

СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ФИЗИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ В СПОРТЕ ВЫСШИХ ДОСТИЖЕНИЙ

*В.В. Епишев, А.П. Исаев, Р.М. Миниахметов, А.В. Мовчан,
А.С. Смирнов, Л.Б. Соколинский, М.Л. Цымблер, В.В. Эрлих*

В работе представлена архитектура системы MedMining, которая предназначена для интеллектуального анализа данных физиологических исследований спортсменов. Система обеспечивает экспорт результатов измерений в хранилище данных. Поддерживается хранение как необработанных результатов измерений (значений, поступающих непосредственно с приборов), так и их обработанных аналогов (получаемых путем усреднения, аппроксимации или других интегрирующих действий над обработанными данными). Интеллектуальный анализ результатов измерений направлен на определение ключевых показателей результативности и эффективности методики тренировок, а также поиск трендов и аномалий в этих показателях для гибкого изменения тренировочного графика.

Ключевые слова: интеллектуальный анализ данных, хранилище данных, физиологические исследования, спорт высших достижений

Введение

Под *интеллектуальным анализом данных (Data Mining)* понимают совокупность методов для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [2]. Технологии интеллектуального анализа данных применяются в широком спектре предметных областей, в том числе в медицине и физиологии [1, 3]. Одной из сфер приложения технологий интеллектуального анализа данных в медицине и физиологии является спорт высших достижений. Мониторинг, накопление и интеллектуальный анализ данных о физической и биохимической активности спортсменов направлены на определение ключевых показателей результативности и эффективности методики тренировок и поиск трендов и аномалий в этих показателях для гибкого изменения тренировочного графика [4, 10–12].

В рамках программы развития Южно-Уральского национального исследовательского государственного университета (НИУ ЮУрГУ) кафедрой теории и методики физической культуры и спорта и кафедрой системного программирования выполняется совместный проект, в рамках которого разрабатывается программная система MedMining, предназначенная для сбора и интеллектуального анализа данных физиологических исследований спортсменов НИУ ЮУрГУ. В данной статье рассматривается контекст данного проекта, архитектура и принципы реализации системы MedMining.

Статья организована следующим образом. Раздел 1 содержит описание предметной области проекта MedMining. В разделе 2 рассмотрена архитектура системы и технологический цикл работы с ней. В заключении суммируются полученные результаты и обсуждаются направления дальнейших исследований.

1. Предметная область проекта MedMining

Модель предметной области описывает организацию и проведение исследований, связи между спортсменами и тренерами, а также используемое оборудование. Диаграмма «сущность-связь», отражающая модель предметной области, представлена на рис. 1.

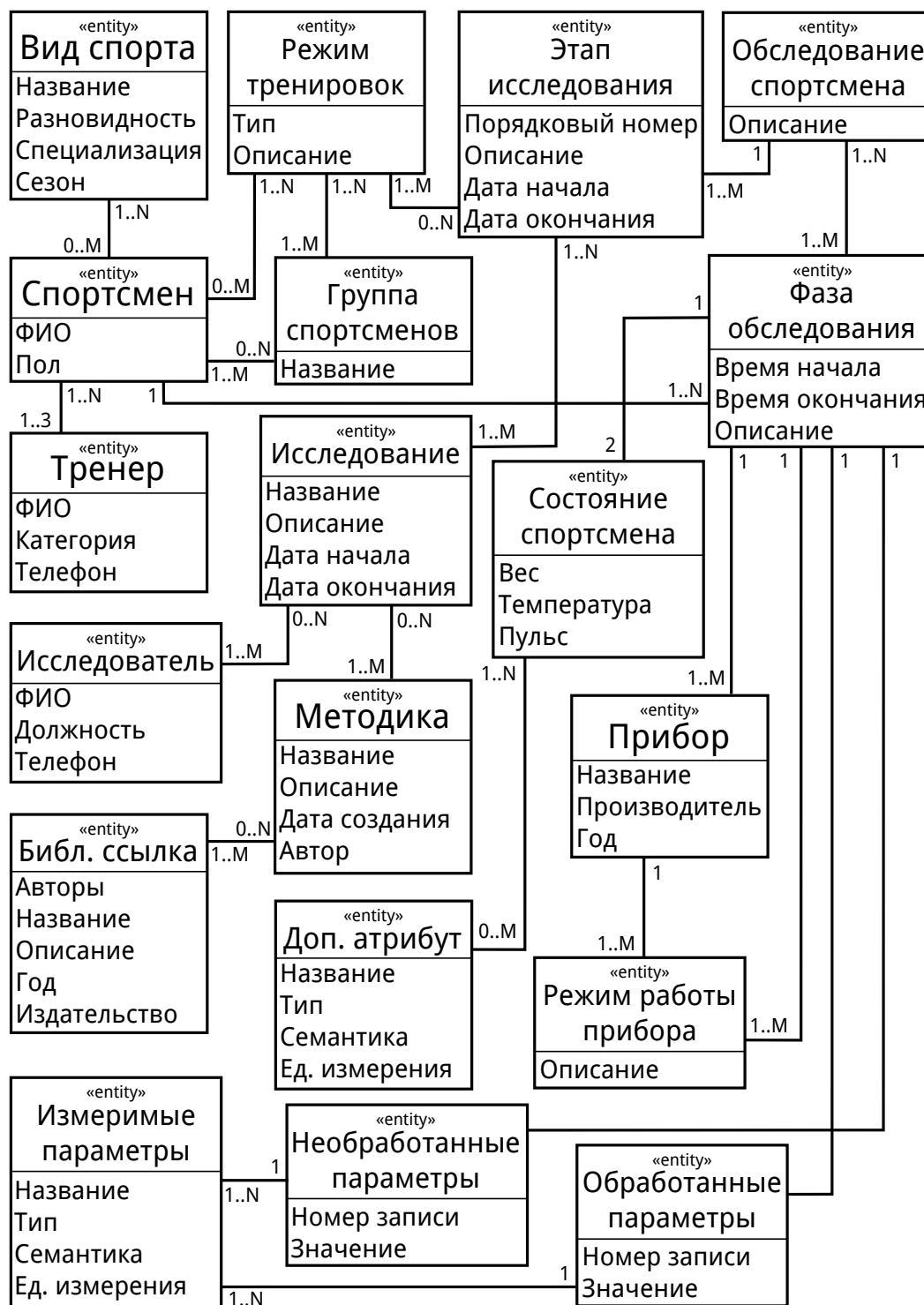


Рис. 1. Модель предметной области

1.1. Организация исследований

Организация исследований характеризуется следующими сущностями: исследование, исследователь, методика, библиографическая ссылка и этап исследования.

Исследование направлено на оценку состояния и подготовленности спортсменов на базе комплекса методик тренировки.

Исследователь — сотрудник, осуществляющий подготовку приборов, наблюдение за спортсменом во время исследования и экспорт данных исследования.

Методика представляет собой описание подхода к проведению исследований и подразумевает проведение набора исследований.

Библиографическая ссылка — библиографический источник, использованный в разработанной методике.

Этап исследования — одна из последовательных задач исследования, которая связана с оценкой физиологического состояния спортсмена.

Лаборант проводит измерение различных показателей различных показателей спортсменов физиологического состояния спортсменов (пульс, давление, ЭКГ и др.) во время и после тренировок с помощью различного оборудования.

1.2. Спортсмены и тренеры

Связи между спортсменами и тренерами характеризуются следующими сущностями: спортсмен, группа спортсменов, режим тренировок, вид спорта, тренер.

Спортсмен — испытуемый в исследованиях. Может заниматься несколькими видами спорта под руководством нескольких тренеров.

Группа спортсменов — спортсмены, которые имеют одинаковый режим тренировок.

Режим тренировок — описание тренировочного процесса спортсмена или группы спортсменов в рамках одного этапа исследования.

Вид спорта — название и характеристики определенного вида спорта.

Тренер — ответственный за подготовку спортсменов. Может тренировать различных спортсменов по разным видам спорта.

1.3. Проведение исследований

Проведение исследований характеризуется следующими сущностями: обследование спортсмена, фаза обследования, состояние спортсмена, дополнительный атрибут, необработанное данное и обработанное данное.

Обследование спортсмена — снятие физиологических показателей спортсмена на приборах в рамках одного этапа исследования.

Фаза обследования — шаг обследования, предполагающий снятие показаний определенного прибора в ходе тренировки спортсмена.

Состояние спортсмена — набор атрибутов, характеризующих динамичное физиологическое состояние спортсмена.

Дополнительный атрибут — атрибут сущности «Состояние спортсмена», название и семантика которого определяется исследователем. Может быть введено неограниченное количество дополнительных атрибутов. Дополнительный атрибут позволяет исследователю

отслеживать влияние различных факторов любой природы на физиологическое состояние спортсмена тренировочный процесс (например, факт приема и количество принятых препаратов до или после тренировки).

Необработанное данные — единичное значение одного из параметров, измеряемых прибором.

Обработанное данные — значение, получаемое из необработанных данных путем усреднения, аппроксимации или других интегрирующих действий.

1.4. Оборудование

Оборудование, используемое при проведении исследований, описывается сущностями прибор, режим работы прибора, измеряемый параметр.

Прибор представляет собой программно-аппаратный комплекс для измерения физиологических параметров спортсмена.

Режим работы прибора — состояние работы прибора, при котором он может измерять определенную группу параметров.

Изменяемый параметр — параметр физиологического состояния спортсмена, который можно измерить данным прибором.

2. Архитектура системы MedMining

Описание архитектуры системы приводится в нотации UML, и включает в себя описание аппаратных компонент и программной архитектуры.

2.1. Аппаратные компоненты системы

На рис. 2 представлена диаграмма развертывания системы MedMining.

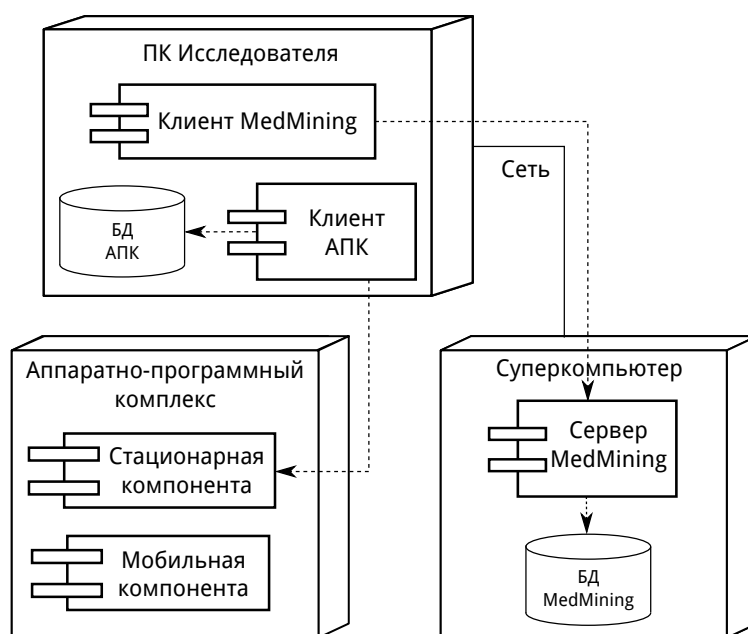


Рис. 2. Размещение системы MedMining

Система MedMining размещается на трех основных компонентах: с помощью аппаратно-программного комплекса (АПК) Исследователь проводит снятие физиологических показателей спортсменов, для взаимодействия с АПК и сервером системы MedMining, исследователь использует обычный персональный компьютер (ПК), для проведения ресурсоемких вычислений, таких как интеллектуальный анализ данных сверхбольших объемов, используется суперкомпьютер «СКИФ-Аврора ЮУрГУ» [5].

На диаграмме деятельности (рис. 3) показано взаимодействие актеров с основными компонентами системы.

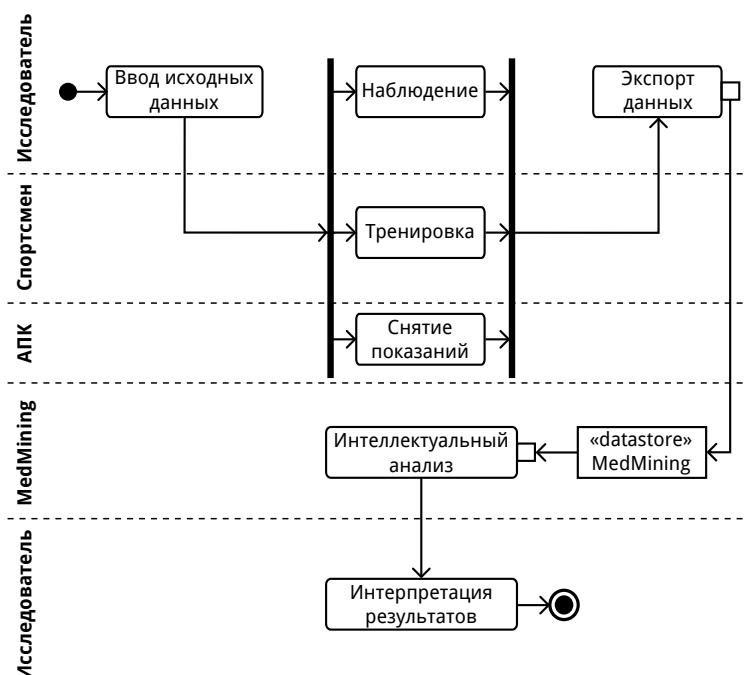


Рис. 3. Технологический цикл проведения исследований

В процессе снятия показаний задействованы следующие актеры: Исследователь, Спортсмен, АПК и система MedMining. Перед началом каждой Фазы обследования (тренировки) Исследователь выполняет снятие показаний Состояния спортсмена и Дополнительных атрибутов, которые относятся к данному Исследованию. Полученные показатели Исследователь вносит в систему MedMining.

После калибровки оборудования АПК спортсмен начинает тренировку. Во время тренировки Исследователь следит за состоянием Спортсмена. Показатели физиологического состояния спортсмена поступают в ПО АПК, на котором проходит данная фаза обследования. По окончании тренировки Исследователь выполняет выгрузку обработанных и необработанных данных исследования и вносит их в систему MedMining. Полученные данные хранятся на сервере системы MedMining — как в виде текстовых файлов, так и в виде таблиц хранилища.

Собранные данные подвергаются интеллектуальному анализу по выбору Исследователя. Система MedMining обеспечивает решение следующих основных задач интеллектуального анализа данных: кластеризация, классификация и анализ ассоциативных правил.

2.2. Программные компоненты системы

Модульная структура системы MedMining представлена на рис. 4 и содержит следующие компоненты. *Внешний интерфейс (FrontEnd)* включает в себя подсистемы, обеспечивающие интерфейс конечного пользователя и передачу полученной от него информации подсистемам *внутреннего интерфейса (BackEnd)*, которые выполняют ее обработку.

В состав BackEnd входят следующие подсистемы. Свободная реляционная СУБД с открытым исходным кодом *PostgreSQL* [6], в которой хранятся все данные системы MedMining. Асинхронная распределенная очередь задача *Celery* [7] позволяет выполнять ресурсоемкие задачи в фоновом режиме, получать информацию о ходе выполнения задач и др. Библиотека параллельных алгоритмов интеллектуального анализа данных *LibDM*, реализующих решения задач кластеризации, классификации и поиска ассоциативных правил.

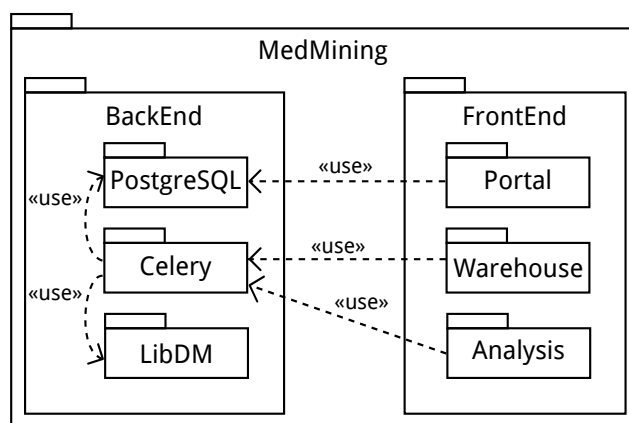


Рис. 4. Модульная структура системы MedMining

Интерфейс FrontEnd включает в себя следующие подсистемы. *Portal* представляет собой веб-приложение, написанное на основе фреймворка Django [8] и является тонким клиентом системы MedMining. Подсистема *Warehouse* отвечает за импорт данных физиологических исследований в хранилище на основе СУБД *PostgreSQL*. Функции импорта и преобразования подсистемы *Warehouse* работают асинхронно в виде заданий *Celery*. Подсистема *Analysis* предназначена для постановки задач интеллектуального анализа и визуализации результатов анализа.

3. Заключение

В работе представлена архитектура системы MedMining, которая предназначена для интеллектуального анализа данных физиологических исследований спортсменов. Система обеспечивает экспорт результатов измерений в хранилище данных. Поддерживается хранение как необработанных результатов измерений (значений, поступающих непосредственно с приборов), так и их обработанных аналогов (получаемых путем усреднения, аппроксимации или других интегрирующих действий над обработанными данными). Интеллектуальный анализ результатов измерений направлен на определение ключевых показателей результативности и эффективности методики тренировок и поиск трендов и аномалий в этих показателях.

Система позволяет Исследователю выполнять следующие основные функции:

- выработка заключения о состоянии конкретного спортсмена;
- проведение корректировок режима тренировок спортсмена;
- выявление наиболее успешной методики тренировок;
- исследование влияния тренировочного процесса на физиологическое состояние спортсмена (включая дополнительные параметры).

Исследования, представленные в данной работе, предполагается продолжить по следующим направлениям. Создание методов интеллектуального анализа данных, которые учитывают специфику данной предметной области (физиология профессиональных спортсменов). Другим направлением является применение параллельных СУБД [9, 13, 14] для обработки сверхбольших объемов данных физиологических исследований методами интеллектуального анализа данных [15].

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 12-07-00443-а.

Литература

1. Дюк, В.А. Предварительные результаты обработки разнотипных биометрических данных методами Data Mining / В.А. Дюк, О.В. Жвалевский, С.Б. Рудницкий, Д.А. Толстоногов // Труды СПИИРАН. – 2009. – № 9. – С. 197–210.
2. Хан, J. Data Mining: Concepts and Techniques / J. Хан, М. Kamber – Morgan Kaufmann, 2006. – 743 p.
3. Yoo, I. Data Mining in Healthcare and Biomedicine: A Survey of the Literature / I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J.-F. Chang, L. Hua // Journal of Medical Systems. – 2012. – Vol. 36, No. 4. – P. 2431–2448.
4. Lo, B.P.L. Pervasive sensing for athletic training / B.P.L. Lo, A. Atallah, B. Crewther, et al. // Delivering London 2012: ICT Enabling the Games, The IET special interest publication. – 2011. – P. 53–62.
5. Московский, А.А. Исследование производительности суперкомпьютеров семейства «СКИФ Аврора» на индустриальных задачах / А.А. Московский, М.П. Перминов, Л.Б. Соколинский, В.В. Черепенников, А.В. Шамакина // Вестник ЮУрГУ. Серия «Математическое моделирование и программирование». – 2010. – № 35(211). – С. 66–78.
6. Stonebraker, M. The POSTGRES next-generation database management system / M. Stonebraker, G. Kemnitz // Communications of the ACM. – 1991. – Vol. 34, No. 10. – P. 78–92.
7. Celery: Distributed Task Queue // Celery Project. URL: <http://celeryproject.org> (дата обращения: 08.11.2012).
8. Ortiz, A. Web Development with Python and Django // Proceedings of the 43rd ACM technical symposium on Computer science education. ACM, – 2012. – P. 686.

9. Пан, К.С. Разработка параллельной СУБД на основе последовательной СУБД PostgreSQL с открытым исходным кодом / К.С. Пан, М.Л. Цымблер // Вестник ЮУрГУ. Серия «Математическое моделирование и программирование». – 2012. – № 18(277), Вып. 12. – С. 112–120.
10. Исаев, А.П. Функциональное состояние кардиореспираторной системы бегунов в первые два дня деакклиматизации после двадцати дней пребывания в верхнем среднегорье / А.П. Исаев, В.В. Эрлих // Вестник ЮУрГУ. Серия «Образование, здравоохранение, физическая культура». – 2012. – № 8 (267). – С. 34–37.
11. Исаев, А.П. Стратегии формирования адаптационных реакций у спортсменов. Основы теории адаптации и закономерности ее формирования в спорте высоких и высших достижений / А.П. Исаев, В.В. Рыбаков, В.В. Эрлих с соавт. // Вестник ЮУрГУ. Серия «Образование, здравоохранение, физическая культура». – 2012. – № 21 (280). – С. 46–56.
12. Эрлих, В.В. Ключевые морфометрические, спирографические и кардиопульмональные портретные характеристики ведущих спортсменов в возрасте 15-16 лет в состоянии покоя и оценка физической работоспособности в период участия в социально значимых соревнованиях / В.В. Эрлих, А.П. Исаев, В.В. Епишев, Ю.Б. Хусаинова // Вестник ЮУрГУ. Серия «Образование, здравоохранение, физическая культура». – 2011. – № 39 (256). – С. 19–21.
13. Соколинский, Л.Б. Обзор архитектур параллельных систем баз данных // Программирование. – 2004. – № 6. – С. 49–63.
14. Костенецкий, П.С. Технологии параллельных систем баз данных для иерархических многопроцессорных сред / П.С. Костенецкий, А.В. Лепихов, Л.Б. Соколинский // Автоматика и телемеханика. – 2007. – № 5. – С. 112–125.
15. Miniakhmetov, R.M. Integrating Fuzzy c-Means Clustering with PostgreSQL // Труды Института системного программирования РАН. – 2011. – Т. 21. – С. 263–276.

Виталий Викторович Епишев, кандидат биологических наук, доцент, кафедра теории и методики физической культуры и спорта, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), epishev74@mail.ru.

Александр Петрович Исаев, доктор биологических наук, профессор, кафедра теории и методики физической культуры и спорта, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), tmfcs@mail.ru.

Руслан Марсович Минахметов, аспирант, кафедра системного программирования, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), tavein@gmail.com.

Александр Вячеславович Мовчан, студент 3 курса, кафедра системного программирования, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), movchan174@gmail.com.

Алексей Сергеевич Смирнов, руководитель информационного отдела спортивного комплекса, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), 2231034@mail.ru.

Леонид Борисович Соколинский, доктор физико-математических наук, профессор, кафедры системного программирования, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), sokolinsky@acm.org.

Михаил Леонидович Цымблер, кандидат физико-математических наук, доцент, кафедры системного программирования, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), zymbler@gmail.com.

Вадим Викторович Эрлих, кандидат биологических наук, доцент, кафедра теории и методики физической культуры и спорта, Южно-Уральский национальный исследовательский государственный университет (г. Челябинск, Российская Федерация), tmfcs@mail.ru.

PHYSIOLOGICAL DATA MINING SYSTEM FOR ELITE SPORTS

V.V. Epishev, South Ural State University (Chelyabinsk, Russian Federation),
A.P. Isaev, South Ural State University (Chelyabinsk, Russian Federation),
R.M. Miniakhmetov, South Ural State University (Chelyabinsk, Russian Federation),
A.V. Movchan, South Ural State University (Chelyabinsk, Russian Federation),
A.S. Smirnov, South Ural State University (Chelyabinsk, Russian Federation),
L.B. Sokolinsky, South Ural State University (Chelyabinsk, Russian Federation),
M.L. Zymbler, South Ural State University (Chelyabinsk, Russian Federation),
V.V. Ehrlich, South Ural State University (Chelyabinsk, Russian Federation)

The paper presents the architecture of MedMining system, which is designed for mining physiological studies of sportsmen. The system provides export of measurements to data warehouse. Storing of raw measurements (coming directly from the devices) and their processed analogues (obtained by averaging, approximation or other integrating operations on the processed data) is supported. Mining of measurements aims to discovering of key performance indicators and performance training techniques, as well as to searching trends and anomalies in these indicators for a flexible changes of training schedule.

Keywords: data mining, data warehouse, physiological research, elite sports.

References

1. Djuk V.A., Zhvalevskij O.V., Rudnickij S.B., Tolstonogov D.A. Predvaritel'nye rezul'taty obrabotki raznotipnyh biometricheskikh dannyh metodami Data Mining [Preliminary Results of the Processing of Different Types of Biometric Data by Means of Data Mining Methods] // Trudy SPIIRAN [SPIIRAN Proceedings]. 2009. No 9. P. 197–210.
2. Han J., Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2006. 743 p.
3. Yoo I., Alafairet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J.-F., Hua L. Data Mining in Healthcare and Biomedicine: A Survey of the Literature // Journal of Medical Systems. 2012. Vol. 36, No. 4. P. 2431–2448.
4. Lo B.P.L., Atallah A., Crewther B., et al. Pervasive Sensing for Athletic Training. Delivering London 2012: ICT Enabling the Games, The IET special interest publication. 2011. P. 53–62.

5. Moskovskij A.A., Perminov M.P., Sokolinskij L.B., Cherepennikov V.V., Shamakina A.V. Issledovanie proizvoditel'nosti superkomp'yuterov semejstva «SKIF Avrora» na industrial'nyh zadachah [Performance Analysis of the Supercomputer Family «SKIF Avrora» on Industrial Tasks] // Vestnik JuUrGU. Serija «Matematicheskoe modelirovanie i programmirovanie» [Bulletin of South Ural State University. Series «Mathematical Modeling, Programming & Computer Software»]. 2010. No 35(211). P. 66–78.
6. Stonebraker M., Kemnitz G. The POSTGRES next-generation database management system // Communications of the ACM. Oct. 1991. Vol. 34, No. 10. P. 78–92.
7. Celery: Distributed Task Queue // Celery Project. URL: <http://celeryproject.org> (дата обращения: 08.11.2012).
8. Ortiz A. Web Development with Python and Django // Proceedings of the 43rd ACM technical symposium on Computer science education. ACM, 2012. P. 686.
9. Pan K.S., Cymbler M.L. Razrabotka parallel'noj SUBD na osnove posledovatel'noj SUBD PostgreSQL s otkryтым ishodnym kodom [Development of a Parallel Database Management System on the Basis of Open-Source PostgreSQL DBMS] // Vestnik JuUrGU. Serija «Matematicheskoe modelirovanie i programmirovanie» [Bulletin of South Ural State University. Series «Mathematical Modeling, Programming & Computer Software»]. 2012. No 18(277), Iss. 12. P. 112–120.
10. Isaev A.P., Jerlih V.V. Funkcional'noe sostojanie kardiorespiratornoj sistemy begunov v pervye dva dnja deakklimatizacii posle dvadcati dnej prebyvanija v verhnem srednegor'e [Functional State of Athletes' Cardiorespiratory System in Two First Days of Deacclimatization after Twenty Days in Higher Uplands] // Vestnik JuUrGU. Serija «Obrazovanie, zdavoohranenie, fizicheskaja kul'tura» [Bulletin of South Ural State University. Series «Education, Healthcare, Physical Culture»]. 2012. No 8 (267). P. 34–37.
11. Isaev A.P., Rybakov V.V., Jerlih V.V. s soavt. Strategii formirovanija adaptacionnyh reakcij u sportmenov. Osnovy teorii adaptacii i zakonomernosti ee formirovanija v sporte vysokih i vysshih dostizhenij [Strategy for the Formation of Adaptive Reactions of Athletes. Adaptation Theory Basics and Its Formation in the Elite Sport] // Vestnik JuUrGU. Serija «Obrazovanie, zdavoohranenie, fizicheskaja kul'tura» [Bulletin of South Ural State University. Series «Education, Healthcare, Physical Culture»]. 2012. No 21 (280). P. 46–56.
12. Jerlih V.V., Isaev A.P., Epishev V.V., Husainova Ju.B. Kljuchevye morfometricheskie, spirograficheskie i kardiopul'monal'nye portretnye harakteristiki vedushhijh sportmenov v vozraste 15-16 let v sostojanii pokoja i ocenka fizicheskoj rabotosposobnosti v period uchastija v social'no znachimyh sorevnovanijah [Key Morphometric, and Cardiopulmonary Spirographic Portrait Characteristics for Leading Athletes in Age 15-16 at Rest and Evaluation of Physical Performance During Participation in Socially Important Events] // Vestnik JuUrGU. Serija «Obrazovanie, zdavoohranenie, fizicheskaja kul'tura» [Bulletin of South Ural State University. Series «Education, Healthcare, Physical Culture»]. 2011. No 39 (256). P. 19–21.
13. Sokolinskij L.B. Obzor arhitektur parallel'nyh sistem baz dannyh [Parallel Databases Architecture Review] // Programmirovanie [Programming]. 2004. No 6. P. 49–63.

14. Kosteneckij P.S., Lepihov A.V., Sokolinskij L.B. Tehnologii parallel'nyh sistem baz dannyh dlja ierarhicheskikh mnogoprocessornyh sred [Parallel Databases Technologies for Hierarchical Multiprocessor Environments] // Avtomatika i telemekhanika [Automatics & Telemechanics]. 2007. No 5. P. 112–125.
15. Miniakhmetov R.M. Integrating Fuzzy c-Means Clustering with PostgreSQL // Trudy Instituta sistemnogo programirovanija RAN [Programming and Computer Software]. 2011. V. 21. P. 263–276.

Поступила в редакцию 1 декабря 2012 г.