

КОНЦЕПЦИЯ СИСТЕМЫ МОДЕЛИРОВАНИЯ СТАТИСТИК, РАЗЛИЧАЮЩИХ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Содержит обзор требований к современному программному обеспечению и описание концепции разработки специализированной программной системы, предназначенной для моделирования статистик, различающих случайные величины.

Ключевые слова: программное обеспечение, мобильность, переносимость, совместимость, интерфейс пользователя, функциональная пригодность.

T.I. Vedernikova
L.A. Dulatova

CONCEPT OF MODELLING SYSTEM FOR STATISTICS THAT DIFFERENTIATE RANDOM VALUES

The article contains a review of requirements for contemporary software, and a description of the concept of a problem-oriented software system work-out for modelling statistics that differentiate random values.

Keywords: software, mobility, transferability, compatibility, user's interface, functional usability.

Система моделирования статистик, различающих случайные величины (СИМОС), предназначена для построения рабочего словаря признаков (РСП) с целью решения задач распознавания образов. РСП — это совокупность характеристик объекта, непосредственно используемых для идентификации. Метод построения РСП основан на функциональном преобразовании многомерных случайных величин в одномерные, аккумулирующие в себе отличительные особенности исходных величин [6].

Современные тенденции и требования к программам. Характеристики качества, общие для всех программных средств, регламентируются стандартом ГОСТ Р ИСО/МЭК 9126-93 [7]. Это функциональные возможности (пригодность, правильность, способность к взаимодействию с другими системами, согласованность и защищенность), надежность, практичность, эффективность, сопровождаемость и мобильность. Кроме того, к программам статистической обработки данных предъявляются специфические требования [1; 2; 4; 10]: корректность и скорость расчетов; наличие свойств открытых систем; ассистирование при выборе способа обработки данных; введение данных пользователя; наличие средств графического представления данных; наличие командного языка и графического интерфейса пользователя.

Большинство современных статистических пакетов позволяют обрабатывать объемы данных, ограниченные только ресурсами компьютера, предоставляют возможность экспорта и импорта данных в популярные форматы. Некоторые являются сетевыми и позволяют организовывать распределенные вычисления. Существующие статистические пакеты часто включают средства для интеллектуального ана-

лиза данных (Data Mining) и возможности интегрирования в корпоративные информационные системы.

Одной из важных тенденций в разработке компьютерных программ является применение концепции открытых систем. Характерные черты открытых систем — изменяемость, переносимость между программно-аппаратными платформами, интероперабельность, использование стандартных интерфейсов пользователей. Концепция открытых систем позволяет повысить экономическую эффективность функционирования, производства и сопровождения программных систем в непрерывно меняющихся условиях, обеспечить повторное использование программ и данных [8].

От современных программ требуют высокого качества взаимодействия с пользователем. Для описания качества взаимодействия программы с пользователем служит характеристика «практичность» (понятность, обучаемость и простота использования), определенная в стандарте ГОСТ Р ИСО/МЭК 9126-93 [7]. Как сказано в [3], «практичность выражает степень сложности выполнения пользователем стоящей перед ним задачи и тип реализованного в системе механизма помощи пользователю».

Требование практичности реализуется с помощью правил построения интерфейсов и стилей взаимодействия с пользователем [3; 11]. Выделяют следующие стили взаимодействия программных систем с пользователем [11]: прямое манипулирование; выбор из меню; заполнение экранных форм; командный язык; естественный язык. Развитые программы для статистической обработки данных реализуют практически все эти стили (за исключением разве что естественного языка). Достоинствами прямого манипулирования является быстрое и интуитивно понятное взаимодействие и легкость в изучении. Однако это стиль взаимодействия подходит только для задач и объектов, имеющих зрительный образ [9]. Прямое манипулирование используется в интерфейсах для Data Mining.

Требования к системе и направления ее реализации. Разработка новой системы должна обеспечить современный уровень взаимодействия с пользователем; совместимость с другими программами; переносимость на различные аппаратно-программные платформы; корректность выполнения существующих функций системы; модифицируемость системы, расширяемость ее функциональных возможностей; работу с наборами данных практически неограниченной длины.

Создание системы возможно в одном из следующих направлений [5]: разработка кодов для интерпретации в средах (статистического или математического) моделирования; разработка компонента какого-либо распространенного статистического пакета; создание самостоятельного продукта.

Первые два варианта привлекательны тем, что в готовых пакетах, как правило, уже имеется расширяемый интерфейс пользователя и модули для импорта и экспорта данных в распространенных форматах, доступно множество математических и статистических процедур, которые были бы полезны для СИМОС. Кроме того, не требуется создавать архитектуру программного пакета, так как она существует. Все основополагающие решения уже приняты, однако эти решения могут вступить в противоречие с основными требованиями к СИМОС.

Коммерческие статистические пакеты дороги, и в этих условиях удачной заменой им могут стать свободно распространяемые программы. Из них большим количеством функций обладают те, что распространяются с открытыми исходными текстами.

Одним из серьезных недостатков программ, распространяемых с открытыми исходными текстами, является сложность контроля их ка-

чества. С одной стороны, благодаря поддержке сообщества быстро устраняются ошибки и выпускаются обновления. С другой стороны, скорость изменения версий и изменения программ ведет к несогласованности изменений. Поэтому (а также потому, что открытое сообщество избегает использовать открытые стандарты и патентованные технологии, даже если они бесплатны) для программ с открытыми исходными текстами остро стоит вопрос совместимости [9].

Препятствием для второго направления является отсутствие необходимой документации на русском языке, так как расширяемые пакеты, как правило, производятся за границей. Большим недостатком является также связывание жизненного цикла СИМОС с жизненным циклом другого продукта со всеми вытекающими отсюда последствиями. Поэтому представляется целесообразной разработка независимого приложения.

Концепция системы. Для достижения целей сопровождаемости и удобства использования СИМОС уместно применить объектно-ориентированный и компонентный подходы, в качестве основного средства реализации взять кроссплатформенную технологию Java, которая обеспечивает переносимость системы в исполняемом виде. Чтобы система была расширяемой в процессе эксплуатации, ее предполагается строить на основе модулей расширения (плагинов).

СИМОС планируется разрабатывать так, чтобы ее функциональная часть при необходимости могла бы использоваться как компонент стороннего продукта. Взаимодействие с какой-либо существующей программой при необходимости будет организовано с помощью дополнительных модулей-адаптеров.

Кроме того, для переносимости и повторного использования в других программных продуктах функций программы, а также для изменяемости интерфейса пользователя систему предполагается разбить на слои так, чтобы функциональность, реализующая интерфейс пользователя, и функциональность, относящаяся к предметной области, были помещены в разные классы, а физически — и в разные файлы. Аналогично от функций обработки данных отделяются функции создания данных и сами данные. Для изменяемости и расширяемости статистических функций интерфейсы проектируются, исходя из требований к организации функций, и отделяются от реализации всюду, где в будущем предполагаются изменения.

Легкость обучения обеспечивается за счет применения подходящих стилей взаимодействия с пользователями, стандартных пользовательских интерфейсов и контекстной справки, раскрывающей суть используемых процедур.

В качестве основных стилей взаимодействия пользователя с системой предполагается использовать меню и формы ввода. Однако по мере усложнения системы может потребоваться применение других способов взаимодействия, и проект должен предусматривать возможность их дальнейшего подключения.

Простота и ввод данных предполагается осуществлять стандартным для статистических программ образом: в виде таблицы и списка объектов пользователя (переменных, отчетов и т.п.).

Простота управления программой может быть достигнута за счет сохранения и повторного использования данных пользователя и его настроек; параллельности или псевдопараллельности при работе с данными и функциями; работы в темпе пользователя; предотвращения возможных ошибок пользователя (откаты, предупреждения и т.п.).

Считывание из файлов и обработку данных предлагается проводить в зависимости от объема данных: данные, не превышающие некоторого заранее оговоренного объема, — целиком, а превышающие — частями. При этом хранение данных в оперативной памяти можно осуществлять в виде динамических структур данных.

Совместимость с другими программными продуктами, с одной стороны, будет реализована как возможность использования функциональной части СИМОС вне самой системы, а с другой — как возможность переноса данных, созданных или обработанных СИМОС, в другие системы.

Доступ к внешним данным планируется осуществлять посредством JDBC и ODBC. Однако для этого требуется наличие на компьютере пользователя программы-сервера, способной обработать данные в нужном формате, и JDBC- или ODBC-драйвера для доступа к ней. Поэтому необходимы средства, позволяющие осуществлять доступ к данным без ODBC. Такие средства должны предоставлять доступ к файлам конкретных форматов.

Для обеспечения переносимости данных между приложениями желательно реализовывать все известные варианты формата CSV. Для доступа к данным в формате BIFF можно использовать существующую свободно распространяемую библиотеку Java Excel API. Формат поддерживается практически всеми программами обработки числовых данных [12].

Предлагаемая концепция позволит:

- обеспечить мобильность программы как на уровне исходных текстов, так и на уровне исполняемого кода между операционными системами семейств Windows, Linux и др.;

- переносить данные между СИМОС и другим программным обеспечением, поддерживающим форматы CSV и BIFF (XLS), и взаимодействовать с СУБД, для которых доступны JDBC- и ODBC-драйверы;

- сделать функциональность и интерфейс пользователя системы СИМОС расширяемыми без перекомпиляции;

- снять ограничение на обрабатываемый объем данных;

- при необходимости организовать непосредственное взаимодействие СИМОС с какой-либо системой;

- сделать интерфейс пользователя легким для изучения и использования.

Список использованной литературы

1. Айвазян С.А. Инструменты статистического анализа данных / С.А. Айвазян, В.С. Степанов. — URL: <http://www.osp.ru/pcworld/1997/08/157830>.

2. Айвазян С.А. Программное обеспечение по статистическому анализу данных: методология сравнительного анализа и выборочный обзор рынка / С.А. Айвазян, В.С. Степанов. — URL: <http://www.cemi.rssi.ru/rus/publicat/e-pubs/ep97001/title.htm>.

3. Басс Л. Архитектура программного обеспечения на практике / Л. Басс, П. Клементс, Р. Кацман. — СПб.: Питер, 2006.

4. Борисова С.Ф. Компьютер и Интернет для социолога [Электронный ресурс] / С.Ф. Борисова. — Н. Новгород, 2002. — URL: <http://www.unn.ru/rus/f14/k2/courses/borisova.htm>.

5. Ведерникова Т.И. Концепция программного средства ДИСИМОС / Т.И. Ведерникова, Л.А. Дулатова // Применение математических методов и информационных технологий в экономике: сб. науч. тр. — Иркутск: Изд-во БГУЭП, 2010. — Вып. 9. — С. 32–43.

6. Ведерникова Т.И. Эффективное использование статистических ресурсов в управлении как задача распознавания образов / Т.И. Ведерникова, В.И. Дибирдеев // Известия Иркутской государственной экономической академии. — 2007. — № 4 (54). — С. 95–97.

7. ГОСТ Р ИСО/МЭК 9126-93. Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению. — Изд. офиц.

8. Липаев В.В. Экономика производства сложных программных продуктов / В.В. Липаев. — М.: СИНТЕГ, 2008.
9. Мамыкин В. Открытые стандарты и совместимость ИС / В. Мамыкин. — URL: <http://www.osp.ru/lan/2006/11/3675867/>.
10. Русаков А.А. Методы математической статистики и анализ данных психолого-педагогических исследований: курс лекций и практических занятий [Электронный ресурс] / А.А. Русаков, Ю.И. Богатырева. — Тула: ТГПУ, 2004. — URL: <http://www.tspu.tula.ru/res/math/mop/lections.htm>.
11. Соммервилл И. Инженерия программного обеспечения / И. Соммервилл. — М.: Изд. дом «Вильямс», 2002.
12. How To: The Comma Separated Value (CSV) File Format. — URL: <http://www.creativyst.com/Doc/Articles/CSV/csv01.htm>.

References

1. Aivazyan S.A. Instrumenty statisticheskogo analiza dannykh / S.A. Aivazyan, V.S. Stepanov. — URL: <http://www.osp.ru/pcworld/1997/08/157830>.
2. Aivazyan S.A. Programmnoe obespechenie po statisticheskomu analizu dannykh: metodologiya sravnitel'nogo analiza i vyborochnyi obzor rynka / S.A. Aivazyan, V.S. Stepanov. — URL: <http://www.cemi.rssi.ru/rus/publicat/e-pubs/ep97001/title.htm>.
3. Bass L. Arkhitektura programmnoho obespecheniya na praktike / L. Bass, P. Klements, R. Katsman. — SPb.: Piter, 2006.
4. Borisova S.F. Komp'yuter i Internet dlya sotsiologa [Elektronnyi resurs] / S.F. Borisova. — N. Novgorod, 2002. — URL: <http://www.unn.ru/rus/f14/k2/courses/borisova.htm>.
5. Vedernikova T.I. Kontseptsiya programmnoho sredstva DISIMOS / T.I. Vedernikova, L.A. Dulatova // Primenenie matematicheskikh metodov i informatsionnykh tekhnologii v ekonomike: sb. nauch. tr. — Irkutsk: Izd-vo BGUEP, 2010. — Вып. 9. — С. 32–43.
6. Vedernikova T.I. Effektivnoe ispol'zovanie statisticheskikh resursov v upravlenii kak zadacha raspoznavaniya obrazov / T.I. Vedernikova, V.I. Dibirdeev // Izvestiya Irkutskoi gosudarstvennoi ekonomicheskoi akademii. — 2007. — № 4 (54). — С. 95–97.
7. GOST R ISO/MEK 9126-93. Informatsionnaya tekhnologiya. Otsenka programmnoi produktsii. Kharakteristiki kachestva i rukovodstva po ikh primeneniyu. — Izd. ofits.
8. Lipaev V.V. Ekonomika proizvodstva slozhnykh programmykh produktov / V.V. Lipaev. — М.: SINTEG, 2008.
9. Mamykin V. Otkrytye standarty i sovместimost' IS / V. Mamykin. — URL: <http://www.osp.ru/lan/2006/11/3675867/>.
10. Rusakov A.A. Metody matematicheskoi statistiki i analiz dannykh psikhologo-pedagogicheskikh issledovaniy: kurs lektstii i prakticheskikh zanyatii [Elektronnyi resurs] / A.A. Rusakov, Yu.I. Bogatyreva. — Tula: TGPU, 2004. — URL: <http://www.tspu.tula.ru/res/math/mop/lections.htm>.
11. Sommervill I. Inzheneriya programmnoho obespecheniya / I. Sommervill. — М.: Изд. дом «Vil'yams», 2002.
12. How To: The Comma Separated Value (CSV) File Format. — URL: <http://www.creativyst.com/Doc/Articles/CSV/csv01.htm>.

Информация об авторах

Ведерникова Татьяна Ивановна — кандидат технических наук, доцент, кафедра информатики и кибернетики, Байкальский государственный университет экономики и права, г. Иркутск, e-mail: vedernikova-ti@isea.ru.

Дулатова Лия Асангалиевна — магистрант, кафедра информатики и кибернетики, Байкальский государственный университет экономики и права, г. Иркутск, e-mail: dulatova_liya@mail.ru.

Authors

Vedernikova Tatiana Ivanovna — PhD in Engineering Science, Associate Professor, Chair of Computer Science and Cybernetics, Baikal State University of Economics and Law, Irkutsk, e-mail: vedernikova-ti@isea.ru.

Dulatova Lia Asangalieвна — Candidate for Master's Degree, Chair of Computer Science and Cybernetics, Baikal State University of Economics and Law, Irkutsk, e-mail: dulatova_liya@mail.ru.