический университет) Россия. – Режим доступа: <http://nit.miem.edu.ru/2003/tezisy/articles/323.htm>.
4. *Антонов А.А., Грушин А.В.* Базовые подсистемы программно-инструментального комплекса МИРЭА «Деканат» // Сб. тр. 5-й науч.-практической конф. – М.: НИИ «Восход». – 2005. – С. 287–293.
5. *Григорьев В.К., Антонов А.А., Грушин А.В., Зорин Д.С.* МИРЭА. Регистрация отраслевого фонда алгоритмов и программ. Программно-инструментальный комплекс МИРЭА «Деканат». Номер отраслевой регистрации № 4192 от 27 декабря 2004 г.
6. *Грушин А.В.* Инструментально-программная поддержка массовых операций в системе «Деканат» // Сб. тр. 53-й науч.-техн. конф. МИРЭА. – М., 2004 – Ч. 4. – С. 72–77.
7. Спецификация штрихкода стандарта Code128. – Режим доступа: http://www.openbarcode.org/wiki/Code_128_specification.

Статья поступила после доработки 2 сентября 2008 г.

Григорьев Виктор Карлович – кандидат технических наук, доцент кафедры математического обеспечения вычислительных систем МИРЭА. *Область научных интересов:* ОС МВК «Эльбрус», моделирование специальных управляющих систем; разработка и исследование ИУС; методы обучения пользователей больших корпоративных ИУС, методы управления информацией в распределенных гетерогенных ИУС.

Грушин Алексей Валерьевич – аспирант кафедры математического обеспечения вычислительных систем МИРЭА. *Область научных интересов:* компьютерное распознавание образов, построение моделей учебно-научной деятельности вуза.