

УДК 33

О программном комплексе обработки данных Всероссийского экономического диктанта

Голов Роман Сергеевич

Доктор экономических наук, профессор,
заведующий кафедрой 501,
Московский авиационный институт (национальный исследовательский университет),
125993, Российская Федерация, Москва, шоссе Волоколамское, 4;
e-mail: roman_golov@rambler.ru

Костиков Юрий Александрович

Кандидат физико-математических наук,
заведующий кафедрой 812,
Московский авиационный институт (национальный исследовательский университет),
125993, Российская Федерация, Москва, шоссе Волоколамское, 4;
e-mail: jkostikov@mail.ru

Романенков Александр Михайлович

Кандидат технических наук,
доцент кафедры 812,
Московский авиационный институт (национальный исследовательский университет),
125993, Российская Федерация, Москва, шоссе Волоколамское, 4;
e-mail: romanaleks@gmail.com

Аннотация

Статья посвящена исследованию особенностей методологии разработки и применения программного комплекса для автоматизированной обработки результатов Всероссийского экономического диктанта. Авторы описывают структуру информационно-аналитической системы и программно-аналитического комплекса, реализующих необходимый функционал для эффективной обработки большого объема данных. Отмечается, что важнейшими задачами программно-аналитического комплекса являются автоматизированная обработка результатов диктанта, обеспечение безбумажного документооборота, представление результатов в аналитической и наглядной графической форме. Программные решения, предлагаемые в данной статье, дают возможность автоматизировать процесс проверки и учета индивидуальных результатов участников диктанта с использованием алгоритмов и технологий распознавания данных и современных программных реализаций этих алгоритмов. Использование современных

информационно-коммуникационных технологий позволяет поднять на качественно новый уровень статистический анализ, социологический анализ экономической грамотности населения России, а также выработать рекомендации по совершенствованию форм и методов проведения диктанта в будущем.

Для цитирования в научных исследованиях

Голов Р.С., Костиков Ю.А., Романенков А.М. Инструментарий эффективной статистической обработки экономических данных // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 9А. С. 278-286.

Ключевые слова

Программный комплекс, обработка данных, Всероссийский экономический диктант, информационно-аналитическая система, статистический анализ.

Введение

Четвертого октября 2018 года во всех регионах Российской Федерации и трех дружественных государствах (Монголия, Приднестровская Молдавская Республика и Республика Таджикистан) Вольное экономическое общество проведет образовательную акцию «Всероссийский экономический диктант на тему "Сильная экономика – процветающая Россия!"», в которой примут участие около 100 тысяч человек. Стоит отметить, что на территории Российской Федерации данное мероприятие будет проводиться уже второй раз и, по сравнению с прошлым годом, заметно существенное увеличение числа участников [Дэйт, 2017]. Планируется, что в дальнейшем Всероссийский экономический диктант станет международным событием, в котором примет участие еще большее количество человек.

Для обеспечения информационно-коммуникационной поддержки при проведении, обработке и анализе результатов Всероссийского экономического диктанта необходима разработка информационно-аналитической системы и программно-аналитического комплекса (ПАК), опирающихся на современные информационные технологии. Важнейшими задачами ПАК являются автоматизированная обработка результатов диктанта, обеспечение безбумажного документооборота, представление результатов в аналитической и наглядной графической форме [Аналитический отчет..., www].

Информационно-аналитическая система предоставляет возможность проводить сравнительный анализ экономической грамотности различных слоев населения, выявлять ее зависимость от регионов, в которых ведут свою деятельность участники диктанта [Гамма, Хелм, Джонсон, Влиссидес, 2016]. Выделение слоев населения и регионов с пониженным уровнем экономических знаний представляется важной социально-экономической задачей, в решении которой разрабатываемая система может оказать заметную помощь.

Система предоставляет возможность быстро и качественно обрабатывать результаты Диктанта, анализировать их зависимость от таких характеристик участников, как возраст, пол, образование, род занятий, сфера деятельности и отношение к экономике.

Как система поддержки принятия решений, данная информационно-аналитическая система позволяет выявлять вопросы и группы вопросов, оказавшиеся излишне простыми или излишне сложными для тех или иных регионов и слоев населения, помогает принимать организационные и методические решения по совершенствованию содержания Диктанта, выработать рекомендации по улучшению учебных программ по экономике и организации учебного процесса.

Описание программно-аналитического комплекса обработки данных Всероссийского экономического диктанта

Всероссийский экономический диктант проходит в форме тестирования на бланках, которые после завершения диктанта сканируются и отправляются в центр обработки данных.

Главными задачами комплекса являются следующие:

1. Обработка результатов Всероссийского экономического диктанта в автоматизированном режиме.
2. Сокращение времени обработки результатов Всероссийского экономического диктанта.
3. Представление результатов Всероссийского экономического диктанта в наглядной графической и табличной форме.

Общая структурная схема программно-аналитического комплекса Всероссийского экономического диктанта (ПАК ВЭД 1.0) представлена на рисунке 1. Далее опишем кратко функционал и области ответственности программных модулей.

Модуль получения данных является стартовым в ПАК. Работа данного модуля начинается с установления соединения с провайдером электронной почты. Программа с заданной периодичностью (один раз в 5 секунд) проверяет почтовый ящик с фиксированным именем на наличие новых (непрочитанных) писем. При нахождении нового письма данные (адрес отправителя, дата и время получения) заносятся в таблицу mails, затем происходит анализ содержимого полученного письма [Прохоренок, Дронов, 2016]. После загрузки данных осуществляется проверка соответствия заданному формату. Стоит обратить внимание на то, что в ПАК реализована возможность скачивания данных с трех различных облачных сервисов (Яндекс.Диск, Облако Mail.ru, Google Drive). Каждый облачный сервис использует свой протокол передачи и шифрования данных, поэтому были разработаны индивидуальные методы загрузки данных.

Модуль получения данных реализует в себе следующий набор функций и процедур:

1. Установление соединения с провайдером электронной почты.
2. Обработка ошибок и протоколирование установления соединения с провайдером электронной почты.
3. Проверка наличия непрочитанных писем.
4. Анализ содержимого прочитанного письма и определение принадлежности ссылок к допустимым облачным сервисам.
5. Сохранение архивов с последующей распаковкой или сохранение файлов на жесткий диск.
6. Установление соединения с первичной базой данных и занесение данных в первичную базу данных.

После загрузки и валидации графические файлы отправляются в систему обмена сообщениями RabbitMQ для последующей обработки и анализа.

Модуль распознавания.

В модуль распознавания передаются графические файлы (скан-копии бланков для написания Диктанта) после работы модуля получения данных. Графическое представление бланка проверяется на наличие опорных графических объектов для подтверждения достоверности бланка. Если бланк не проходит проверку, то такие бланки сохраняются в папке «Некорректные файлы».

При успешном прохождении проверки изображения программа распознает информацию на

бланке. В результате успешного распознавания бланка для написания Диктанта модуль заносит полученную информацию в таблицу `blank_buf`. В случае ошибки распознавания нераспознанный файл сохраняется в папке «Нераспознанные файлы».

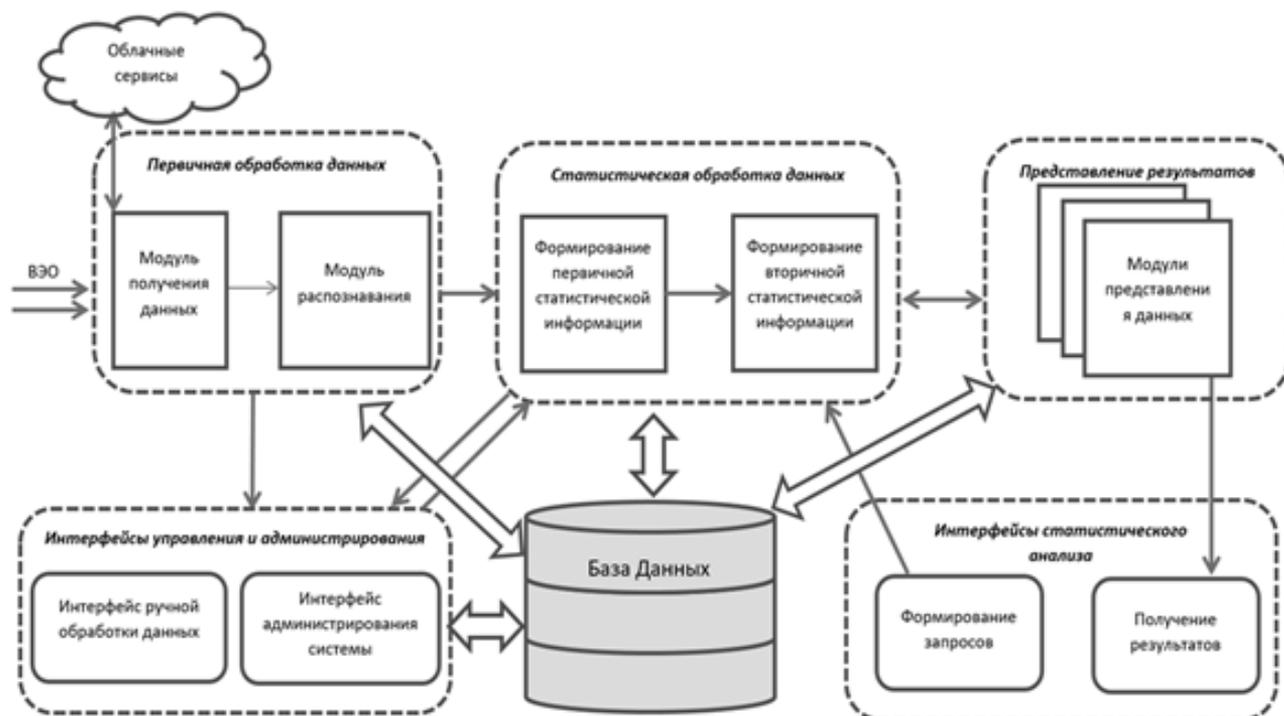


Рисунок 1- Общая структурная схема программно-аналитического комплекса

Перечислим базовый набор функции модуля распознавания данных:

1. Проверка бланка Всероссийского экономического диктанта.
2. Детектирование номера бланка и его распознавание.
3. Обработка неправильных и нераспознанных номеров бланка.
4. Распознавание ответов.
5. Проверка и протоколирование формальной правильности заполнения бланка ответов.
6. Заполнение структуры первичной информации бланка (номер бланка (номер региона, номер площадки, номер участника), таблица номеров ответов, таблица баллов за каждый вопрос, количество правильных ответов, количество набранных баллов).

База данных для хранения результатов Всероссийского экономического диктанта имеет схему, представленную на рисунке 2. Выполним описание некоторых таблиц этой базы данных.

- 1) `access_level` – таблица с полями уровня доступа к программе, которая содержит в себе названия уровней доступа, на основании которого разделяется функционал и области ответственности пользователей:
 - 0 – Администратор. Имеет доступ к полному функционалу программы. Может создавать и удалять пользователей.
 - 1 – Аналитик. Имеет доступ ко вкладке «Статистический анализ».
 - 2 – Оператор подтверждения номера бланка. Имеет доступ ко вкладке «Подтверждение № бланка».
 - 3 – Оператор-аналитик (расширенный доступ). Имеет доступ ко вкладкам «Статистический

анализ» и «Подтверждение № бланка».

- 2) blank – таблица с информацией о бланках, в которой установлено взаимно-однозначное соответствие каждого поля данной таблицы полю на бумажном бланке.
- 3) blank_buf – буферная таблица, заполняемая временными данными, которые поступают от модуля распознавания и модуля получения данных;
- 4) blank_info – таблица с информацией о бланках, заполняемая модулем распознавания и модулем получения данных;
- 5) links_cloud – таблица с информацией о ссылках на архивы с бланками, заполняемая модулем получения данных;
- 6) mails – таблица с информацией о письмах, заполняемая модулем получения данных;
- 7) result_distant – таблица с информацией о бланках с их номерами, заполняемая процедурой move;
- 8) right_answers – таблица с правильными ответами на каждый вопрос;
- 9) users – таблица с пользователями программы;
- 10) statistic1 – таблица со статистическими данными.

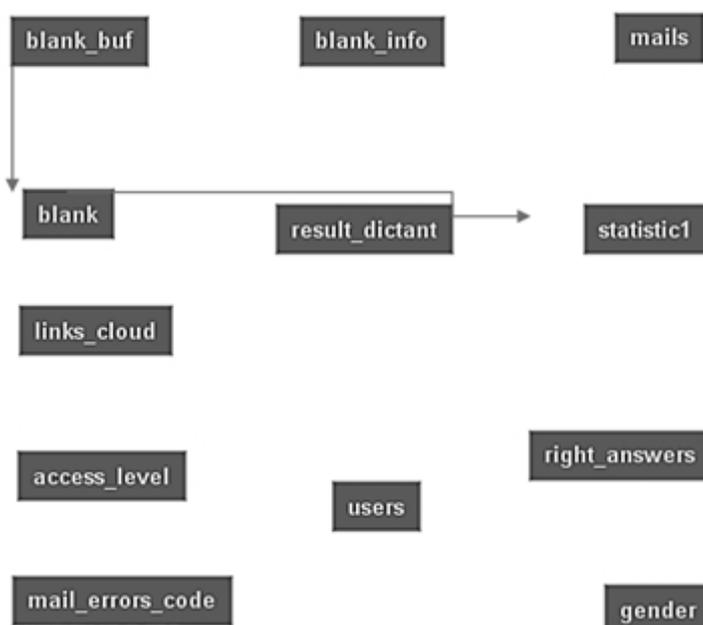


Рисунок 2 - Схема базы данных экономического диктанта (свернутая)

Модуль формирования и обработки статистических данных представляет собой формирование и обработку первичных данных, полученных после распознавания заполненных бланков для написания Всероссийского экономического диктанта. Модуль распознавания устанавливает связь с буфером первичной базы данных. В базу данных с помощью встроенных процедур на языке MySQL заносятся данные заполненных бланков в буфер.

Таблица blank_buf состоит из полей, в которых хранятся анкетные данные участников Всероссийского экономического диктанта, а также ответы участников на каждый вопрос. Из таблицы формируется первичная статистическая таблица statistic 1. Таблица объединяет данные всех участников, имеющих одинаковые характеристики (номер региона, номер площадки, количество участников мужского пола, количество участников женского пола, количество участников по возрасту, количество участников по полям (образование, сфера деятельности, род

занятий, вуз, отношение к экономике)), таблицу распределения номеров ответов, таблицу распределения баллов по каждому вопросу, количество набранных баллов и содержит количество участников, имеющих одинаковые характеристики.

Таблица blank_buf устанавливает связь с первичной статистической таблицей statistic 1, и происходит формирование статистических таблиц. Формирование осуществляется только по статистически значимым точкам.

Используя встроенные процедуры MySQL, программа создает отчет на основе стандартных запросов пользователя, также основываясь на данных в таблицах базы данных информационно-аналитической системы, аналитик может запросить распределения.

Модуль представления результатов формирует наглядные и удобные для пользователя графические представления статистической информации, полученной из базы данных информационно-аналитической системы Всероссийского экономического диктанта. Во вложенной вкладке «Выбор распределения и диаграммы» пользователю необходимо выбрать распределение из приведенного списка.

В блоке «Выбор диаграммы» пользователь может выбрать вид графического представления.

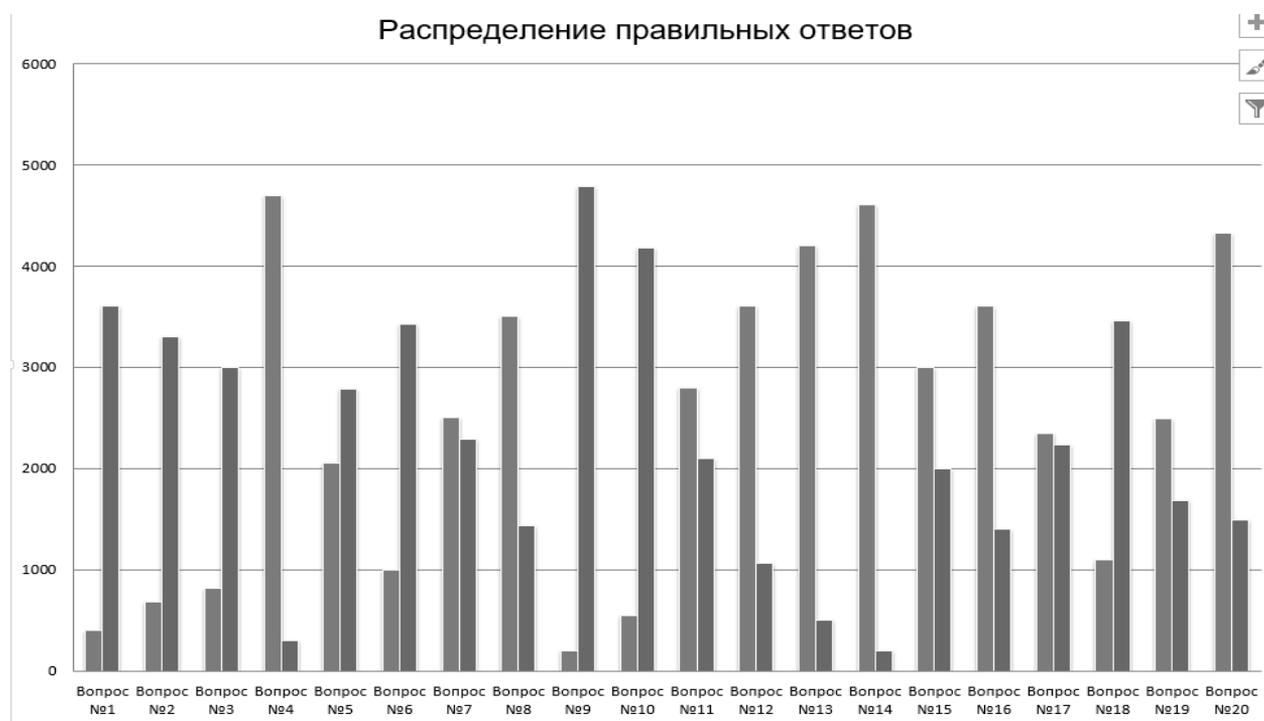


Рисунок 3 - Пример столбчатой гистограммы

Интерфейс получения результатов предназначен для отображения результатов статистического анализа в Microsoft Excel. После выбора нужного вида диаграммы (пример вывода диаграммы представлен на рисунке 3) необходимо нажать кнопку «Вывести данные в Excel». Примечание: после нажатия кнопки «Вывести данные в Excel» программа формирует SQL-запрос и отправляет его в базу данных; данные выгружаются; создается и открывается Excel-файл; данные, полученные из базы данных, записываются в файл; строится диаграмма по полученным данным, и файл сохраняется. Форматирование диаграммы (масштабирование, изменение проекции и угла наклона и т.д.) возможно с помощью стандартных операций в Excel после создания графика в файле.

Заключение

Информационно-аналитическая система Всероссийского экономического диктанта значительно повышает эффективность обработки результатов экономического диктанта, сокращает время проверки, повышает удобство и скорость работы на первоначальном этапе сбора и внесения индивидуальных ответов участников диктанта, минимизирует непосредственную работу с бумажными носителями информации. Использование современных информационно-коммуникационных технологий дает возможность поднять на качественно новый уровень статистический анализ, социологический анализ экономической грамотности населения России. Как система поддержки принятия решений, данная информационно-аналитическая система позволяет вырабатывать рекомендации по совершенствованию форм и методов проведения диктанта в будущем. Наглядное интуитивно понятное представление результатов обеспечивает повышение эффективности обработки данных.

Библиография

1. Аналитический отчет по итогам Общероссийской образовательной акции «Всероссийский экономический диктант». URL: <https://clck.ru/DaKdu>
2. Боровков А.А. Математическая статистика. М.: Физматлит, 2007. 703 с.
3. Вентцель Е.С. Теория вероятностей и ее инженерные приложения. М.: Высшая школа, 2007. 491 с.
4. Гамма Э., Хелм Р., Джонсон Р., Влиссидес Дж. Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2016. 366 с.
5. Дэйт К.Дж. Введение в системы баз данных. М.: Вильямс, 2017. 1316 с.
6. Костиков Ю.А., Павлов В.Ю., Романенков А.М., Терновсков В.Б. Адаптивная архитектура программно-аппаратного комплекса хранения и обработки данных // Экономика: вчера, сегодня, завтра. 2017. Том 7. № 9А. С. 192-207.
7. Прохоренко Н., Дронов В. Python 3 и PyQt 5. Разработка приложений. СПб.: БХВ-Петербург, 2016. 832 с.
8. Тюрин Ю.Н. Теория вероятностей и статистика. М.: МЦНМО, 2009. 256 с.

On the software package for processing of the data of the All-Russian economic dictation

Roman S. Golov

Doctor of Economics, Professor,
Head of the Department 501,
Moscow Aviation Institute (National Research University),
125993, 4 Volokolamskoe highway, Moscow, Russian Federation;
e-mail: roman_golov@rambler.ru

Yurii A. Kostikov

PhD in Physical and Mathematical Sciences,
Head of the Department 812,
Moscow Aviation Institute (National Research University),
125993, 4 Volokolamskoe highway, Moscow, Russian Federation;
e-mail: jkostikov@mail.ru

Aleksandr M. Romanenkov

PhD in Technical Sciences, Associate Professor,
Department 812,
Moscow Aviation Institute (National Research University),
125993, 4 Volokolamskoe highway, Moscow, Russian Federation;
e-mail: romanaleks@gmail.com

Abstract

The article is devoted to the study of the methodology of development and application of software package for automated processing of the results of the All-Russian economic dictation. The authors describe the structure of the information-analytical system and the software-analytical package that implement the necessary functionality for effective processing of large amounts of data. The software solutions offered in this article allow to provide paperless document flow, to automate the process of verification and accounting of individual results of the participants of the dictation using algorithms and technologies of data recognition and modern software implementations of these algorithms. The information and analytical system of the All-Russian economic dictation allows to significantly improve the efficiency of processing the results of the economic dictation, to reduce the time of verification, to increase the convenience and speed of work at the initial stage of collection and introduction of individual answers of the dictation of participants, to minimize the direct work with paper. The use of modern information and communication technologies makes it possible to raise the statistical analysis and sociological analysis of the economic literacy of the Russian population to a qualitatively new level. A visual and intuitive presentation of the results improves the efficiency of data processing.

For citation

Golov R.S., Kostikov Yu.A., Romanenkov A.M. (2018) Instrumentarii ef-fektivnoi statisticheskoi obrabotki ekonomicheskikh dannykh [On the software package for processing of the data of the All-Russian economic dictation]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (9A), pp. 278-286.

Keywords

Software package, data processing, All-Russian economic dictation, analysis and information system, statistical analysis.

References

1. *Analiticheskii otchet po itogam Obshcherossiiskoi obrazovatel'noi aktsii "Vserossiiskii ekonomicheskii diktant"* [Analytical report on the results of the All-Russian educational campaign "All-Russian economic dictation"]. Available at: <https://clck.ru/DaKdu> [Accessed 14/09/18].
2. Borovkov A.A. (2007) *Matematicheskaya statistika* [Mathematical statistics]. Moscow: Fizmatlit Publ.
3. Deit K.Dzh. (2017) *Vvedenie v sistemy baz dannykh* [Introduction to database systems]. Moscow. Vill'yams Publ.
4. Gamma E., Khelm R., Dzhonson R., Vlissides Dzh. (2016) *Priemy ob"ektno-orientirovannogo proektirovaniya. Patterny proektirovaniya* [Techniques of object-oriented design. Design patterns]. Saint Petersburg: Piter Publ.
5. Kostikov Yu.A., Pavlov V.Yu., Romanenkov A.M., Ternovskov V.B. (2017) *Adaptivnaya arkhitektura programmno-apparatnogo kompleksa khraneniya i obrabotki dannykh* [Adaptive architecture of software and hardware complex for data storage and processing]. *Ekonomika: vchera, segodnya, zavtra* [Economics: yesterday, today, tomorrow], 7 (9A), pp. 192-207.

6. Prokhorenok N., Dronov V. (2016) *Python 3 i PyQt 5. Razrabotka prilozhenii* [Python 3 and PyQt 5. Application development]. Saint Petersburg: BKhV-Peterburg Publ.
7. Tyurin Yu.N. (2009) *Teoriya veroyatnostei i statistika* [Probability theory and statistics]. Moscow: Moscow Center For Continuous Mathematical Education.
8. Venttsel' E.S. (2007) *Teoriya veroyatnostei i ee inzhenernye prilozheniya* [Probability theory and its engineering applications]. Moscow: Vysshaya shkol Publ.