

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года

Журнал выходит ежеквартально

Учредители:

Российская академия наук

Федеральный исследовательский центр

«Информатика и управление» Российской академии наук

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета

академик РАН Г. И. Савин

академик РАН А. Л. Стемпковский

член-корреспондент РАН Ю. Б. Зубарев

профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)

профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)

профессор М. Никиulin (M. Nikulin, Bordeaux, France)

профессор В. Ротарь (V. Rotar, San-Diego, USA)

профессор М. Финкельштейн (M. Finkelstein, Rostok, Germany)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор

профессор, д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора

д.т.н. В. Н. Захаров проф., д.г.-м.н. Р. Б. Сейфуль-Мулюков

проф., д.т.н. В. Д. Ильин д.ф.-м.н. В. И. Синицын

проф., д.ф.-м.н. Л. А. Калиниченко проф., д.т.н. И. Н. Синицын

проф., д.т.н. К. К. Колин к.ф.-м.н. А. К. Горшенин

проф., д.ф.-м.н. В. Ю. Королев к.ф.-м.н. С. А. Христочевский

Редакция

профессор, д.г.-м.н. Р. Б. Сейфуль-Мулюков

к.ф.-м.н. Е. Н. Арутюнов

С. Н. Стригина (ответственный секретарь)

© Федеральный исследовательский центр «Информатика
и управление» Российской академии наук, 2017

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science

Журнал входит в систему Российского индекса научного цитирования (РИНЦ)

Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory

и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНТИ
и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных
изданий, в которых должны быть опубликованы основные научные результаты диссертаций
на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 27 № 2 Год 2017

СОДЕРЖАНИЕ

Аналитическое моделирование нормальных процессов
в стохастических системах с интегральными нелинейностями

И. Н. Синицын 3

Статистический анализ максимум-линейных случайных процессов

А. В. Лебедев 16

Классификация ошибочных состояний в распределенных
вычислительных системах и источники их возникновения

**А. А. Грушо, М. И. Забежайло, А. А. Зацаринный, А. В. Николаев,
В. О. Писковский, Е. Е. Тимонина** 29

Возможность выявления инсайдера статистическими методами

Е. А. Мартынянов 41

Имитационная модель поиска инсайдера статистическими методами

Е. А. Мартынянов 48

Система управления обучением ELIS. Архитектурные решения

А. К. Горшенин, Е. С. Данилович, Д. Р. Хромов 60

Система управления обучением ELIS. Пользовательский интерфейс
и функциональные возможности

А. К. Горшенин, Е. С. Данилович, Д. Р. Хромов 70

Синтез самосинхронных комбинационных секций функциональным
методом

Л. П. Плеханов 85

Научно-методические подходы к совершенствованию
нормативной базы для создания и развития
информационно-телекоммуникационных систем

С. А. Головин, А. А. Зацаринный, С. В. Козлов 98

Информационная структура угроз национальной безопасности

А. П. Сучков 113

Обратимость и альтернативность генерализации моделей перевода
коннекторов в параллельных текстах

И. М. Зацман, О. С. Мамонова, А. Ю. Щурова 125

Об одном подходе к моделированию процесса развития семантической
сети

И. М. Адамович, О. И. Волков 143

Табс-представление задач и задачных графов

В. Д. Ильин 155

Об авторах 170

Правила подготовки рукописей статей 172

Requirements for manuscripts 176

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НОРМАЛЬНЫХ ПРОЦЕССОВ В СТОХАСТИЧЕСКИХ СИСТЕМАХ С ИНТЕГРАЛЬНЫМИ НЕЛИНЕЙНОСТЯМИ (I)

И. Н. Синицын¹

Аннотация: Разработано общее методическое и алгоритмическое обеспечение аналитического моделирования нормальных (гауссовских) процессов в дифференциальных стохастических системах (СтС) с интегральными нелинейностями (ИН) на основе методов статистической линеаризации (МСЛ) и нормальной аппроксимации (МНА). При этом ИН аппроксимировались степенными и эрмитовыми представлениями. Для ИН, описываемых функциями Лапласа, интегралами Френеля и интегральным синусом, получены выражения для коэффициентов МСЛ и МНА. Алгоритмы метода аналитического моделирования (МАМ) положены в основу разрабатываемого инструментального программного обеспечения StS-Analysis.2017. Приведены необходимые сведения из теории функций Лапласа, интегралов Френеля и интегрального синуса, а также коэффициенты МСЛ ИН по методу степенных разложений. В качестве тестового примера рассмотрена стохастическая динамика интегрального осциллятора (ИО).

Keywords: интегралы Френеля; интегральная нелинейность (ИН); интегральный синус; метод аналитического моделирования (МАМ); метод нормальной аппроксимации (МНА); метод статистической линеаризации (МСЛ); полином Эрмита; степенное разложение; стохастическая система (СтС); функция Лапласа

DOI: 10.14357/08696527170201

1 Введение

В цикле статей [1–3] изложены методы и инструментальные программные средства МАМ нормальных процессов в СтС со сложными нелинейностями, описываемых такими специальными функциями, как функции Бесселя целого и дробного порядка, эллиптические функции Якоби и Вейерштрасса, а также связанными с ними функциями.

Продолжим цикл [1–3] и рассмотрим типовые сложные ИН, МСЛ и МНА для дифференциальных СтС со сложными ИН. В качестве тестового примера рассмотрим динамику интегрального осциллятора в стохастической среде. Особое внимание уделим ИН, описываемым функциями Лапласа, интегралами Френеля и интегральным синусом.

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sinitsin@dol.ru

2 Интегральные нелинейности и их статистическая линеаризация

Рассмотрим сначала скалярную ИН вида

$$z = \int_0^y \psi^{\text{ин}}(y, \eta) d\eta, \quad (1)$$

где $\psi^{\text{ин}}(y, \eta)$ — детерминированная нелинейная функция отмеченных аргументов.

В физико-технических приложениях примерами могут служить: функция Лапласа (см. приложение П1), функции, выражаемые интегралами Френеля (см. приложение П2), и интегральный синус (см. приложение П3).

Примерами сложных ИН могут служить многочленные и дробно-рациональные функции из типовых ИН, а также функции, получаемые путем соответствующего преобразования аргумента в ИН [1–3].

В случае векторных и матричных сложных ИН (1) имеют место соответствующие представления для компонент.

Согласно методам численного функционального анализа [4–6], наиболее распространенными методами аппроксимации сложных ИН являются следующие:

- степенные разложения;
- многочленные приближения по ортогональными полиномам Эрмита и др.;
- рациональные приближения;
- разложения в цепные дроби;
- асимптотические приближения;
- итеративные процессы.

Итак, пусть задана нестационарная ИН вида

$$Z_t = \int_0^{Y_t} \psi^{\text{ин}}(Y_t, t) dY_t.$$

В случаях степенного, многочленного эрмитова представления ИН может быть аппроксимирована следующими выражениями:

$$\begin{aligned} Z_t &= \varphi^{\text{ct}}(Y_t, t) + R_n^{\text{ct}}, \\ \varphi^{\text{ct}}(Y_t, t) &= \varphi^+(Y_t, t) \varphi_n^-(Y_t, t), \quad \varphi_n^-(Y_t, t) = \sum_{k=0}^{n_{\text{ct}}} a_k^-(t) Y_t^k; \end{aligned} \quad (2)$$

$$\begin{aligned} Z_t &= \varphi^{\text{sp}}(Y_t, t) + R_n^{\text{sp}}, \\ \varphi^{\text{sp}}(Y_t, t) &= \varphi^+(Y_t, t) \varphi_n^-(Y_t, t), \quad \varphi_n^-(Y_t, t) = \sum_{k=0}^{n_{\text{sp}}} a_k H_k(Y_t) Y_t^k, \end{aligned} \quad (3)$$

где R_n — остаточный член; H_k — полином Эрмита. Применим МСЛ [7–9] к аппроксимированному выражению ИН:

$$Z_t \approx \varphi(Y_t, t) . \quad (4)$$

Здесь функция (4) определяется (2) или (3).

Далее будем считать стохастический процесс (СтП) Y_t действительным, несимметричным ($m_y \neq 0$) и с конечной дисперсией D_y . Тогда, согласно МСЛ [7–9], (4) будет, в смысле минимума средней квадратической ошибки, статистически эквивалентно выражению вида:

$$Z_t \approx \varphi_0(m_y, D_y, t) + k_1(m_y, D_y, t) Y_t^0 \quad (Y_t^0 = Y_t - m_y) , \quad (5)$$

где

$$\varphi_0 = \varphi_0(m_y, D_y, t) = \frac{1}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} \varphi(\eta, t) \exp \left[-\frac{(\eta - m_y)^2}{2D_y} \right] d\eta ; \quad (6)$$

$$\begin{aligned} k_1 = k_1(m_y, D_y, t) &= \frac{1}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} (\eta - m_y) \varphi(\eta, t) \exp \left[-\frac{(\eta - m_y)^2}{2D_y} \right] d\eta = \\ &= \frac{\partial \varphi_0(m_y, D_y, t)}{\partial D_y} . \end{aligned}$$

Для нечетных (4) формулы (5) и (6) можно представить в виде:

$$Z_t \approx k_0(m_y, D_y, t) m_y + k_1(m_y, D_y, t) Y_t^0 ,$$

где

$$k_0 = k_0(m_y, D_y, t) = \frac{1}{m_y \sqrt{2\pi D_y}} \int_{-\infty}^{\infty} \varphi(\eta, t) \exp \left[-\frac{(\eta - m_y)^2}{2D_y} \right] d\eta .$$

Коэффициенты МСЛ для лапласовых, френелевых и интегрального синуса ИН приведены в приложениях П1–П3.

3 Алгоритмы аналитического моделирования нормальных процессов в стохастических системах с интегральными нелинейностями

Уравнения функционирования конечномерных непрерывных нелинейных систем со стохастическими возмущениями путем расширения вектора состояния СтС могут быть записаны в виде следующего векторного стохастического дифференциального уравнения Ито [7–9]:

$$dY_t = a^{\text{иин}}(Y_t, t) dt + b^{\text{иин}}(Y_t, t) dW_0 + \int_{R_0} c^{\text{иин}}(Y_t, t, v) P^0(dt, dv), \\ Y(t_0) = Y_0. \quad (7)$$

Здесь Y_t — $(p \times 1)$ -мерный вектор состояния, $Y_t \in \Delta_y$ (Δ_y — многообразие состояний); $a^{\text{иин}} = a^{\text{иин}}(Y_t, t)$ и $b^{\text{иин}} = b^{\text{иин}}(Y_t, t)$ — известные $(p \times 1)$ -мерная и $(p \times m)$ -мерная функции Y_t и t ; $W_0 = W_0(t)$ — $(r \times 1)$ -мерный винеровский СтП интенсивности $\nu_0 = \nu_0(t)$; $c^{\text{иин}} = c^{\text{иин}}(Y_t, t, v)$ — $(p \times 1)$ -мерная функция Y_t, t и вспомогательного $(q \times 1)$ -мерного параметра v ; $\int_{\Delta} dP^0(t, A)$ — центрированная пуассоновская мера, определяемая

$$\int_{\Delta} dP^0(t, A) = \int_{\Delta} dP(t, A) = \int_{\Delta} \nu_P(t, A) dt.$$

При этом принято: \int_{Δ} — число скачков пуассоновского СтП в интервале времени $\Delta = (t_1, t_2]$; $\nu_P(t, A)$ — интенсивность пуассоновского СтП $P(t, A)$; A — некоторое борелевское множество пространства R_0^q с выколотым началом. Начальное значение Y_0 представляет собой случайную величину, не зависящую от приращений $W_0(t)$ и $P(t, A)$ на интервалах времени, следующих за t_0 , $t_0 \leq t_1 \leq t_2$, для любого множества A .

В случае аддитивных гауссовских (нормальных) и обобщенных пуассоновских возмущений уравнение (7) принимает вид [7–9]:

$$\dot{Y} = a^{\text{иин}}(Y_t, t) + b_0(t)V, \quad V = \dot{W}, \quad Y(t_0) = Y_0. \quad (8)$$

Здесь W — СтП с независимыми приращениями, представляющий собой смесь нормального и обобщенного пуассоновского СтП.

Если существуют конечные вероятностные моменты второго порядка для моментов времени t_1 и t_2 , то уравнения МНА примут следующий вид [7–9]:

— для характеристических функций:

$$\left. \begin{aligned} g_1^N(\lambda; t) &= \exp \left[i\lambda^T m_t - \frac{1}{2} \lambda^T K_t \lambda \right]; \\ g_{t_1, t_2}^N(\lambda_1, \lambda_2; t_1, t_2) &= \exp \left[i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \bar{\lambda} \right], \end{aligned} \right\} \quad (9)$$

где

$$\bar{\lambda} = \left[\lambda_1^T \lambda_2^T \right]^T ; \quad \bar{m}_2 = \left[m_{t_1}^T m_{t_2}^T \right]^T ; \quad \bar{K}_2 = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix} ;$$

- для математических ожиданий m_t , ковариационной матрицы K_t и матрицы ковариационных функций $K(t_1, t_2)$:

$$\left. \begin{aligned} \dot{m}_t &= a_1^{ин} (m_t, K_t, t) , \quad m_0 = m(t_0) ; \\ \dot{K}_t &= a_2^{ин} (m_t, K_t, t) , \quad K_0 = K(t_0) ; \\ \frac{\partial K(t_1, t_2)}{\partial t_2} &= K(t_1, t_2) a_{21}^{ин} (m_{t_2}, K_{t_2}, t_2)^T , \quad K(t_1, t_1) = K_{t_1} . \end{aligned} \right\} \quad (10)$$

Здесь приняты следующие обозначения:

$$\left. \begin{aligned} m_t &= M_{\Delta_y}^N [Y_t] , \quad Y_t^0 = Y_t - m_t ; \\ K_t &= M_{\Delta_y}^N [Y_t^0 Y_t^{0T}] , \quad K(t_1, t_2) = M_{\Delta_y}^N [Y_{t_1}^0 Y_{t_2}^{0T}] ; \\ a_1^{ин} &= a_1^{ин} (m_t, K_t, t) = M_{\Delta_y}^N [a^{ин} (Y_t, t)] ; \\ a_2^{ин} &= a_2^{ин} (m_t, K_t, t) = \\ &= a_{21}^{ин} (m_t, K_t, t) + a_{21}^{ин} (m_t, K_t, t)^T + a_{22}^{ин} (m_t, K_t, t) ; \\ a_{21}^{ин} &= a_{21}^{ин} (m_t, K_t, t) = M_{\Delta_y}^N [a^{ин} (Y_t, t) Y_t^{0T}] ; \\ a_{22}^{ин} &= a_{22}^{ин} (m_t, K_t, t) = M_{\Delta_y}^N [\bar{\sigma}^{ин} (Y_t, t)] ; \\ \sigma^{ин} (Y_t, t) &= b^{ин} (Y_t, t) \nu_0(t) b^{ин} (Y_t, t)^T ; \\ \bar{\sigma} (Y_t, t) &= \sigma^{ин} (Y_t, t) + \int_{R_0^q} c^{ин} (Y_t, t, v) c^{ин} (Y_t, t, v)^T \nu_P(t, dv) , \end{aligned} \right\} \quad (11)$$

где $M_{\Delta_y}^N$ — символ вычисления математического ожидания для нормальных распределений (9) на гладком многообразии Δ_y .

Для стационарных СтС нормальные стационарные СтП — если они существуют, то $m_t = m^*$, $K_t = K^*$, $K(t_1, t_2) = k(\tau)$ ($\tau = t_1 - t_2$) — определяются уравнениями [6–8]:

$$\left. \begin{aligned} a_1^{ин} (m^*, K^*) &= 0 ; \quad a_2^{ин} (m^*, K^*) = 0 ; \\ \dot{k}_\tau (\tau) &= a_{21}^{ин} (m^*, K^*) K^{*-1} k(\tau) ; \\ k(0) &= K^* \quad (\forall \tau > 0) ; \quad k(\tau) = k(-\tau)^T \quad (\forall \tau < 0) . \end{aligned} \right\} \quad (12)$$

При этом необходимо, чтобы матрица $a_{21}^{\text{иин}}(m^*, K^*) = a_{21}^{\text{иин*}}$ была асимптотически устойчивой.

В случае СтС (8) уравнения МНА переходят в известные уравнения МСЛ [7–9]:

$$\left. \begin{aligned} \dot{m}_t &= a_1^{\text{иин}}(m_t, K_t, t), \quad m_0 = m(t_0); \\ \dot{K}_t &= k_1^{\text{иина}}(m_t, K_t, t) K_t + K_t k_1^{\text{иина}}(m_t, K_t, t)^T + \sigma_0(t), \quad K_0 = K(t_0); \\ \frac{\partial K(t_1, t_2)}{\partial t_2} &= K(t_1, t_2) K_{t_2} k_1^{\text{иина}}(m_{t_2}, K_{t_2}, t_2)^T, \quad K(t_1, t_2) = K_{t_1}, \end{aligned} \right\} \quad (13)$$

где

$$\begin{aligned} a^{\text{иин}}(Y_t, t) &= a_0^{\text{иин}}(m_t, K_t) + k_1^{\text{иина}}(m_t, K_t) Y_t^0; \\ k_1^{\text{иина}}(m_t, K_t, t) &= \left[\left(\frac{\partial}{\partial m_t} \right) a_0^{\text{иин}}(m_t, K_t, t)^T \right]^T; \\ b^{\text{иин}}(Y_t, t) &= b_0^{\text{иин}}(t); \quad \sigma^{\text{иин}}(Y_t, t) = b_0^{\text{иин}}(t) \nu(t) b_0^{\text{иин}}(t)^T = \sigma_0^{\text{иин}}(t). \end{aligned}$$

При условии асимптотической устойчивости матрицы $k_1^a(m^*, K^*)$ для стационарных СтС (8) в основе МСЛ лежат уравнения (12), записанные в виде:

$$\left. \begin{aligned} a_0^{\text{иин}}(m^*, K^*) &= 0; \quad k_1^{\text{иина}}(m^*, K^*) K^* + K^* k_1^{\text{иина}}(m^*, K^*)^T + \bar{\sigma}_0^{\text{иин*}} = 0; \\ \dot{k}_\tau(\tau) &= k_1^{\text{иина}}(m^*, K^*) k(\tau), \quad k(0) = K^* \quad (\forall \tau > 0); \\ k(\tau) &= k(-\tau)^T \quad (\forall \tau < 0). \end{aligned} \right\} \quad (14)$$

Таким образом, в основе МАМ лежат следующие утверждения.

Теорема 3.1. Если существуют интегралы (11), то уравнения (9), (10) лежат в основе нестационарных алгоритмов МАМ для негауссовских СтС (7), а уравнения (13) — для негауссовских СтС (8).

Теорема 3.2. Если СтС (7) и (8) стационарны и существует стационарный нормальный процесс и матрица a_{21}^* асимптотически устойчива, то уравнения (12) и (14) лежат в основе стационарных алгоритмов МАМ.

Замечание. Для гауссовских СтС алгоритмы упрощаются, если принять $c^{\text{иин}}(Y_t, t, v) \equiv 0$ в (7) и $V = V_0$, $\nu^V = v^{V_0}$ в (8).

Для алгоритмизации МНА необходимо уметь вычислять следующие интегралы:

$$\begin{aligned} I_0^{\text{иина}} &= I_0^{\text{иина}}(m_t, K_t, t) = a_1^{\text{иин}}(m_t, K_t, t) = \mathbb{M}_{\Delta_y}^{\mathcal{N}} [a^{\text{иин}}(Y_t, t)]; \\ I_1^{\text{иина}} &= I_1^{\text{иина}}(m_t, K_t, t) = a_{21}^{\text{иин}}(m_t, K_t, t) = \mathbb{M}_{\Delta_y}^{\mathcal{N}} \left[a^{\text{иин}}(Y_t, t) Y_t^{0T} \right]; \end{aligned} \quad (15)$$

$$I_0^{\text{ин}\bar{\sigma}} = I_0^{\text{ин}\bar{\sigma}}(m_t, K_t, t) = a_{22}^{\text{ин}}(m_t, K_t, t) = M^{\mathcal{N}}[\bar{\sigma}^{\text{ин}}(Y_t, t)] ,$$

а для МСЛ достаточно вычислить первый интеграл в (15), причем интеграл $I_1^{\text{ин}}$ вычисляется по формуле [7–9]:

$$k_1^{\text{ин}a} = k_1^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) I_0^{\text{ин}a}(m_t, K_t, t)^T \right]^T .$$

Уравнения МНА (МСЛ) содержат интегралы $I_0^{\text{ин}a}$, $I_1^{\text{ин}a}$ и $I_0^{\text{ин}\sigma}$ в виде соответствующих коэффициентов, поэтому процедура вычисления интегралов должна быть согласована с методом численного решения обыкновенных дифференциальных уравнений для m_t , K_t и $K(t_1, t_2)$. Эти коэффициенты допускают дифференцирование по m_t и K_t , так как под интегралом стоит гладкая нормальная плотность.

В [10] изложены алгоритмы дискретного аналитического и статистического моделирования типовых распределений (в том числе нормальных) в нелинейных СтС на многообразиях. Алгоритмы дискретного аналитического и статистического моделирования для СтС с ИН, а также смешанные алгоритмы различной степени точности относительно шага интегрирования также представлены в [10].

4 Интегральный осциллятор в стохастической среде

Рассмотрим нелинейную двумерную СтС, описываемую следующими уравнениями:

$$\begin{aligned} \dot{Y}_1 &= Y_2, & \dot{Y}_2 &= -\omega_0^2 \int_0^{Y_1} \psi^{\text{ин}}(Y_1) dY_1 + L_0 - 2\varepsilon\omega_0 Y_2 + h^2 V; \\ Y_1(t_0) &= Y_{10}; & Y_2(t_0) &= Y_{20}. \end{aligned}$$

Здесь Y_1 и Y_2 — координата и скорость; $\psi^{\text{ин}}(Y_1)$ — интегральная функция, определяющая ИО; L_0 — постоянная или медленно меняющаяся на интервале времени $T = 2\pi\omega_0^{-1}$ величина; $-2\varepsilon\omega_0 Y_2$ — линейная диссипативная составляющая; $h^2 V$ — стохастическая составляющая, представляющая собой нормальный белый шум интенсивности $h^2\nu$; ε , h и ω_0^2 — постоянные параметры; Y_{10} и Y_{20} — нормальные независимые начальные координата и скорость.

Применяя уравнения разд. 3 для $D_1 = K_{11}$, $D_2 = K_{22}$, K_{12} и различных ИН, получим искомые уравнения аналитического моделирования нормальных процессов в ИО:

$$\left. \begin{aligned} \dot{m}_1 &= 0, & m_1(t_0) &= m_{10}; \\ \dot{m}_2 &= -\omega_0^2 \varphi_0(m_1, D_1) + L_0 - 2\varepsilon\omega_0 m_2, & m_2(t_0) &= m_{20}; \end{aligned} \right\} \quad (16)$$

$$\left. \begin{aligned} \dot{D}_1 &= 2K_{12}, \quad D_1(t_0) = D_{10}; \\ \dot{D}_2 &= -2[\omega_0^2\varphi_0(m_1, D_1)K_{12} + 2\varepsilon\omega_0 D_2] + h^2\nu, \quad D_2(t_0) = D_{20}; \\ \dot{K}_{12} &= D_2 - \omega_0^2 k_1^\varphi(m_1, D_1)D_1 - 2\varepsilon\omega_0 K_{12}, \quad K_{12}(t_0) = K_{120}. \end{aligned} \right\} \quad (17)$$

Коэффициенты $\varphi_0 = \varphi_0(m_1, D_1)$ и $k_1^\varphi = k_1^\varphi(m_1, D_1)$ для типовых ИН приведены в приложениях П1–П3.

При $\varepsilon > 0$ для стационарного случая уравнения (16) и (17) принимают вид:

$$\begin{aligned} m_1 &= m_1^*; \quad m_2 = m_2^* = 0; \quad -\omega_0^2\varphi_0(m_1^*, D_1^*) + L_0 = 0; \\ K_{12} &= K_{12}^* = 0; \quad D_2^* = \frac{h^2\nu}{4\varepsilon\omega_0}; \quad \omega_0^2 D_1^* k_1^{\text{иН}}(m_1^*, D_1^*) = D_2^*. \end{aligned}$$

Находим

$$m_1^* = 0; \quad m_2^* = 0; \quad K_{12}^* = 0; \quad D_2^* = \frac{h^2\nu}{4\varepsilon\omega_0},$$

при этом D_1^* определяется из уравнения

$$D_1^* k_1^{\text{иН}}(0, D_1^*) = \gamma \quad \left(\gamma = \frac{h^2\nu}{4\varepsilon\omega_0^3} \right).$$

5 Заключение

Представлено общее методическое и алгоритмическое обеспечение аналитического моделирования нормальных (гауссовских) процессов в дифференциальных СтС с ИН на основе МСЛ и МНА.

Для ИН, описываемых функциями Лапласа, интегралами Френеля и интегральным синусом, получены выражения для коэффициентов МСЛ и МНА. Алгоритмы МАМ положены в основу модуля разрабатываемого инструментального программного обеспечения StS-Analysis.2017.

В приложениях П1–П3 приведены необходимые сведения из теории функций Лапласа, интегралов Френеля и интегрального синуса, а также коэффициенты МСЛ ИН по методу степенных разложений. В качестве тестового примера рассмотрена стохастическая динамика ИО. Проведенные вычислительные эксперименты показывают достаточную точность моделирования для технических приложений в задачах безопасности.

Результаты допускают обобщения на случай аппроксимации ИН эрмитовыми, чебышевскими и другими представлениями. Особый интерес представляют вероятностные ИН в виде функций распределения от случайных величин.

Приложения

П1. Нелинейности Лапласа

$$z = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_0^y e^{-\eta^2/2} d\eta = \frac{1}{2} \operatorname{erf}\left(\frac{y}{\sqrt{2}}\right); \quad z = \operatorname{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-\eta^2} d\eta; \quad (18)$$

$$\Phi(-y) = -\Phi(-y), \quad \Phi(0) = 0, \quad \Phi(\infty) = \frac{1}{2}; \quad (19)$$

$$\operatorname{erf}(y) = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k y^{2k+1}}{k!(2k+1)} = \frac{2}{\sqrt{\pi}} e^{-y^2} \sum_{k=0}^{\infty} \frac{2^k y^{2k+1}}{(2k+1)!!}. \quad (20)$$

При $m_y = m = 0$ и $D_y = D \neq 0$, пользуясь (18)–(20) и табличными интегралами [4, 6]

$$\int_0^\infty x^{2n} e^{-px^2} dx = \frac{(2n-1)!!}{2(2p)^n} \sqrt{\frac{\pi}{p}}; \quad \int_0^\infty x^{2n+1} e^{-px^2} dx = \frac{n!}{2p^{n+1}} \left(p = \frac{1}{2D} > 0\right), \quad (21)$$

получим

$$\varphi_0^\Phi(0, D) = 0, \quad k_1^\Phi(0, D) = \sum_{k=0}^n A_k^\Phi D^{k+1}; \quad A_k^\Phi = \frac{1}{\sqrt{2\pi}} \frac{(-1)^k (2k+1)!!}{k!(2k+1)} \cdot 2^{-k}.$$

При $m_y = m \neq 0$ и $D_y = D \neq 0$, пользуясь (20) и табличными интегралами [4–6]

$$\int_{-\infty}^\infty x^n \exp(-px^2 + 2qx) dx = n! \exp\left(\frac{q^2}{p}\right) \sqrt{\frac{\pi}{p}} \sum_{l=0}^{\text{E}[n/2]} \frac{1}{(n-2l)!!} \left(\frac{p}{4q^2}\right)^l \quad (22)$$

($\text{E}[\cdot]$ — символ целой части), получаем

$$\begin{aligned} \varphi_0^\Phi(m, D) &= \sum_{k=0}^n (-1)^k \varphi_{0k}(\zeta^2); \quad k_1^\varphi(m, D) = \frac{\partial \varphi_0^\Phi}{\partial m}; \\ \varphi_{0k}(\zeta^2) &= \sum_{l=0}^{\text{E}[(2k+1)/2]} \frac{2}{[2(k-l)+1]!!} \zeta^{2l}, \quad \zeta^2 = \frac{D}{2m^2}. \end{aligned}$$

П2. Нелинейности, выражаемые интегралами Френеля

$$z = C(y) = \int_0^y \cos \frac{\pi}{2} \eta^2 d\eta; \quad z = S(y) = \int_0^y \sin \frac{\pi}{2} \eta^2 d\eta;$$

$$C(-y) = -C(y); \quad S(-y) = -S(y); \quad C(\infty) = S(\infty) = \frac{1}{2};$$

$$\begin{aligned} C(y) &= \sum_{k=0}^{\infty} \frac{(-1)^k (\pi/2)^{2k}}{(2k)!(4k+1)} y^{4k+1} = \\ &= \cos \frac{\pi}{2} y^2 \sum_{k=0}^{\infty} \frac{(-1)^k \pi^{2k}}{(4k+1)!!} y^{4k+1} + \sin \frac{\pi}{2} y^2 \sum_{k=0}^{\infty} \frac{(-1)^k \pi^{2k+1}}{(4k+1)!!} y^{4k+3}; \end{aligned}$$

$$\begin{aligned} S(y) &= \sum_{k=0}^{\infty} \frac{(-1)^k (\pi/2)^{2k+1}}{(2k+1)!(4k+3)} y^{4k+1} = \\ &= -\cos \frac{\pi}{2} y^2 \sum_{k=0}^{\infty} \frac{(-1)^k \pi^{2k+1}}{(4k+3)!!} y^{4k+3} + \sin \frac{\pi}{2} y^2 \sum_{k=0}^{\infty} \frac{(-1)^k \pi^{2k}}{(4k+1)!!} y^{4k+1}. \end{aligned}$$

При $m_y = m = 0$ и $D_y = D \neq 0$, пользуясь интегралами (21), получаем:

$$\begin{aligned} \varphi_0^C(0, D) &= 0; \quad k_1^C(0, D) = \sum_{k=0}^n A_k^C D^{2k+1}, \quad A_k^C = \frac{(-1)^k (\pi/2)^{2k} (4k+1)!!}{(2k)!(4k+1)}; \\ \varphi_0^S(0, D) &= 0; \quad k_1^S(0, D) = \sum_{k=0}^n A_k^S D^{2(k+1)}, \quad A_k^S = \frac{(-1)^k \pi^{2k+1} (4k+3)!!}{(4k+3)(2k+1)!}. \end{aligned}$$

При $m_y = m \neq 0$ и $D_y = D \neq 0$, пользуясь табличными интегралами (22), находим

$$\begin{aligned} \varphi_0^C(m, D) &= \frac{(-1)^k (\pi/2)^{2k} (4k+1)!}{(2k)!(4k+1)} \chi_k^C(\zeta^2); \\ \varphi_0^S(m, D) &= \frac{(-1)^k (\pi/2)^{2k+1} (4k+1)!}{(2k+1)!(4k+3)} \chi_k^S(\zeta^2); \\ k_1^C(m, D) &= \frac{\partial \varphi_0^C}{\partial m}; \quad k_1^S(m, D) = \frac{\partial \varphi_0^S}{\partial m}, \end{aligned}$$

где

$$\chi_k^C(\zeta^2) = \chi_k^S(\zeta^2) = \sum_{l=0}^{\lfloor (4k+1)/2 \rfloor} \frac{1}{(4k-2l+1)!! l!} \zeta^{2l}, \quad \zeta^2 = \frac{D}{2m^2}.$$

П3. Интегральный синус

$$\begin{aligned} z = \text{Si}(y) &= \int_0^y \frac{\sin \eta}{\eta} d\eta, \quad \text{Si}(-y) = -\text{Si}(y), \quad \text{Si}(\infty) = \frac{\pi}{2}; \\ z = \text{Si}(y) &= \sum_{k=0}^{\infty} \frac{(-1)^k y^{2k+1}}{(2k+1)(2k+1)!}. \end{aligned} \tag{23}$$

При $m_y = m = 0$ и $D_y = D \neq 0$, пользуясь (23) и интегралами (21), получаем

$$\varphi_0^{\text{Si}}(0, D) = 0; \quad k_1^{\text{Si}}(0, D) = \sum_{k=0}^n A_k^{\text{Si}} D^{k+1}, \quad A_k^{\text{Si}} = \frac{(-1)^k (2k+1)!!}{(2k+1)(2k+1)!}.$$

При $m_y = m \neq 0$ и $D_y = 0$, пользуясь (23) и интегралами (22), находим

$$\varphi_0^{\text{Si}}(m, D) = \sum_{k=0}^n \frac{(-1)^k}{(2k+1)} \chi_k^{\text{Si}}(\zeta^2); \quad k_1^{\text{Si}}(m, D) = \frac{\partial \varphi_0^{\text{Si}}}{\partial m},$$

где

$$\zeta^2 = \frac{D}{2m^2}; \quad \chi_k^{\text{Si}}(\zeta^2) = \sum_{l=0}^{\lfloor (2k+1)/2 \rfloor} \frac{1}{[2(k-l)+1]!!} \zeta^{2l}.$$

Литература

1. Синицын И. Н. Аналитическое моделирование распределений в динамических системах с бесселевыми нелинейностями // Информатика и её применения, 2015. Т. 9. Вып. 4. С. 37–47.
2. Синицын И. Н. Аналитическое моделирование нормальных процессов в стохастических системах со сложными бесселевыми нелинейностями дробного порядка // Информатика и её применения, 2016. Т. 10. Вып. 3. С. 25–35.
3. Синицын И. Н. Аналитическое моделирование нормальных процессов в стохастических системах с эллиптическими нелинейностями // Системы и средства информатики, 2017. Т. 27. № 1. С. 3–19.
4. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. — М.: ГИФМЛ, 1963. 1100 с.
5. Справочник по специальным функциям / Под ред. М. Абрамовича, И. Стигана. — М.: Наука, 1979. 832 с.
6. Попов Б. А., Теслер Г. С. Вычисление функций на ЭВМ: Справочник. — Киев: Наукова думка, 1984. 599 с.
7. Пугачёв В. С., Синицын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с.
8. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
9. Синицын И. Н., Синицын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
10. Синицын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.

Поступила в редакцию 02.03.17

ANALYTICAL MODELING OF NORMAL PROCESSES IN STOCHASTIC SYSTEMS WITH INTEGRAL NONLINEARITIES (I)

I. N. Sinitsyn

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: General methodological and algorithmical support for analytical modeling of normal processes in differential stochastic systems (StS) with integral nonlinearities (IN) and Wiener and Poisson noises is presented. Support is based on the methods of normal approximation (MNA) and of statistical linearization (MSL). Integral nonlinearities were approximated by power and Hermite series. The MSL and MNA coefficients for IN described by Laplace, Fresnel integrals, and sine integrals are given. Necessary information about IN is given and the software tool StS-Analysis.2017 is described. Stochastic dynamic of an integral oscillator is used as a test example. Some generalizations are mentioned.

Keywords: analytical modeling; Fresnel integral; Hermite polynomial power expansion; integral sine; Laplace function; method of normal approximation (MNA); method of statistical linearization (MSL)

DOI: 10.14357/08696527170201

References

1. Sinitsyn, I. N. 2015. Analiticheskoe modelirovaniye raspredeleniy v dinamicheskikh sistemakh s besselevymi nelineynostyami [Analytical modeling of distributions in dynamical systems with Bessel nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):37–47.
2. Sinitsyn, I. N. 2016. Analiticheskoe modelirovaniye normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi besselevymi nelineynostyami drobnogo poryadka [Analytical modeling of normal processes in stochastic systems with complex Bessel nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 10(3):25–35.
3. Sinitsyn, I. N. 2017. Analiticheskoe modelirovaniye normal'nykh protsessov v stokhasticheskikh sistemakh s ellipticheskimi nelineynostyami [Analytical modeling of normal processes in stochastic systems with elliptic nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(1):3–19.
4. Gradshteyn, I. S., and I. M. Ryzhik. 1963. *Tablitsy integralov, summ, ryadov i proizvedenyi* [Tables of integrals, sums, and series]. Moscow: GIFML. 1100 p.
5. Abramovich, M., and I. Stigan, eds. *Spravochnik po spetsial'nym funktsiyam* [Computing of functions: Handbook]. Moscow: Nauka. 1979. 832 p.
6. Popov, B. A., and G. S. Tesler. 1984. *Vychislenie funktsiy na EVM: Spravochnik* [Handbook for special functions]. Kiev: Naukova Dumka. 599 p.

7. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York, NY: John Wiley. 549 p.
8. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
9. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation in stochastic systems]. Moscow: TORUS PRESS. 488 p.
10. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovaniye raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraziyakh [Parametric statistical and analytical modeling of distributions in stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.

Received March 02, 2017

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

СТАТИСТИЧЕСКИЙ АНАЛИЗ МАКСИМУМ-ЛИНЕЙНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ*

A. V. Лебедев¹

Аннотация: Рассматриваются максимум-линейные случайные процессы. Предполагается, что инновации и значения процесса имеют стандартное распределение Фреше. Предлагаются простые статистические оценки параметров. Проводится сравнение некоторых оценок параметра процесса максимум-авторегрессии первого порядка по дисперсии.

Ключевые слова: максимум-линейные процессы; процессы максимум-авторегрессии; распределение Фреше; статистический анализ; тяжелые хвосты

DOI: 10.14357/08696527170202

1 Введение

В последние несколько десятилетий в прикладной теории вероятностей и статистике произошел существенный сдвиг в понимании того, что далеко не все можно описать нормальным распределением. Напротив, как выяснилось, многие явления в природе и обществе, технике и экономике требуют рассмотрения тяжелых хвостов.

Часто дело не только в тяжелых хвостах, но и в структуре зависимости случайных величин, нехарактерной для гауссовских моделей. Например, экстремальные события (превышения высокого уровня) могут иметь тенденцию происходить не по одиночке, а образовывать группы (кластеры) [1, гл. 8]. Заметим, что превышения высокого уровня для компонент невырожденного многомерного нормального распределения оказываются асимптотически независимы [2, с. 211], что может не вполне отражать действительность.

Между тем явление кластеризации возникает (и уже довольно хорошо изучено) в ARMA-процессах с инновациями, распределения которых имеют тяжелые хвосты [1, гл. 5, 7], в ARCH-процессах, известных в финансовой математике [1, § 8.4], и др.

В современной теории случайных процессов интересным классом объектов, позволяющим моделировать как тяжелые хвосты, так и экстремальные структуры

*Работа поддержана РФФИ (проект 14-01-00075).

¹Московский государственный университет имени М. В. Ломоносова, механико-математический факультет, кафедра теории вероятностей, avlebed@yandex.ru

зависимости, являются max-ARMA или MARMA-процессы, введенные в [3, 4] по аналогии с ARMA путем замены сложения на максимум¹:

$$X_n = \phi_1 X_{n-1} \vee \cdots \vee \phi_p X_{n-p} \vee \theta_0 Z_n \vee \cdots \vee \theta_q Z_{n-q}, \quad (1)$$

где $\phi_i, \theta_j \geq 0$, $1 \leq i \leq p$, $0 \leq j \leq q$; инновации Z_n независимы и одинаково распределены на \mathbf{R}_+ .

MARMA-процессы оказываются удобней для анализа, чем ARMA, поскольку конечномерные распределения для них легко выписываются в явном виде. Как показано в [3] путем компьютерного моделирования, их траектории ведут себя сходным образом в случае неотрицательных инноваций с тяжелыми (степенными) хвостами.

Отмечены приложения подобных процессов и их обобщений для описания различных явлений в природе, технике и экономике, например: оценки полезности оборудования, очередей с приоритетами [3], обмена воды во фьордах [5], накопления тепловой солнечной энергии [6], выпадения осадков [7], финансовых рядов [8]. Процессам максимум-авторегрессии посвящены, например, работы [9, 10], а также относительно недавние статьи [11, 12].

Удобно рассматривать максимум-устойчивые MARMA-процессы, в которых все величины имеют распределение Фреше $\Phi_{\alpha,\sigma}(x) = \exp\{-(x/\sigma)^{-\alpha}\}$, $\alpha, \sigma > 0$, $x > 0$. Оно играет здесь такую же роль, как и нормальное распределение в классическом случае. Для определенности далее будем полагать $\alpha = 1$, а под «стандартным» распределением Фреше будем иметь в виду $\Phi_{1,1}$.

Для независимых случайных величин Z_1, Z_2 и Z с таким распределением верно соотношение²:

$$c_1 Z_1 \vee c_2 Z_2 \stackrel{d}{=} (c_1 + c_2) Z, \quad c_{1,2} \geq 0. \quad (2)$$

Одной из проблем на пути широкого применения MARMA-процессов является нехватка эффективных статистических процедур для них (по оценке параметров, прогнозированию и т. п.). Понятно, что классические методы для временных рядов [13] здесь не применимы (поскольку случайные величины не имеют конечных моментов первого и второго порядка).

Работы [4, 12, 14, 15] посвящены статистическому анализу MARMA-процессов в простейших случаях.

В разд. 2 рассмотрим более общие максимум-линейные процессы

$$X_n = \bigvee_{k=0}^{+\infty} \psi_k Z_{n-k}, \quad \sum_{k=0}^{+\infty} \psi_k = 1, \quad (3)$$

¹Знаком \vee далее обозначен максимум.

²Через $\stackrel{d}{=}$ обозначено равенство по распределению.

где последовательность Z_n продолжена в область $n \leq 0$, причем все X_n и Z_n имеют стандартное распределение Фреше. Очевидно, любой нормированный стационарный MARMA-процесс можно представить в виде (3). Это свойство называют причинностью или каузальностью (causality) [3]. Коэффициенты ψ_k , $k \geq 0$, выражаются через ϕ_i , θ_j , $1 \leq i \leq p$, $0 \leq j \leq q$. Построим для них статистические оценки.

Отметим, что в работе [16] изучалась модель топа новостей на основе экстремального дробового шума. Дискретизация этого шума по времени приводит к (3).

В разд. 3 рассмотрим еще более общие процессы

$$X_n = \bigvee_{k=-\infty}^{+\infty} \psi_k Z_{n-k}, \quad \sum_{k=-\infty}^{+\infty} \psi_k = 1, \quad (4)$$

зависящие от будущего (некаузальные). Впрочем, эту зависимость не следует понимать буквально. Речь может идти о тенденциях, которые развиваются в настоящем и максимально проявляются в будущем, хотя их причины лежат в прошлом. Или имеется в виду не временной, а пространственный параметр n . Для таких процессов построим оценки коэффициентов ψ_k , $k \in \mathbf{Z}$, в предположении их симметричности относительно нуля и убывания в обе стороны.

В разд. 4 обсудим оценки параметра максимум-авторегрессии первого порядка, предложенные в [12], с точки зрения минимизации дисперсии.

2 Основные результаты (каузальные процессы)

Введем дополнительное условие (*): последовательность $\{\psi_k\}$ не возрастает. Если инновации описывают независимые факторы, возникающие в последовательные моменты времени, то это условие означает, что со временем их влияние на процесс может только уменьшаться или какое-то время оставаться неизменным, но не возрастать. Такое поведение свойственно устойчивым системам. Оно аналогично свойству обратимости для классических временных рядов [13, гл. 3].

Обозначим $p_k = \mathbf{P}(X_{k+1} \leq X_1)$, $k \geq 1$. Для таких вероятностей существуют простые оценки

$$\hat{p}_k = \frac{1}{n-k} \sum_{m=k+1}^n \mathbf{I}(X_m \leq X_{m-k}), \quad n > k. \quad (5)$$

Теорема 1. Для процесса (3) при условии (*) верно

$$p_k = \left(1 + \sum_{l=0}^{k-1} \psi_l \right)^{-1}, \quad k \geq 1.$$

Доказательство. В доказательстве используем соотношение (2) и лемму [14]: для независимых случайных величин Z_1 и Z_2 со стандартным распределением Фреше верно равенство

$$\mathbf{P}(\alpha Z_1 \leq Z_2) = \frac{1}{\alpha + 1}, \quad \alpha > 0. \quad (6)$$

Имеем

$$\begin{aligned} X_1 &= \bigvee_{l=0}^{\infty} \psi_l Z_{1-l}; \\ X_{k+1} &= \bigvee_{l=0}^{\infty} \psi_l Z_{k+1-l} = \left(\bigvee_{l=0}^{k-1} \psi_l Z_{k+1-l} \right) \vee \left(\bigvee_{l=k}^{\infty} \psi_l Z_{k+1-l} \right) = \\ &= \left(\bigvee_{l=0}^{k-1} \psi_l Z_{k+1-l} \right) \vee \left(\bigvee_{j=0}^{\infty} \psi_{j+k} Z_{1-j} \right), \end{aligned}$$

где

$$\bigvee_{j=0}^{\infty} \psi_{j+k} Z_{1-j} \leq X_1.$$

В силу (2) и (6) получаем:

$$p_k = \mathbf{P}(X_{k+1} \leq X_1) = \mathbf{P}\left(\bigvee_{l=0}^{k-1} \psi_l Z_{k+1-l} \leq X_1\right) = \left(1 + \sum_{l=0}^{k-1} \psi_l\right)^{-1}. \quad \square$$

Следствие 1. Для процесса (3) при условии (*) верно

$$\psi_0 = \frac{1}{p_1} - 1, \quad \psi_k = \frac{1}{p_{k+1}} - \frac{1}{p_k}, \quad k \geq 1.$$

Таким образом, с помощью следствия 1, исходя из оценок вероятностей (5), можно построить оценки коэффициентов $\hat{\psi}_k$, $k \geq 0$.

Для процесса скользящего максимума произвольного порядка q это сразу дает оценки всех его параметров, поскольку $\theta_j = \psi_j$, $0 \leq j \leq q$. Для процессов, включающих максимум-авторегрессию, необходимо сначала выразить коэффициенты ψ_k через параметры ϕ_i , θ_j (что достаточно просто при небольших p и q), а затем, наоборот, ϕ_i , θ_j через ψ_k (что можно сделать, как правило, не единственным образом).

Пример 1. Рассмотрим процесс скользящего максимума MMA(1). В силу условия (*) имеем $\theta_0 \in [1/2, 1]$, $p_1 \in [1/2, 2/3]$. Получаем $\theta_0 = \psi_0 = 1/p_1 - 1$, что совпадает с результатом [14, следствие 2].

Пример 2. Рассмотрим процесс максимум-авторегрессии¹ MAR(1). В силу условия нормировки $\phi_1 + \theta_0 = 1$ получаем $\phi_1 = 1 - \theta_0 = 1 - \psi_0 = 2 - 1/p_1$, что совпадает с результатом [14, следствие 1]. Однако если воспользоваться равенством $\psi_k = \phi_1^k \theta_0$, $k \geq 1$, в частности $\psi_1 = \phi_1 \theta_0$, получаем:

$$\phi_1 = \frac{\psi_1}{\psi_0} = \frac{p_1 - p_2}{p_2(1 - p_1)},$$

что дает уже другую оценку, и подобных можно придумать еще много.

Например, в [12] изучались оценки величин ϕ_1^m , $m \geq 1$, исходя из соотношения $\phi_1^m = 2 - 1/p_m$, откуда $\phi_1 = (2 - 1/p_m)^{1/m}$.

В следующих примерах приведем лишь наиболее очевидные оценки.

Пример 3. Рассмотрим процесс максимум-авторегрессии MAR(2). Получаем

$$\psi_0 = \theta_0, \quad \psi_1 = \phi_1 \theta_0, \quad \psi_2 = (\phi_1^2 \vee \phi_2) \theta_0, \quad \dots$$

Как показано в [3], этот процесс не сводим к MAR(1), если и только если $\phi_2 > \phi_1^2$; тогда $\psi_2 = \phi_2 \theta_0$. Условие (*) принимает вид: $\phi_2 \leq \phi_1$. В этих предположениях получаем

$$\theta_0 = \psi_0, \quad \phi_1 = \frac{\psi_1}{\psi_0}, \quad \phi_2 = \frac{\psi_2}{\psi_0}.$$

Пример 4. Рассмотрим процесс MARMA(1,1). Получаем

$$\psi_0 = \theta_0, \quad \psi_1 = \theta_1 \vee \phi_1 \theta_0, \quad \psi_2 = \phi_1(\theta_1 \vee \phi_1 \theta_0), \quad \dots$$

Этот процесс не сводим к MAR(1) при $\theta_1 > \phi_1 \theta_0$, тогда $\psi_2 = \phi_1 \theta_1$, а условие (*) дает $\theta_1 \leq \theta_0$. В этих предположениях получаем:

$$\theta_0 = \psi_0; \quad \theta_1 = \psi_1; \quad \phi_1 = \frac{\psi_2}{\psi_1}.$$

Заметим, что оценки $\hat{\psi}_k$, $k \geq 0$, можно использовать не только для оценивания параметров MARMA-процессов (при известных p и q), но и для идентификации модели, подобно выборочным автокорреляциям для классических временных рядов [13, гл. 6]. Например, если $\hat{\psi}_k$ убывают экспоненциально, это указывает на процесс максимум-авторегрессии; если же конечное их число (при малых k) заметно отличается от нуля, а остальные близки к нему, то разумно выбрать модель скользящего максимума. Построение конкретных статистических критериев остается для будущих исследований.

¹ В литературе эти процессы обозначаются также через ARMAX.

Оценки $\hat{\psi}_k$, $k \geq 0$, очевидно, асимптотически нормальны, поскольку выражаются через асимптотически нормальные оценки \hat{p}_k , $k \geq 1$. Однако вычисление асимптотических дисперсий для $\hat{\psi}_k$, $k \geq 0$, в явном виде затруднительно в силу зависимости между собой оценок \hat{p}_k , $k \geq 1$, и нелинейности выражений.

Заметим, что поскольку отношения «больше–меньше» инвариантны относительно любого строго монотонного преобразования данных, предлагаемый метод работает и в случае, когда наблюдаются не сами значения максимум-линейного процесса, а произвольные строго монотонные функции от них. На эти отношения не влияет и амплитуда (масштабный фактор) колебаний. Таким образом, метод должен действовать и для процессов, у которых амплитуда меняется медленно или редко.

3 Основные результаты (некаузальные процессы)

Введем дополнительное условие (**): коэффициенты ψ_k , $k \in \mathbf{Z}$, симметричны относительно нуля и строго убывают в обе стороны, когда положительны, т. е.

- (1) $\psi_{-k} = \psi_k$, $k \geq 0$;
- (2) $\psi_k > \psi_l$ при $0 \leq k < l$ и $\psi_k > 0$.

В данном случае процесс стохастически симметричен относительно замены направления времени и $\mathbf{P}(X_{k+1} < X_1) = \mathbf{P}(X_{k+1} > X_1)$, так что вероятности p_k , $k \geq 1$, для его исследования практически бесполезны.

Введем вероятности $r_k = \mathbf{P}(X_n < X_{n-k} \vee X_{n+k})$, $k \geq 1$, $n > k$. Для таких вероятностей существуют простые оценки:

$$\hat{r}_k = \frac{1}{n-2k} \sum_{m=k+1}^{n-k} \mathbf{I}(X_m \leq X_{m-k} \vee X_{m+k}), \quad n > 2k. \quad (7)$$

Используем эквивалентное (4) представление

$$X_n = \bigvee_{l=-\infty}^{+\infty} \psi_{n-l} Z_l. \quad (8)$$

Введем обозначения:

$$X_n^{(i,j)} = \bigvee_{l=i}^j \psi_{n-l} Z_l; \quad S_{i,j} = \sum_{l=i}^j \psi_l, \quad i \leq j,$$

тогда в силу (2) верно

$$X_n^{(i,j)} \stackrel{d}{=} S_{n-j,n-i} Z_1.$$

При $i > j$ положим $X_n^{(i,j)} = 0$, $S_{i,j} = 0$.

Для удобства сравнения случайных величин докажем следующую лемму.

Лемма 1. Если $b < c$ и $d < a$, то $a \vee b < c \vee d$ эквивалентно $a < c$.

Доказательство. Утверждение $a \vee b < c \vee d$ эквивалентно системе неравенств $a < c \vee d$ и $b < c \vee d$, но второе из них заведомо верно в силу $b < c$. Утверждение $a < c \vee d$ эквивалентно совокупности неравенств $a < c$ и $a < d$, но второе из них невозможно в силу $d < a$. Таким образом, эквивалентность доказана. \square

Теорема 2. Для процесса (4) при условии (**) верно

$$r_{2k-1} = \frac{1}{2} \left(1 + \frac{1}{1 + 2(2S_{0,k-1} - \psi_0)} \right), \quad k \geq 1.$$

Доказательство. Используем разложения:

$$\begin{aligned} X_{n-(2k-1)} &= X_{n-(2k-1)}^{(-\infty, n-k)} \vee X_{n-(2k-1)}^{(n-k+1, n-1)} \vee \psi_{2k-1} Z_n \vee X_{n-(2k-1)}^{(n+1, +\infty)}; \\ X_{n+(2k-1)} &= X_{n+(2k-1)}^{(-\infty, n-1)} \vee \psi_{2k-1} Z_n \vee X_{n+(2k-1)}^{(n+1, n+k-1)} \vee X_{n+(2k-1)}^{(n+k, +\infty)}; \\ X_n &= X_n^{(-\infty, n-k)} \vee X_n^{(n-k+1, n+k-1)} \vee X_n^{(n+k, +\infty)}. \end{aligned}$$

В силу (**) верно

$$X_{n-(2k-1)}^{(n+1, +\infty)} < X_{n+(2k-1)}^{(n+1, +\infty)}; \quad X_{n-(2k-1)}^{(-\infty, n-1)} < X_{n+(2k-1)}^{(-\infty, n-1)},$$

поэтому

$$\begin{aligned} X_{n-(2k-1)} \vee X_{n+(2k-1)} &= \\ &= X_{n-(2k-1)}^{(-\infty, n-k)} \vee X_{n-(2k-1)}^{(n-k+1, n-1)} \vee \psi_{2k-1} Z_n \vee X_{n+(2k-1)}^{(n+1, n+k-1)} \vee X_{n+(2k-1)}^{(n+k, +\infty)}. \end{aligned}$$

В силу (**) верно

$$\begin{aligned} X_n^{(-\infty, n-k)} &< X_{n-(2k-1)}^{(-\infty, n-k)}; \quad X_n^{(n+k, +\infty)} < X_{n+(2k-1)}^{(n+k, +\infty)}; \\ X_{n-(2k-1)}^{(n-k+1, n-1)} \vee \psi_{2k-1} Z_n \vee X_{n+(2k-1)}^{(n+1, n+k-1)} &< X_n^{(n-k+1, n+k-1)}, \end{aligned}$$

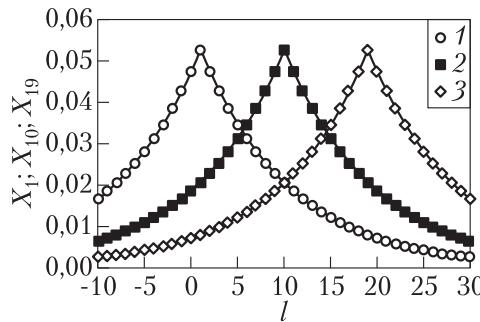
отсюда по лемме 1 заключаем, что неравенство $X_n < X_{n-(2k-1)} \vee X_{n+(2k-1)}$ эквивалентно $X_n^{(n-k+1, n+k-1)} < X_{n-(2k-1)}^{(-\infty, n-k)} \vee X_{n+(2k-1)}^{(n+k, +\infty)}$.

С помощью (2) и (6) получаем:

$$\begin{aligned} r_{2k-1} &= \mathbf{P}(X_n < X_{n-(2k-1)} \vee X_{n+(2k-1)}) = \\ &= \mathbf{P}\left(X_n^{(n-k+1, n+k-1)} < X_{n-(2k-1)}^{(-\infty, n-k)} \vee X_{n+(2k-1)}^{(n+k, +\infty)}\right) = \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{P}(S_{-(k-1), k-1} Z_1 < 2S_{-\infty, k-1} Z_2) = \frac{2S_{-\infty, k-1}}{S_{-(k-1), k-1} + 2S_{-\infty, k-1}} = \\
 &= \frac{1 + (2S_{0, k-1} - \psi_0)}{1 + 2(2S_{0, k-1} - \psi_0)} = \frac{1}{2} \left(1 + \frac{1}{1 + 2(2S_{0, k-1} - \psi_0)} \right). \square
 \end{aligned}$$

В качестве иллюстрации к доказательству теоремы 2 на рисунке представлены коэффициенты в разложениях вида (8) к Z_l , $-10 \leq l \leq 30$, для X_1 , X_{10} и X_{19} при $\psi_k = (1/19) \cdot 0,9^{|k|}$, $k \in \mathbf{Z}$. В данном случае $n = 10$, $k = 5$, $2k - 1 = 9$.



Коэффициенты в разложении (8) к Z_l для X_1 (1), X_{10} (2) и X_{19} (3)

Следствие 2. Для процесса (4) при условии (**) верно

$$\psi_0 = \frac{1}{2} \left(\frac{1}{2r_1 - 1} - 1 \right), \quad \psi_k = \frac{1}{4} \left(\frac{1}{2r_{2k+1} - 1} - \frac{1}{2r_{2k-1} - 1} \right), \quad k \geq 1.$$

Доказательство. По теореме 2 получаем:

$$2S_{0, k-1} - \psi_0 = \frac{1}{2} \left(\frac{1}{2r_{2k-1} - 1} - 1 \right),$$

что при $k = 1$ дает выражение для ψ_0 , а при $k \geq 1$ имеем

$$\psi_k = \frac{2S_{0, k} - \psi_0 - (2S_{0, k-1} - \psi_0)}{2} = \frac{1}{4} \left(\frac{1}{2r_{2k+1} - 1} - \frac{1}{2r_{2k-1} - 1} \right).$$

Таким образом, с помощью следствия 2, исходя из оценок вероятностей (7), можно построить оценки коэффициентов $\hat{\psi}_k$, $k \in \mathbf{Z}$.

Сопоставим теперь процессы (3) и (4), коэффициенты ψ_k которых убывают в геометрической прогрессии с показателем $0 < \phi < 1$.

Для процесса (3) с $\psi_k = \psi_0\phi^k$, $k \geq 0$, из условия нормировки следует $\psi_0 = 1 - \phi$. Этот процесс эквивалентен MAR(1) и допускает представление

$$X_n = \phi X_{n-1} \vee (1 - \phi)Z_n, \quad n \geq 1,$$

из которого следует также, что он является марковским процессом.

Как было отмечено выше, имеет место равенство $\phi = 2 - 1/p_1$, откуда получаем оценку $\hat{\phi} = 2 - 1/\hat{p}_1$.

Для процесса (4) с $\psi_k = \psi_0\phi^{|k|}$, $k \in \mathbf{Z}$, из условия нормировки следует

$$\psi_0 = \frac{1 - \phi}{1 + \phi}. \quad (9)$$

Этот процесс, разумеется, немарковский, но допускает представление

$$X_n = \phi X_{n-1} \vee \phi X_{n+1} \vee \frac{1 - \phi}{1 + \phi} Z_n, \quad n \in \mathbf{Z},$$

откуда следует, что он представляет собой пример марковского случайного поля на всей целочисленной оси в том смысле, что значения поля условно независимы при известных значениях в соседних точках (слева и справа).

Что касается оценки параметра ϕ , то с помощью следствия 2 и (9) получаем равенство:

$$\phi = \frac{1 - \psi_0}{1 + \psi_0} = 3 - \frac{2}{r_1},$$

которое приводит к оценке $\hat{\phi} = 3 - 2/\hat{r}_1$. Как и в предыдущем случае, эта оценка не единственная, которую можно получить из коэффициентов ψ_k , но самая простая.

4 Сравнение оценок параметра максимум-авторегрессии первого порядка

Сосредоточимся теперь на процессе MAR(1). Обозначим для простоты $\phi = \phi_1$. В [12] были предложены оценки $\widehat{\phi}^m = 2 - 1/\hat{p}_m$, $m \geq 1$. Из них получается следующее семейство оценок параметра ϕ :

$$\hat{\phi}^{(m)} = \begin{cases} \left(2 - \frac{1}{\hat{p}_m}\right)^{1/m}, & p_m > \frac{1}{2}, \\ 0, & p_m < \frac{1}{2}, \end{cases} \quad m \geq 1.$$

Возникает вопрос, какая из них лучше.

В [12, предложение 2.2] показано (с точностью до замены обозначений), что оценки $\widehat{\phi}^m$ асимптотически нормальны и

$$n^{1/2} \left(\widehat{\phi}^m - \phi^m \right) \xrightarrow{d} N(0, \Delta_m(\phi)) , \quad n \rightarrow \infty ,$$

где

$$\Delta_m(\phi) = -\frac{(\phi^m - 2)^2(\phi^m - 1)^2(\phi^{2m} - 2\phi - 2\phi^m(1 + \phi))}{2\phi + \phi^{3m} + \phi^m(2 + \phi + 6\phi^2 + 4\phi^3) - \phi^{2m}(3 + 2\phi(1 + \phi))} .$$

Следовательно, оценки $\hat{\phi}^{(m)}$ также асимптотически нормальны и

$$n^{1/2} \left(\hat{\phi}^{(m)} - \phi \right) \xrightarrow{d} N(0, \Delta^{(m)}(\phi)) , \quad n \rightarrow \infty ,$$

где

$$\Delta^{(m)}(\phi) = \frac{1}{m^2} \phi^{2(1-m)} \Delta_m(\phi) .$$

Легко проверить, что при любом $m \geq 1$ верно $\Delta_m(\phi) \rightarrow 4$ при $\phi \rightarrow 0$ и $\Delta_m(\phi) \rightarrow 0$ при $\phi \rightarrow 1$. Имеем $\Delta^{(1)}(\phi) = \Delta_1(\phi)$. При $m \geq 2$ имеет место асимптотика $\Delta^{(m)}(\phi) \sim 4m^{-2}\phi^{2(1-m)} \rightarrow +\infty$ при $\phi \rightarrow 0$. Кроме того, при любом $0 < \phi < 1$ верно $\Delta_m(\phi) \rightarrow 4$ при $m \rightarrow \infty$, откуда $\Delta^{(m)}(\phi) \sim 4m^{-2}\phi^{2(1-m)} \rightarrow +\infty$ при $m \rightarrow \infty$.

Аналитическое исследование поведения величины $\Delta^{(m)}(\phi)$ затруднительно, поэтому были проведены ее вычисления, результаты которых приведены в таблице.

Видно, что значения $\Delta^{(m)}(\phi)$ убывают по ϕ и возрастают по m . Это дает основания полагать, что нет смысла пользоваться оценками $\hat{\phi}^{(m)}$ с $m \geq 2$ (если только действительно не нужно оценить именно m -ю степень параметра ϕ , а не сам параметр), а лучше пользоваться простейшей оценкой $\hat{\phi} = 2 - 1/\hat{p}_1$. В последнем случае получаем

$$\Delta^{(1)}(\phi) = \frac{(\phi - 2)^2(\phi - 1)^2(\phi + 4)}{2\phi^3 + 5\phi^2 - 2\phi + 4} .$$

Понятно, что построение асимптотических доверительных интервалов для ϕ^m по $\widehat{\phi}^m$ с последующим их пересчетом в интервалы для ϕ дает асимптотически ту же точность, что и сразу использование оценок $\hat{\phi}^{(m)}$, поэтому не может улучшить ситуацию.

Вычисленные значения величины $\Delta^{(m)}(\phi)$

ϕ	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
0,05	3,55	397,38	71 087,5	$1,6 \cdot 10^7$	$4,1 \cdot 10^9$
0,1	3,11	97,29	4431,96	249 930	$1,6 \cdot 10^7$
0,15	2,69	41,66	869,15	21914,7	624 153
0,2	2,28	22,17	270,9	3886,5	62 436,5
0,25	1,90	13,16	108,07	1010,7	10 451,3
0,3	1,56	8,30	50,00	333,26	2417,6
0,35	1,25	5,40	25,40	128,64	696,62
0,4	0,98	3,57	13,70	55,28	234,45
0,45	0,75	2,37	7,65	25,54	88,17
0,5	0,56	1,56	4,35	12,36	35,82
0,55	0,41	1,01	2,49	6,13	15,30
0,6	0,29	0,65	1,41	3,07	6,72
0,65	0,19	0,40	0,78	1,52	2,96
0,7	0,12	0,23	0,42	0,74	1,29
0,75	0,08	0,13	0,21	0,34	0,54
0,8	0,04	0,07	0,10	0,14	0,21
0,85	0,02	0,03	0,04	0,05	0,07
0,9	0,0077	0,0098	0,0124	0,0154	0,0188
0,95	0,0016	0,0019	0,0021	0,0024	0,0027

5 Заключение

Рассмотрены некоторые максимум-линейные случайные процессы со стандартным распределением Фреше в качестве маргинального. Предложены простые статистические оценки параметров, основанные на сравнении значений процесса в различные моменты времени. Проведено сравнение ранее известных оценок параметра максимум-авторегрессии первого порядка по дисперсии.

Литература

1. *Embrechts P., Klüppelberg C., Mikosh T.* Modelling extremal events for insurance and finance. — New York, NY, USA: Springer-Verlag, 2003. 648 p.
2. *McNeil A. J., Frey R., Embrechts P.* Quantitative risk management. — Princeton, NJ, USA: Princeton University Press, 2005. 538 p.
3. *Davis R. A., Resnick S. I.* Basic properties and prediction of max-ARMA processes // *Adv. Appl. Probab.*, 1989. Vol. 21. No. 4. P. 781–803.
4. *Davis R. A., Resnick S. I.* Prediction of stationary max-stable processes // *Ann. Appl. Probab.*, 1993. Vol. 3. No. 2. P. 497–525.
5. *Holland I., Nilsen T.* On a general random exchange model // *J. Appl. Probab.*, 1976. Vol. 13. No. 4. P. 781–790.
6. *Daley D., Haslet J.* A thermal energy storage with controlled input // *Adv. Appl. Probab.*, 1982. Vol. 14. No. 2. P. 257–271.

7. Coles S. G. Regional modelling of extreme storms via max-stable processes // J. R. Stat. Soc. B, 1993. Vol. 55. No. 4. P. 797–816.
8. Zhang Z., Smith R. L. Modelling financial time series data as moving maxima processes. Department of Statistics, University of North Carolina, 2001. Technical Report. 19 p.
9. Alpuim M. T. An extremal Markovian sequence // J. Appl. Probab., 1989. Vol. 26. No. 2. P. 219–232.
10. Alpuim M. T., Catkan N. A., Hüsler J. Extremes and clustering of nonstationary max-AR(1) sequences // Stoch. Proc. Appl., 1995. Vol. 56. No. 1. P. 174–184.
11. Феррейра М. О зависимости от поведения хвостов: описание характера процессов максимум-авторегрессии первого порядка // Математические заметки, 2011. Т. 58. № 6. С. 902–917.
12. Ferreira H. Parameter estimation and dependence characterization of the MAR(1) process // ProbStat Forum, 2012. Vol. 5. No. 12. P. 107–111.
13. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Т. 1 / Пер. с англ. — М.: Мир, 1974. 406 с. (Box G., Jenkins G. Time series analysis. Forecasting and control. — San Francisco, CA, USA: Holden-day, 1970. 537 р.)
14. Лебедев А. В. Статистический анализ MARMA-процессов первого порядка // Математические заметки, 2008. Т. 83. № 4. С. 552–558.
15. Лебедев А. В. Нелинейное прогнозирование процессов максимум-авторегрессии // Математические заметки, 2009. Т. 85. № 4. С. 636–640.
16. Лебедев А. В. Об одной модели топа новостей // Проблемы передачи информации, 2009. Т. 45. № 3. С. 98–105.

Поступила в редакцию 30.10.16

STATISTICAL ANALYSIS OF MAX-LINEAR PROCESSES

A. V. Lebedev

Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, Main Building, 1 Leninskiye Gory, Moscow 119991, Russian Federation

Abstract: Max-linear processes are considered. It is assumed that innovations and process values have the standard Fréchet distribution. Simple statistical estimators for parameters are proposed. Some estimators for parameters of the first-order max-autoregressive processes are compared on variances.

Keywords: max-linear processes; max-autoregressive processes; Fréchet distribution; statistical analysis; heavy tails

DOI: 10.14357/08696527170202

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 14-01-00075).

References

1. Embrechts, P., C. Klüppelberg, and T. Mikosh. 2003. *Modelling extremal events for insurance and finance*. New York, NY: Springer-Verlag. 648 p.
2. McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative risk management*. Princeton, NJ: Princeton University Press. 538 p.
3. Davis, R. A., and S. I. Resnick. 1989. Basic properties and prediction of max-ARMA processes. *Adv. Appl. Probab.* 21(4):781–803.
4. Davis, R. A., and S. I. Resnick. 1993. Prediction of stationary max-stable processes. *Ann. Appl. Probab.* 3(2):497–525.
5. Helland, I., and T. Nilsen. 1976. On a general random exchange model. *J. Appl. Probab.* 13(4):781–790.
6. Daley, D., and J. Haslet. 1982. A thermal energy storage with controlled input. *Adv. Appl. Probab.* 14(2):257–271.
7. Coles, S. G. 1993. Regional modelling of extreme storms via max-stable processes. *J. R. Stat. Soc. B* 55(4):797–816.
8. Zhang, Z., and R. L. Smith. 2001. Modelling financial time series data as moving maxima processes. Department of Statistics, University of North Carolina. Technical Report. 19 p.
9. Alpuim, M. T. 1989. An extremal Markovian sequence. *J. Appl. Probab.* 26(2):219–232.
10. Alpuim, M. T., N. A. Catkan, and J. Hüsler. 1995. Extremes and clustering of nonstationary max-AR(1) sequences. *Stoch. Proc. Appl.* 56(1):174–184.
11. Ferreira, M. 2011. On tail dependence: A characterization for first-order max-autoregressive processes. *Math. Notes* 90(5):882–893.
12. Ferreira, H. 2012. Parameter estimation and dependence characterization of the MAR(1) process. *ProbStat Forum* 5(12):107–111.
13. Box, G., and G. Jenkins. 1970. *Time series analysis. Forecasting and control*. San Francisco, CA: Holden-day. 537 p.
14. Lebedev, A. V. 2008. Statistical analysis of first-order MARMA processes. *Math. Notes* 83(3):506–511.
15. Lebedev, A. V. 2009. Nonlinear prediction in max-autoregressive processes. *Math. Notes* 85(3):602–606.
16. Lebedev, A. V. 2009. On one top news list model. *Probl. Inf. Transm.* 45(3):282–288.

Received October 30, 2016

Contributor

Lebedev Alexey V. (b. 1971) — Doctor of Sciences in physics and mathematics, associate professor, Department of Probability Theory, Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, Main Building, 1 Leninskiye Gory, Moscow 119991, Russian Federation; avlebed@yandex.ru

КЛАССИФИКАЦИЯ ОШИБОЧНЫХ СОСТОЯНИЙ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ И ИСТОЧНИКИ ИХ ВОЗНИКОВЕНИЯ*

*A. A. Грушо¹, М. И. Забежайл², А. А. Зацаринный³, А. В. Николаев⁴,
В. О. Писковский⁵, Е. Е. Тимонина⁶*

Аннотация: Определены классы ошибочных состояний распределенных вычислительных систем и источники их возникновения. Основные классы ошибочных состояний проиллюстрированы экспериментально на примере сбоя работы платформы RabbitMQ, используемой в качестве системы обмена сообщениями на основе стандарта AMQP (Advanced Message Queuing Protocol) между компонентами программной платформы OpenStack. Моделируемое ошибочное состояние неоднократно наблюдалось на практике при эксплуатации платформы в условиях, близких к исчерпанию ресурсов. Для эмуляции сбоя во всей распределенной системе в конфигурационный файл платформы обмена сообщений RabbitMQ в качестве эксперимента были внесены изменения с целью получить нехватку ресурсов. Также представлен результат анализа инцидента в качестве реализации классов ошибочных состояний в традиционных информационных сетевых инфраструктурах.

Ключевые слова: распределенные вычислительные системы; облачные вычислительные среды; OpenStack; RabbitMQ; классы ошибочных состояний; информационные сетевые инфраструктуры

DOI: 10.14357/08696527170203

1 Введение

В статье [1] рассмотрены аномалии, возникающие в процессах реконфигурации облачных вычислительных сред (ОВС), показано существование методов

*Работа поддержана РФФИ (проект 15-29-07981 офи-м).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabzhailo@yandex.ru

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, alex250451@mail.ru

⁴Институт химической физики им. Н. Н. Семёнова Российской академии наук, gentoorion@mail.ru

⁵Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, urvpr80@yandex.ru

⁶Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

защиты от уязвимостей, вызванных конфликтами на уровне процессов реконфигурации ОВС. В предлагаемой работе рассмотрены наблюдаемые на практике ошибочные состояния, к которым может привести эксплуатация таких уязвимостей.

Рассмотрим, из чего состоят ОВС, и обсудим уязвимые места ОВС. Сегодня уже никого не удивишь наличием компьютера, «умного телефона», планшета или другого гаджета. Гораздо важнее не наличие устройств и даже не установленные непосредственно на них приложения, а возможность этих устройств соединяться в информационные инфраструктуры, где отдельные узлы выполняют свою, присущую только им роль, предоставляя самые разнообразные услуги по обработке данных. От этого ценность использования инфраструктуры вместо автономных компьютерных устройств возрастает многократно. Таким образом, можно говорить, что на смену автономным системам приходят распределенные системы (РС). Современные РС являются технической основой облачных сервисов и ОВС. Как правило, современные РС реализованы как совокупность отдельных компьютеров, управляемых автономными операционными системами (ОС) и связанных друг с другом посредством разнообразных сетевых сервисов. Эта совокупность сервисов предоставляет клиентам функционал облачной среды по выделению ресурсов, таких как виртуальные машины, объемы хранилищ данных, сетевые каналы с теми или иными свойствами, различные вспомогательные службы и т. д. Со стороны клиента такая система выглядит в чем-то похожей на традиционную ОС: та же задача управления с тем же набором ресурсов, что и в обычной ОС; в основном те же подходы и методы. Однако назвать их аналогами ОС нельзя, равно как и невозможно использовать многие традиционные для ОС подходы при работе с ними. Основными причинами отличий между ОС и РС являются отсутствие каналов синхронизации между ОС, участниками РС, за исключением сетевых, а значит значительно менее надежных, чем внутренние шины современных компьютеров, а также исключительная сложность самих систем и фактическое отсутствие единого проекта для всех компонентов, реализующих РС. Ведь не секрет, что даже ведущие мировые гиганты в области коммерческого программного обеспечения (ПО) давно потеряли возможность разработки всего спектра нужных приложений в жестких условиях коммерческой окупаемости и темпов развития современного рынка ИТ. В результате в РС критически важные компоненты ПО зачастую либо полностью заимствованы из других проектов, либо являются минимально модифицированными и адаптированными фрагментами такового ПО. Следствием является крайне проблематичная система обработки ошибочных ситуаций, тайм-аутов, маркирования и соотнесения сообщений в журналах различных сервисов. В результате на современном этапе развития РС одним из факторов, сильно удорожающих их работу, становится проблема диагностирования состояния РС, локализации, устранения сбоев, проверки исправности систем, а также повышения их надежности. При этом возникает целый ряд новых типов ошибочных состояний, не характерных или редких для традиционных ОС.

Многие ошибки работы РС характерны как для систем, представляющих собой традиционные сетевые инфраструктуры, построенные на автономных устройствах, так и для ОВС, когда на одном физическом устройстве размещаются десятки и сотни виртуальных сетевых узлов. В последнем случае ошибки конфигурирования вызваны не столько действиями администратора, сколько уязвимостями программных средств, применяемых для управления ОВС.

В работе [1] представлена адаптация концепции непротиворечивости и управления транзакциями системы управления базами данных в применении к динамической реконфигурации облачных вычислительных сред.

В разд. 2 рассмотрены классы ошибочных состояний РС и источники их возникновения, а в разд. 3 эти рассуждения проиллюстрированы экспериментально и приведен пример анализа инцидента на основе записей из системных журналов РС.

2 Классы ошибочных состояний облачных вычислительных сред и источники их возникновения

Практический опыт работы с РС OpenStack указывает на следующие основные классы новых ошибочных состояний и источники их возникновения:

- потеря сетевого соединения, распад облачной структуры на «подоблака». В простейшем случае система теряет функциональность. В более нетривиальном случае система теряет часть функционала, но ядро управления остается работоспособным. Такие ситуации похожи на проблемы традиционных ОС при неполадках аппаратного обеспечения компьютеров, особенно периферии. Однако в современных РС существует возможность того, что при сетевых неполадках в РС возникнет два и более ядер управления, которые смогут функционировать самостоятельно. Особенно это касается систем, сконфигурированных с опциями «высокой доступности» (high availability) и географически распределенных. Естественно, что в отсутствие возможности синхронизации такие системы будут принимать различные состояния, приводящие к невозможности последующей совместной работы. При устранении неполадок с сетью возникает абсолютно новая задача «объединения» РС, работавших какое-то время автономно, но имевших общее управление ранее;
- частичное повреждение (например, из-за преднамеренных действий) или неправильная конфигурация, недостаток производительности одного из служебных сетевых компонентов, что приводит либо к вероятностным отказам, либо к включению настроек по умолчанию у других компонентов, которые лишились важного, но не критичного сервиса. Примером такой ситуации могут служить сбои в сервисе передачи служебных сообщений RabbitMQ [2, 3]. Частичная функциональность RabbitMQ, например из-за недостатков ресурсов, таких как память, приводит к потере файловых дескрипторов

и к резкому падению надежности передачи служебных сообщений между всеми компонентами OpenStack [3], что, в свою очередь, приводит в лучшем случае к сильному замедлению работы, а в худшем — к аварийному останову отдельных служб, некорректным записям в сервисных базах данных, сбоям при операциях запуска или удаления виртуальных машин, виртуальных сетей и т. д. При этом в диагностике ошибочных состояний сбоящих подсистем упоминание о проблемах с сервисом служебных сообщений либо не встречается вообще, либо встречается в виде предупреждения о том, что соединение закрыто по тайм-ауту и открыто новое. То, что при этом теряется определенная доля служебных сообщений, можно узнать только на максимально детальном уровне журналирования или в отладчике. Ошибки этого класса приводят к непредсказуемым сбоям: от так называемых мерцающих ошибок до периодических процессов раз渲ала и сбора облака, что относится к первому классу ошибок согласно представленной здесь классификации. Маркером такого события, как удалось установить, является некритическое предупреждение о нехватке ресурсов. В разд. 3, описывающем эксперимент, представлен опыт моделирования такой ситуации;

- неактуальные или некорректные данные в служебных базах данных, нарушение политик безопасности. Разворачивание новых ресурсов, или восстановление из резервного архива (backup) старых конфигураций, или перенастройка подсистем РС без их полной остановки являются крайне плохо детерминированными процессами из-за того, что компоненты, утратившие или никогда не синхронизированные по конфигурации с текущей системой, тем не менее, при использовании настройки по умолчанию или старых настроек, регистрируются как сервисы, службы, ресурсы РС. Это приводит к фантомным записям в служебных базах данных и, как следствие, к серьезным непредсказуемым сбоям;
- отметим, что в условиях отсутствия в РС централизованной службы журналирования с надежной синхронизацией меток определить, какая из аномалий [4] стала причиной перехода сети в ошибочное состояние, практически невозможно. Чтобы восстановить точную последовательность событий, записанных в журналы, приходится отрабатывать несколько вариантов сбоев, что, в свою очередь, сильно замедляет починку и восстановление консистентного состояния РС.

3 Описание эксперимента

В качестве эксперимента на работающей РС, использующей платформу¹ OpenStack и платформу доставки сообщений RabbitMQ, было внесено частич-

¹Авторы выражают благодарность компании НРСHub Research (ООО «ЭйчПиСи Хаб Ри-серч») за предоставленную возможность проведения экспериментов и данные для анализа.

ное повреждение в виде ограничения предела числа файловых дескрипторов и последующей перезагрузки системы доставки сообщений RabbitMQ.

Цель эксперимента — продемонстрировать классы ошибочных состояний ОВС на примере указанной платформы при кратковременном сбое, вызвавшем перезагрузку одного из служебных сервисов.

В 12:03 был изменен файл `/etc/systemd/system/rabbitmq-server.service.d/limits.conf`, параметр `LimitNOFILE = 500` в разделе `[Service]`, и тут же перезапущен RabbitMQ (рис. 1).

```
systemctl daemon-reload; systemctl restart rabbitmq-server
```

Рис. 1 Команда перезапуска RabbitMQ

Это событие отразилось в журнале:

- в 12:04:05 — журнал начинается с информационного сообщения о запуске сервиса RabbitMQ (рис. 2);
- далее, вплоть до 12:04:17, сообщения не содержат ничего необычного, в журнале зарегистрированы записи в основном об установленных соединениях и единичных сбоях;
- в 12:04:17 — появляется предупреждение о превышении предела числа файловых дескрипторов, после чего практически каждая запись сообщает об ошибке соединений (рис. 3).

Остальные журналы в зависимости от политики аудита также содержат сообщения о потере соединений начиная примерно с 12:04:17. Так, в журнале `nova-conductor.log.ostack-head` находятся повторяющиеся записи (рис. 4).

```
=INFO REPORT==== ***-***-***::12:04:05 ===
Starting RabbitMQ 3.6.2 on Erlang 18.3.3
Copyright (C) 2007-2016 Pivotal Software, Inc.
Licensed under the MPL. See http://www.rabbitmq.com/

=INFO REPORT==== ***-***-***::12:04:05 ===
node          : rabbit@ostack-head
home dir      : /var/lib/rabbitmq
config file(s) : /etc/rabbitmq/rabbitmq.config
cookie hash   : JyDl+/mb74rGdAh+Ds9Ldg==
log           : /var/log/rabbitmq/rabbit@ostack-head.log
sasl log      : /var/log/rabbitmq/rabbit@ostack-head-sasl.log
database dir  : /var/lib/rabbitmq/mnesia/rabbit@ostack-head
...
```

Рис. 2 Информационное сообщение о запуске сервиса RabbitMQ

```
...
=WARNING REPORT==== 27-Jan-2017::12:04:17 ===
file descriptor limit alarm set.
...
```

Рис. 3 Предупреждение о превышении предела числа файловых дескрипторов

```
...
12:04:16.586 5206 INFO oslo.messaging._drivers.impl_rabbit [-] A
recoverable connection/channel error occurred, trying to
reconnect: Socket closed

12:04:16.586 5206 INFO oslo.messaging._drivers.impl_rabbit [-] A
recoverable connection/channel error occurred, trying to
reconnect: Socket closed

12:04:16.672 5203 INFO oslo.messaging._drivers.impl_rabbit [-] A
recoverable connection/channel error occurred, trying to
reconnect: Socket closed

...
```

Рис. 4 Записи журнала nova-conductor.log.ostack-head

Таким образом, дальнейшая работа платформы была частично парализована внесенным сбоем. Например, neutron linuxbridge-агент каждую минуту пытался сообщить свой статус через RabbitMQ (рис. 5).

А вот журнал агента одного из узлов платформы linuxbridge-agent.log.n008 содержит только одно такое сообщение (рис. 6).

Это сообщение следует практически сразу за перезапуском RabbitMQ, т. е. все остальное время узел мог рапортовать сообщения.

Далее на стенде были предприняты попытки протестировать некоторые функции. Многократная попытка удалить старый стек (рис. 7) в конце концов увенчалась успехом, а попытка создать новый стек на 5 виртуальных машин не удалась.

Таким образом, в результате внесенного ограничения числа открытых файловых дескрипторов смоделирована частичная потеря функциональности RabbitMQ: при 500 дескрипторах против обычных с запасом 8192 у фермы не работали neutron-агенты, что, в свою очередь, повлекло частичную потерю функциональности платформы и вероятностным отказам.

Необходимо также отметить, что перечисленные в предыдущем разделе классы ошибочных состояний ОВС могут возникать не только в современных ОВС, но и в традиционных сетевых инфраструктурах. Так, несколько лет назад произошла остановка работы одного из крупных предприятий обслуживания на

```
12:05:17.517 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:06:47.524 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:07:47.528 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:08:47.533 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:09:47.537 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:10:47.541 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
12:11:47.546 2376 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
```

Рис. 5 Записи журнала linuxbridge-agent.log.n001

```
...
12:05:03.166 2414 ERROR
neutron.plugins.ml2.drivers.agent._common_agent [-] Failed
reporting state!
...
```

Рис. 6 Записи журнала linuxbridge-agent.log.n008

```
heat stack-delete 4a9c8420-bd00-4e72-8b2b-6e0d80b336ee
```

Рис. 7 Попытка удалить старый стек

несколько дней из-за серьезных нарушений в работе локальной вычислительной сети предприятия. Причиной стало нарушение корректности данных в одном из управляющих компонентов сетевой инфраструктуры. Нарушению корректности данных предшествовали эксперименты по внесению тонких настроек и изменений в служебную базу данных, управляющую доступом к части сетевых ресурсов, включая рабочие места сотрудников. Работы проводились вне операционного периода обслуживания клиентов и носили, как казалось их исполнителям, локальный характер. Анализ журналов и действий обслуживающего персонала показал, что работы совпали с проведением автоматизированных регламентных работ, что привело к частичному повреждению данных в конфигурационных файлах сетевой инфраструктуры и в результате завершилось лавинообразным вероятностным распадом сетевой инфраструктуры на отдельные сегменты.

4 Заключение

Актуальность поиска нарушений корректности работы инфокоммуникационной сети отмечалась в работах [5–7]. В продолжение этих исследований в статье приведены новые следующие результаты:

- приведена классификация сбоев РС, наблюдаемых при эксплуатации сетевых инфраструктур;
- представлен пример со сбоем RabbitMQ. Сбой относится к некоторому классу ошибочных ситуаций, а именно: частичному повреждению конфигурации по причине преднамеренных действий. При эксплуатации платформы Open-Stack было отмечено некритическое предупреждение о нехватке ресурсов службы RabbitMQ, после которого наблюдался вероятностный обрыв и восстановление соединений с виртуальным машинами. В результате проведенных экспериментов удалось смоделировать данную ситуацию;
- как показано на примере анализа инцидента в традиционных сетевых инфраструктурах, ошибки персонала при конфигурации службы или недостаток вычислительных ресурсов одного из служебных сетевых компонентов могут привести к частичному повреждению облака, что по представленной классификации относится к первому классу ошибочных состояний. Отметим, что пример с наложением автоматических и ручных конфигураций скорее относится к третьему классу, когда неактуальные и / или некорректные данные соседствуют в служебных базах с актуальными, нарушая политики безопасности. Разворачивание новых ресурсов или восстановление из резервного архива (backup) старых конфигураций, а также перенастройка подсистем РС без их полной остановки приводят к лавинообразному необратимому процессу развала инфраструктуры.

Отметим, что одна из основных проблем при анализе ошибочных состояний вызвана отсутствием в РС централизованной службы журнализирования.

Представленные выше три примера классов (однородных по своей природе) ошибочных состояний ОВС (состояний, характеризуемых теми или иными «патологиями» в их архитектуре) можно проиндексировать набором характерных для этих состояний признаков, принимающих те или иные конкретные значения. При этом каждое ошибочное состояние характеризуется определенной комбинацией значений этих признаков. В свою очередь, состояния, находящиеся за пределами множеств значений для ошибочных конфигураций, будут соответствовать тем или иным допустимым состояниям соответствующей ОВС. Опираясь на precedents подобного рода описаний конкретных состояний ОВС (в простейшем случае рассматривая результаты предлагаемого «индексирования» как множество значений соответствующих признаков, а в более продвинутом случае — как множество значений соответствующих признаков с учетом определенных отношений¹ на некоторых признаках, в том числе отношений типа РАНЬШЕ–ПОЗЖЕ²), можно сформировать механизм обучения на коллекциях накапливаемых precedents, позволяющий в ряде случаев диагностировать (причем на ранних стадиях развития соответствующей архитектурной «патологии») новые конфигурации как ошибочные. Процедурная конструкция подобного индуктивного обучения и последующих выводов «по аналогии» (разумеется, с учетом достаточности оснований для соответствующих заключений «по аналогии», оцениваемых путем формирования специальных *абдуктивных объяснений*) может быть основана, в частности, на математической технике DCM (Джона Стюарта Милля) метода, развивающего школой проф. В. К. Финна (см., например, работы [8, 9]). Таким образом, развивая предлагаемую процедурную конструкцию, можно создавать интеллектуальные компьютерные системы — «ассистентов» администратора ОВС, которые способны взять на себя поддержку принятия решений по обеспечению корректного конфигурирования ОВС с учетом постоянно накапливаемого опыта выявления ее ошибочных состояний. К более подробному обсуждению обозначенной проблематики применения специализированных методов искусственного интеллекта и анализа неструктурированных данных для диагностики состояния РС авторы планируют обратиться в одной из следующих публикаций.

Литература

- Грушо А. А., Забежайло М. И., Зацаринный А. А., Писковский В. О. Безопасная автоматическая переконфигурация облачных вычислительных сред // Системы и средства информатики, 2016. Т. 26. № 3. С. 83–92.
- RabbitMQ documentation. <http://www.rabbitmq.com/documentation.html>.

¹Например, значения признака Р1 и признака Р2 в случае ошибочной конфигурации являются зависимыми (в частности, свойства Р1 и Р2 либо одновременно присутствуют, либо одновременно отсутствуют).

²Что позволяет описывать и некоторые элементы динамики изменений анализируемой текущей конфигурации ОВС.

3. AMQP and Nova documentation. <https://docs.openstack.org/developer/nova/rpc.html>.
4. Грушо А. А., Грушо Н. А., Тимонина Е. Е., Шоргин С. Я. Архитектурные уязвимости распределенных информационно-вычислительных систем // Системы и средства информатики, 2016. Т. 26. № 3. С. 74–82.
5. Ju X., Soares L., Shin K. G., Ryu K. D., Da Silva D. On fault resilience of OpenStack // 4th Annual Symposium on Cloud Computing Proceedings. — New York, NY, USA: ACM, 2013. Article No. 2. <https://kabru.eecs.umich.edu/papers/publications/2013/socc2013.ju.pdf>.
6. Guo Zh., McDirmid S., Yang M., Zhuang Li, Zhang Pu, Luo Y., Bergan T., Bodik P., Musuvathi M., Zhang Zh., Zhou L. Failure recovery: when the cure is worse than the disease // 14th USENIX Conference on Hot Topics in Operating Systems Proceedings. — Berkeley, CA, USA: USENIX Association, 2013. Article No. 8. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/FailureRecoveryBeEvil.pdf>.
7. Sharma D., Poddar R., Mahajan K., Dhawan M., Mann V. HANSEL: Diagnosing faults in OpenStack // 11th ACM Conference on Emerging Networking Experiments and Technologies Proceedings. — New York, NY, USA: ACM, 2015. Article No. 23. <http://dl.acm.org/citation.cfm?doid=2716281.2836108>.
8. Финн В. К. Индуктивные методы Д. С. Милля в системах искусственного интеллекта // Искусственный интеллект и принятие решений, 2010. Ч. I: № 3. С. 3–21; Ч. II: № 4. С. 14–40.
9. Забежайло М. И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений, 2014. Ч. I: № 1. С. 95–110; Ч. II: № 3. С. 3–21.

Поступила в редакцию 05.03.17

ERRONEOUS STATES CLASSIFICATION IN DISTRIBUTED COMPUTING SYSTEMS AND SOURCES OF THEIR OCCURENCE

**A. A. Grusho¹, M. I. Zabeshailo¹, A. A. Zatsarinnyy², A. V. Nikolaev³,
V. O. Piskovski¹, and E. E. Timonina¹**

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³N. N. Semenov Institute of Chemical Physics of the Russian Academy of Sciences, 4 Kosygin Str., Moscow 119991, Russian Federation

Abstract: The paper defines the classes of erroneous states in distributed computing systems and sources of their occurrence. Some classes of these states

are illustrated experimentally by inducing RabbitMQ abnormal behavior in the OpenStack software platform. The simulated erroneous states were observed actually in conditions close to the exhaustion of resources of the OpenStack software platform operation. To get erroneous states in the OpenStack platform, the RabbitMQ configuration file was tuned in order to simulate shortage of resources. The paper also contains the analysis of the incident, which occurred in the traditional information network infrastructures and can be treated as the implementation of the classes of erroneous states as well.

Keywords: distributed computing system; cloudy computing environments; OpenStack; RabbitMQ; erroneous states; informational infrastructure

DOI: 10.14357/08696527170203

Acknowledgments

The paper was supported by the Russian Foundation for Basic Research (project 15-29-07981 ofi-m).

References

1. Grusho, A. A., M. I. Zabeshailo, A. A. Zatsarinnyy, and V. O. Piskovski. 2016. Bezopasnaya avtomaticheskaya rekonfiguratsiya oblacnykh vychislitel'nykh sred [Secure automatic reconfiguration of cloudy computing]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):83–92.
2. RabbitMQ documentation. Available at: <http://www.rabbitmq.com/documentation.html> (accessed March 5, 2017).
3. AMQP and Nova documentation. Available at: <https://docs.openstack.org/developer/nova/rpc.html> (accessed March 5, 2017).
4. Grusho, A., N. Grusho, E. Timonina, and S. Shorgin. 2016. Arkhitekturnye uyazvimosti raspredelennykh informatsionno-vychislitel'nykh sistem [Architectural vulnerabilities of the distributed information systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):74–82.
5. Ju, X., L. Soares, K. G. Shin, K. D. Ryu, and D. Da Silva. 2013. On fault resilience of OpenStack. *4th Annual Symposium on Cloud Computing Proceedings*. New York, NY: ACM. Article No. 2. Available at: <https://kabru.eecs.umich.edu/papers/publications/2013/socc2013.ju.pdf> (accessed March 5, 2017).
6. Guo, Zh., S. McDirmid, M. Yang, Li Zhuang, Pu Zhang, Y. Luo, T. Bergan, P. Bodik, M. Musuvathi, Zh. Zhang, and L. Zhou. 2013. Failure recovery: When the cure is worse than the disease. *14th USENIX Conference on Hot Topics in Operating Systems Proceedings*. Berkeley, CA: USENIX Association. Article No. 8. Available at: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/FailureRecoveryBeEvil.pdf> (accessed March 5, 2017).
7. Sharma, D., R. Poddar, K. Mahajan, M. Dhawan, and V. Mann. 2015. HANSEL: Diagnosing faults in Open Stack. *11th ACM Conference on Emerging Networking Experiments and Technologies Proceedings*. New York, NY: ACM. Article No. 23. Available at: <http://dl.acm.org/citation.cfm?doid=2716281.2836108> (accessed March 5, 2017).

8. Finn, V. K. 2011–2012. Mill’s inductive methods in artificial intelligence systems. *Scientific and Technical Information Processing*. Pt. I (2011): 38(6):385–402; Pt. II (2012): 39(5):241–261.
9. Zabezhailo, M. I. 2014. O nekotorykh vozmozhnostyakh upravleniya pereborom v DSM-metode [About some opportunities of management of search in the DSM-method]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision-Making]. Pt. I: 1:95–110; Pt. II: 3:3–21.

Received March 05, 2017

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, head of laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; grusho@yandex.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, head of laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Zatsarinnyy Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; alex240451@mail.ru

Nikolaev Andrei V. (b. 1973) — Candidate of Science (PhD) in physics and mathematics, senior scientist, N. N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, 4 Kosygin Str., Moscow 119991, Russian Federation; gentoorion@mail.ru

Piskovski Viktor O. (b. 1963) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vpvp80@yandex.ru

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; eltimon@yandex.ru

ВОЗМОЖНОСТЬ ВЫЯВЛЕНИЯ ИНСАЙДЕРА СТАТИСТИЧЕСКИМИ МЕТОДАМИ

E. A. Мартынов¹

Аннотация: Рассматривается задача выявления инсайдера в множестве аналитиков, работающих с хранилищем данных, которое представлено в виде плоской таблицы с большим числом атрибутов. Ключевым отличием поведения честного аналитика от инсайдера заключается в том, что последний в процессе работы с данными накапливает избыточную информацию для осуществления враждебных действий. Отсюда следует, что для определения инсайдера необходимо выявить факт регулярного сбора информации, не входящей в его функциональные обязанности. Описана математическая модель работы аналитиков с хранилищем данных и приводятся условия, при которых вероятность ложной тревоги равна 0, а функция мощности критерия проверки статистических гипотез принимает значение 1 за конечное число шагов. Таким образом доказано, что предложенный подход, использующий теорию вероятностей и математическую статистику для выявления факта накопления инсайдером избыточной информации, дает ответ на поставленный вопрос.

Ключевые слова: поиск инсайдера; обнаружение сбора избыточной информации; статистические методы; запрет меры; проверка статистических гипотез

DOI: 10.14357/08696527170204

1 Введение

Задача поиска инсайдера среди сотрудников определенного коллектива изучается уже довольно длительное время, с тех самых пор, когда объемы информации в корпоративных хранилищах стали такими, что контроль их распространения перестал быть тривиальной задачей.

Пожалуй, самым серьезным шагом в сторону решения этой проблемы можно считать проект ADAMS от DARPA [1]. Одно из ключевых мест в нем занимает задача выявления инсайдера [2] (в том числе потенциального) с помощью регистрации аномальных действий как в самой информационной системе предприятия, так и в социальных сетях [3] и их анализа. Одним из результатов этого проекта стал выход в свет программного комплекса IBM System G, состоящего из различных модулей выявления и анализа предпосылок враждебных действий и самих действий [4]. Центральное место в алгоритме выявления инсайдера в IBM Insider Threat Solution занимает предположение о том, что инсайдер

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, zhenya388@yandex.ru

выдает свою грядущую атаку через предшествующие враждебным действиям «слабые сигналы», которые наблюдаются на протяжении длительного периода времени [4] и анализируются с применением теории графов.

В данной работе не будут рассматриваться предпосылки враждебных действий, а задача ставится таким образом, что необходимо выявить инсайдера в множестве аналитиков, работающих с хранилищем данных (ХД), основываясь только на его работе с ним. Хранилище данных представляется в виде плоской таблицы с большим числом атрибутов (более 1000), а сами данные являются строками этой таблицы, при этом допускается неполное заполнение строк, т. е. значения некоторых атрибутов в строке могут отсутствовать. Это допущение позволяет гарантировать, что большинство реляционных баз данных можно привести к указанному виду.

Будем считать, что каждый аналитик в своей повседневной работе нацелен на решение определенной задачи. Это означает, что он использует вполне определенный набор атрибутов (на практике, порядка 20–30), при этом допускается, что он может также обращаться и к другим атрибутам. А выбор конкретных строк в данных атрибутах определяется текущими потребностями аналитика.

Предполагается, что среди данного конечного множества аналитиков, каждый из которых решает свою задачу, возможен инсайдер, имеющий враждебные цели. Инсайдер собирает и использует информацию из ХД в криминальных целях. Разумеется, он может использовать данные, которые получает и для решения своей профессиональной задачи. Однако в этом случае выявление самого факта враждебного использования информации приведет к факту выявления инсайдера.

Отсюда следует вывод, что враждебные действия инсайдера должны быть основаны на информации, не связанной с его функциональными обязанностями. Такой подход можно реализовать только с помощью сбора большей информации, чем это необходимо для выполнения функциональных обязанностей.

Следовательно, чтобы выявить инсайдера в множестве аналитиков, в рамках описанной модели необходимо выявить факт сбора избыточной информации на регулярной основе. Для решения поставленной задачи предлагается использовать средства теории вероятностей и математической статистики. Однако, чтобы это можно было сделать, необходимо выполнение важного требования — сходимости. Далее сформулируем ограничения модели, чтобы это требование выполнялось.

2 Математическая модель выявления инсайдера

Пусть $A = \{a_1, \dots, a_m\}$ — множество атрибутов данных, собираемых в хранилище. Обозначим через \mathfrak{R} — множество всех подмножеств множества A . Функциональные обязанности аналитика U характеризуются множеством $X = \{X_1, \dots, X_k\}$, $X_i \subseteq \mathfrak{R}$. Будем считать, что подмножества из X содержат не только атрибуты, которые обычно использует аналитик U , но и возможные дополнительные атрибуты, которые могут потребоваться при решении его

профессиональной задачи и $\bigcup_{i=1}^k X_i = B \subset A$. Обозначим декартово произведение множества \mathfrak{R} через \mathfrak{R}^n , а через \mathfrak{R}^∞ — множество последовательностей с элементами из \mathfrak{R} .

Определим \mathcal{A} как σ -алгебру на \mathfrak{R}^∞ , порожденную цилиндрическими множествами. \mathcal{A} также является борелевской σ -алгеброй в тихоновском произведении \mathfrak{R}^∞ , где \mathfrak{R} имеет дискретную топологию [5, 6].

Работу аналитика можно представить в виде последовательности элементов X_i , $i = 1, 2, \dots$, множества X . Множество $X = \{X_1, \dots, X_k\}$ — конечное множество; X^n — декартово произведение X ; X^∞ — множество всех последовательностей, в которых i -й элемент принадлежит X .

Определим на $(\mathfrak{R}^\infty, \mathcal{A})$ вероятностную меру P . Предположим, что для любого $n = 1, 2, \dots$ вероятностное распределение P_n есть проекция меры P на первые n координат случайной последовательности из \mathfrak{R}^∞ . Ясно, что для любого $C_n \subseteq \mathfrak{R}^n$:

$$P_n(C_n) = P_n(C_n \times \mathfrak{R}^\infty). \quad (1)$$

Пусть $D_n(P)$ — носитель меры P_n [6]:

$$D_n(P) = \{\bar{\omega}_n \in X^n, P_n(\bar{\omega}_n) > 0\}.$$

Определим цилиндрическое множество $\Delta_n(P)$ следующим образом:

$$\Delta_n(P) = D_n(P) \times \mathfrak{R}^\infty.$$

Последовательность цилиндрических множеств $\Delta_n(P)$, $n = 1, 2, \dots$, — не возрастающая последовательность [7] и

$$\Delta(P) = \lim_{n \rightarrow \infty} \Delta_n(P) = \bigcap_{n=1}^{\infty} \Delta_n(P).$$

Здесь $\Delta(P)$ является замкнутым множеством в тихоновском произведении и является носителем меры P [7]. Для честного аналитика $\Delta(P) \subseteq X^\infty$.

Пусть $\bar{\omega}_s \in \mathfrak{R}^s$ и $\tilde{\omega}_{s-1}$ получен из $\bar{\omega}_s$ отбрасыванием последней координаты.

Определение 1. Запретом меры P_n [7, 8] называется вектор $\bar{\omega}_s \in \mathfrak{R}^s$, $s \leq n$, такой, что

$$P_n(\bar{\omega}_s \times X^{n-s}) = 0.$$

Определение 2. Запрет $\bar{\omega}_s$ меры P_n является наименьшим запретом меры P [7, 8], если

$$P_{s-1}(\tilde{\omega}_{s-1}) > 0.$$

Если $\bar{\omega}_s$ — запрет меры P_n , тогда для всех $s \leq l \leq n$ и всех последовательностей $\bar{\omega}_l$, начинающихся с последовательности $\bar{\omega}_s$, имеем

$$P_l(\bar{\omega}_l) = 0.$$

Действительно, если $P_s(\bar{\omega}_s) = 0$, то из (1) следует, что

$$\begin{aligned} P(\bar{\omega}_s \times X^\infty) &= 0; \\ P\left(\bar{\omega}_s \times X^{l-s} \times X^\infty\right) &= 0. \end{aligned}$$

Следовательно,

$$P_l(\bar{\omega}_l) = P(\bar{\omega}_l \times \mathfrak{R}^\infty) \leq P\left(\bar{\omega}_s \times \mathfrak{R}^{l-s} \times \mathfrak{R}^\infty\right) = 0.$$

Если существует вектор $\bar{\omega}_n \in \mathfrak{R}^n$ такой, что $P_n(\bar{\omega}_n) = 0$, тогда существует наименьший запрет, который определяется значениями первых координат вектора $\bar{\omega}_n$.

Пусть y — это атрибут, не принадлежащий множеству B . Тогда добавление y к любому множеству из X образует множество, не принадлежащее X . Такое множество является запретом, так как честный аналитик не может выбрать атрибуты, не принадлежащие B . Значит, если последовательность $\bar{\omega} \in \mathfrak{R}^\infty$ содержит хотя бы одно множество, содержащее y , то $\bar{\omega} \notin \Delta(P)$. Это следует из того, что $\bar{\omega} \notin X^\infty$.

По сделанному ранее предположению инсайдер в своей деятельности должен набирать информацию, не относящуюся к его профессиональной задаче. Обозначим через Q вероятностную меру, описывающую враждебную деятельность инсайдера. Это распределение характеризуется тем, что в последовательности выбора данных этим инсайдером обязательно найдется множество, содержащее дополнительные атрибуты, не входящие в B . Обозначим через $\Delta(Q)$ носитель меры Q . Как было отмечено выше, ни одна последовательность из $\Delta(Q)$ не может принадлежать $\Delta(P)$. Отсюда

$$\Delta(Q) \cap \Delta(P) = \emptyset.$$

Рассмотрим последовательность критериев для проверки гипотез $H_{0,n} : P_n$ против альтернатив $H_{1,n} : Q_n$, $n = 1, 2, \dots$. Статистический критерий проверки $H_{0,n}$ против $H_{1,n}$ определяется критическим множеством S_n векторов длины n .

Если наблюденный вектор принадлежит S_n , то это приводит к принятию альтернативы $H_{1,n}$. Если все векторы из S_n являются запретами меры P_n , то говорят, что критерий определяется запретами меры P . Заметим, что для всех n $P_n(S_n) = 0$, если S_n определяется запретами.

Пусть W_n — функция мощности критерия проверки $H_{0,n}$ против $H_{1,n}$. Известно, что $W_n = Q_n(S_n)$.

Тогда из работы [9] следует, что существует конечное число N такое, что существует последовательность критериев проверки $H_{0,n}$ против $H_{1,n}$ с критическими множествами S_n , $n = 1, 2, \dots$, определяемыми запретами, для которой существует N такое, что для всех $n \geq N$ функция мощности $W_n = 1$.

Пусть в работе участвуют аналитики U_1, \dots, U_M . Для каждого из них существует N_i , $i = 1, \dots, M$, для которого мощность соответствующего критерия принимает значение 1. Тогда при $N = \max_{i=1;M} N_i$ инсайдер будет определен однозначно.

3 Заключение

Из высказанного следует, что для решения задачи выявления инсайдера в множестве аналитиков можно применять статистические методы, так как они позволяют однозначно выявить инсайдера за конечное число шагов. В описанной модели существует последовательность критериев проверки $H_{0,n}$ против $H_{1,n}$ такая, что начиная с какого-то конечного момента функция мощности будет принимать значение 1. Это достигается за счет использования запретов, что по определению предотвращает ложные срабатывания. А это значит, что за конечное число шагов можно обеспечить вероятность выявления инсайдера, равную 1.

Литература

1. General Services Administration. Anomaly detection at multiple scales (ADAMS). 22.10.2010. Retrieved 2011-12-05. 40 p. https://www.fbo.gov/download/2f6/2f6289e99a0c04942bbd89ccf242fb4c/DARPA-BAA-11-04_ADAMS.pdf.
2. Senator T., Bader D., Chow E., et al. Detecting insider threats in a real corporate database of computer usage activity // 19th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings. — New York, NY, USA: ACM, 2013. P. 1393–1401.
3. Yu R., He X., Liu Y. GLAD: Group anomaly detection in social media analysis // arXiv.org e-Print Archive. 2014. arXiv:1410.1940.
4. System G: Developed graph computing industry solutions. <http://systemg.research.ibm.com/solutions.html>.
5. Бурбаки Н. Общая топология. Основные структуры / Пер. с фр. — М.: Наука, 1968. 272 с. (Bourbaki N. Topologie générale. — Paris: Hermann, 1971. 356 p.)
6. Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — М.: Наука, 1993. 496 с.
7. Grusho A., Grusho N., Timonina E. Consistent sequences of tests defined by bans // Optimization theory, decision making, and operation research applications. — Springer proceedings in mathematics & statistics ser. — New York – Heidelberg – Dordrecht, London: Springer-Verlag, 2013. Vol. 31. P. 281–291.
8. Грушо А., Тимонина Е. Запреты в дискретных вероятностно-статистических задачах // Дискретная математика, 2011. Т. 23. Вып. 2. С. 53–58.

9. Grusho A., Grusho N., Timonina E. Power functions of statistical criteria defined by bans // 29th European Conference on Modelling and Simulation Proceedings. — Germany: Digitaldruck Pirrot GmbH, 2015. P. 617–621.

Поступила в редакцию 15.03.17

POSSIBILITY OF INSIDER DETECTION BY STATISTICAL TECHNIQUES

E. A. Martyanov

M. V. Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

Abstract: The paper considers the task of insider detection in a group of analytics, who work with a data warehouse, presented as a raw table with a huge amount of attributes. The difference between a legal analyst and an insider is that an insider collects redundant data during his regular work to perform a threat. Therefore, in order to detect an insider, it is necessary to detect the fact of continuously collecting redundant data during a work cycle with a data warehouse. A mathematical model is defined. The author suggests to use statistical techniques with probability of false alarms equal to zero. The author found conditions, under which the power of statistical criteria reaches the value of 1 after a finite number of steps, which means that an insider can be detected definitely.

Keywords: insider threat; anomaly detection; bans of probability measures; statistical criteria; power of criteria

DOI: 10.14357/08696527170204

References

1. General Services Administration. 22.10.2010. Retrieved 05.12.2011. Anomaly detection at multiple scales (ADAMS). 40 p. Available at: https://www.fbo.gov/download/2f6/2f6289e99a0c04942bbd89ccf242fb4c/DARPA-BAA-11-04_ADAMS.pdf (accessed April 1, 2017).
2. Senator, T., D. Bader, E. Chow, *et al.* 2013. Detecting insider threats in a real corporate database of computer usage activity. *19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY: ACM. Available at: <http://dl.acm.org/citation.cfm?id=2488213> (accessed April 1, 2017).
3. Yu, R., X. He, and Y. Liu. 2014. GLAD: Group anomaly detection in social media analysis. eprint arXiv:1410.1940. Available at: <https://arxiv.org/abs/1410.1940> (accessed April 1, 2017).
4. System G: Developed graph computing industry solutions. Available at: <http://systemg.research.ibm.com/solutions.html> (accessed April 1, 2017).

5. Bourbaki, N. 1971. Topologie générale. Paris: Hermann. 356 p.
6. Prokhorov, Yu. V., and Yu. A. Rozanov. 1993. *Teoriya veroyatnostey* [Theory of probabilities]. Moscow: Nauka. 496 p.
7. Grusho, A., N. Grusho, and E. Timonina. 2013. Consistent sequences of tests defined by bans. *Optimization theory, decision making, and operation research applications*. Springer proceedings in mathematics & statistics ser. New York – Heidelberg – Dordrecht – London: Springer-Verlag. 31:281–291.
8. Grusho, A., and E. Timonina. 2011. Prohibitions in discrete probabilistic statistical problems. *Discrete Math. Appl.* 21(3):275–281.
9. Grusho, A., N. Grusho, and E. Timonina. 2015. Power functions of statistical criteria defined by bans. *29th European Conference on Modelling and Simulation Proceedings*. Germany: Digitaldruck Pirrot GmbH. 617–621.

Received March 15, 2017

Contributor

Martyanov Evgeny A. (b. 1991) — PhD Student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; zhenya388@yandex.ru

ИМИТАЦИОННАЯ МОДЕЛЬ ПОИСКА ИНСАЙДЕРА СТАТИСТИЧЕСКИМИ МЕТОДАМИ

E. A. Мартынов¹

Аннотация: Рассматривается задача выявления инсайдера в множестве аналитиков, работающих с хранилищем данных, которое представлено в виде плоской таблицы с большим числом атрибутов. Ключевым отличием поведения честного аналитика от инсайдера является то, что последний в процессе работы с данными либо напрямую собирает не предназначенные для него данные, либо небольшими порциями накапливает избыточную для своих функциональных обязанностей информацию. Отсюда следует, что для определения инсайдера необходимо детектировать факт регулярного сбора избыточной ценной информации, которую затем можно будет использовать для причинения ущерба. Описывается математическая модель поведения аналитика и инсайдера, проводится имитационное моделирование с использованием этой модели. На основе полученных результатов моделирования делаются выводы относительно применимости методов математической статистики для выявления инсайдера и выводятся необходимые для этого ограничения.

Ключевые слова: поиск инсайдера; сбор избыточной информации; статистические методы; математическая модель; имитационное моделирование

DOI: 10.14357/08696527170205

1 Введение

Задача своевременного выявления инсайдера, работающего с хранилищем данных, и заблаговременного предотвращения негативных последствий его действий, например продажи или передачи конфиденциальной информации третьим лицам, является важной и актуальной для всех компаний, чьи данные представляют ценность. В особенности это относится к банкам и государственным организациям. Однако даже с учетом современных методов и средств она остается очень сложной, так как инсайдер обладает всеми правами легального пользователя и выполняет свои функциональные обязанности, работает с информационной системой предприятия изнутри и часто знаком с ее структурой и механизмами обеспечения безопасности. Пользуясь этими знаниями, он может действовать очень осторожно и собирать информацию из разных источников малыми порциями, которые сами по себе могут не представлять большой ценности и не вызывать подозрений у службы внутренней безопасности. Но, собрав их вместе, он может получить по-настоящему ценный набор сведений.

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, zhenya388@yandex.ru

Одной из самых крупных попыток полностью решить эту задачу можно назвать проект ADAMS от DARPA [1], который начал развиваться в 2010 г. На сегодняшний день к основным его результатам можно отнести, во-первых, разработанный обобщенный алгоритм, сочетающий в себе различные подходы к обнаружению инсайдера и элементы машинного обучения, и, во-вторых, программное обеспечение, позволяющее определять потенциальную угрозу (по приведенной классификации) со стороны сотрудников организации.

В рамках этого проекта было выполнено несколько работ, в основе которых лежали разные подходы. В одной из них было собрано большое число алгоритмов и консолидировались их результаты [2]. Проводились симуляции различных сценариев работы на реальных наборах данных, после чего сравнивались результаты работы как отдельных алгоритмов, так и их комбинаций. Проанализировав результаты расследований реальных преступлений, было выявлено, что большинству из них предшествовали «слабые сигналы» (файлы на компьютере, переписка, публикации) [2–4], заблаговременный анализ которых помог бы предсказать и предотвратить правонарушения. В результате этой работы был создан программный комплекс IBM System G [4], позволяющий автоматизировать все используемые алгоритмы и вычисления применительно к «слабым сигналам». Более подробно об используемых алгоритмах и результатах исследований можно прочитать в [2]. В работе [3] внимание фокусировалось на поиске скрытых аномалий в поведении в выделенных группах в социальных сетях, после чего проводилось моделирование распределения по группам на синтетических и на реальных данных. В работе [5] для выявления инсайдера использовалась специально разработанная система сенсоров, которые встраивались в файлы документов для дальнейшего сбора информации о действиях с ним и ее анализа.

К недостаткам первого и второго подходов можно отнести необходимость собирать и анализировать большое количество не связанных друг с другом характеристик. Во-первых, в процессе эксплуатации реальной системы их тяжело отслеживать. Во-вторых, их полноту невозможно доказать, так как «слабый сигнал» может появиться в любом аспекте жизни человека. Его всегда можно упустить, что сильно повлияет на полученный ответ. В-третьих, если инсайдер целенаправленно собирает информацию, он может вовсе не подавать сигналов. Недостатком третьего подхода можно назвать необходимость отслеживать весь цикл работы с документом, тем самым можно выявить утечку информации из него, но могут возникнуть сложности в определении источника утечки.

В данной статье описывается математическая модель поведения аналитика, приводятся необходимые отличия в поведении аналитика и инсайдера, а также подтверждающие их результаты имитационного моделирования [6, 7] работы с хранилищем данных. При этом анализ основывается только на метаинформации о данных, с которыми он работает, без необходимости собирать и анализировать информацию, напрямую не относящуюся к профессиональной деятельности аналитика, что упрощает применение данного подхода в реальных хранилищах данных. На базе полученных результатов делаются выводы относительно

применимости методов математической статистики для выявления инсайдера и выводятся необходимые для этого ограничения.

2 Описание модели

Для решения поставленной задачи была разработана модель поведения аналитика при работе с хранилищем данных. Компонентами модели [8] являются множество аналитиков $U = \{U_1, \dots, U_M\}$ и само хранилище, которое определяется множеством всех его атрибутов $A = \{a_1, \dots, a_N\}$. Параметрами модели [8] являются M, N, T , p_n и p_m , где M — число аналитиков, работающих с системой, $M \in \mathbb{N}$; N и T — соответственно количество атрибутов и строк в хранилище данных, $N, T \in \mathbb{N}$; p_n и p_m — размер выборки, с которой работает аналитик, $p_n \in \mathbb{N}$, $p_n \leq N$, $p_m \in \mathbb{N}$, $p_m \leq T$.

Переменной модели [8] будет множество номеров атрибутов, попавших в выборку аналитика. К ограничениям модели относится период времени, на протяжении которого аналитик работает с данными и во время которого проводилось наблюдение.

Будем считать, что аналитик в своей повседневной работе нацелен на решение определенной задачи, поэтому он, как правило, активно использует 20–30 атрибутов, но может обращаться и к другим атрибутам. Иногда такие выборки могут иметь большой объем. Этот случай будет рассмотрен в задаче 1. Затем в задачах 2 и 3 будет рассмотрен случай, когда инсайдер чаще выбирает значимые атрибуты и накапливает их значения. После чего в задаче 4 будет предполагаться, что инсайдер в процессе работы ведет себя как честный аналитик и лишь с определенной частотой выбирает не предназначенные для него атрибуты.

3 Имитационное моделирование

В данном разделе приводятся результаты испытаний, при которых аналитик или получает из хранилища атрибут с номером n , или не получает его. Выбор всех атрибутов производится независимо друг от друга. Такие отклики модели называют переменными Бернулли [8]. Для определения того, попал ли атрибут в выборку аналитика, будет использован генератор псевдослучайных чисел с большим периодом $(2^{19937} - 1)$ [9] и метод Монте Карло [8].

Все испытания будут повторяться q раз для достижения требуемой в каждой задаче точности. Все наблюдения независимы.

Задача 1. Как было сказано ранее, в процессе работы с хранилищем данных бывают случаи, когда аналитик делает выборки большого объема. Можно ли в этом случае однозначно отличить его от инсайдера по информации о тех атрибутах и их значениям, которые он выбирает?

Пусть хранилище данных имеет размер $N \times T$ и аналитик делает из него выборки размером $p_n \times p_m$ случайно равновероятно. Все атрибуты и их значения классифицированы как «важные» и «обычные».

Для параметров модели выбраны следующие значения параметров: $N = 2000$, $T = 50\,000$, $pn = 400$ и $pm = 400$. Все значения близки к реальным объемам данных в хранилище. Подматрица размером $pn \times pm$, которую выбрал аналитик, называется его «профилем».

Основываясь на изучении существующих хранилищ данных, можно предположить, что в нем порядка 0,5% «важных» атрибутов и 5% «важных» значений атрибутов, поэтому на первом шаге имитационного моделирования матрица для модели заполняется с использованием генератора псевдослучайных чисел [9].

Для всех атрибутов из A вводится функция веса атрибута v , принимающая значения

$$v(a_i) = \begin{cases} 100, & \text{если атрибут } a_i \text{ «важный»;} \\ 1 & \text{в ином случае} \end{cases}$$

и весовая функция значений атрибутов (обозначим так же) v , принимающая следующие значения:

$$v(a_{ij}) = \begin{cases} 100, & \text{если значение } a_{ij} \text{ «важное»;} \\ 1 & \text{в ином случае.} \end{cases}$$

Для профиля также вводится функция

$$v(P) = \sum_{j=1}^{pm} \left(\sum_{i=1}^{pn} \left(\frac{v(a_i)v(a_{ij})}{pn} h \right)^2 \right) p,$$

где

$$h = \begin{cases} 1000, & \text{если одновременно } a_{ij} \text{ и } a_i \text{ «важные»;} \\ 1 & \text{в ином случае;} \end{cases}$$

$$p = \begin{cases} 1000, & \text{если в одной строке профиля более чем один раз встречаются} \\ & \text{одновременно важные } a_{ij} \text{ и } a_i; \\ 1 & \text{в ином случае.} \end{cases}$$

Существуют различные способы определения «важности» атрибута через подбор значений весовых функций, но для больших объемов данных существенна разница в несколько порядков для определения значимости. Проведя моделирование работы аналитика с хранилищем данных, в процессе которой он сделал 20 выборок размера $pn \times pm$, можно наблюдать картину для весов профилей, отраженную на рис. 1.

На гистограмме по оси абсцисс расположены профили аналитиков, упорядоченные по убыванию их веса. По оси ординат — вес соответствующего профиля. Над каждым столбцом гистограммы расположено число, обозначающее количество значимых атрибутов в соответствующем профиле.

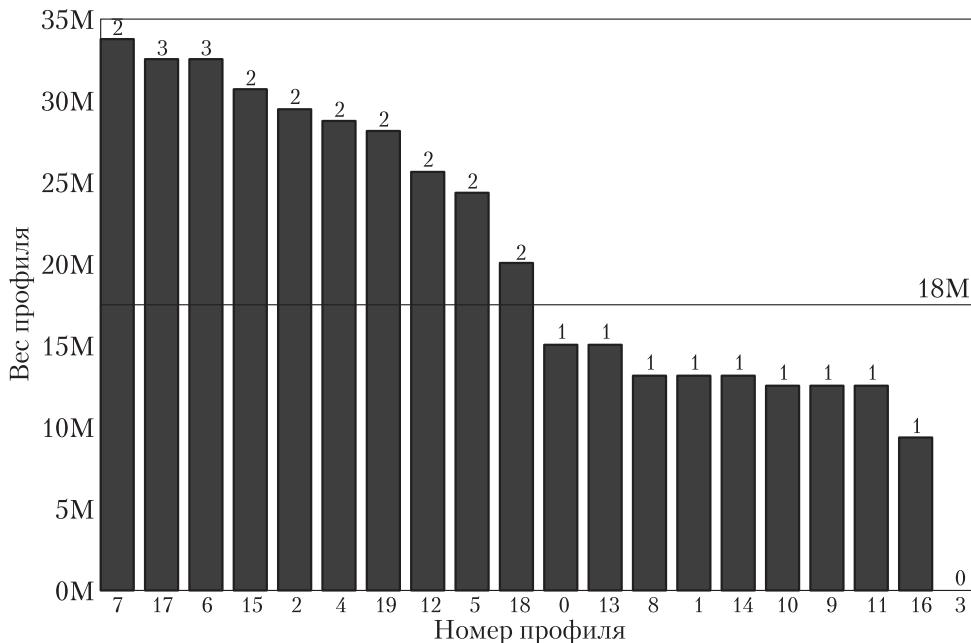


Рис. 1 Веса профилей аналитиков для выборок большого объема. Параметры модели: размер хранилища $2000 \times 50\,000$; размер профиля 400×400 ; функции веса профиля $\text{sum}(\text{sum}((v(a_i) * v(a_{ij})) * h/pn)^2) * p$; $\max v(a_i) = 100,00$; $\max v(a_{ij}) = 100,00$; количество важных атрибутов 8; доля важных значений 5,00%; доля важных атрибутов 0,50%

Нетрудно показать, что модель устойчива для данных значений параметров и их незначительное изменение не повлечет за собой заметного изменения результата моделирования. Сделать это можно, например, проведя повторное моделирование для новых значений.

Как видно из результатов симуляции, значение весовой функции варьируется для разных профилей, но практически каждый из них содержит хотя бы один значимый атрибут, а почти в половине из них встречаются пары и тройки значимых атрибутов, которые являются еще более ценными для инсайдера (например, в банковской сфере это пара {ФИО, номер счета} или тройка {ФИО, номер счета, сумма счета}). Отсюда можно сделать вывод, что для больших выборок любой аналитик (как честный, так и инсайдер) получит ценную информацию из хранилища данных в описанной модели и при схожем поведении однозначно отличить инсайдера от аналитика нельзя.

Задача 2. Пусть теперь поведение инсайдера отличается от поведения аналитика, т. е. вероятность выбрать ценный атрибут, не входящий в его функциональные обязанности, у него выше. Необходимо ответить на вопрос: можно ли в этом

случае отличить их? Для этого упорядочим атрибуты в хранилище по убыванию вероятности их попадания в выборку аналитика U_α . Саму вероятность положим равной $P_\alpha(n) = c_\alpha/n^\alpha$, где

$$c_\alpha = \frac{1}{\sum_{n=1}^{\text{pm}} 1/n^\alpha},$$

n — порядковый номер атрибута.

Выберем параметры модели таким образом, чтобы отразить ситуацию, когда аналитик делает $\text{pm} = 1000$ выборок из хранилища данных по одной строке с вероятностью $P_\alpha(n) = c_\alpha/n^\alpha$. Тогда для разных значений α ожидаемая вероятность попадания атрибута в выборку аналитика имеет вид, представленный на рис. 2.

Инсайдер, в отличие от аналитика, накапливает ценную информацию, не относящуюся к его функциональным обязанностям, т. е. атрибутов с большим номером. Обозначим вероятность выбора атрибута n для инсайдера как $P_\gamma(n) = c_\gamma/n^\gamma$, $\gamma < \alpha$, т. е., начиная с определенного номера (координата на оси абсцисс места пересечения графиков), вероятность попадания атрибута в выборку у инсайдера выше, чем у аналитика. В табл. 1 приведены рассчитанные точки пересечения графиков $P_\alpha(n)$ и $P_\gamma(n)$ для разных значений α и γ . Из нее заранее убраны те значения, где графики пересекаются ранее 20-го атрибута. В столбцах табл. 1 расположены значения параметра γ , а в строках — α . На пересечении расположены номер атрибута, начиная с которого аналитик с параметром γ (который в описанной модели является инсайдером) накапливает лишнюю информацию.

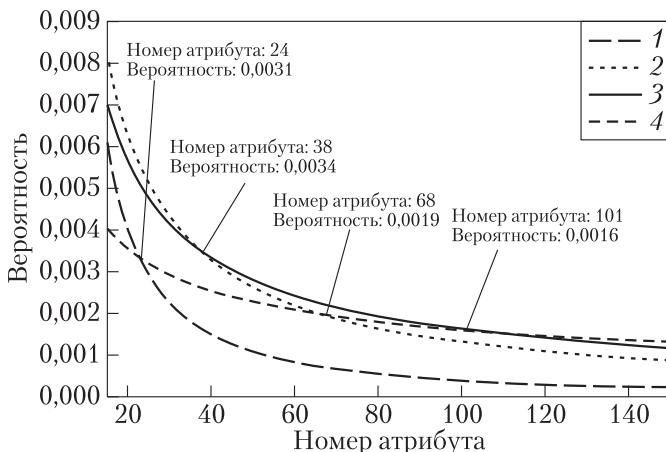


Рис. 2 Ожидаемая вероятность попадания атрибута в выборку аналитика: 1 — $\alpha = 1,5$; 2 — $1,0$; 3 — $0,8$; 4 — $\alpha = 0,5$

Таблица 1 Номера атрибутов места пересечения графиков $P_\alpha(n)$ и $P_\gamma(n)$

α	γ																			
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0,1	312	290	268	243	218	193	167	143	120	100	83	68	56	47	39	33	28	25	21	
0,2	312	270	248	224	200	175	151	128	106	88	72	59	49	40	34	29	24			
0,3	290	270	227	204	180	157	134	113	93	76	62	51	42	35	29	24	21			
0,4	268	248	227	183	161	139	117	98	80	65	53	43	35	29	24	21				
0,5	243	224	204	183	141	120	101	84	68	55	45	36	29	24	20					
0,6	218	200	180	161	141	103	86	70	57	46	37	30	24							
0,7	193	175	157	139	120	103		71	58	47	37	30	24							
0,8	167	151	134	117	101	86	71		47	38	30	24								
0,9	143	128	113	98	84	70	58	47												
1,0	120	106	93	80	68	57	47	38	30											
1,1	100	88	76	65	55	46	37	30	24											
1,2	83	72	62	53	45	37	30	24												
1,3	68	59	51	43	36	30	24													
1,4	56	49	42	35	29	24														
1,5	47	40	35																	
1,6	39	34	29	24	21															
1,7	33	29	24																	
1,8	28	24	21																	
1,9	25	21																		
2,0	21																			

Таким образом, в табл. 1 для каждой пары параметров приведена нижняя граница интервала, на котором возможен поиск инсайдера статистическими методами.

Задача 3. Необходимо методом имитационного моделирования определить условия, в которых можно однозначно выявить инсайдера в множестве аналитиков.

Для моделирования будет использоваться генератор псевдослучайных чисел аналогично задаче 1. Потребуем, чтобы наблюдаемая частота выборки атрибутов θ отличалась от истинной частоты p не более чем на $d = 0,02$ с вероятностью $a = 0,95$. Для оценки числа испытаний воспользуемся формулой $n = Z_{a/2}^2 / (4d^2)$ [8], где $Z_{a/2}$ — стандартная нормальная статистика для искомой вероятности, $Z_{a/2} = 1,96$; $d = 0,02$ — заданная положительная величина, такая что $P\{|\theta/n - p| \leq d\} = 1 - a$.

Тогда для вышеописанных требований получается $n = 1,96^2 / (4 \cdot 0,02^2) \geq 2400$.

Вводятся дополнительные ограничения на модель:

- (1) в задаче моделируется работа аналитика с хранилищем данных компании, поэтому должно быть ограничено время наблюдения за его работой. Доста-

точным будет потребовать, чтобы вероятность попадания атрибута в выборку превышала $1/1000$. Поэтому будут рассматриваться только первые $h(\alpha)$ атрибутов, где $h^\alpha = 1000c_\alpha$. Значения параметра h для рассматриваемых значений α приведены в табл. 2;

- (2) на практике аналитик работает с 20–30 атрибутами, поэтому случаи, где графики пересекаются до 20-го номера, тоже не рассматриваются.

Таким образом, для всех пар α и γ получаются отрезки, на которых необходимо проводить моделирование.

Рассмотрим случай, когда $\alpha = 0,8$ и $\gamma = 0,5$. Тогда интересующий отрезок равен [101; 183].

На рис. 3 приведены графики $P_\alpha(n)$ для значений $\alpha \in \{0,5; 0,8; 1,0; 1,5\}$.

Как видно из графиков, результат моделирования показал, что отличить аналитика от инсайдера можно на меньшем отрезке, чем на основе ожидаемой вероятности попадания атрибута в выборку (см. табл. 1), из-за чего количество подходящих для использования пар α и γ меньше. Для взятых ранее значений параметров отрезок на основе табл. 1 и 2 равен [101; 183], а на основе наблюдений — [120; 183]. Для остальных пар α и γ можно получить аналогичные результаты.

Согласно приведенной выше оценке, 2400 наблюдений достаточно для обеспечения требуемой близости частот. При меньшем числе наблюдений отрезок будет все больше сокращаться, что иллюстрирует рис. 4, на котором приведены данные 100 наблюдений.

Как несложно заметить, из-за сильного разброса наблюдаемых частот при таком числе наблюдений нельзя различить аналитика и инсайдера со значениями параметров $\alpha = 0,8$ и $\gamma = 0,5$ соответственно.

Задача 4. Необходимо найти условия, при которых можно однозначно выявить инсайдера в множестве аналитиков, если инсайдер с вероятностью $1 - p$ ведет себя как честный аналитик, т. е. вероятность попадания атрибута в его выборку равна c_α/n^α , но с вероятностью p он делает выборку атрибутов с c_γ/n^γ , $\gamma < \alpha$. Данный случай наиболее точно описывает реальное поведение инсайдера — как правило, он проявляет осторожность при работе с данными и собирает их порциями. Ниже приводятся результаты моделирования данной стратегии поведения с теми же требованиями к точности, что предъявились и ранее. Значения параметров модели выбираются такими же: пусть $\alpha = 0,8$ и $\gamma = 0,5$. Симуляции проводятся для всех $p \in (0; 1)$ с шагом 0,1 с целью приближенно

Таблица 2 Расчетные ограничения на количество используемых атрибутов в модели

α	h
0,1	351
0,2	331
0,3	309
0,4	286
0,5	262
0,6	236
0,7	210
0,8	183
0,9	158
1,0	134
1,1	112
1,2	93
1,3	77
1,4	64
1,5	54
1,6	45
1,7	38
1,8	33
1,9	28
2,0	25

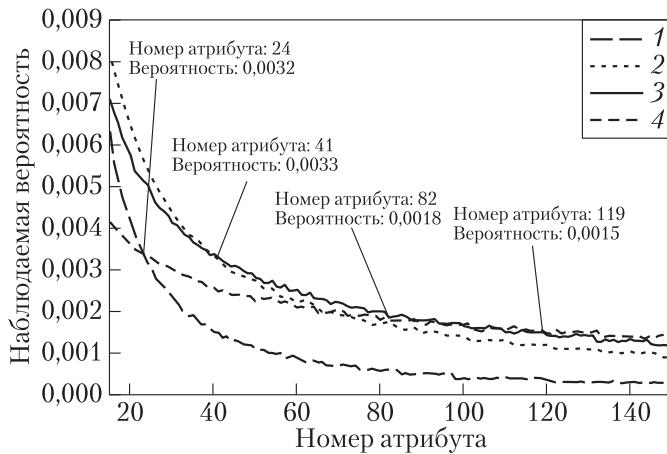


Рис. 3 Зависимости $P_\alpha(n)$ для 3000 наблюдений: 1 — $\alpha = 1,5$; 2 — $1,0$; 3 — $0,8$; 4 — $\alpha = 0,5$

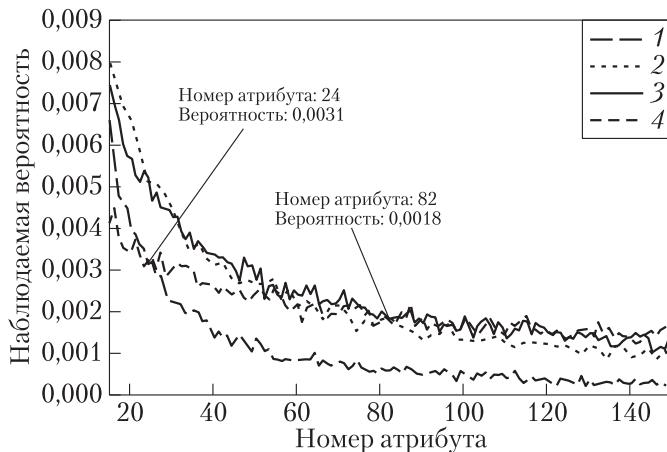


Рис. 4 Зависимости $P_\alpha(n)$ для 100 наблюдений: 1 — $\alpha = 1,5$; 2 — $1,0$; 3 — $0,8$; 4 — $\alpha = 0,5$

найти то значение p , когда среднюю вероятность выбора атрибутов до 180-го (см. задачу 3) инсайдером невозможно однозначно отличить от соответствующей вероятности для честного аналитика.

В результате моделирования получается картина, отраженная на рис. 5, и вывод: если $p < 0,2$, т. е. инсайдер с вероятностью не более 20% делает выбор «лишних» для него атрибутов, невозможно отличить его от честного аналитика за рассматриваемое время наблюдения. При этом если $p = 0,3$, то существует интервал, на котором статистическими методами можно их отличить.

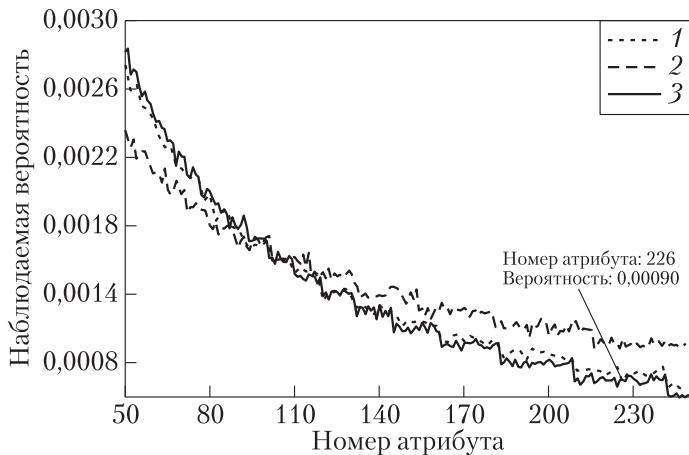


Рис. 5 Инсайдер с вероятностью 20% (1) выбирает лишние для него атрибуты: 2 — γ ; 3 — α

Аналогичные эксперименты можно провести и для других пар α и γ и получить приближенные граничные значения для p .

4 Заключение

Задача поиска инсайдера в множестве аналитиков сложна и при использовании любого подхода к решению требует продолжительного сбора информации о поведении сотрудника и ее анализа. Практически все способы ее решения придерживаются принципа, что чем больше информации из разных аспектов жизни и работы аналитика получается собрать, тем раньше и точнее можно определить, является ли он инсайдером или же работает честно. При этом увеличение объема данных приводит к необходимости работы с нечисловыми данными или с данными высокой размерности. Эта проблема рассматривалась в работе [10], где показана ее разрешимость на конечном шаге. Однако в определенных случаях достаточно наблюдать, к каким данным в хранилище обращается инсайдер.

В статье описана модель поведения инсайдера в множестве аналитиков, рассмотрены различные сценарии его поведения и их формализация, а также с помощью математического моделирования получены условия, при которых статистическими методами за конечное число шагов можно однозначно выявить инсайдера.

Литература

1. General Services Administration. Anomaly detection at multiple scales (ADAMS). 22.10.2010. Retrieved 05.12.2011. 40 p. https://www.fbo.gov/download/2f6/2f6289e99a0c04942bbd89ccf242fb4c/DARPA-BAA-11-04_ADAMS.pdf.

2. Senator T., Bader D., Chow E., et al. Detecting insider threats in a real corporate database of computer usage activity // 19th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings. — New York, NY, USA: ACM, 2013. P. 1393–1401.
3. Yu R., He X., Liu Y. GLAD: Group anomaly detection in social media analysis // arXiv.org, October 7, 2014. arXiv:1410.1940.
4. System G: Developed graph computing industry solutions. <http://systemg.research.ibm.com/solutions.html>.
5. Allure Security Technology Inc. Anomaly detection at multiple scales (ADAMS). November 9, 2011. Final Report. <https://info.publicintelligence.net/DARPA-ADAMS.pdf>.
6. Клейнен Дж. Статистические методы в имитационном моделировании. Часть 1 / Пер. с англ.— М.: Статистика, 1974. 222 с. (Kleijnen J. P. C. Statistical techniques in simulation. Part 1. — New York, NY, USA: Marcel Dekker, Inc., 1974. 285 р.)
7. Клейнен Дж. Статистические методы в имитационном моделировании. Часть 2 / Пер. с англ.— М.: Статистика, 1978. 221 с. (Kleijnen J. P. C. Statistical techniques in simulation. Part 2. — New York, NY, USA: Marcel Dekker, Inc., 1975. 488 р.)
8. Шеннон Р. Имитационное моделирование систем — искусство и наука / Пер. с англ.— М.: Наука, 1978. 420 с. (Shannon R. E. Systems simulation — the art and science. — Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1975. 387 р.)
9. Matsumoto M., Nishimura T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator // ACM Trans. Model. Comput., 1998. Vol. 8. No. 1. P. 3–30.
10. Grusho A., Grusho N., Timonina E. Detection of anomalies in non-numerical data // 8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2016. P. 273–276.

Поступила в редакцию 20.03.17

IMITATION MODEL OF INSIDER DETECTION BY STATISTICAL TECHNIQUES

E. A. Martyanov

M. V. Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

Abstract: The paper considers the task of insider detection in a group of analysts who work with a data warehouse, presented as a raw table with a huge amount of attributes. The main difference in the behavior of a legitimate analyst and an insider is that the latter collects data redundant for his/her functionality during his/her work cycle. Thus, to detect an insider, it is enough to detect the regular fact of redundancy on his/her requests of data, which he/she can consider and use to damage a company. The paper presents the mathematical model of

insider behavior, the formal definition of the main difference in the behavior of a legitimate analyst and an insider, and the results of modeling. The conditions when it is possible to use statistical criteria to solve the task are found.

Keywords: insider threat; redundant data collection; statistical criteria; mathematical model; systems simulation

DOI: 10.14357/08696527170205

References

1. General Services Administration. 22.10.2010. Retrieved 05.12.2011. Anomaly detection at multiple scales (ADAMS). 40 p. Available at: <https://www.fbo.gov/download/2f6/2f6289e99a0c04942bbd89ccf242fb4c/DARPA-BAA-11-04-ADAMS.pdf> (accessed April 27, 2017).
2. Senator, T., D. Bader, E. Chow, et al. 2013. Detecting insider threats in a real corporate database of computer usage activity. *19th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings*. New York, NY: ACM. 1393–1401.
3. Yu, R., X. He, and Y. Liu. 07.10.2014. GLAD: Group anomaly detection in social media analysis. *arXiv.org*. arXiv:1410.1940.
4. System G: Developed graph computing industry solutions. 2014. Available at: <http://systemg.research.ibm.com/solutions.html> (accessed April 27, 2017).
5. Allure Security Technology Inc. November 9, 2011. Anomaly detection at multiple scales (ADAMS). Final Report. Available at: <https://info.publicintelligence.net/DARPA-ADAMS.pdf> (accessed April 27, 2017).
6. Kleijen, J. P. C. 1974. *Statistical techniques in simulation. Part 1*. New York, NY: Marcel Dekker, Inc. 285 p.
7. Kleijen, J. P. C. 1975. *Statistical techniques in simulation. Part 2*. New York, NY: Marcel Dekker, Inc. 488 p.
8. Shennon, R. E. 1975. *Systems simulation — the art and science*. Englewood Cliffs, NJ: Prentice-Hall, Inc. 387 p.
9. Matsumoto, M., and T. Nishimura. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput.* 8(1):3–30.
10. Grusho, A., N. Grusho, and E. Timonina. 2016. Detection of anomalies in non-numerical data. *8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 273–276.

Received March 20, 2017

Contributor

Martyanov Evgeny A. (b. 1991) — PhD Student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskije Gory, Moscow 119991, Russian Federation; zhenya388@yandex.ru

СИСТЕМА УПРАВЛЕНИЯ ОБУЧЕНИЕМ ELIS. АРХИТЕКТУРНЫЕ РЕШЕНИЯ

A. K. Горшенин¹, Е. С. Данилович², Д. Р. Хромов³

Аннотация: Развитие информационных технологий оказывает значительное влияние на различные сферы человеческой деятельности, в том числе приводит к формированию новых образовательных требований и стандартов. Электронные формы обучения вызывают повышенный интерес у учащихся и предоставляют новые возможности преподавателям. Для автоматизации различных аспектов современного образовательного процесса используются так называемые системы управления обучением. В статье описываются основные архитектурные решения, разработанные для системы управления обучением ELIS (Electronic Learning Intelligent System), которая может служить единой образовательной площадкой как для высших учебных заведений, так и для различных частных и государственных компаний.

Ключевые слова: система управления обучением; электронное обучение; SaaS; облачные вычисления

DOI: 10.14357/08696527170206

1 Введение

Развитие информационных технологий оказывает значительное влияние на различные сферы человеческой деятельности [1], в том числе приводит к формированию новых образовательных требований и стандартов. В качестве примеров можно упомянуть систему школьных электронных дневников, автоматизированные средства оценивания знаний и учета успеваемости в университетах, решения для обучения персонала в коммерческих компаниях и др. Стоит отметить, что электронные формы обучения вызывают повышенный интерес у студентов, так как позволяют использовать привычные для них технологии или получить некоторый новый опыт. Преподаватели получают возможности интерактивной демонстрации различных аспектов курса и автоматизации контроля образовательного процесса, в частности при проведении промежуточных и итоговых видов аттестации.

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский технологический университет (МИРЭА), agorshenin@frccsc.ru

²Московский технологический университет (МИРЭА), cio@andromeda-inc.com

³Московский технологический университет (МИРЭА), ceo@andromeda-inc.com

При разработке электронных решений возникают задачи эффективной реализации взаимодействия между образовательными системами и пользователями, обеспечения быстрого сбора дополнительной информации для дальнейшего извлечения знаний и получения аналитических материалов. Для автоматизации современного образовательного процесса принято разрабатывать и внедрять различные системы управления обучением (**LMS**, *Learning Management System*), которые позволяют осуществлять общее администрирование, формирование отчетности по образовательным курсам и учебным программам, координацию взаимодействия преподавателей и обучающихся, мониторинг показателей их деятельности и т. п. При этом в ряде работ отмечено (см., например, [2]), что наибольшую заинтересованность пользователей таких систем вызывают прежде всего возможности эффективной коммуникации, а не только реализованные интерактивные решения в образовательном процессе. Поэтому зачастую [3] организации предпочитают разработку собственных систем электронного обучения для наиболее точного учета всех специфических требований, в частности для реализации адаптивных решений [4].

В данной работе описываются основные архитектурные решения для системы управления обучением **ELIS**, которая может служить единой образовательной площадкой как для высших учебных заведений, так и для различных частных и государственных компаний.

2 Основные требования к системе ELIS

В статье [5] была предложена реализация автоматизированного средства аттестации с учетом основных требований работы [6] для электронной поддержки образовательного процесса в рамках курса «Программирование» в Московском технологическом университете (МИРЭА) [7–10]. Данное решение разрабатывалось в рамках парадигмы «клиент–сервер», которая предоставляет следующие преимущества: отсутствие необходимости дублирования кода на стороне клиента, умеренные требования к пользовательскому оборудованию и программному обеспечению, повышение уровня безопасности за счет организации политик доступа на сервере. Однако хорошо известны и недостатки данной архитектуры: зависимость работоспособности и производительности системы от состояния сервера, сложности с администрированием, необходимость в финансовых затратах на серверное оборудование, увеличение трафика при усложнении функционала клиентских приложений.

Современные системы управления обучением должны предоставлять возможности масштабирования, обеспечивать отказоустойчивость и т. д. Поэтому решения для системы **ELIS** построены на облачной платформе с моделью обслуживания **SaaS** (*Software-as-a-Service* — программное обеспечение как услуга), в рамках которой обучающиеся и преподаватели получают возможность доступа к реализованным в системе решениям с помощью тонкого клиента. Использование облачных решений может также существенно снизить инфра-

структурные расходы, а также позволяет гибко реагировать на увеличение потребностей в вычислительных мощностях (например, при масштабировании — в случае увеличения числа пользователей). Необходимо отметить, что использование данной парадигмы весьма востребовано в образовательной сфере (см., например, статью [11]).

Коротко рассмотрим основные требования к системе управления обучением, которые были учтены при разработке **ELIS**:

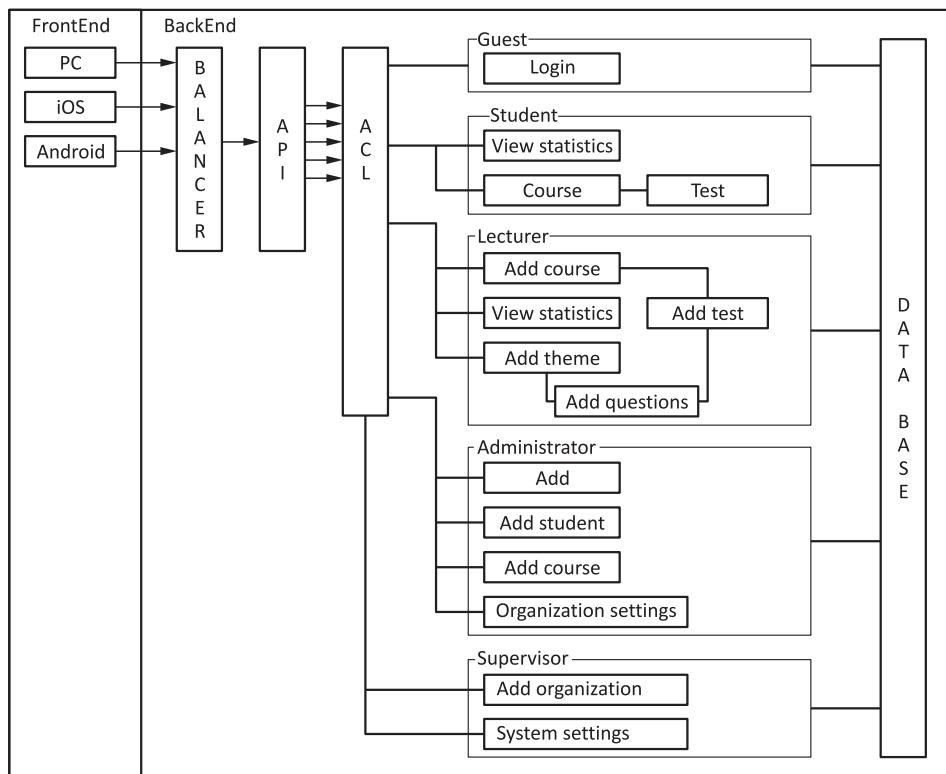
- система должна быть ориентирована на облачные технологии для обеспечения одновременного доступа большого числа пользователей;
- необходима поддержка кроссплатформенности системы и реализация возможности доступа через специальные мобильные приложения;
- сервис должен обладать достаточной степенью универсальности для предоставления образовательных услуг как в высших учебных заведениях, так и в государственных и частных компаниях;
- все данные, включая учебные материалы, проверочные задания, правильные ответы и т. д., хранятся на сервере; доступ предоставляется исключительно в защищенном режиме;
- каждый пользователь имеет уникальную учетную запись (личный кабинет) с реализацией различных уровней прав доступа в зависимости от роли в системе;
- администратор может добавлять представителей учебных заведений или компаний, подавших заявку на подключение их организации к системе, а также осуществляет общие функции поддержки корректности работы;
- представитель может изменить информацию о компании (например, название, список предметов, расписание и т. д.), добавляет преподавателей и закрепляет за ними предметы и учебные группы, делегирует права доступа к статистике всего учебного заведения, структурного подразделения, предмета и т. д.;
- преподавателям доступно создание и редактирование учебных и тестовых материалов, отслеживание выполнения домашнего задания. Тесты можно создавать на основе разделов курса, в частности без привязки к какой-либо конкретной теме. Должна быть предусмотрена опция «случайного» формирования вопросов из заданных разделов и порядка вывода возможных вариантов ответов для них;
- обучающиеся должны иметь возможность просматривать учебные материалы по любому доступному предмету, выполнять домашние задания, проходить тесты и сдавать контрольные работы, иметь обратную связь с преподавателями и другими учениками внутри системы с помощью решений, реализованных в их личном кабинете.

В следующих разделах рассмотрим основные архитектурные решения и использованные средства разработки для реализации описанной в данном разделе концепции.

3 Архитектурные компоненты системы ELIS

Система управления обучением ELIS состоит из двух основных архитектурных частей — FrontEnd и BackEnd (см. рисунок), функционирование которых рассмотрено в данном разделе.

Все данные и основная часть логики бизнес-процессов размещается на серверах системы, при этом мобильные приложения и сайт выполняют роль тонких клиентов. Взаимодействие между компонентами системы осуществляется посредством единого программного интерфейса (API, Application Programming Interface). Это позволяет создавать тонкие клиенты для различных платформ без модификации BackEnd-компоненты. Обращение к API для осуществления взаимодействия происходит посредством GET/POST HTTP-методов, ответ реализуется в виде объектов JSON (JavaScript Object Notation). Использование подобного формата позволяет передавать структуры произвольной сложности, при этом пакеты данных имеют гораздо меньший размер по



Архитектура FrontEnd- и BackEnd-компонентов системы ELIS

сравнению с XML (*eXtensible Markup Language*), что положительно сказывается на скорости передачи информации по сети. Обмен данными происходит через шифрованный SSL-канал для уменьшения возможности перехвата или подмены данных в процессе передачи. При обработке каждого запроса предварительно происходит проверка прав доступа пользователя к запрошенному объекту.

В начале работы необходимо пройти авторизацию и получить идентификатор сессии, который будет использоваться во всех дальнейших запросах к API. Авторизация осуществляется посредством отправки пары значений логин / пароль, которая передается на сервер, где производится хеширование пароля, его проверка и поиск пользователя. В случае, если пользователь был найден, происходит генерация сессионного токена, который передается клиентскому приложению. Данный токен используется во всех последующих запросах для авторизации пользователя. Предусмотрена возможность ассоциации сессионного токена с IP-адресом пользователя для предотвращения его использования в случае перехвата третьими лицами. В случае обнаружения пользователя и корректности введенного пароля производится авторизация и осуществляется загрузка списков контроля доступа ACL (*Access Control List*) для конкретного пользователя. Права каждого пользователя основаны на его персональной роли, а также возможностях, доступных для группы, в которой он состоит. Отметим, что проверка наличия разрешения у авторизованного пользователя на выполнение действия или доступа к запрошенным данным проводится при каждом обращении к API. В случае, если разрешение имеется, соответствующий запрос выполняется успешно. Все обращения к API осуществляются посредством балансировщика нагрузки, который направляет запросы к наименее загруженному узлу, что позволяет эффективно решать задачу масштабирования системы для изменяющегося числа пользователей.

В системе ELIS предусмотрена возможность автоматического тестирования введенных пользователями программ. При загрузке кода программы он попадает в очередь на проверку. При достижении очереди происходит компиляция программы и запуск в виртуальной среде выполнения. Для каждого набора входных данных, заданных преподавателем, происходит запуск программы и проверка результата. Общая оценка за программу выставляется на основании количества успешно пройденных тестов. Тестирование программ производится в отдельной виртуальной машине KVM (*Kernel-based Virtual Machine*), что исключает возможность проникновения на основные серверы системы. Помимо этого, внутри виртуальной машины производится запуск в виртуальном контейнере LXC (*Linux Containers*).

В качестве технологической основы при реализации были также использованы:

- MySQL — для работы с базами данных;
- xPDO — как интерфейс программного взаимодействия с базой данных (на основе технологии Object-Relational Mapping, ORM);

- `nginx` — как веб-сервер;
- `Kleine` — как библиотека роутинга на PHP (используется для серверной части).

Детали реализации FrontEnd-компоненты системы, включая разработку приложений для мобильных платформ (см. рисунок), будут рассмотрены в дальнейших работах. Отметим, что в данном случае речь идет о реализации тонких клиентов, в то время как основной функционал системы реализован в серверной части.

4 Пользовательские роли в системе ELIS

Внутри системы **ELIS** предусмотрено разделение пользователей на следующие группы с различными политиками доступа (см. рисунок):

- Супервизор;
- Администратор;
- Преподаватель;
- Студент;
- Гость.

Далее рассмотрим каждую из групп подробнее.

4.1 Супервизор

Пользователи из этой группы имеют неограниченные права в системе. Их основные функции:

- поддержание работы сервиса;
- регистрация новых учебных заведений;
- управление администраторами.

4.2 Администратор

Пользователи этой группы являются представителями организаций и осуществляют техническое управление образовательными процессами в своих компаниях на базе площадки **ELIS**. Администраторы назначаются при регистрации учебных заведений, их список может быть изменен супервизором. Основные функции:

- регистрация преподавателей;
- регистрация студентов (включая выдачу разрешений на самостоятельную регистрацию);
- добавление предметов и курсов (включая делегирование соответствующих функций преподавателям);
- изменение настроек и сведений об учебном заведении в системе.

4.3 Преподаватель

Пользователи этой группы назначаются администратором учебного учреждения. Основные функции:

- создание тестов на основе собственных или публичных баз вопросов других преподавателей;
- формирование баз вопросов и их публикация в общесистемном каталоге;
- контроль результатов тестов обучающихся из своих групп с добавлением комментариев по пройденным тестам;
- размещение учебных материалов по преподаваемым курсам;
- взаимодействие с обучающимися посредством инструментов системы.

4.4 Студент

Наиболее многочисленная группа пользователей, права делегируются администратором соответствующего учреждения. Основные доступные функции:

- прохождение тестов;
- контроль собственной успеваемости по изучаемым дисциплинам;
- взаимодействие с преподавателями и другими обучающимися посредством инструментов системы.

4.5 Гость

Базовая группа для всех неавторизированных пользователей для реализации возможности просмотра публичных материалов и первичного знакомства с основными функциональными элементами в тестовом режиме.

5 Заключение

В работе представлено описание основных архитектурных решений разрабатываемой авторами системы управления обучением **ELIS** на базе облачной архитектуры. Описаны пять основных типов пользовательских профилей (**Супервизор**, **Администратор**, **Преподаватель**, **Студент**, **Гость**), отражающих принципиальное различие ролей, соответствующих им в системе. Разработанные архитектурные решения будут использованы в дальнейшем для реализации основных функциональных возможностей, прежде всего для проведения различных видов аттестации и поддержки способов взаимодействия между преподавателями и обучающимися. Безусловно, при развертывании реального решения потребуется уделять повышенное внимание вопросам безопасности (в частности, ряд основных угроз для облачной среды рассмотрен в работе [12]), которые будут изучены в дальнейших исследованиях.

Литература

1. Зацаринный А. А., Горшенин А. К. О социальных аспектах развития современных информационных систем // Консолидация общества: аналитика обеспечения развития России и ее национальной безопасности: Сб. мат-лов III Всеросс. конф. — М.: Когито-Центр, 2016. С. 65–68.
2. Lonn S., Teasley S. D. Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems // Computers Education, 2009. Vol. 53. Iss. 3. P. 686–694.
3. Paulsen M. F. Experiences with Learning Management Systems in 113 European institutions // Educational Technology Society, 2003. Vol. 6. Iss. 4. P. 134–148.
4. Shute V., Towle B. Adaptive E-learning // Educational Psychologist, 2003. Vol. 38. Iss. 2. P. 105–114.
5. Горшенин А. К., Кучин А. С. Об одной реализации автоматизированного средства аттестации // Системы и средства информатики, 2016. Т. 26. № 1. С. 192–205.
6. Горшенин А. К. О принципах разработки электронных средств аттестации учащихся по курсам направления «Программирование» // Информатизация инженерного образования: Труды Междунар. научно-методической конф. ИНФОРИНО-2014. — М.: МЭИ, 2014. С. 529–530.
7. Горшенин А. К. Основы языка программирования С: Курс лекций в форме компьютерных презентаций. — М.: ФИЦ ИУ РАН, 2015. 100 с.
8. Горшенин А. К. Основы алгоритмов и структур данных: Курс лекций в форме компьютерных презентаций. — М.: ФИЦ ИУ РАН, 2015. 88 с.
9. Горшенин А. К. Принципы разработки программного обеспечения и основы языка С. — М.: ФИЦ ИУ РАН, 2016. 100 с.
10. Горшенин А. К. Методы реализации абстрактных типов данных. — М.: ФИЦ ИУ РАН, 2016. 100 с.
11. Sultan N. Cloud computing for education: A new dawn? // Int. J. Inform. Manage., 2010. Vol. 30. Iss. 2. P. 109–116.
12. Subashini S., Kavitha V. A survey on security issues in service delivery models of cloud computing // J. Netw. Comput. Appl., 2011. Vol. 34. Iss. 1. P. 1–11.

Поступила в редакцию 15.03.17

LEARNING MANAGEMENT SYSTEM ELIS. ARCHITECTURE SOLUTIONS

A. K. Gorshenin^{1,2}, E. S. Danilovich², and D. R. Khromov²

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation

Abstract: The progress in information technology has a significant impact on various areas of human activity, including the formation of new educational requirements and standards. Electronic forms of learning provoke increased interest among students and provide new opportunities for teachers. To automate various aspects of the modern educational process, so-called learning management systems are used. The article describes the main architecture solutions developed for the learning management system **ELIS** (**E**lectronic **L**earning **I**ntelligent **S**ystem), which can serve as a common educational platform for both higher education institutions and various private and state companies.

Keywords: learning management system; e-Learning; SaaS; cloud computing

DOI: 10.14357/08696527170206

References

1. Zatsarinny, A. A., and A. K. Gorshenin. 2016. O sotsial'nykh aspektakh razvitiya sovremennoykh informatsionnykh sistem [On social aspects of development of modern information systems]. *Konsolidatsiya obshchestva: Analitika obespecheniya razvitiya Rossii i ee natsional'noy bezopasnosti. Sbornik mat-lov III Vseross. konf.* [3rd All-Russian Conference “Consolidation of Society: Analytics for Ensuring the Development of Russia and Its National Security” Proceedings]. Moscow. 65–68.
2. Lonn, S., and S. D. Teasley. 2009. Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems. *Computers Education* 53(3):686–694.
3. Paulsen, M. F. 2003. Experiences with Learning Management Systems in 113 European institutions. *Educational Technology Society* 6(4):134–148.
4. Shute, V., and B. Towle. 2003. Adaptive E-learning. *Educational Psychologist* 38(2):105–114.
5. Gorshenin, A. K., and A. S. Kuchin. 2016. Ob odnoy realizatsii avtomatizirovannogo sredstva attestatsii [On a realization of an automated testing service]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(1):192–205.
6. Gorshenin, A. K. 2014. O printsipakh razrabotki elektronnykh sredstv attestatsii uchashchikhsya po kursam napravleniya “Programmirovaniye” [On principles of development of the electronic certifications for students in programming courses]. *Conference (International) INFORINO-2014 Proceedings*. Moscow. 529–530.

7. Gorshenin, A. K. 2015. *Osnovy yazyka programmirovaniya C: Kurs lektsiy v forme komp'yuternykh prezentatsiy* [Basic aspects of C Programming language: Lectures in a form of presentation slides]. Moscow: FRC CSC RAS. 100 p.
8. Gorshenin, A. K. 2015. *Osnovy algoritmov i struktur dannykh: Kurs lektsiy v forme komp'yuternykh prezentatsiy* [Basic aspects of algorithms and data structures: Lectures in a form of presentation slides]. Moscow: FRC CSC RAS. 88 p.
9. Gorshenin, A. K. 2016. *Printsipy razrabotki programmnogo obespecheniya i osnovy yazyka C* [Foundations of software development and C programming language]. Moscow: FRC CSC RAS. 100 p.
10. Gorshenin, A. K. 2016. *Metody realizatsii abstraktnykh tipov dannykh* [Methods of implementing abstract data types]. Moscow: FRC CSC RAS. 100 p.
11. Sultan, N. 2010. Cloud computing for education: A new dawn? *Int. J. Inform. Manage.* 30(2):109–116.
12. Subashini, S., and V. Kavitha. 2011. A survey on security issues in service delivery models of cloud computing. *J. Netw. Comput. Appl.* 34(1):1–11.

Received March 15, 2017

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; agorshenin@frccsc.ru

Danilovich Evgeniy S. (b. 1996) — student, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; cio@andromeda-inc.com

Khromov Dmitry R. (b. 1998) — student, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; ceo@andromeda-inc.com

СИСТЕМА УПРАВЛЕНИЯ ОБУЧЕНИЕМ ELIS. ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС И ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ

A. K. Горшенин¹, Е. С. Данилович², Д. Р. Хромов³

Аннотация: Для автоматизации современного образовательного процесса используются различные системы управления обучением, которые позволяют осуществлять общее администрирование, формирование отчетности по образовательным курсам и учебным программам, координацию взаимодействия преподавателей и обучающихся, мониторинг показателей их деятельности. При этом необходимо учитывать качество системы управления обучением, реализованных сервисов и внедренного контента, так как данные аспекты оказывают значительное влияние на удовлетворенность учащихся и успешность использования информационных технологий поддержки электронного обучения. В статье описаны основные функциональные возможности, предоставляемые пользовательским интерфейсом разрабатываемой авторами системы управления обучением ELIS (Electronic Learning Intelligent System). Продемонстрированы результаты пробного тестирования, проведенного с ее помощью в студенческих фокус-группах Московского технологического университета (МИРЭА) и Национального исследовательского университета «Высшая школа экономики».

Ключевые слова: система управления обучением; электронное обучение; SaaS; облачные вычисления

DOI: 10.14357/08696527170207

1 Введение

В статье [1] была предложена реализация автоматизированного средства аттестации с учетом основных требований работы [2] на основе архитектуры «клиент–сервер». Однако хорошо известны недостатки данной парадигмы, поэтому для соответствия требованиям к разработке современных информационных комплексов для системы управления обучением ELIS была использована облачная платформа с моделью обслуживания SaaS (программное обеспечение как

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский технологический университет (МИРЭА), agorshenin@frccsc.ru

²Московский технологический университет (МИРЭА), cio@andromeda-inc.com

³Московский технологический университет (МИРЭА), ceo@andromeda-inc.com

услуга, **Software-as-a-Service**), в рамках которой обучающиеся и преподаватели получают возможность использования реализованных в системе решений с помощью тонкого клиента. В частности, это упрощает решение чрезвычайно актуальной задачи масштабирования, так как число пользователей систем управления обучением может значительно возрастать в ограниченные временные сроки. Более подробно основные архитектурные решения для **ELIS** представлены в статье [3].

Как отмечено в работе [4], при разработке и внедрении систем управления обучением необходимо учитывать качество программного комплекса, реализованных сервисов и используемого контента, так как данные аспекты оказывают значительное влияние на удовлетворенность учащихся (см., например, исследования [5, 6]), а значит, во многом и на оценку успешности использования информационных технологий поддержки электронного обучения. При этом из-за различий в понимании данных требований в каждом конкретном случае зачастую организации предпочитают разработку собственных систем электронного обучения (см., например, исследование [7] с анализом опыта 113 европейских институтов). Отметим, что осуществляются попытки создания унифицированных подходов к разработке систем управления обучением, один из них предложен в работе [8]. Весьма перспективным представляется создание достаточно универсальных решений, которые носили бы междисциплинарный и надорганизационный характер, а также могли обеспечивать некоторые специальные возможности (например, удаленный доступ к научному оборудованию для проведения лабораторных работ, как описано в статье [9] для LMS (Learning Management System) или в исследовании [10] для классической архитектуры «клиент–сервер») для максимально широкого охвата потребностей различных дисциплин.

В данной статье описаны основные функциональные возможности, предоставляемые пользовательским интерфейсом системы управления обучением **ELIS**. Продемонстрированы результаты пробного тестирования, проведенного с ее помощью в студенческих фокус-группах Московского технологического университета (МИРЭА) и Национального исследовательского университета «Высшая школа экономики».

2 Интерфейс системы управления обучением **ELIS**

В рамках системы **ELIS** предусмотрено разделение пользователей на группы с различными политиками доступа (подробнее см. статью [3]). В данном разделе рассмотрим функциональные возможности и примеры представления интерфейса для следующих основных групп: **Студент**, **Преподаватель**, **Администратор**.

Для начала работы необходимо пройти процедуру авторизации. Для входа в систему пользователь вводит свой логин и пароль. После успешной проверки правильности введенных данных происходит определение принадлежности пользователя к одной из описанных выше системных групп и соответствующего

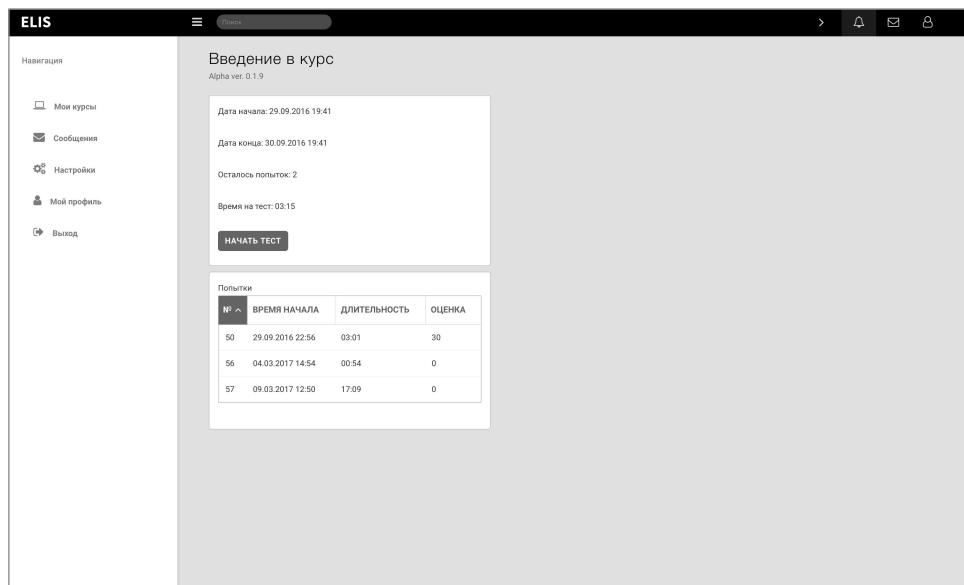


Рис. 1 Личный кабинет обучающегося. Пример

уровня доступа, а затем осуществляется переход на начальную страницу в личном кабинете.

В личном кабинете обучающегося (см. пример на рис. 1) слева расположена панель меню с разделами «Мои курсы», «Сообщения», «Настройки», «Мой профиль». В основной части интерфейса продемонстрировано отображение названия теста, заданного диапазона для его прохождения, количества доступных попыток, значения временного лимита для выполнения заданий, а также результаты (включая предысторию). Для тестов с несколькими допустимыми попытками возможно несколько вариантов выставления итоговой оценки: максимальная из всех попыток, среднее значение по всем прохождениям и т. д. Данный параметр задается преподавателем при создании теста. Если лимит попыток не был израсходован, то тест может быть пройден. Здесь же могут быть указаны доступные лекции на соответствующую тему или иные загруженные преподавателем материалы.

На рис. 2 отображается процедура прохождения некоторого испытания. Продемонстрированный тест состоит из набора вопросов с фиксированными вариантами ответов. Ряд заданий дополнен картинками, формулами, набранными в *TeX*, или примерами программного кода с подсвеченными ключевыми словами. По мере прохождения теста ответы сохраняются — это позволяет прервать выполнение теста и вернуться к нему позже, если такая возможность была преду-

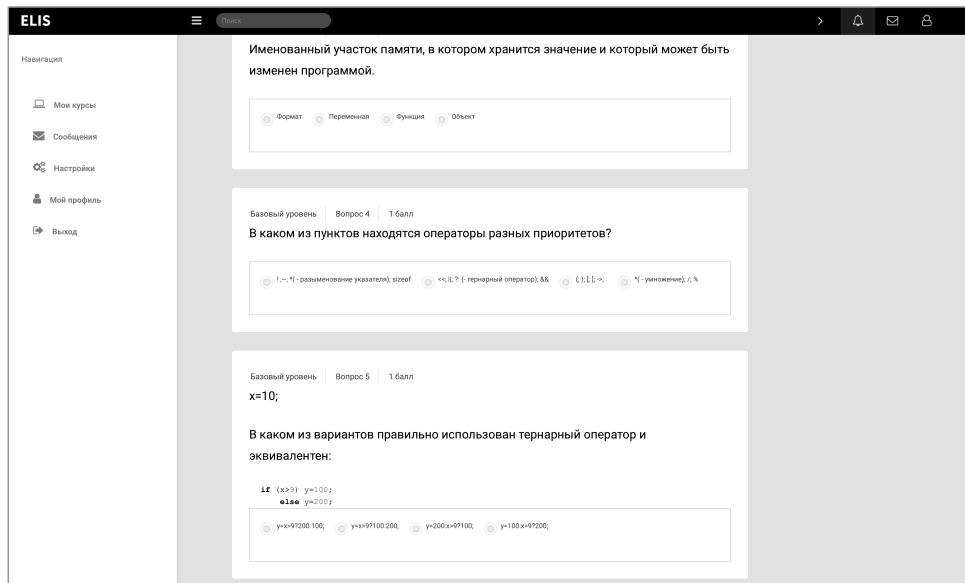


Рис. 2 Тестовые задания. Пример

смотрена преподавателем (в рамках заданного лимита времени). При исчерпании лимита, отведенного на одну попытку прохождения теста, выполнение теста автоматически прекращается, а оценка выставляется на основе полученных к этому моменту ответов.

На начальной странице личного кабинета преподавателя отображается сводная статистика по его курсам. Здесь можно просмотреть сведения о прохождении тестов обучающимися. Преподаватель имеет возможность создавать свои базы вопросов (см. пример на рис. 3), на основе которых выполняется автоматическое формирование вариантов заданий. Все вопросы делятся на темы, каждая из которых относится к определенному предмету. Для добавления новых вопросов в базу в первую очередь необходимо выбрать для редактирования тему, причем как одну из существующих, так и создать новую. Для поиска предусмотрены фильтры по предмету и названию. Для добавления новой темы нужно ввести ее название, выбрать предмет, к которому тема будет относиться, а также указать, будет ли этот набор вопросов публичным, т. е. доступным для использования другими преподавателям.

На рис. 4 представлен пример создания теста из нескольких тем, относящихся к разным разделам курса «Программирование» [11, 12] с указанием количества вопросов, относящихся к соответствующей теме. Здесь же могут быть заданы параметры прохождения теста, которые будут отображаться в лич-

The screenshot shows the ELIS platform's interface for managing topics. On the left is a navigation sidebar with links for 'My courses', 'Topics', and 'Logout'. The main area has a header 'Topics' and a search bar. Below is a table of existing topics:

Nº	TOPIC	QUESTIONS	ADDED	UPDATED	EDIT	DELETE
20	Operators of the C language	0	25.09.2015 09:46	13.11.2015 09:46	EDIT	DELETE
21	Data types	0	25.09.2015	13.11.2015	EDIT	DELETE
22	Another topic				EDIT	DELETE
25	ЭВМ и ПО				EDIT	DELETE
26	Язык С				EDIT	DELETE
27	Algorithms. Forms of representation.				EDIT	DELETE
28	Variables. Operators. Loops.				EDIT	DELETE
29	Arrays. Specifiers of memory classes.				EDIT	DELETE
30	Functions. Special features.	1	25.09.2015 09:46	13.11.2015 09:46	EDIT	DELETE
31	Pointers	0	25.09.2015	13.11.2015 09:46	EDIT	DELETE
32	Language C. Methods of testing. MT.	3	25.09.2015	13.11.2015 09:46	EDIT	DELETE
33	Methods of sorting	0	25.09.2015	13.11.2015 09:46	EDIT	DELETE
34	Structures of data.	0	25.09.2015	13.11.2015 09:46	EDIT	DELETE
35	Files	0	25.09.2015	13.11.2015 09:46	EDIT	DELETE

A modal window titled 'Добавление темы' (Add topic) is open, showing a text input field with 'Новая тема' (New topic) and a checked checkbox for 'разрешить другим преподавателям (из других вузов) использовать вопросы из этой темы' (Allow other teachers from other universities to use questions from this topic). There is a 'SAVE' button at the bottom.

Рис. 3 База тем вопросов. Пример

The screenshot shows the ELIS platform's interface for creating a test. On the left is a navigation sidebar with links for 'My courses', 'Topics', and 'Logout'. The main area has a header 'Create test' and a search bar. Below is a form for setting test parameters:

Start of test:	End of test:	Time:	Attempts:
26.03.2017 01:23	26.03.2017 01:23	01:23	

Below this is a section for adding categories:

NAME OF TOPIC	DISPLAY QUESTIONS:	REMOVE
Variables. Operators. Loops.	7	REMOVE
Language C. Methods of testing. MT.	3	REMOVE
Functions. Special features.	1	REMOVE

At the bottom is a 'SAVE' button.

Рис. 4 Создание теста. Пример

ном кабинете обучающегося. После выбора (или создания) нужной темы можно перейти к редактированию вопросов. Для добавления вопроса необходимо заполнить специальную форму. Сначала задается уровень сложности, всего их предусмотрено три: базовый, средний, сложный. Далее производится ввод текста вопроса, при этом форматирование задается с помощью средств языка разметки **Markdown**. Поддерживается оформление заданий с помощью элементов программного кода на различных языках — при указании его названия производится автоматическое выделение цветом и полужирным шрифтом ключевых слов (см., например, вопрос 5 на рис. 2). Также доступен ввод формул в формате **TeX** для поддержки широкого спектра курсов естественно-научных дисциплин. Все эти средства позволяют оформить текст вопроса максимально наглядным образом. Для удобства в правой части экрана доступен предпросмотр верстки задания.

Поддерживаются следующие допустимые типы ответов:

- выбор единственного корректного варианта из списка заданных;
- выбор нескольких правильных опций из доступных;
- произвольный ввод — ответ представляется в некоторой предопределенной текстовой форме (например, такой вариант удобно использовать для определений, формулировок теорем); предусмотрена возможность указать несколько вариантов написания правильного ответа, в таком случае корректным будет считаться любой из них;
- числовой ввод — ответом является число (либо некоторый диапазон);
- восстановление порядка — создается несколько частей ответа, которые пользователь должен расположить в правильном порядке: восстановление хронологии исторических событий, последовательности шагов алгоритма и т. п.;
- развернутый текст — ответ представляется в свободной форме (например, эссе), оценка осуществляется в ручном режиме преподавателем;
- программа — обучающийся вводит программный код, записанный на одном из поддерживаемых языков; затем он исполняется на заранее заданном преподавателем наборе тестов.

После ввода текста вопроса необходимо выбрать тип ответа и максимальное число баллов, которые можно получить в случае успеха. Затем вопрос сохраняется и добавляется в базу.

В разделе личного кабинета преподавателя «Мои курсы» отображается список всех курсов, к которым преподаватель имеет доступ. Для каждого курса указывается название, описание, соответствующая дисциплина из учебного плана. На публичные курсы может записаться любой студент, на остальные — только после одобрения преподавателем. При выборе определенного курса открывается соответствующий раздел (рис. 5), в котором отображаются интерфейсные элементы «Добавить тест», «Добавить лекцию», «Добавить файл».

The screenshot shows the ELIS application interface. On the left is a sidebar with 'Навигация' (Navigation) containing 'Мои курсы' (My Courses), 'Темы вопросов' (Question Topics), and 'Выход' (Logout). The main area is titled 'Мои курсы' (My Courses) with 'Alpha ver. 0.1.9'. It features three buttons: 'Добавить тест' (Add Test), 'Добавить лекцию' (Add Lecture), and 'Добавить файл' (Add File). Below is a table listing courses:

ID	НАЗВАНИЕ	ТИП	ДАТА	ЗАВЕРШЕНО
7	Первый тест	test	16.06.2016	18.06.2016
8	Второй тест	test	16.06.2016	18.06.2016
9	Тестовая тема теста	test	22.06.2016	15.06.2016
10	Test	test	30.06.2016	10.07.2016
11	Международный Инвестиционный Банк	test	23.09.2016	28.09.2016
12	Aspire One	test	20.09.2016	20.09.2016

Рис. 5 Раздел личного кабинета преподавателя «Мои курсы». Пример

При добавлении теста преподаватель вводит название (тему) теста, дату, время начала и окончания доступности заданий, устанавливает ограничение по времени одной попытки, определяет количество попыток прохождения теста. После этого происходит формирование содержимого теста. При нажатии кнопки «Добавить категорию» открывается окно со списком всех доступных категорий из баз вопросов с указанием количества в каждой. Возможно добавление вопросов как из заданных категорий, так и введение новых. После сохранения тест попадает в элементы курса. Для каждого преподавателя есть возможность прикрепления произвольных файлов в качестве элементов курса. При добавлении файла также указывается его название и краткое описание. При добавлении лекции доступны те же методы форматирования, что и в тексте вопросов.

Администраторы осуществляют техническое управление образовательными процессами в своей компании на базе площадки ELIS. На главной странице администратора отображаются новости организации с возможностью публикации новых элементов, а также сводная статистика по образовательному учреждению: число преподавателей, студентов, процент успеваемости студентов по каждой дисциплине и т. д. Администратор может добавить нового преподавателя посредством заполнения формы с подробной информацией о нем, включая дисциплины, которые он ведет. Предусмотрена возможность импорта сведений о преподавателях из внешнего файла в формате CSV или XML. После добавления преподавателя на заданный адрес электронной почты высыпается письмо с логином и паролем. На рис. 6 продемонстрирован пример редактирования списков контроля доступа в личном кабинете администратора.

В настройках образовательного учреждения указывается регистрационная информация, а также опция самостоятельной регистрации для студентов. Если эта возможность отключена, обучающихся необходимо добавлять вручную. На экране управления студентами отображаются имя, фамилия, группа обучающе-

ID	НАЗВАНИЕ
1	do
3	open_/student/
4	open_/student/course/
5	open_/student/test/run/
6	open_/student/test/view/
7	open_/student/test/attempt/
8	api/student/course/list
9	api/student/course/view
10	open_/api/auth/logout.dse

Рис. 6 Личный кабинет администратора. Списки контроля доступа. Пример

гося, доступны фильтры для поиска. Здесь же доступна полная информация о студенте и его успеваемости. Добавление студентов также возможно в форме импорта из файла.

В качестве технологической основы при реализации были использованы:

- HTML5 (*HyperText Markup Language, version 5*) — верстка веб-интерфейса с использованием наиболее современных стандартов;
- CSS3 (*Cascading Style Sheets*) — стилевое оформления элементов веб-интерфейса;
- DSUI — фреймворк собственной разработки для создания интерфейса и реализации взаимодействия его элементов (с использованием *JavaScript*);
- Highcharts — библиотека для построения диаграмм и графиков;
- Markdown — язык разметки для оформления текстов на сайте;
- поддержка выражений в формате *TeX* для отображения формул;
- Ajax (*Asynchronous Javascript and XML*) — фоновый обмен клиента (браузера) и веб-сервера для обновления контента без полной перезагрузки страницы;
- WebSocket — мгновенная передача сообщений и уведомлений;
- GCC (*GNU Compiler Collection*) — компилятор для проверки пользовательских программ.

3 Тестирование системы ELIS на фокус-группах

В данном разделе продемонстрированы результаты пробного тестирования, проведенного с помощью системы управления обучением ELIS в студенческих

фокус-группах Московского технологического университета (МИРЭА) и Национального исследовательского университета «Высшая школа экономики».

На основе материалов учебных пособий [11–14] был разработан набор тестовых вопросов по языку Си, основным алгоритмам и структурам данных. В качестве проверочной была выбрана тема «Введение в С», состоящая из 17 вопросов. Для каждого студента из фокус-группы бакалавров первого курса Института информационных технологий Московского технологического университета (МИРЭА) в автоматическом режиме был сформирован индивидуальный вариант из десяти заданий. Всего в пробном тестировании приняли участие 28 студентов для оценки работоспособности и удобства пользования системой. Время на прохождение теста в среднем составило 6 мин 15 с. Средний балл ответов — 6,1, максимальный — 10, минимальный — 3. В системе ELIS в разделе для преподавателей доступна полная статистика по каждому тесту (подробнее см. рис. 7). Отметим возможность сортировки обучающихся по полученной оценке, вывод круговой и столбчатой диаграмм для анализа успеваемости студентов.

На рис. 7 продемонстрирован еще один важный элемент интерфейса — панель уведомлений в правой части. Она позволяет реализовать контроль преподавателя



Рис. 7 Результаты тестирования на фокус-группе Московского технологического университета (МИРЭА)



Рис. 8 Результаты тестирования на фокус-группе Национального исследовательского университета «Высшая школа экономики»

за событиями, происходящими в системе: выполнением студентами домашних заданий, прохождением тестов, получением сообщений и т. п. Подобный инструмент чрезвычайно важен для систем управления обучением.

Для тестирования функциональных возможностей были также привлечены студенты Национального исследовательского университета «Высшая школа экономики». Для них предлагался тест на тему «Основы экономики». Всего в данном пробном тестировании приняли участие 17 студентов для оценки работоспособности и удобства пользования системой. Время на прохождение в среднем составило 4 мин 33 с. Средний балл — 7, максимальный — 10, минимальный — 5. На рис. 8 соответствующие аналитические материалы представлены в формате системы ELIS.

Необходимо отметить, что для фокус-групп не ставилась задача максимально быстрого или успешного прохождения испытания. Наибольший интерес представляло именно комплексное взаимодействие с системой, оценка пользовательского опыта и удобства интерфейса. По результатам тестирования студенты были опрошены на предмет удовлетворенности качеством работы системы управления обучением ELIS. В частности, был отмечен значительный интерес к такой

форме прохождения испытаний, взаимодействия с преподавателями. Все студенты проявили желание участвовать в повторном тестировании, в том числе при увеличении числа предметов. Потенциальная реализация тонких клиентов в виде мобильных приложений для различных платформ была признана чрезвычайно привлекательной.

4 Заключение

В работе представлено описание основных функциональных возможностей разрабатываемой авторами системы управления обучением **ELIS**, а также продемонстрированы результаты пробного тестирования на фокус-группах Московского технологического университета (МИРЭА) и Национального исследовательского университета «Высшая школа экономики». В настоящий момент весь функционал доступен исключительно через средства портала, однако предполагается разработка приложений для популярных мобильных платформ на базе архитектуры «клиент–сервер». Необходимо отметить актуальность использования такого рода решений в процессах электронного обучения, в том числе из-за возможности привлечения дополнительной аудитории, проявляющей повышенный интерес к новым информационным технологиям. Кроме того, наблюдается и положительный эффект в качестве усвоения знаний, полученных при работе с мобильными решениями (см., например, [15]). При создании клиентских приложений необходимо ориентироваться на требования, описанные в работе [16], с учетом особого внимания к уровню безопасности (некоторые интересные криптографические и биометрические решения описаны, например, в статьях [17–19]). Данные задачи составляют предмет дальнейших исследований.

Литература

1. *Горшенин А. К., Кучин А. С.* Об одной реализации автоматизированного средства аттестации // Системы и средства информатики, 2016. Т. 26. № 1. С. 192–205.
2. *Горшенин А. К.* О принципах разработки электронных средств аттестации учащихся по курсам направления «Программирование» // Информатизация инженерного образования: Труды Междунар. научно-методической конф. ИНФОРИНО–2014. — М.: МЭИ, 2014. С. 529–530.
3. *Горшенин А. К., Данилович Е. С., Хромов Д. Р.* Система управления обучением **ELIS**. Архитектурные решения // Системы и средства информатики, 2017. Т. 27. № 2. С. 60–69.
4. *Ozkan S., Koseler R.* Multi-dimensional students' evaluation of e-Learning systems in the higher education context: An empirical investigation // Comput. Educ., 2009. Vol. 53. Iss. 4. P. 1285–1296.
5. *Sun P.-Ch., Tsai R.J., Finger G., Chen Y.-Y., Yeh D.* What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction // Comput. Educ., 2008. Vol. 50. Iss. 4. P. 1183–1202.

6. *Liau S.-S.* Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-Learning: A case study of the Blackboard system // Comput. Educ., 2008. Vol. 51. Iss. 2. P. 864–873.
7. *Paulsen M. F.* Experiences with Learning Management Systems in 113 European institutions // Educ. Technol. Soc., 2003. Vol. 6. Iss. 4. P. 134–148.
8. *Avgeriou P., Papasalouros A., Retalis S., Skordalakis M.* Towards a pattern language for learning management systems // Educ. Technol. Soc., 2003. Vol. 6. Iss. 2. P. 11–24.
9. *Rapuano S., Zoino F.* A learning management system including laboratory experiments on measurement instrumentation // IEEE Trans. Instrum. Meas., 2006. Vol. 55. Iss. 5. P. 1757–1766.
10. *Callaghan M. J., Harkin J., McColgan E., McGinnity T. M., Maguire L. P.* Client–server architecture for collaborative remote experimentation // J. Netw. Comput. Appl., 2007. Vol. 30. Iss. 4. P. 1295–1308.
11. *Горшенин А. К.* Принципы разработки программного обеспечения и основы языка С. — М.: ФИЦ ИУ РАН, 2016. 100 с.
12. *Горшенин А. К.* Методы реализации абстрактных типов данных. — М.: ФИЦ ИУ РАН, 2016. 100 с.
13. *Горшенин А. К.* Основы языка программирования С: Курс лекций в форме компьютерных презентаций. — М.: ФИЦ ИУ РАН, 2015. 100 с.
14. *Горшенин А. К.* Основы алгоритмов и структур данных: Курс лекций в форме компьютерных презентаций. — М.: ФИЦ ИУ РАН, 2015. 88 с.
15. *Chen H. R., Huang H. L.* User acceptance of Mobile Knowledge Management Learning System: Design and analysis // Educ. Technol. Soc., 2010. Vol. 13. Iss. 3. P. 70–77.
16. *Горшенин А. К.* Некоторые аспекты разработки мобильных приложений для аттестации учащихся // Информатизация инженерного образования: Труды Междунар. научно-методической конф. ИНФОРИНО-2016. — М.: МЭИ, 2016. С. 92–95.
17. *Wu T.-Y., Tseng Y.-M.* An efficient user authentication and key exchange protocol for mobile client–server environment // Comput. Netw., 2010. Vol. 54. Iss. 9. P. 1520–1530.
18. *Xi K., Ahmad T., Han F., Hu J.* A fingerprint based bio-cryptographic security protocol designed for client/server authentication in mobile computing environment // Secur. Commun. Netw., 2011. Vol. 4. Iss. 5. P. 487–499.
19. *He D. B., Chen J. H., Hu J.* An ID-based client authentication with key agreement protocol for mobile client–server environment on ECC with provable security // Inform. Fusion, 2012. Vol. 13. Iss. 3. P. 223–230.

Поступила в редакцию 15.03.17

LEARNING MANAGEMENT SYSTEM ELIS. USER INTERFACE AND FUNCTIONAL CAPABILITIES

A. K. Gorshenin^{1,2}, E. S. Danilovich², and D. R. Khromov²

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation

Abstract: For automation of the modern educational process, various learning management systems are currently being used. They implement the functionality for general administration, reporting on educational courses and curricula, coordination of interaction between lecturers and students, and monitoring of their progress. At the same time, it is necessary to take into account the quality of a learning management system, its services, and content since these aspects have a significant impact on student satisfaction and the success of using information technologies to support e-Learning. The article describes main functional capabilities provided by the user interface of the learning management system **ELIS** (**E**lectronic **L**earning **I**ntelligent **S**ystem) developed by the authors. The results of trial testing in student focus groups of Moscow Technological University (MIREA) and National Research University Higher School of Economics are demonstrated.

Keywords: learning management system; e-Learning; SaaS; cloud computing

DOI: 10.14357/08696527170207

References

1. Gorshenin, A. K., and A. S. Kuchin. 2016. Ob odnoy realizatsii avtomatizirovannogo sredstva attestatsii [On a realization of an automated testing service]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(1):192–205.
2. Gorshenin, A. K. 2014. O printsipakh razrabotki elektronnykh sredstv attestatsii uchashchikhsya po kursam napravleniya “Programmirovaniye” [On principles of development of the electronic certifications for students in programming courses]. *Conference (International) INFORINO-2014 Proceedings*. Moscow. 529–530.
3. Gorshenin, A. K., E. S. Danilovich, and D. R. Khromov. 2017. Sistema upravleniya obucheniem ELIS. Arkhitekturnye resheniya [Learning management system ELIS. Architecture solutions]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):60–69.
4. Ozkan, S., and R. Koseler. 2009. Multi-dimensional students’ evaluation of e-Learning systems in the higher education context: An empirical investigation. *Comput. Educ.* 53(4):1285–1296.

5. Sun, P.-Ch., R. J. Tsai, G Finger., Y.-Y. Chen, and D. Yeh. 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Comput. Educ.* 50(4):1183–1202.
6. Liaw, S.-S. 2008. Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Comput. Educ.* 51(2):864–873.
7. Paulsen, M. F. 2003. Experiences with Learning Management Systems in 113 European institutions. *Educ. Technol. Soc.* 6(4):134–148.
8. Avgeriou, P., A. Papasalouros, S. Retalis, and M. Skordalakis. 2003. Towards a pattern language for learning management systems. *Educ. Technol. Soc.* 6(2):11–24.
9. Rapuano, S., and F. Zoino. 2006. A learning management system including laboratory experiments on measurement instrumentation. *IEEE Trans. Instrum. Meas.* 55(5):1757–1766.
10. Callaghan, M. J., J. Harkin, E. McColgan, T. M. McGinnity, and L. P. Maguire. 2007. Client-server architecture for collaborative remote experimentation. *J. Netw. Comput. Appl.* 30(4):1295–1308.
11. Gorshenin, A. K. 2016. *Printsypr razrabotki programmnogo obespecheniya i osnovy yazyka C* [Foundations of software development and C programming language]. Moscow: FRC CSC RAS. 100 p.
12. Gorshenin, A. K. 2016. *Metody realizatsii abstraktnykh tipov dannykh* [Methods of implementing abstract data types]. Moscow: FRC CSC RAS. 100 p.
13. Gorshenin, A. K. 2015. *Osnovy yazyka programmirovaniya C: Kurs lektsiy v forme kompyuternykh prezentatsiy* [Basic aspects of C programming language: Lectures in a form of presentation slides]. Moscow: FRC CSC RAS. 100 p.
14. Gorshenin, A. K. 2015. *Osnovy algoritmov i struktur dannykh: Kurs lektsiy v forme kompyuternykh prezentatsiy* [Basic aspects of algorithms and data structures: Lectures in a form of presentation slides]. Moscow: FRC CSC RAS. 88 p.
15. Chen, H. R., and H. L. Huang. 2010. User acceptance of Mobile Knowledge Management Learning System: Design and analysis. *Educ. Technol. Soc.* 13(3):70–77.
16. Gorshenin, A. K. 2016. Nekotorye aspekty razrabotki mobil'nykh prilozheniy dlya attestatsii uchashchikhsya [Some issues of development of mobile applications for student testing]. *Conference (International) INFORINO-2016 Proceedings*. Moscow. 92–95.
17. Wu, T.-Y., and Y.-M. Tseng. 2010. An efficient user authentication and key exchange protocol for mobile client-server environment. *Comput. Netw.* 54(9):1520–1530.
18. Xi, K., T. Ahmad, F. Han, and J. Hu. 2011. A fingerprint based bio-cryptographic security protocol designed for client/server authentication in mobile computing environment. *Secur. Commun. Netw.* 4(5):487–499.
19. He, D. B., J. H. Chen, and J. Hu. 2012. An ID-based client authentication with key agreement protocol for mobile client-server environment on ECC with provable security. *Inform. Fusion* 13(3):223–230.

Received March 15, 2017

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; agorshenin@frccsc.ru

Danilovich Evgeniy S. (b. 1996) — student, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; cio@andromeda-inc.com

Khromov Dmitry R. (b. 1998) — student, Federal State Budget Educational Institution of Higher Education “Moscow Technological University,” 78 Vernadskogo Ave., Moscow 119454, Russian Federation; ceo@andromeda-inc.com

СИНТЕЗ САМОСИНХРОННЫХ КОМБИНАЦИОННЫХ СЕКЦИЙ ФУНКЦИОНАЛЬНЫМ МЕТОДОМ*

Л. П. Плеханов¹

Аннотация: Для разработки самосинхронных (СС) электронных схем, обладающих уникальными свойствами, необходимы специальные методы. Комбинационные схемы могут быть построены из нескольких секций, представляющих собой множество элементов, выполняющих параллельные функциональные преобразования. Известный событийный метод синтеза СС-схем имеет лишь теоретическое значение и непригоден для реального проектирования. В предлагаемом функциональном методе работа секции описывается не событиями (переключениями элементов), а логическими функциями и уравнениями. Метод заключается в составлении и решении логических уравнений. Метод решает задачу синтеза в самом общем виде и позволяет подбирать элементы для реализации из полузаизданных или заказных библиотек.

Ключевые слова: самосинхронные схемы; асинхронные схемы; синтез схем

DOI: 10.14357/08696527170208

1 Введение

Самосинхронные схемы обладают уникальными свойствами по сравнению с другими типами схем, синхронными или асинхронными. К ним относятся отсутствие состязаний, отказобезопасность, правильность функционирования в максимально широком диапазоне внешних условий (температуры и напряжения питания) и некоторые другие. Эти свойства были обоснованы теоретически и проверены экспериментально [1–3].

Многие схемы в цифровой схемотехнике относятся к комбинационным. Это различные варианты арифметико-логических устройств (АЛУ), сумматоров, мультиплексоров, умножителей, делителей и др. Схемы с памятью часто тоже имеют комбинационную часть, поэтому синтез комбинационных схем в СС-исполнении является актуальным.

В классическом событийном подходе известен метод синтеза СС-схем по диаграммам изменения (ДИ) [4]. Как исходные данные в нем разработчик должен подготовить замкнутую ДИ без входов и выходов, описывающую процесс переключений элементов. Метод заключается в проверке корректности ДИ и пе-

* Исследование выполнено при частичной финансовой поддержке по Программам фундаментальных исследований 2017 г. Президиума РАН (проект 0063-2016-0015) и подпрограмме № 4 ОНИТ РАН на 2016 г. (проект 0063-2016-0018).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, lplekhanov@inbox.ru

реводе ее описания в логические функции элементов. Построение корректной ДИ для реальных схем — практически невыполнимая задача, так как надо описать все возможные состояния схемы в замкнутом виде с учетом того, что процессы, связанные с операцией «ИЛИ», особенно громоздки. Корректность ДИ в методе эквивалентна свойству полумодулярности (теоретическая база самосинхронности), а результирующая схема получается также замкнутой. Метод ДИ логично назвать не синтезом, а конверсией, так как основное свойство схемы — полумодулярность (как и самосинхронность) не появляется в результате синтеза, а должно быть заранее обеспечено.

В функциональном подходе [5, 6] предлагается построить СС-схему по нескольким исходным логическим функциям, описывающим требуемое функциональное преобразование. Описание имеет традиционный вид: выходы в зависимости от входов.

Секция представляет собой часть комбинационной схемы и включает множество функций / элементов, выполняющих *параллельные* функциональные преобразования.

В рамках такого подхода синтез комбинационных СС-схем и, в частности, СС-секций описан в [6]. Для СС-секций там предложен табличный метод. Табличные методы имеют известный недостаток быстрого возрастания размера таблиц при увеличении числа аргументов (входов элементов). Кроме того, табличные методы оперируют конкретными значениями — нулями и единицами, что не позволяет выявить какие-либо общие закономерности.

В статье предлагается новый метод синтеза СС-секций на основе функционального подхода — описания задачи в логических функциях и уравнениях и решения логических уравнений. Метод позволяет подойти к синтезу с более общих позиций и сделать процедуру решения более формальной и в итоге более простой.

2 Исходные определения

Важным аспектом для синтеза является способ построения элементов и технология изготовления схем. В данной статье будут рассматриваться схемы, изготавливаемые по наиболее распространенной КМДП (комплементарная структура металл–диэлектрик–полупроводник) технологии.

В соответствии с теорией [7] элементом СС-схем КМДП-технологии будем считать однокаскадный элемент с одним выходом. В таком элементе цепи перезаряда нагрузки на выходе управляются транзисторами, затворы которых непосредственно соединены со входами элемента.

Самосинхронной секцией будем называть комбинационную СС-схему, реализованную на однокаскадных элементах КМДП-технологии с одним выходом.

Из-за особенностей построения СС-схем: парафазного кодирования информации, индикации и некоторых других — целесообразно синтезировать парал-

лельные функциональные преобразования не по одной функции, а по нескольку одновременно, что позволяет получать более оптимальные решения.

На практике для СС-схем используется самосинхронное кодирование, называемое *парафазным со спейсером* (ПФС) [1]. В этом кодировании каждый исходный информационный бит представляется двумя битами: бит 0 — битами 01, бит 1 — битами 10, спейсерное (промежуточное) состояние кодируется одинаковыми битами 00 или 11, оставшееся состояние, противоположное спейсерному (11 или 00), запрещено. Соответственно каждая исходная переменная (сигнал в схеме) представляется двумя переменными, которые будем называть *ПФС-переменными*.

Функциональное преобразование для реализации в СС-секции задается множеством исходных (некодированных) логических функций, которые будем называть монофункциями, или *M-функциями*. Множество M-функций назовем *M-секцией*.

Функции, полученные после ПФС-кодирования, будем называть *ПФС-функциями*; ПФС-функций будет вдвое больше, чем M-функций.

В СС-схемах для обеспечения отказобезопасности необходима индикация сигналов на выходах схемы [1]. Применительно к СС-секциям это значит, что на ее выходах можно индицировать часть ее входов, а оставшаяся часть будет индицирована в других местах содержащей эту секцию СС-схемы. Подразумевается также, что выходы СС-секции будут индицироваться в содержащей схеме. В задаче синтеза СС-секции поэтому задаются некоторые входы для индикации (от ни одного до всех).

В функциональном подходе [6] индикация означает, что переход секции в очередную (спейсерную или рабочую) фазу своей работы должен остановиться при возникновении константной неисправности типа залипания на 0 или 1 любого заданного для индикации входа секции. Для выполнения этого условия в СС-секции кроме элементов, реализующих ПФС-функции, могут появиться дополнительные элементы — индикаторы.

Обозначим: M — число M-функций; N — число входов M-секции (число различных аргументов M-функций); K — количество индикаторов; x_i — входная переменная M-секции, $i = 1, \dots, N$, она же первая ПФС-переменная после ПФС-кодирования; y_i — вторая ПФС-переменная после кодирования, сопряженная с X_i ; $F_m(x_1, \dots, x_N)$ — M-функция, $m = 1, \dots, M$; $U_m(x_i, y_i)$ — первая функция выхода СС-секции, соответствующая F_m ; $V_m(x_i, y_i)$ — вторая функция выхода, сопряженная с U_m ; $I_k(x_i, y_i)$ — функция индикатора, $k = 1, \dots, K$; s_1 и s_2 — значения входного и выходного спейсеров всей СС-секции соответственно.

Одно из основных требований соблюдения самосинхронности состоит в том, что все функции, выполняемые в СС-секции, должны быть монотонными по Варшавскому [1] по каждому аргументу. *Монотонная* функция по Варшавскому — это либо изотонная, либо антитонная функция. Функция F будет *изотонной* по аргументу z , если $F_{z=1} \geq F_{z=0}$, и *антитонной*, если $F_{z=1} \leq F_{z=0}$.

Однокаскадный элемент КМДП-технологии всегда выполняет антитонную функцию по всем аргументам, такой элемент будем называть *антитонным*.

Антитонный элемент, к выходу которого подключен инвертор, будет выполнять изотонную функцию по всем аргументам, и его будем называть *изотонным*.

И тот и другой тип элемента допустим для реализации СС-схем. Если значения входного и выходного спейсеров различны, СС-секция будет реализована на антитонных элементах, если одинаковы — то на изотонных.

Критерием оптимальности при синтезе может служить либо быстродействие, либо затраты в транзисторах.

Быстродействие антитонного элемента зависит от его индекса задержки (ИЗ): чем он больше, тем меньше быстродействие.

Индекс задержки антитонного элемента есть максимальное число последовательно соединенных транзисторов в цепи перезаряда его нагрузки. Для изотонного элемента этот индекс принимается равным ИЗ антитонной части плюс 1.

Быстродействие всей СС-секции будем оценивать как максимальный ИЗ ее элементов.

Затраты в транзисторах складываются из суммы транзисторов реализации СС-секции, включая индикаторы, и затрат на дальнейшую индикацию при наличии в ней индикаторов. Вторая составляющая зависит от последующих схем индикации и в реальных схемах «стбит» в среднем не менее 7 транзисторов на один индикатор [6].

Далее приняты такие обозначения логических операций: « \wedge » — отрицание; « \vee » — операция ИЛИ; « \oplus » — сумма по модулю 2; отсутствие символа — операция И.

Следуя [1], будем отождествлять имя элемента с именем его выходного сигнала.

3 Постановка задачи

Процедура синтеза будет аналогична при различных сочетаниях входного и выходного спейсеров s_1 и s_2 . Разница будет состоять в изотонности или антитонности получаемых функций элементов. Поэтому для определенности и простоты примем значения спейсеров $s_1 = 0$ и $s_2 = 1$. В этом случае функции элементов должны быть антитонными.

Таким образом, исходными данными для синтеза будут:

- М-секция;
- значения спейсеров $s_1 = 0$ и $s_2 = 1$;
- указание индикации входов СС-секции: какие должны индицироваться на ее выходах;
- критерий оптимизации: быстродействие или затраты в транзисторах.

В результате синтеза должна быть получена СС-секция, удовлетворяющая следующим условиям:

- СС-секция должна содержать $2M$ элементов, реализующих ПФС-функции, и, возможно, несколько индикаторов;
- все выходные функции секции должны быть антитонными;
- должен быть оптимизирован заданный критерий.

В классическом событийном подходе задача синтеза в подобной постановке не рассматривалась.

4 Метод синтеза

Для наглядности процедура решения будет сопровождаться простым примером со следующими исходными данными.

Пример:

- М-секция из двух функций ($M = 2, N = 3$):

$$F_1 = x_1x_2; \quad F_2 = x_2x_3;$$

- спейсеры: $s_1 = 0, s_2 = 1$;
- на выходах секции должны индицироваться входы x_1 и y_1 .

Теоретическим обоснованием метода является утверждение 4.4 в [1, с. 92] о том, что двухфазная комбинационная схема является СС тогда и только тогда, когда она индицируема.

Таким образом, в процедуре синтеза необходимо и достаточно выполнить ПФС-кодирование и обеспечить индикацию заданных входов.

Общий порядок решения задачи таков.

На первом этапе из М-функций необходимо получить ПФС-функции. Данное действие будем называть *ПФС-преобразованием*.

Далее будут составляться и решаться логические уравнения для получения общих функций выходов секции.

На последнем этапе для каждого выхода секции подбираются подходящие элементы из принятой библиотеки с учетом заданного критерия оптимизации.

4.1 Парофазное со спейсером преобразование

В соответствии с ПФС-кодированием [1] каждой М-функции F_m соответствуют две ПФС-функции P_m и G_m . Для спейсеров $s_1 = 0$ и $s_2 = 1$ эти функции должны быть антитонными.

Назовем *ПФС-заменой* замену в выражениях ДНФ (дизъюнктивных нормальных форм) функций вхождений переменных x_i на $\wedge y_i$ и замену вхождений y_i на $\wedge x_i$ с целью получения антитонной функции. Эта замена правильна, так как переменные x_i и y_i в рабочем состоянии взаимно инверсны, а в спейсере обе равны нулю.

Функция P_m получается путем ПФС-замены из F_m , функция G_m получается из $\wedge F_m$ путем такой же замены.

В примере это будут:

$$P_1 = \wedge y_1 \wedge y_2; \quad G_1 = \wedge x_1 \vee \wedge x_2; \quad P_2 = \wedge y_2 \wedge y_3; \quad G_2 = \wedge x_2 \vee \wedge x_3.$$

Парафазная со спейсером замена позволяет также заменять целиком функцию $\wedge G_m$ на P_m и функцию $\wedge P_m$ на G_m .

4.2 Общие уравнения задачи

Уравнения задачи в общем случае зависят от количества индикаторов в секции. Индикация может быть обеспечена частично за счет информационных выходов секции и частично за счет индикаторов. Поэтому число индикаторов заранее не известно — оно будет определяться в процессе синтеза. При увеличении числа индикаторов упрощаются функции элементов, но увеличиваются затраты секции в транзисторах и может снизиться быстродействие.

Начинать следует с отсутствия индикаторов. Если не удается решить задачу с текущим количеством индикаторов, то надо увеличить их число и повторить процедуру.

Предельное количество индикаторов $K = N$. В этом случае все входы секции индицируются попарно на индикаторах — двухходовых логических элементах типа 2И-НЕ или 2ИЛИ-НЕ независимо от информационных выходов. Дальнейшее увеличение K не имеет смысла.

Для примера примем один индикатор: $K = 1$.

Задача будет решаться с помощью систем логических уравнений [8].

Уравнения должны отражать специфику СС-схем и требования по индикации.

В дальнейшем считается, что каждое логическое уравнение истинно, т. е. имеет логическое значение 1.

Назовем *набором* совокупность булевых значений входов секции.

Введем специальные функции и уравнения для решения задачи.

1. Входы СС-секции.

Функция входного спейсера:

$$S = \wedge x_1 \wedge x_2 \cdots \wedge x_N \wedge y_1 \wedge y_2 \cdots \wedge y_N. \quad (1)$$

Входная рабочая функция:

$$W = (x_1 \oplus y_1) \cdots (x_N \oplus y_N). \quad (2)$$

Сопряженные входы секции не могут одновременно принимать значение, инверсное входному спейсеру, что дает уравнение

$$x_1 y_1 \vee x_2 y_2 \vee \cdots \vee x_N y_N = 0. \quad (3)$$

2. Выходы СС-секции.

Функция выходного спейсера:

$$H = U_1 V_1 U_2 V_2 \cdots U_M V_M I_1 \cdots I_K . \quad (4)$$

Выходная рабочая функция:

$$\begin{aligned} R = (U_1 \oplus V_1) \cdots (U_M \oplus V_M) & (U_1 = P_1) (V_1 = G_1) \cdots \\ & \cdots (U_M = P_M) (V_M = G_M) {}^{\wedge} I_1 \cdots {}^{\wedge} I_K . \end{aligned} \quad (5)$$

Последняя функция получена из условий, что в рабочей фазе сопряженные выходы секции должны быть взаимно обратны, а также быть равны ПФС-функциям. Индикаторы должны принимать обратные выходному спейсеру значения.

Сопряженные выходы U_m и V_m СС-секции не должны одновременно принимать значение, инверсное выходному спейсеру:

$${}^{\wedge} U_1 {}^{\wedge} V_1 \vee {}^{\wedge} U_2 {}^{\wedge} V_2 \vee \cdots \vee {}^{\wedge} U_M {}^{\wedge} V_M = 0 . \quad (6)$$

3. Функционирование СС-секции.

Работа секции описывается двумя уравнениями: при появлении спейсерного набора на входе должен появиться спейсерный набор на выходе, а при появлении рабочего набора на входе — рабочий набор на выходе:

$$S \rightarrow H ; \quad W \rightarrow R , \quad (7)$$

где стрелка означает логическое следование (импликацию).

4. Уравнение индикации.

Для индикации одного входа необходимо обеспечить остановку работы СС-секции при его залипании отдельно на 0 и 1. Это требует выполнения двух условий: если вход залипает на значение спейсера s_1 (в данном случае это 0), то выходы секции не должны попасть в рабочую фазу; если залипает на ${}^{\wedge} s_1$ (здесь это 1), то выходы секции не должны попасть в спейсер. Условия выражаются уравнением:

$$(S_Z \rightarrow {}^{\wedge} R) \ (W_Z \rightarrow {}^{\wedge} H) , \quad (8)$$

где S_Z и W_Z — функции залипания соответственно в спейсере и в рабочей фазе.

Функции залипания выражают состояние, когда неиндицируемые входы уже находятся в соответствующей фазе, а индицируемые — нет.

Для рассматриваемого примера это будут функции:

$$S_Z = (x_1 \vee y_1) {}^{\wedge} x_2 {}^{\wedge} x_3 {}^{\wedge} y_2 {}^{\wedge} y_3 ; \quad W_Z = {}^{\wedge} x_1 {}^{\wedge} y_1 (x_2 \oplus y_2) (x_3 \oplus y_3) . \quad (9)$$

4.3 Решение уравнений

Для решения логических уравнений необходимо все переменные, которые в них присутствуют, разделить на три категории:

- (1) неизвестные — переменные (функции), которые требуется определить;
- (2) аргументы — переменные, от которых будут зависеть неизвестные;
- (3) исключаемые — переменные, которые надо исключить из решения.

Решения логических уравнений имеют ряд особенностей [8].

1. Вычисляется условие совместности, т. е. условие существования решения. Это условие накладывается на аргументы. Оно может быть трех типов:
 - (а) тождество (типа $0 = 0$), которое означает, что решение существует при любых аргументах;
 - (б) противоречие (типа $1 = 0$), которое означает, что решения нет ни при каких аргументах;
 - (в) логическое условие, означающее, что решение существует только при выполнении данного условия.
2. Решения (функции неизвестных) могут зависеть от произвольных переменных. Произвольным переменным можно придавать любые значения: константы 0 или 1 или любые функции от аргументов. Наличие произвольных переменных показывает неоднозначность (множественность) решения. Произвольные переменные возникают при вычислении конкретной неизвестной. Поэтому для определенности они будут обозначаться в виде $\$Q$, где Q — имя неизвестной, которая «породила» произвольную переменную. Однако никакой связи по значениям между неизвестными и произвольными переменными нет.
3. Вид логических функций для неизвестных зависит от очередности их определения. В частности, каждая следующая по порядку решения неизвестная может зависеть от произвольных переменных всех предыдущих неизвестных.

Рассмотрим систему уравнений общего вида (4)–(8), в которых отсутствуют входные переменные. Конечные решения можно получать из общих решений по мере необходимости подстановкой в них (1), (2) и (9) с учетом (3).

Неизвестными в системе будут функции выходов $U_1, V_1, \dots, U_M, V_M$, а аргументами — $S, W, S_Z, W_Z, P_1, G_1, \dots, P_M, G_M$, исключаемых нет.

Вычисление условия совместности этих уравнений дает

$$SW \vee SS_Z \vee WW_Z \vee W((P_1 = G_1) \vee \dots \vee (P_M = G_M)) = 0. \quad (10)$$

Это условие должно быть проверено упомянутыми выше подстановками. По факту такая проверка означает проверку правильности составления функций S, W, S_Z, W_Z и всех ПФС-функций.

Произведение SW при подстановке должно обращаться в 0. Это означает, что спейсерные и рабочие наборы не должны пересекаться, что является необходимым условием для СС-схем. Два следующих произведения (10) требуют, чтобы функции фаз и их залипаний не пересекались, что необходимо для индикации. Наконец, оставшаяся левая часть (10) требует, чтобы в рабочей фазе ПФС-функции P_m и G_m были взаимно инверсны. Это условие должно выполняться по построению ПФС-функций.

Из-за специфики решения логических уравнений искомые функции удобно представлять последовательно, когда каждая следующая функция зависит от предыдущих.

Например, решение для исходных данных примера после упрощений по (10) и ПФС-замен примет вид:

$$\left. \begin{aligned} U_1 &= S \vee WP_1 \vee {}^W\!U_1; \\ V_1 &= S \vee {}^W\!U_1 \vee WG_1 \vee {}^W\!V_1; \\ U_2 &= S \vee W(P_2 \vee U_1V_1 \vee {}^W\!U_1P_1 \vee {}^W\!V_1G_1) \vee {}^W\!U_2; \\ V_2 &= S \vee {}^W\!U_2 \vee W(G_2 \vee U_1V_1 \vee {}^W\!U_1P_1 \vee {}^W\!V_1G_1) \vee {}^W\!V_2; \\ I_1 &= I_1(S, W, S_Z, W_Z, P_1, G_1, P_2, G_2, U_1, V_1, U_2, V_2, \$I_1). \end{aligned} \right\} \quad (11)$$

(Выражение для индикатора не приводится ввиду громоздкости.)

После подстановки всех зависимых переменных образуются немонотонные функции аргументов. Монотонные решения (в данном случае антитонкие) легко получаются из них путем ПФС-замен.

Далее задача состоит в подборе элементов, реализующих монотонные решения.

4.4 Получение функций элементов

Элементы для реализации секции можно подбирать из некоторого множества полузаизной или заизной библиотеки элементов.

После подстановок (1), (2), (9), функций P_m и G_m , а также подстановок предыдущих функций каждая из найденных функций представляется в форме неполностью определенной функции (НОФ):

$$D = A \vee B\$, \quad (12)$$

где $\$$ — произвольная переменная.

Назовем *единицами функции* наборы, на которых функция получает значение 1, и *нулями функции* — наборы, на которых она равна 0.

Единицами функции (12) будут корни уравнения $A = 1$, а нулями — корни уравнения ${}^W\!A \vee {}^W\!B = 1$.

Для реализации НОФ типа (12) элементом, имеющим функцию E , необходимо и достаточно выполнения двух условий:

- (1) единицы функции D должны содержаться в единицах функции E ;
- (2) нули функции D должны содержаться в нулях функции E .

Эти условия задаются уравнениями:

$$EA = A; \quad {}^{\wedge}E {}^{\wedge}A {}^{\wedge}B = {}^{\wedge}A {}^{\wedge}B.$$

После преобразований оба уравнения можно свести в одно:

$$A {}^{\wedge}E \vee {}^{\wedge}A {}^{\wedge}BE = 0. \quad (13)$$

Это уравнение является основным для подбора библиотечных элементов.

Функции из решения (11) и функции элементов E имеют разные аргументы, поэтому перед проверкой по условию (13) их необходимо привести к единым аргументам: библиотечным или ПФС-переменным — путем перекодировки.

Поскольку искомые функции зависят друг от друга, подбор элементов для всей СС-секции — задача переборная. Можно предложить следующий пошаговый алгоритм.

1. Для первой/очередной найденной функции с помощью условия (13) подобрать и запомнить ряд подходящих элементов в порядке неубывания критерия.
2. Взять первый/очередной элемент из предыдущего ряда и подставить во все последующие функции секции.
3. Выбрать ряд элементов для следующей функции.
4. Подобным образом найти множества элементов для всей секции.
5. Подсчитать общий критерий по первым элементам в рядах.
6. Далее можно вернуться к любой предыдущей функции, выбрать другой элемент и повторить вычисления от нее до конца секции.

Во многих случаях решения получаются более просто.

Обратимся к примеру.

Если в (11) выбрать произвольную переменную $\$U_1 = P_1$, то

$$U_1 = P_1 = {}^{\wedge}y_1 {}^{\wedge}y_2,$$

так как функция S , состоящая из одной импликанты со всеми аргументами, всегда поглощается.

Подставляя функцию U_1 в выражение для V_1 , делая ПФС-замены и полагая $\$V_1 = G_1$, получим:

$$V_1 = G_1 = {}^{\wedge}x_1 \vee {}^{\wedge}x_2.$$

Аналогично можно найти простые решения:

$$U_2 = P_2 = {}^{\wedge}y_2 {}^{\wedge}y_3; \quad V_2 = G_2 = {}^{\wedge}x_2 \vee {}^{\wedge}x_3.$$

Подставим найденные функции в решение для индикатора:

$$I_1 = A \vee B\$I_1.$$

В этом выражении обозначено:

$$A = {}^{\wedge}x_1 {}^{\wedge}y_1 {}^{\wedge}x_2 ({}^{\wedge}x_3 \vee {}^{\wedge}y_3);$$

$$B = {}^{\wedge}x_1 {}^{\wedge}y_1 \vee {}^{\wedge}x_2 {}^{\wedge}x_3 {}^{\wedge}y_3 \vee {}^{\wedge}x_2 {}^{\wedge}y_2 x_3 \vee {}^{\wedge}x_2 {}^{\wedge}y_2 y_3 \vee {}^{\wedge}y_2 {}^{\wedge}x_3 {}^{\wedge}y_3 \vee {}^{\wedge}x_1 {}^{\wedge}x_3 {}^{\wedge}y_3.$$

Уравнение для индикаторного элемента получится:

$$\begin{aligned} ({}^{\wedge}x_1 {}^{\wedge}y_1 {}^{\wedge}x_2 ({}^{\wedge}x_3 \vee {}^{\wedge}y_3)) {}^{\wedge}E \vee (x_1 x_2 x_3 \vee x_1 x_2 y_3 \vee x_1 y_2 y_3 \vee x_1 y_2 x_3 \vee \\ \vee y_1 y_2 x_3 \vee y_1 y_2 y_3 \vee y_1 x_2 y_3 \vee y_1 x_2 x_3 \vee x_1 {}^{\wedge}x_2 {}^{\wedge}y_2 {}^{\wedge}x_3 {}^{\wedge}y_3) E = 0. \end{aligned}$$

Этому уравнению удовлетворяет, например, функция

$$I_1 = E = {}^{\wedge}x_1 {}^{\wedge}y_1.$$

Таким образом, определены все функции (элементы) примера.

5 Заключение

Предложен метод синтеза комбинационных СС-секций — схем, осуществляющих параллельные функциональные преобразования. В отличие от существующих методов: событийного по диаграммам изменений и ранее предложенного автором табличного метода — данный метод основан на функциональном подходе. В методе для описания и решения задачи используются только логические функции и логические уравнения.

Метод исходит из самых общих уравнений, отражающих свойства СС-схем. Процедура решения позволяет более формально и более просто проводить синтез.

Изложенный подход к синтезу СС-схем и метод решения предлагаются впервые.

Литература

1. Автоматное управление асинхронными процессами в ЭВМ и дискретных системах / Под ред. В. И. Варшавского. — М.: Наука, 1986. 400 с.
2. Плеханов Л. П., Степченков Ю. А. Экспериментальная проверка некоторых свойств строго самосинхронных схем // Системы и средства информатики, 2006. Вып. 16. С. 476–485.

3. Соколов И. А., Степченков Ю. А., Бобков С. Г. и др. Базис реализации супер-ЭВМ эксафлопного класса // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 45–70.
4. Варшавский В. И., Кишиневский М. А., Кондратьев А. Ю., Розенблюм Л. Я., Таубин А. Р. Модели для спецификации и анализа процессов в асинхронных схемах // Техническая кибернетика, 1988. № 2. С. 171–190.
5. Плеханов Л. П. Проектирование самосинхронных схем: функциональный подход // Проблемы разработки перспективных микро- и наноэлектронных систем: IV Всеросс. научн.-технич. конф. (МЭС-2010): Сб. научных тр. — М.: ИППМ РАН, 2010. С. 424–429.
6. Плеханов Л. П. Основы самосинхронных электронных схем. — М.: Бином. Лаборатория знаний, 2013. 208 с.
7. Muller D. E., Bartky W. C. A theory of asynchronous circuits // Symposium (International) on the Theory of Switching Proceedings. — Cambridge, MA, USA: Harvard University Press, 1959. Part 1. P. 204–243.
8. Плеханов Л. П. Логические уравнения в разработке цифровых микроэлектронных устройств // Электронная техника. Сер. 10. Микроэлектронные устройства, 1989. Вып. 5(77). С. 25–28.

Поступила в редакцию 27.01.17

SYNTHESIS OF SELF-TIMED COMBINATIONAL SECTIONS USING THE FUNCTIONAL METHOD

L. P. Plekhanov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Designing of self-timed electronic circuits having unique properties requires special methods. Combinational circuits can consist of several sections representing a variety of elements performing parallel functional transformations. The known event method of self-timed circuit synthesis has theoretical significance only and is unsuitable for real design. The author proposes a new method, which describes functional sections not by events (switches of elements) but by logical functions and equations. The method consists in generating and solving logical equations. The method solves the problem in the most general form and allows choosing elements for circuit implementation from semicustom or custom libraries.

Keywords: self-timed circuit; asynchronous circuit; self-timed circuit synthesis

DOI: 10.14357/08696527170208

Acknowledgments

The research was performed under partial financial support of Program of fundamental research 2017 of the Presidium of RAS (project 0063-2016-0015) and subprogram No. 4 of the RAS Department for Nanotechnologies and Information Technologies (ONIT) for 2016 (project 0063-2016-0018).

References

1. Varshavsky, V. I., ed. 1986. *Avtomatnoe upravlenie asinkhronnymi protsessami v EVM i diskretnykh sistemakh* [Automata control of asynchronous processes in computers and discrete systems]. Moscow: Nauka. 398 p.
2. Plekhanov, L. P., and Yu. A. Stepchenkov. 2006. *Eksperimental'naya proverka nekotorykh svoystv strogo samosinkhronnykh skhem* [Experimental verification of some properties of strictly self-timed circuits]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:476–485.
3. Sokolov, I. A., Y. A. Stepchenkov, S. G. Bobkov, et al. 2014. Bazis realizatsii super-EVM eksaflopnogo klassa [Implementation basis of exaflops class supercomputer]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):45–70.
4. Varshavskiy, V. I., M. A. Kishinevskiy, A. Yu. Kondrat'ev, L. Ya. Rozenblyum, and A. R. Taubin. 1988. Modeli dlya spetsifikatsii i analiza protsessov v asinkhronnykh skhemakh [Models for specification and analysis of processes in asynchronous circuits]. *Tekhnicheskaya kibernetika* [Technical Cybernetics] 2:171–190.
5. Plekhanov, L. P. 2010. Proektirovanie samosinkhronnykh skhem: funktsional'nyy podkhod [Designing of self-timed circuits: A functional approach]. *Vseross. nauchn.-tehnich. konf. "Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem (MES-2010)"* [Russian Scientific-Technical Conference "Problems of Development of Perspective Micro and Nanoelectronic Systems (MES-2010)". Moscow. 1:424–429.
6. Plekhanov, L. P. 2013. *Osnovy samosinkhronnykh elektronnykh skhem* [The basis of self-timed electronic circuits]. Moscow: Binom. Laboratoriya znaniy. 208 p.
7. Muller, D. E., and W. C. Bartky. 1959. A theory of asynchronous circuits. *Symposium (International) on the Theory of Switching Proceedings*. Cambridge, MA: Harvard University Press. 1:204–243.
8. Plekhanov, L. P. 1989. Logicheskie uravneniya v razrabotke tsifrovyykh mikroelektronnykh ustroystv [Logical equations in designing of digital microelectronic devices]. *Electronic Technics. Ser. 10. Microelectronic devices* [Electronic Equipment. Ser. 10. Microelectronic Devices] 5(77):25–28.

Received January 27, 2017

Contributor

Plekhanov Leonid P. (b. 1943)—Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; lplekhanov@inbox.ru

НАУЧНО-МЕТОДИЧЕСКИЕ ПОДХОДЫ К СОВЕРШЕНСТВОВАНИЮ НОРМАТИВНОЙ БАЗЫ ДЛЯ СОЗДАНИЯ И РАЗВИТИЯ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ

С. А. Головин¹, А. А. Зацаринный², С. В. Козлов³

Аннотация: Рассматриваются научно-технические проблемы развития нормативной базы для создания информационно-телекоммуникационных систем (ИТС). Отмечены факторы, определяющие актуальность их решения. Показано, что существующая нормативная база отстает от потребностей разработчиков и пользователей ИТС. Предложен комплексный подход к развитию нормативной базы с учетом условий применения перспективных ИТС. Проведено обоснование направлений совершенствования нормативной базы для их создания и развития на процессной основе взаимодействия ИТС и других элементов в рамках системы управления как высокотехнологичной организационной системы.

Ключевые слова: информационно-телекоммуникационные системы; информационные технологии; нормативная база; интероперабельность; профили стандартов; системный подход; процессный подход

DOI: 10.14357/08696527170209

1 Введение

В условиях нового технологического уклада, содержание которого определяется приоритетным развитием информационных, когнитивных и других технологий с одновременным расширением интеграционных процессов [1], особую актуальность приобретает своевременное и качественное решение проблемы нормативного регулирования при создании и развитии ИТС. Вместе с тем существующая отечественная нормативная база, регламентирующая применение информационных технологий (ИТ) при создании высокотехнологичных организационных систем различного назначения, отстала от мирового уровня развития и значимости ИТ и не соответствует основным положениям принятых в 2015–2016 гг. в нашей стране Стратегии национальной безопасности [2] и Стратегии

¹Московский технологический университет, sgolovin@itstandard.ru

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, azatsarinny@ipiran.ru

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sv_kozlov@mail.ru

научно-технологического развития [3]. Текущее состояние отечественной нормативной базы в области ИТ сдерживает создание современных ИТС различного назначения.

В статье предлагаются научно-методические и организационные подходы к совершенствованию и развитию нормативной базы в области ИТС.

2 Анализ существующего состояния нормативного регулирования в области информационно-телекоммуникационных систем

Исходя из основных принципов системного подхода, ИТС как объект разработки в интересах государственных нужд необходимо рассматривать в качестве технико-технологической основы перспективных систем управления, представляющих в общем виде совокупность органов, пунктов и средств управления [4]. При этом ИТС является одним из элементов средств управления и в значительной мере определяет общий облик, структуру системы управления и топологию ее территориально распределенных элементов.

С учетом ограничений системного подхода [5] нормативное регулирование вопросов создания и развития ИТС определяется следующими особенностями:

- многозначность их целевого предназначения и условий применения, связанная с интенсивными процессами интеграции ИТС в составе различного рода многофункциональных систем;
- противоречивость стоящих перед ИТС задач, обусловленных как на стадии создания, так и в процессе эксплуатации конфликтностью основных и противодействующих процессов внешней среды;
- ресурсные ограничения на разных стадиях жизненного цикла ИТС.

Поскольку перечисленные особенности, по существу, определяют граничные условия применения системного подхода, от полноты и качества их анализа, оценки и учета во многом зависит инвариантность ИТС по отношению к новым угрозам и формирующимся опасностям в инфокоммуникационной сфере на протяжении жизненного цикла системы. В этой связи при обосновании направлений совершенствования нормативной базы создания и развития ИТС представляется целесообразным учитывать в комплексе основные положения системного и процессного подходов. В настоящее время вопросы совершенствования нормативного регулирования в различных областях науки, техники и технологий приобретают особую актуальность и практическую важность [6, 7].

Анализ и обобщение вопросов нормативного регулирования в области информационно-телекоммуникационных технологий на всех стадиях жизненного цикла ИТС показывает, что в настоящее время в нормативном плане можно выделить следующие направления разработки и поддержания в актуальном состоянии нормативных правовых и методических документов, определяющих:

- разработку, создание, испытания, ввод в эксплуатацию и обеспечение эксплуатации ИТС;
- доступ пользователей к информационным системам, в том числе посредством телекоммуникационных сетей, и порядок использования информационных ресурсов;
- обеспечение защиты информации и безопасности связи в ИТС.

Важно отметить, что вопросы применения ИТС в настоящее время регламентированы на уровне девяти федеральных законов. Обширная нормативная база создана в отношении защиты информации и безопасности связи. В то же время для стадии создания ИТС по решению заказчика в зависимости от особенностей объекта разработки и необходимости его тиражирования могут применяться две группы отечественных стандартов: серии ГОСТ Р(РВ) 15.XXX и ГОСТ 34.XXX. В интересах создания объектов на основе применения ИТ в течение длительного времени заказчики и разработчики руководствуются ГОСТом 34.XXX. При всей привлекательности этих стандартов имеются в недостаточной мере регламентированные как стадии жизненного цикла, так и сегменты деятельности субъектов государственного заказа. Как отмечено в [8], ГОСТ 34.XXX:

- регламентирует только деятельность по созданию ИТС, при этом остаются открытыми вопросы ее сопровождения, эксплуатации и вывода из эксплуатации;
- относится только к заказу разработки ИТС и не обладает процессно-ориентированной направленностью, вследствие чего организация процесса выбора, приобретения и внедрения систем и компонентов для ИТС остается за рамками этого ГОСТА;
- не соотносится с международными стандартами в области управления качеством, управления проектами, управления информационной безопасностью и др.

В последние 15–20 лет в условиях принципиально новой глобальной системы на основе углубляющейся взаимосвязи и взаимозависимости во всех сферах международного сообщества с интенсивным освоением современных информационных и телекоммуникационных технологий получили активное развитие работы в области стандартизации ИТ [9]. Это подтверждает статистика развития международной ИТ-стандартизации. Только совместный технический комитет (JTC1) Международной организации по стандартизации (ISO) и Международной электротехнической комиссии (IEC) разработал на настоящий момент 3002 ИТ-стандарта и 583 ИТ-стандарта находятся в стадии разработки. В их создании активное участие принимают 95 стран.

Если еще учесть ИТ-стандарты, разрабатываемые такими организациями, как Международный союз по электросвязи (ITU), Институт инженеров электротехники и электроники (IEEE) и др., то их число превысит 4000. Такое

количество ИТ-стандартов в значительной мере превышает число стандартов, действующих и разрабатываемых в других отраслях. Очевидно, что в таком количестве ИТ-стандартов есть объективная необходимость.

JTC1 ISO/IEC создан в 1987 г. За это время было выделено более 40 приоритетных направлений. Некоторые из них сохранили свою значимость и к настоящему моменту, а некоторые потеряли свою актуальность. Анализ тенденций изменения таких приоритетов позволяет выявить основные тенденции развития ИТ в преддверии шестого технологического уклада.

В табл. 1 представлены действующие в настоящее время подкомитеты JTC1 и показано соотношение приоритетов по различным направлениям (подкомитетам). Представленное в табл. 1 распределение ИТ-стандартов по подкомитетам JTC1 характеризует актуальность конкретных направлений ИТ-стандартизации, интегрально сформировавшееся за последние 10–15 лет. Естественно, что больший интерес представляет динамика изменения приоритетов за более короткий промежуток времени.

Более детальный анализ динамики изменения числа стандартов по направлениям ИТ показывает, что снизились темпы роста и сократилось количество ИТ-стандартов по следующим направлениям:

- (а) компьютерная графика, обработка изображений и подготовка данных;
- (б) кодированные наборы символов;
- (в) цифровые носители для обмена информацией и ее хранения;
- (г) оборудование офисов.

Повышение динамики развития некоторых направлений прежде всего связано с ярко выраженной тенденцией развития единого информационного пространства, предшественницей цифровой трансформации экономики развитых стран, которая является одной из важнейших составляющих перехода страны к новому технологическому укладу. Получает развитие стандартизация по направлению Интернета вещей. Информационные технологии в последние 5–10 лет развиваются так стремительно, что почти каждые 5–10 лет возникает какая-либо новая принципиальная технология. Только что затихли дискуссии о перспективности облачных технологий, как все большее внимание привлекает концепция Интернета вещей (Internet of Things, IoT) как устойчивая тенденция в ИТ. Интернет вещей — это прежде всего взаимодействие различных технических устройств, в основном без участия человека. И количество таких устройств растет в геометрической прогрессии. По оценкам ISO, через 2–3 года будут объединены около 50 млрд вещей. При этом многие из них, объединенные в сети, имеют определенные сенсорные устройства, что предопределяет новые реалии и угрозы, связанные с выходом «интеллектуальных вещей» из-под контроля человека.

Сейчас стандартизация в области Интернета вещей находится на начальной стадии. В основном этой проблематикой занимаются две международные организации: совместный технический комитет ISO и IEC (JTC1) и исследовательская комиссия ITU-T № 20.

Таблица 1 Распределение количества опубликованных стандартов по подкомитетам JTC1

Направления стандартизации ИТ	Распределение количества стандартов					
	до 50	51–100	101–150	151–200	350–400	500–600
Кодированное представление видео / аудио						*
Телекоммуникации					*	
Взаимосвязь оборудования для ИТ				*		
Системная и программная инженерия				*		
Безопасность ИТ				*		
Цифровые носители			*			
Автоматическая идентификация			*			
Биометрия			*			
Идентификационные карты			*			
Языки программирования			*			
Описание документа и языки обработки		*				
Компьютерная графика		*				
Менеджмент данных и обмен данными		*				
Пользовательские интерфейсы		*				
Кодированные наборы символов	*					
Оборудование офисов	*					
Информационные технологии в образовании	*					
Управление ИТ и услугами ИТ	*					
Распределенные приложения	*					
Информационные технологии для Интернета вещей	*					

В JTC1 на текущий момент все стандарты находятся в стадии разработки и носят основополагающий характер. Вот некоторые из них:

- ИСО / МЭК 30141 Интернет вещей. Эталонная архитектура;
- ИСО / МЭК 20924 Информационные технологии. Интернет вещей (IoT). Определение и словарный запас;
- ИСО / МЭК 18577 Информационные технологии. Интернет вещей (IoT) в цепи поставок. Транспортные единицы;
- ИСО / МЭК 21823-1 Интернет вещей (IoT). Совместимость для Интернета вещей и систем. Часть 1.

Всего в настоящий момент в процессе разработки в области Интернета вещей находятся 15 проектов стандартов. Серьезная программа работ в области Интернета вещей принята в ITU. Она включает:

- словарь для «умных» городов и общин;
- поддержку в сетях Интернета вещей приложений для глобальных процессов на земле;
- требования к описанию вещей в Интернете вещей;
- требования к транспортной безопасности, включая варианты использования и сценарии обслуживания;
- основы оценки эффективности системы электронного здравоохранения при использовании Интернета вещей.

Всего соответствующая программа ITU содержит 57 пунктов.

При всех достоинствах такой технологии должна быть предусмотрена защита, которая должна быть комплексной, и ИТ-стандартам в этой сфере также отводится важная роль. Прежде всего это должны быть стандарты, обеспечивающие суверенитет и безопасность нашей страны в этой области. Следует заметить, что вопросам безопасности в РФ и ЕАЭС уделяется значительное внимание. За последние годы разработано и принято определенное количество технических регламентов, среди которых такие, как «О безопасности железнодорожного подвижного состава», «О безопасности низковольтного оборудования», «О безопасности упаковки» и др. Однако нет ни действующего, ни запланированного к разработке технического регламента безопасности ИТ, несмотря на то что возможный суммарный ущерб от ошибок в применении ИТ может превзойти все ущербы, которые должны предотвратить все действующие технические регламенты.

В нашей стране нарастающим итогом отмечается отставание в стандартизации ИТ. Доля современных национальных ИТ-стандартов составляет не более 5% от числа международных. В настоящее время база иностранных стандартов в области ИТ насчитывает около 2700 утвержденных документов. В год в нашей стране принимается до 30–40 стандартов в области ИТ. Низкие темпы приводят

к нарастанию отставания от международного уровня. Такая ситуация привела к тому, что отечественные разработчики практически не имеют информации о конкретных международных ИТ-стандартах, а доступная база данных по национальным ИТ-стандартам в количественном и качественном виде существенно отстает от мирового уровня.

3 Нормативная база — важнейший компонент системы обеспечения процесса создания информационно-телекоммуникационных систем

Важной особенностью пятого и шестого технологических укладов применительно к классу ИТС является их гетерогенность, причем по мере развития ИТ и расширения их перечня проявляется тенденция к углублению интеграции разнородных информационных ресурсов, средств их обработки, передачи и распределения. В этой связи происходит насыщение создаваемых ИТС разнородными средствами вычислительной техники, которые должны взаимодействовать с использованием средств коммуникации, что привело к возникновению гетерогенной среды, в которой неизбежно возникает проблема совместимости входящих в нее компонентов, т. е. проблема интероперабельности. Так, например, в Программе фундаментальных исследований государственных Академий наук (Постановление Правительства РФ от 31 октября 2015 г. № 2217-р) в рамках направления фундаментальных исследований в части научных основ информационно-вычислительных систем в перечне ожидаемых результатов отмечены:

- разработка научно-методологических основ информатизации общества и инновационных видов деятельности, направленных на обеспечение социально-экономического развития и национальной безопасности Российской Федерации;
- развитие принципов интероперабельности и технологий открытых информационных систем.

Оценивая в целом современное состояние и темпы расширения сферы стандартизации по тематике ИТ, обострение проблемы интероперабельности в гетерогенной среде создания и развития ИТС, необходимо отметить, что применительно к комплексной стандартизации систем, создаваемых с использованием достижений ИТ в условиях нового технологического уклада, потребуется выполнить большой объем работ, связанных с взаимоувязанным выбором необходимых стандартов для регламентирования разработок ИТС в отечественной практике как сложных технических систем. Исходя из основных положений процессного подхода к постановке, организации выполнения и контроля разработок перспективных ИТС, представляется целесообразным выделить следующие основные направления и этапы работ по развитию нормативной базы:

- (1) анализ перечня и содержания современных международных стандартов, отбор, перевод и адаптация необходимых международных ИТ-стандартов

к отечественным условиям нормативного регулирования сферы создания и развития ИТС;

- (2) разработка профилей стандартов применительно к организации и проведению комплексных разработок ИТС как сложных технических систем;
- (3) мониторинг угроз и опасностей, связанных с развитием информационно-телекоммуникационных технологий при создании и эксплуатации ИТС. Обоснование комплекса необходимых мер по обеспечению защиты информации и безопасности связи в целях их нейтрализации с обеспечением нормативного регулирования процесса дальнейшего развития ИТ в комплексе с безопасностью их применения;
- (4) совершенствование нормативного регулирования процесса создания ИТС на основе перспективных ИТ с учетом обеспечения информационной безопасности.

Учитывая многогранный характер проблемы нормативного регулирования в области техники и технологий, при создании и развитии на их основе соответствующих систем целесообразно использовать комплексный подход к их развитию с учетом следующих предпосылок:

- создаваемая ИТС как сложная техническая система является элементом системы вышестоящего уровня;
- нормативное регулирование процесса создания и развития сложных технических систем на любой стадии их жизненного цикла не может быть осуществлено в полном объеме только на основе комплекса стандартов, при этом всегда остаются области, в рамках которых по различным объективным и субъективным причинам не представляется возможным урегулировать, в первую очередь, процессы организационного и организационно-технического характера. В меньшей степени эта проблема характерна для технико-технологических процессов, реализуемых при создании сложных технических систем, поскольку они регламентированы большим количеством стандартов.

4 Предложения и рекомендации по совершенствованию процесса нормативного регулирования работ по созданию информационно-телекоммуникационных систем

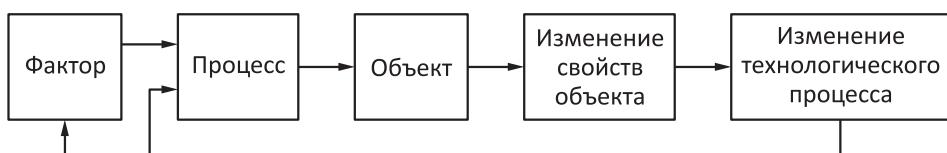
Исходя из изложенных предпосылок, ИТС рассматривается как элемент метасистемы, т. е. является одним из элементов системы управления. Как известно [4], система управления — это взаимоувязанная совокупность органов, пунктов и средств управления, где к средствам управления обычно относят автоматизированные системы управления и системы связи (телеинформационные системы), создаваемых в целях выполнения задач по мониторингу обстановки в заданной сфере ответственности органов управления, своевременному выявлению, идентификации и нейтрализации угроз и опасностей.

При этом под конкретные задачи с учетом прогнозируемых направлений их расширения создаются органы управления, в распоряжение которых выделяются необходимые ресурсы различных видов для выполнения поставленных задач. Для обеспечения деятельности органов управления создаются пункты управления, оснащаемые необходимыми средствами управления (автоматизированными системами управления и телекоммуникационными системами). Зачастую на основе широко применяемого функционального подхода в интересах организации и обеспечения деятельности органов управления производится выбор необходимых ресурсов из имеющихся способов и средств, а также их комплексирование на существующей нормативно-методической, технической и технологической основе. При этом совместное функционирование различных компонентов системы управления в процессе их интеграции обеспечивается посредством создания интерфейсов. С точки зрения организационного процесса, обеспечивающего расширение функционала, цель достигается. В то же время с технико-технологической точки зрения единый процесс выполнения интегрированных функций реализуется нерациональным образом.

Такой подход при своей простоте, оперативности реализации и экономичности обуславливает существенные ограничения в различных ситуациях применения системы управления, поскольку не в полной мере учитывает не только динамику развития угроз и опасностей в сфере ее ответственности и выполняемых задач, но и не рассматривает на процессном уровне природу развития негативных факторов и их взаимодействие с внутренними процессами в рамках системы управления. В свою очередь, в условиях интенсивной конкуренции в области создания перспективных ИТС на основе ИТ это оказывает сдерживающее влияние.

В этой связи целесообразно рассмотреть обобщенный механизм влияния внешних факторов на состояние системы управления и ее компонентов. Так, например, используя разработанный в [10] подход к описанию взаимосвязи внешних факторов, действующих на объект, запускаемых ими процессов в рамках объекта, приводящих к изменению технологических свойств объекта, применительно к системе управления схема их взаимосвязи может быть представлена в следующем виде, как показано на рисунке.

Конкретный фактор внешних условий функционирования системы управления объектами запускает организационный процесс в виде формирования дополнительных требований к ее объектам, либо изменения условий ее функциониро-



Обобщенная схема взаимосвязи факторов, процессов и свойств объектов

вания, в результате чего под воздействием фактора изменяются функциональные свойства системы управления на организационно-техническом уровне, либо на технико-технологическом уровне. Анализ и реализация этих требований должны оцениваться и проводиться с учетом характера их влияния на организационный процесс.

Технико-технологические процессы представляют особый интерес с точки зрения их анализа, оптимизации и обоснования рациональных решений по их интеграции с организационно-техническими процессами в рамках системы управления с учетом достижения синергетического эффекта. В течение длительного периода в отечественных разработках средств управления (средств связи и автоматизации управления) широко применяется принцип функциональной интеграции, реализация которого зачастую сопровождается элементами технической избыточности, приводящей в конечном счете к снижению их надежности и экономичности.

Анализ подхода к классификации процессов по их принадлежности к элементам системы управления и по их роли в едином процессе позволяет отметить, что такая декомпозиция процессов создает условия для выявления приоритетности тех или иных процессов при определении основы для интеграции. При этом выбор направления интеграции процессов может осуществляться с учетом направлений взаимосвязи видов процессов по компонентам системы управления, как показано в табл. 2:

- от организационных к технико-технологическим через организационно-технические процессы (например, объединение органов управления и их реорганизация требуют изменения средств управления в направлении расширения их возможностей);
- от технико-технологических процессов через организационно-технические к организационным (появление новых технологий и средств расширяет возможности средств управления, что обусловливает необходимость реорганизации органов управления, расширения их функциональных задач).

Таблица 2 Направления взаимной связи видов процессов по компонентам системы управления

Компоненты системы управления	Виды процессов в системе управления		
	Организационные	Организационно-технические	Технико-технологические
Органы управления	↓	↑	→
Пункты управления		↓	↑
Средства управления	↓	←	↑

Принимая во внимание широкое разнообразие технологических решений в информационно-телекоммуникационной области, пригодных для реализации систем управления, следует отметить, что несмотря на первенство организационных процессов, определяемых поставленными задачами перед системой управления, в ближайший период в рамках нового технологического уклада нано-, био-, информационных и когнитивных технологий определяющими станут технико-технологические процессы. Технологический уклад, как отмечено в [11], представляет собой совокупность сопряженных производств, имеющих единый технический уровень и развивающихся синхронно. Поэтому представляется весьма актуальным в сущности такого определения выделить сопряженность разнородных процессов и рассматривать ее изменение в рамках всего жизненного цикла систем управления.

Учитывая необходимость всестороннего анализа условий, ограничений и возможностей по созданию и развитию перспективной системы управления, необходимо рассматривать в комплексе указанные выше направления интеграции процессов, а обоснование рациональной структуры разнородных процессов производить на основе итерационной процедуры последовательно по обоим направлениям. При этом в поисках рациональной структуры процессов необходимо исходить из приоритетности поставленных перед системой управления задач. Затем с учетом имеющихся ограничений по реализации необходимых технико-технологических процессов в системе управления целесообразно производить поиск компромиссного решения между организацией и реализацией организационных и организационно-технических процессов с учетом поиска парето-оптимальных альтернатив.

Предложенный подход к совершенствованию нормативного регулирования процесса создания ИТС как элемента системы управления предполагает необходимость реализации следующих действий:

- описание на процессном уровне взаимосвязи ИТС с другими элементами системы управления, анализ места технико-технологических процессов в сети организационных и организационно-технических процессов в рамках единого процесса в жизненном цикле ИТС;
- анализ неурегулированных в нормативно-методическом плане областей взаимодействия разнородных процессов и обоснование предложений по выбору механизма регулирования на уровне стандартов, положений, руководств и других документов;
- разработка профилей нормативно-методических документов различного вида, необходимых для формирования нормативного поля, на основе которого конкретная ИТС может планироваться к разработке и последующему применению по предназначению.

Вариант нормативного регулирования процесса создания ИТС на основе комплексного применения существующих стандартов серии ГОСТ Р(ПБ) 15.XXX и ГОСТ 34.XXX изложен в [12].

5 Заключение

Для восстановления утерянных позиций в области отечественной ИТ-стандартизации предлагается проведение комплекса работ по следующим направлениям:

- совершенствование законодательной и нормативной базы;
- рациональное применение международного опыта в области ИТ-стандартизации;
- активизация взаимодействия с международными организациями по стандартизации в области ИТ;
- совершенствование процедуры рассмотрения и принятия ГОСТ Р В;
- подготовка и переподготовка кадров;
- развитие обеспечивающих информационных систем в области ИТ-стандартизации;
- обеспечение необходимого финансирования развития ИТ-стандартизации.

Предложенные научно-методические подходы и направления работы целесообразно реализовать поэтапно:

- (1) разработка профилей нормативно-методических документов по созданию ИТС;
- (2) разработка концепции технического регулирования ИТС на основе информационных и телекоммуникационных технологий (после апробации профильей);
- (3) разработка программы работ по реализации концепции, включающей работы по созданию комплекса отечественных нормативно-методических документов по созданию ИТС с учетом лучших международных практик.

Литература

1. Гринин Л. Е. Кондратьевские волны, технологические уклады и теория производственных революций // Кондратьевские волны: Аспекты и перспективы: Ежегодник / Отв. ред. А. А. Акаев, Р. С. Гринберг, Л. Е. Гринин, А. В. Коротаев, С. Ю. Малков. — Волгоград: Учитель, 2012. С. 222–262.
2. О стратегии национальной безопасности Российской Федерации: Указ Президента Российской Федерации от 31 декабря 2015 года № 683.
3. О стратегии научно-технологического развития Российской Федерации: Указ Президента РФ от 01.12.2016 № 642.
4. Основные органы управления, силы и средства ГСЧС. http://texts.news/chrezvyichaynyie-situatsii_1537/osnovnyie-organyi-upravleniya-silyi-sredstva-63396.html.
5. Пригожин А. И. Методы развития организаций. — М.: МЦФЭР, 2003. 864 с.

6. Зацаринный А. А., Киселев Э. В. Некоторые подходы к формированию нормативно-технической базы в части требований к архитектурному построению информационных систем организаций — участников единого информационного пространства России // Системы и средства информатики, 2015. Т. 25. № 3. С. 179–194.
7. Зацаринный А. А. Проблемные вопросы развития информационных технологий в условиях современной парадигмы научных исследований // Радиолокация, навигация, связь: Сб. трудов XXII Междунар. научн.-технич. конф. — Воронеж: САКВОЕЕ, 2016. С. 110–120.
8. Бирюков А. Н. Лекции о процессах управления информационными технологиями. — М.: БИНОМ, Лаборатория знаний, 2010. 215 с.
9. Головин С. А. Законодательство в сфере инновационной политики оборонно-промышленного комплекса Российской Федерации // Менеджмент. Вооружение. Качество (Бюллетень Центрального органа Системы добровольной сертификации «Военный Регистр»), 2016. № 2(48). С. 30–37.
10. Акусова А. А., Топоров А. А. Процессный подход к исследованию изменения технического состояния оборудования химических производств. <http://masters.donntu.org/2013/fimm/vypiraiko/library/article7.htm>.
11. Технологические уклады. Краткая справка. <http://www.savproject.ru/articles/tekhnologicheskie-uklady-kratkaya-spravka.php>.
12. Зацаринный А. А., Козлов С. В., Ионенков Ю. С. Проблемы применения нормативно-технической базы, регламентирующей процесс разработки автоматизированных информационно-телекоммуникационных систем специального назначения // Системы и средства информатики, 2007. Спец. вып. «Системы и средства информатики специального назначения». С. 155–166.

Поступила в редакцию 15.03.17

ON SCIENTIFIC AND METHODOLOGICAL APPROACHES TO REGULATORY BASE ENHANCEMENT FOR INFORMATION AND TELECOMMUNICATION SYSTEMS CREATION AND DEVELOPMENT

S. A. Golovin¹, A. A. Zatsarinnyy², and S. V. Kozlov²

¹Moscow Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper considers scientific and technical problems of regulatory base development for information and telecommunication systems (ITS) creation. The factors that determine urgency of their solution are described. It is shown that the current regulatory base falls behind the demands of ITS developers and users. The authors propose a complex approach to regulatory base enhancement

considering conditions of application of promising ITS. The justification of regulatory base enhancement directions is presented. These directions are aimed at ITS creation and development on the process base of interaction of ITS with other components of a management system as a high-tech organizational system.

Keywords: information and telecommunication system; information technologies; regulatory base; interoperability; standards profile; system approach; process approach

DOI: 10.14357/08696527170209

References

1. Grinin, L. E. 2012. Kondrat'evskie volny, tekhnologicheskie uklady i teoriya proizvodstvennykh revolyutsiy [Kondratiev's waves, technological structures and production revolution theory]. *Kondrat'evskie volny. Aspekty i perspektivy* [Kondratiev's waves. Aspects and perspectives]. Eds. A. A. Akaev, R. S. Grinberg, L. E. Grinin, A. V. Korotaev, and S. Yu. Malkov. Volgograd: Uchitel. 222–262.
2. Ukaz Prezidenta Rossiyskoy Federatsii ot 31 dekabrya 2015 goda No. 683 “O strategii natsional'noy bezopasnosti Rossiyskoy Federatsii” [Russian Federation President Decree #683 dated December 31, 2015 “Regarding strategy of Russian Federation national security”].
3. Ukaz Prezidenta RF ot 01.12.2016 No. 642 “O strategii nauchno-tehnologicheskogo razvitiya Rossiyskoy Federatsii” [Russian Federation President Decree #642 dated December 01, 2015 “Regarding strategy of Russian Federation technological development”].
4. Osnovnye organy upravleniya, sily i sredstva GSChS [General authorities, forces and means of GSE]. Available at: http://texts.news/chrezvyichaynyie-situatsii_1537/osnovnye-organyi-upravleniya-silyi-sredstva-63396.html (accessed April 15, 2017).
5. Prigozhin, A. I. 2003. *Metody razvitiya organizatsiy* [The methods of organizations development]. Moscow: MCFER. 864 p.
6. Zatsarinny, A. A., and E. V. Kiselev. 2015. Nekotorye podkhody k formirovaniyu normativno-tehnicheskoy bazy v chasti trebovaniy k arkhitekturnomu postroeniyu informatsionnykh sistem organizatsiy — uchastnikov edinogo informatsionnogo protsessov Rossii [Regarding some approaches to creation of the regulatory and technical base concerning requirements to architectural development of information systems of Russian information space members]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):179–194.
7. Zatsarinnyy, A. A. 2016. Problemye voprosy razvitiya informatsionnykh tekhnologiy v usloviyakh sovremennoy paradigm nauchnykh issledovaniy [On the problems of information technologies development in conditions of science investigations modern paradigm]. *Scientific and Technical Conference (International) on Radiolocation, Navigation, Communication Proceedings*. 110–120.
8. Biryukov, A. N. 2010. *Lektsii o protsessakh upravleniya informatsionnymi tekhnologiyami* [Lectures on information technologies management processes]. Moscow: Internet University of Information Technologies: BINOM. Laboratoriya znanij. 215 p.

9. Golovin, S. A. 2016. Zakonodatel'stvo v sfere innovatsionnoy politiki oboronno-promyshlennogo kompleksa Rossiyskoy Federatsii [Legislation in a field of Russian Federation defense and industrial complex innovation policy]. *Menedzhment. Vooruzhenie. Kachestvo (Byulleten' Tsentral'nogo organa Sistemy dobrovol'noy sertifikatsii "Voennyi Registr")* [Management. Armament. Quality (Bulletin of Central authority of voluntary certification system "Military Register")] 2(48):30–37.
10. Akusova, A. A., and A. A. Toporov. 2013. Protsessnyy podkhod k issledovaniyu izmeneniya tekhnicheskogo sostoyaniya oborudovaniya khimicheskikh proizvodstv [A process approach to investigation of chemical industries facilities technical conditions changes]. Available at: <http://masters.donntu.org/2013/fimm/vypiraiko/library/article7.htm> (accessed April 15, 2017).
11. Tekhnologicheskie uklady. Kratkaya spravka [Technological structures. Brief information]. Available at: <http://www.savproject.ru/articles/tekhnologicheskie-uklady-kratkaya-spravka.php> (accessed April 15, 2017).
12. Zatsarinnyy, A. A., S. V. Kozlov, and Yu. S. Ionenkov. 2007. Problemy primeneniya normativno-tehnicheskoy bazy, reglamentiruyushchey protsess razrabotki avtomatizirovannykh informatsionno-telekommunikatsionnykh sistem spetsial'nogo naznacheniya [Regarding problems of application of regulatory and technical base regulating the process of special purposes automated information and telecommunication systems development]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Special issue:155–166.

Received March 15, 2017

Contributors

Golovin Sergey A. (b. 1950) — Doctor of Science in technology, professor, Head of Department, Moscow Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; sgolovin@itstandard.ru

Zatsarinnyy Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinnyy@ipiran.ru

Kozlov Sergey V. (b. 1955) — Candidate of Science (PhD) in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sv_kozlov@mail.ru

ИНФОРМАЦИОННАЯ СТРУКТУРА УГРОЗ НАЦИОНАЛЬНОЙ БЕЗОПАСНОСТИ

А. П. Сучков¹

Аннотация: Обсуждаются вопросы формирования информационной модели угроз национальной безопасности для создания системы мониторинга и нейтрализации этих угроз. Обоснован выбор представления данных мониторинга в виде динамической семантической сети (ДСС). Рассмотрен состав такой сети и основные операции над ней — идентификация, слияние и включение. Конкретизирован содержательный состав основных объектов информационной модели угроз и связей между ними. Рассмотрен пример фрагмента ДСС для представления данных мониторинга научно-технических и технологических факторов, потенциально пригодных для создания новых средств и способов реализации угроз. Обоснован состав аналитических методов для идентификации и оценки угроз, основанных на методах дискретной математики и статистического анализа. Полученные результаты могут найти применение при формировании требований к разрабатываемым системам мониторинга и оценки угроз.

Ключевые слова: информационная модель угроз; динамические семантические сети; аналитические методы

DOI: 10.14357/08696527170210

1 Введение

В Стратегии национальной безопасности Российской Федерации, утвержденной Указом Президента Российской Федерации от 31 декабря 2015 г. № 683 (далее — Стратегия), дается следующее определение угрозы: «Угроза безопасности — это совокупность условий и факторов, создающих опасность жизненно важным интересам личности, общества и государства». В соответствии со Стратегией основными угрозами государственной и общественной безопасности являются:

- разведывательная и иная деятельность специальных служб и организаций иностранных государств, отдельных лиц, наносящая ущерб национальным интересам;
- деятельность террористических и экстремистских организаций, направленная на насилиственное изменение конституционного строя Российской Федерации.

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ASuchkov@ipiran.ru

ции, дестабилизацию работы органов государственной власти, уничтожение или нарушение функционирования военных и промышленных объектов, объектов жизнеобеспечения населения, транспортной инфраструктуры, устрашение населения, в том числе путем завладения оружием массового уничтожения, радиоактивными, отравляющими, токсичными, химически и биологически опасными веществами, совершения актов ядерного терроризма, нарушения безопасности и устойчивости функционирования критической информационной инфраструктуры Российской Федерации;

- деятельность радикальных общественных объединений и группировок, использующих националистическую и религиозно-экстремистскую идеологию, иностранных и международных неправительственных организаций, финансовых и экономических структур, а также частных лиц, направленная на нарушение единства и территориальной целостности Российской Федерации, дестабилизацию внутриполитической и социальной ситуации в стране, включая инспирирование «цветных революций», разрушение традиционных российских духовно-нравственных ценностей;
- деятельность преступных организаций и группировок, в том числе транснациональных, связанная с незаконным оборотом наркотических средств и психотропных веществ, оружия, боеприпасов, взрывчатых веществ, организацией незаконной миграции и торговлей людьми;
- деятельность, связанная с использованием информационных и коммуникационных технологий для распространения и пропаганды идеологии фашизма, экстремизма, терроризма и сепаратизма, нанесения ущерба гражданскому миру, политической и социальной стабильности в обществе;
- преступные посягательства, направленные против личности, собственности, государственной власти, общественной и экономической безопасности;
- коррупция;
- стихийные бедствия, аварии и катастрофы, в том числе связанные с глобальным изменением климата, ухудшением технического состояния объектов инфраструктуры и возникновением пожаров.

В статье предлагается формализация данной в Стратегии национальной безопасности содержательной трактовки перечня угроз под углом зрения проблемы мониторинга, анализа текущего состояния и прогнозирования угроз с целью управления процессами противодействия им.

2 Проблемно ориентированный подход к формированию информационной структуры

Несомненно, что угроза в своем развитии представляет собой совокупность процессов, проявление которых можно наблюдать с помощью объективных

и субъективных методов мониторинга. В связи с этим информационная модель угроз должна быть проблемно ориентирована и позволять накапливать данные о текущем состоянии угроз, быть достаточной для проведения анализа с целью идентификации угроз, определения стадии их развития и прогнозирования их состояния, разработки планов их нейтрализации.

Накопление данных об угрозах осуществляется путем мониторинга состояния взаимосвязанных элементов *обстановки* в контролируемом пространстве, изменения обстановки определяются *событиями*, образующими некоторые разворачивающиеся во времени, наблюдаемые и регистрируемые потоки. *Ситуация* определяется состоянием составляющих ее элементов обстановки, характеризующих конкретный этап развития угрозы. На основе событийного мониторинга обстановки осуществляется идентификация ситуаций, слежение за состоянием ситуаций, прогноз сценариев их развития, а также поддержка процесса выработки управляющих воздействий для достижения целевой ситуации (нормализация ситуации).

Таким образом, можно конкретизировать понятие угрозы национальной безопасности с информационной точки зрения как совокупность элементов обстановки (ситуаций), текущее состояние которых наносит или может нанести в перспективе ущерб материальному, экономическому, социальному состоянию объекта угрозы. Данная совокупность элементов обстановки должна учитывать широкий спектр научно-технических, технологических, информационных, организационных и иных факторов, определяющих состояние угроз. В зависимости от состава элементов обстановки и уровня воздействия угрозы могут быть глобальными, региональными, местными. Величина угрозы (с точки зрения ранжирования угроз по важности) количественно оценивается по величине наносимого экономического ущерба, качественно могут оцениваться и другие виды ущерба.

Наиболее адекватной моделью представления данных мониторинга угроз, с точки зрения автора, является ДСС (см., например, [1]).

Динамическая семантическая сеть определяется как $W(t) = \langle S, R, t \rangle$, где S — множество узлов (в данном случае это информационные объекты предметной области); R — множество связей (отношений) на S : $S_1 R_{12} S_2$ (например, «объект угрозы» — «содержит» — «критически важные элементы инфраструктуры»); t — параметр, отражающий время (обычно дискретный: $t = t_1, t_2, \dots, t_n, \dots$). Информационные объекты семантической сети обладают атрибутами, одним из основных атрибутов является «состояние объекта», например «террористическая группа» — «состояние: высокая степень готовности». Тогда событие можно определить как изменение состояния ДСС и, в частности, состава и состояния объектов. Также атрибутами объектов являются их идентификационные характеристики, позволяющие их однозначно определять.

В процессе мониторинга в каждый момент времени t_i возможны следующие изменения $W(t_i)$ по сравнению с $W(t_{i-1})$ (события):

- $S(t_i) = S(t_{i-1}) \cup S^+$ и $R(t_i) = R(t_{i-1}) \cup R^+$ — появление новых узлов и связей;
- $S(t_i) = S(t_{i-1}) \setminus S^-$ и $R(t_i) = R(t_{i-1}) \setminus R^-$ — исчезновение существовавших узлов и связей;
- $S(t_i) = S'(t_{i-1})$ и $R(t_i) = R'(t_{i-1})$ — изменение состояния существующих узлов и связей.

Важнейшими операциями над ДСС в процессе мониторинга являются операции *идентификации*, *слияния* и *включения* узлов и связей в процессе динамического развития обстановки. В случае появления новой информации о состоянии объекта мониторинга на основе идентификационных характеристик можно определить, является ли указанный объект уже зарегистрированным в системе, тогда происходит слияние узлов и изменение состояния объекта, если объект новый — происходит его включение в ДСС. Аналогично происходит идентификация и слияние связей. Таким образом:

$$W(t_i) = \bigcup_{j=1}^{i-1} W(t_j) \cup W'_t(t_i),$$

где $W'_t(t_i) = W(t_i) \cup S'_i(t_i) \cup R'_i(t_i) \cup S_i^+(t_i) \cup R_i^+(t_i) \setminus S_i^-(t_i) \setminus R_i^-(t_i)$.

Динамика изменения ДСС с течением времени обсуждалась в [2], где приведен пример использования динамических операций в ДСС (рис. 1) и демонстрируется, как событие (в смысле трех видов изменения $W(t_i)$) порождает ситуацию, влекущую необходимость либо изменения состава общих классов (операция *включения*), либо изменения свойств существующих сущностей (операции *идентификации* и *слияния*).

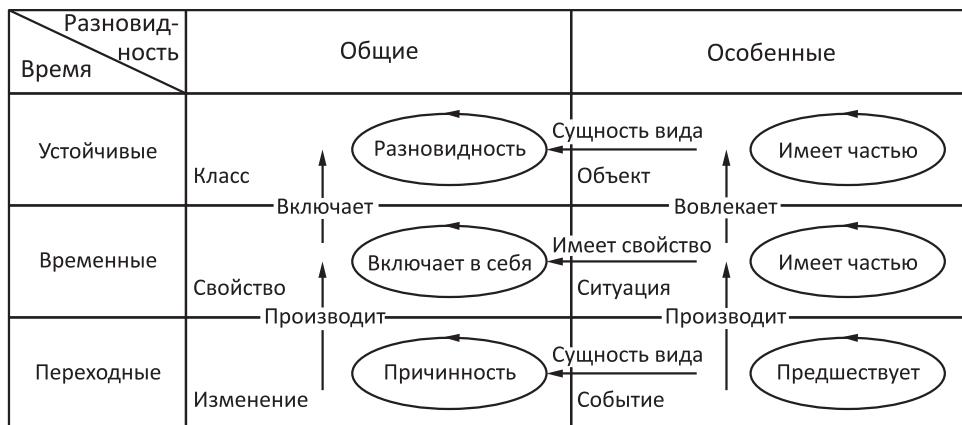


Рис. 1 Пример реализации динамических операций в ДСС

Процесс динамики изменения ДСС в ходе мониторинга обстановки в условиях стабильности предметной области, когда мала доля появляющихся новых сущностей, как правило, должен стабилизироваться:

$$W(t_i) \rightarrow W^0, \quad i \rightarrow \infty.$$

Стабилизация ДСС позволяет сделать следующие выводы.

1. В результате мониторинга данных предметной области получена представительная и объемлющая выборка контролируемых объектов и связей между ними.
2. Дальнейшие, даже небольшие, изменения ДСС могут быть с большой долей вероятности связаны со значимыми изменениями состава и опасности угроз.

3 Содержательные аспекты

Перейдем к обсуждению содержания проблемно ориентированной информационной модели угроз. Если рассматривать угрозу как сложную динамическую систему, то в общем виде угроза в каждый момент своего существования характеризуется рядом базовых сущностей:

- **объект угрозы** (чаще всего объектом угрозы является некоторая уязвимость на государственном, региональном или местном уровне — существующая или потенциальная);
- **субъект угрозы** (организационная система, включающая организованные или неорганизованные группы людей, обладающие ресурсами для создания предпосылок и для реализации угрозы);
- **средства и способы реализации угрозы** (технологии и материальные ресурсы для реализации угрозы);
- **ущерб объекту угрозы** (реальный или возможный);
- **силы и средства нейтрализации угрозы**.

Несомненно, что каждая составляющая является сложным объектом, имеет свою структуру и может быть детализирована.

В своем развитии угроза проходит определенные стадии:

- замысел;
- подготовка;
- начало и завершение реализации.

В связи с этим адекватная информационная структура угроз должна являться динамической системой, в которой состав компонент на каждой стадии может кардинально меняться (см. таблицу).

Это и обуславливает выбор ДСС как наиболее адекватного инструмента описания информационной структуры угрозы.

Компоненты угрозы на разных стадиях развития (НИР — научно-исследовательская работа)

Компоненты угрозы	Стадии		
	Замысел, концепция угрозы	Подготовка: НИР, проектирование, разработка, испытание, производство	Реализация
Объект угрозы	Существующая или потенциальная уязвимость	Существующая уязвимость. Конкретизация объекта	Воздействие на уязвимость конкретного объекта
Субъект угрозы	Политическое руководство, эксперты	Заказчик, научно-исследовательская организация, конструкторское бюро, промышленность	Госслужбы, террористические организации
Средства реализации угрозы	Существующие или перспективные технологии	Финансы	Информационные, технологические
Ущерб объекта угрозы	Прогноз ущерба	Оценка ущерба конкретного объекта	Нанесенный ущерб
Силы и средства нейтрализации угрозы	Оценка возможностей противодействия	Оценка возможностей противодействия конкретного объекта	Результаты противодействия

С точки зрения формулировок Стратегии и проблемы создания системы мониторинга и нормализации угроз можно определить следующие объекты мониторинга S , подлежащие анализу [3]:

- **угрозы национальной безопасности** (идентификационные характеристики, стадия, статус, опасность), связанные:
 - с научно-техническими и технологическими факторами, потенциально пригодными для создания новых средств и способов реализации угроз;
 - разведывательной и иной деятельностью специальных служб и организаций иностранных государств, отдельных лиц;
 - деятельностью террористических, радикальных, экстремистских организаций, преступных организаций и группировок, в том числе транснациональных, отдельных лиц;
 - стихийными бедствиями, авариями и катастрофами;

- **целевые показатели** (например, величина ущерба, интегральные оценки опасности);
- **существующие и перспективные средства и способы реализации угроз:**
 - научно-технические факторы;
 - технологические факторы;
 - средства разведки;
 - средства террористического нападения, включая оружие массового уничтожения;
 - средства информационной войны;
 - способы организации:
 - * незаконного оборота наркотических средств и психотропных веществ, оружия, боеприпасов, взрывчатых веществ;
 - * незаконной миграции и торговли людьми;
 - * коррупционной деятельности;
- **объекты угроз:**
 - конституционный строй Российской Федерации;
 - единство и территориальная целостность Российской Федерации;
 - население;
 - внутриполитическая и социальная ситуация в стране;
 - общественная и экономическая безопасность;
 - органы государственной власти;
 - военные и промышленные объекты;
 - объекты жизнеобеспечения населения;
 - традиционные российские духовно-нравственные ценности;
 - гражданский мир, политическая и социальная стабильность в обществе;
 - объекты критической информационной инфраструктуры Российской Федерации;
 - собственность;
 - личность;
 - климат;
- **силы и средства нейтрализации угроз;**
- **неконтролируемые факторы реализации угроз:**
 - противоборствующие ресурсы (силы, средства);
 - специальные службы;
 - организации иностранных государств;

- террористические и экстремистские организации;
- радикальные общественные объединения и группировки;
- иностранные и международные неправительственные организации;
- финансовые и экономические структуры;
- национальные и транснациональные преступные организации и группировки;
- отдельные лица;
- элементы окружающей среды (природные, техногенные);
- социальные, политические и экономические факторы контролируемого пространства.

Для формирования ДСС мониторинга угроз можно использовать следующие отношения R на множестве узлов S семантической сети:

- связи типа «часть–целое» или «is-a», «part-of»;
- связи типа «имеет» или «has-a» (атрибутивные связи — иметь свойство, иметь значение и т. п.);
- ассоциативные связи (определяемые обычно глаголами «производит», «влияет» и т. п.);
- количественные (больше, меньше, равно и т. п.);
- временные (раньше, позже, в течение и т. п.);
- логические связи (и, или, не) и др.

Рассмотрим пример представления данных мониторинга научно-технических и технологических факторов, потенциально пригодных для создания новых средств и способов реализации угроз.

По сообщению Reuters, Институт Френсиса Крика получил разрешение на применение техники «редактирования генов» человеческого эмбриона. Эмбрионы будут использовать исключительно в научных целях, хотя критики уже беспокоятся, что этот шаг приведет к созданию «генетически модифицированных детей».

Генетики Калифорнийского университета в Дейвисе соединили стволовые клетки человека и ДНК свиньи. Эмбрион развивался 28 дней, после чего эксперимент был искусственно прекращен для дальнейшего изучения тканей. Закон пока запрещает выращивать эмбрионы химер (гибридов человека и животных) более 28 дней, после чего эксперимент требуется прекратить [4].

На рис. 2 представлен фрагмент ДСС потенциальной угрозы населению РФ, связанной с модификацией генома человека и находящейся на начальной стадии разработки. Дальнейшими целями мониторинга этой угрозы могут быть оценка способов реализации этой угрозы, оценка связанного с ними ущерба и меры возможного противодействия.

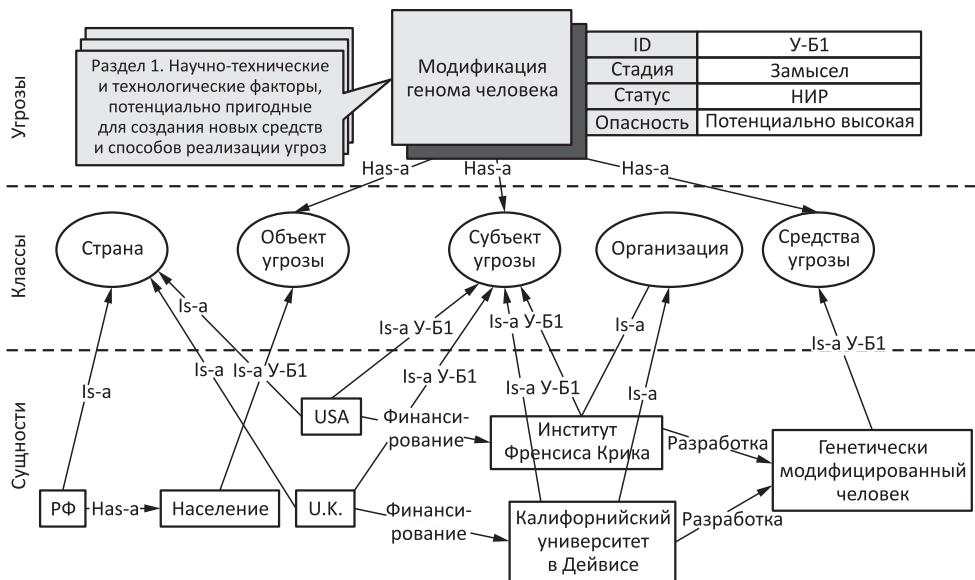


Рис. 2 Пример формализации результатов мониторинга угроз

4 Аналитические возможности предлагаемой информационной модели

Как уже упоминалось, основными характеристиками информационной модели угроз является:

- сетевая организация данных о динамических объектах и связях между ними (ДСС);
- объектно-ориентированный подход к описанию иерархии классов и наследование свойств объектов (транзитивность связей «is-a», «has-a», «part-of»);
- наличие компонент для идентификации и регистрации сложных объектов предметной области;
- специальная организация данных для осуществления дискретных методов анализа.

Дискретная структура информационной модели угроз в виде ДСС позволяет применять методы дискретной математики, связанные с теорией графов, математической логикой и лингвистическим анализом:

- выделение фактов и формализация фактографических данных на основе лингвистического анализа слабоструктурированной информации;
- идентификация и регистрация объектов, слияние подсетей;

- поиск подобных пространственно-временных конфигураций методами теории графов (изоморфизм и изоморфное вложение графов);
- логические выводы (поиск решения) на семантической сети;
- поиск прямых и ассоциативных связей (путей на графе);
- расчет интегральных и целевых показателей на графах.

С другой стороны, наличие в модели такого параметра, как время, дает возможность применять различные статистические методы анализа:

- анализ временных рядов, характеризующих изменение количественных и качественных атрибутов узлов и связей ДСС (анализ трендов, сезонных колебаний, тенденций и аномалий);
- прогнозирование изменения параметров с учетом выявленных трендов и анализируемых сценариев развития обстановки;
- динамическое моделирование ситуаций;
- статистическая оценка количественных и качественных характеристик потоков событий.

В [3] определены пять стадий анализа обстановки в системе управления, которые могут быть применены к системе мониторинга и нейтрализации угроз:

- (1) оценка параметров ненаблюдаемых (скрытых) элементов обстановки на основе выборочных или косвенных данных по результатам мониторинга, выявление и формализация фактов;
- (2) оперативный анализ обстановки путем ее сравнения с прошедшим периодом (без изменений, хуже, лучше, аномалия) с целью выявления ситуаций, требующих немедленного реагирования;
- (3) оценка ситуации с целью определения необходимости выработки решений по ее нормализации и степени сложности ситуации — штатная, критическая, чрезвычайная;
- (4) прогнозирование изменения обстановки — без управляющего воздействия, с управляющим воздействием, сценарное прогнозирование с учетом внешних факторов;
- (5) поддержка процессов принятия управленческих решений — адаптация типовых решений и выработка нетиповых решений (с учетом прогнозирования).

Представление данных мониторинга угроз, пригодное для применения сложных методов анализа, предполагает тщательную проработку вопросов классификации и стандартизации терминологии в рамках единого информационно-лингвистического обеспечения.

Литература

1. Осипов Г. С., Жилякова Л. Ю. Динамические семантические сети // Научная сессия МИФИ, 2004. Т. 3. С. 16–21.
2. Смолин Д. В. Введение в искусственный интеллект: Конспект лекций. – М.: Физматлит, 2004. 208 с.
3. Сучков А. П. Некоторые подходы к интеграции аналитических данных существующих и перспективных систем поддержки принятия решений // Системы и средства информатики, 2015. Т. 25. № 3. С. 201–211.
4. Британским генетикам разрешили модифицировать человеческие эмбрионы // Интерфакс, 01.02.2016. <http://www.interfax.ru/world/492641>.

Поступила в редакцию 09.02.17

THE INFORMATION STRUCTURE OF THREATS TO NATIONAL SECURITY

A. P. Suchkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper discusses the creation of the information model of the system of monitoring and neutralizing threats to national security. The author justifies the choice of dynamic semantic networks for representing data. The author discusses the composition of such network and the basic operations on it—identification, merge, and inclusion. The meaningful part of the main objects of the threats information model and the connections between them are specified. The author describes an example of a fragment of the dynamic semantic network used to represent the data of monitoring of scientific-technical and technological factors, potentially suitable for creation of new means and methods of threats realization. The author justifies the composition of analytical methods used to identify and assess threats based on methods of discrete mathematics and statistical analysis. The obtained results can be used to form the requirements for the systems of monitoring and assessment of threats that are being developed by the authors.

Keywords: information threat model; dynamic semantic network; analytical methods

DOI: 10.14357/08696527170210

References

1. Osipov, G. S., and L. Ju. Zhilyakova. 2004. Dinamicheskie semanticheskie seti [Dynamic semantic network]. *Nauchnaya sessiya MIFI-2004* [Scientific Session MEPhI-2004] 3:16–21.

2. Smolin, D. V. 2004. *Vvedenie v iskusstvennyy intellekt: Konspekt lektsiy* [Introduction to artificial intelligence: Lecture notes]. Moscow: Fizmatlit. 208 p.
3. Suchkov, A. P. 2015. Nekotorye podkhody k integratsii analiticheskikh dannykh sushchestvuyushchikh i perspektivnykh sistem podderzhki prinyatiya resheniy [Some approaches to the analytical data integration of the existing and future decision support systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):201–211.
4. British scientists granted permission to genetically modify human embryos. 2017. Available at: <http://www.telegraph.co.uk/science/2016/03/12/british-scientists-granted-permission-to-genetically-modify-huma/> (accessed February 8, 2017).

Received February 9, 2017

Contributor

Suchkov Alexander P. (b. 1954) — Doctor of Science in technology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ASuchkov@frccsc.ru

ОБРАТИМОСТЬ И АЛЬТЕРНАТИВНОСТЬ ГЕНЕРАЛИЗАЦИИ МОДЕЛЕЙ ПЕРЕВОДА КОННЕКТОРОВ В ПАРАЛЛЕЛЬНЫХ ТЕКСТАХ*

И. М. Зацман¹, О. С. Мамонова², А. Ю. Щурова³

Аннотация: Рассматривается задача аннотирования коннекторов русского языка и их переводов с помощью надкорпусной базы данных (НБД). Первая характерная черта НБД состоит в возможности формирования лингвистами двуязычных аннотаций, включающих одновременно как рубрики исследуемых языковых единиц (ЯЕ), в данном случае коннекторов, так и рубрики их переводов. Вторая черта заключается в том, что проставляемые лингвистами рубрики принадлежат к фасетным классификациям. Их реализация в НБД обеспечивает альтернативность генерализации аннотаций, которые являются конкретными информационными сущностями НБД. В процессе их генерализации формируются абстрактные модели перевода разной степени обобщения. Эти модели сохраняют ряд общих черт (аспектов) обобщаемых аннотаций. Поддержка в НБД фасетных классификаций обеспечивает возможность проведения многоаспектного статистического анализа сформированных аннотаций и моделей перевода коннекторов. При этом получаемые статистические данные являются верифицируемыми, так как динамически формируются ссылки от вычисленных данных к спискам соответствующих им аннотаций. Основная цель статьи заключается в описании обратимости и альтернативности процесса генерализации в НБД, что является основой многоаспектного и верифицируемого статистического анализа аннотаций и моделей перевода коннекторов в параллельных текстах.

Ключевые слова: надкорпусная база данных; аннотирование коннекторов; фасетные классификации; корпусная лингвистика; генерализация аннотаций

DOI: 10.14357/08696527170211

1 Введение

Одна из актуальных задач компьютерной лингвистики состоит в аннотировании ЯЕ, называемых коннекторами, которые обеспечивают связность текста.

* Работа выполнена при поддержке РФФИ (проект 16-24-41002) и ШННФ (проект IZL-RZ1_164059).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, izatsman@yandex.ru

² Факультет иностранных языков и регионоведения Московского государственного университета им. М. В. Ломоносова, matonovaoks@mail.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ann.shurova@gmail.com

Об актуальности этой задачи говорит тот факт, что в настоящее время разрабатывается стандарт ISO по методике аннотирования дискурсивных отношений и выражающих их коннекторов. В процессе разработки этого стандарта сопоставляются общие черты и отличия существующих подходов к аннотированию [1, 2].

Аннотируемый коннектор может состоять из одного (*хотя, притом, кстати*) и большего числа слов (*не только . . . но и . . . ; скорее . . . чем . . .*); первые являются по своему составу однокомпонентными, вторые — многокомпонентными [3]. Получение статистических данных о коннекторах русского языка и моделях их перевода является одной из задач совместного российско-швейцарского проекта «Контрастивное корпусное исследование коннекторов русского языка», который в настоящее время выполняется в Институте проблем информатики ФИЦ ИУ РАН и на филологическом факультете Женевского университета.

Основой многоаспектного статистического анализа являются прямые и реверсивные двуязычные аннотации коннекторов и их переводов. Аннотации формируются лингвистами в НБД в процессе контрастивного анализа текстов на русском и французском языках [4–7]. Прямые аннотации формируются в процессе сопоставления текстов на русском языке и их переводов на французский язык, а реверсивные аннотации — текстов на французском языке и их переводов на русский язык. Примеры русско-французских аннотаций приводятся в следующих разделах статьи.

В процессе сопоставления текстов при формировании прямой аннотации коннектору русского языка ставится в соответствие функционально эквивалентный фрагмент (ФЭФ — термин Д. Добровольского [8]) из текста перевода. При формировании реверсивной аннотации в переводе на русский язык ищется исследуемый коннектор. Затем для него определяется ФЭФ из оригинального текста на иностранном языке, который является стимулом появления этого коннектора в переводе на русский язык. Описание понятия «стимул» в контексте исследования моделей перевода дано в [9–11]. Отличительной чертой всех моделей перевода является их обратимость в НБД, т. е. возможность обратного перехода ко всем аннотациям, в результате обобщения которых была сформирована каждая из моделей. Обычно в процессе моделирования текстов и их ЯЕ сохраняются только общие черты обобщаемых объектов без возможности обратного к ним перехода [12].

Целью статьи является описание двуязычных аннотаций, сформированных лингвистами в процессе сопоставления параллельных русско-французских и французско-русских текстов, а также многоуровневой обратимой и альтернативной генерализации аннотаций, которые являются конкретными информационными объектами НБД.

2 Параллельные тексты

На январь 2017 г. НБД содержала 1 248 956 слов в оригинальных текстах на русском языке, 1 745 405 слов в их переводах на французский язык, 258 062 слова

Таблица 1 Предложения и их профессиональные переводы в НБД

Оригинальный текст	Перевод
Вместо того чтобы идти брить чиновничьи подбородки, он отправился в заведение с надписью «Кушанье и чай» спросить стакан пуншу, как вдруг заметил в конце моста квартального надзирателя благородной наружности, с широкими бакенбардами, в треугольной шляпе, со шпагою.	Et au lieu d'aller raser les mentons bureaucratiques, il se dirigea vers un établissement dont l'enseigne indiquait: Repas et thé, pour y prendre un verre de punch; mais il aperçut soudain à l'autre bout du pont l'officier de police du quartier d'aspect majestueux: larges favoris, tricorne sur la tête et épée au côté.
Он обмер; а между тем квартальный кивал ему пальцем и говорил: — А подойди сюда, любезный!	Ivan Iakovlevitch se sentit défaillir. Or l'officier de police lui fit signe d'approcher et dit: "Viens un peu ici, l'ami!"
Иван Яковлевич, зная форму, снял издали еще картуз и, подошвши проворно, сказал: — Желаю здравия вашему благородию!	Connaissant les usages, Ivan Iakovlevitch enleva de loin sa casquette et accourut prestement: "Je souhaite, bonne santé à Votre Noblesse."

в оригинальных текстах на французском языке и 205 000 слов в их переводах на русский язык. Основная причина различия в количестве слов в переводах на французский язык почти в 40% заключается в том, что отдельным текстам на русском языке в НБД соответствует несколько их переводов на французский язык. В НБД предложения на русском языке поставлены в соответствие их переводам на французский язык (табл. 1). Всего образовано 104 471 русско-французских пары и 13 402 французско-русских пары. Как правило, пары формируются таким образом, что одному предложению оригинального текста соответствует одно предложение в его переводе. Общий объем НБД составляет около 3,5 млн слов, что дает возможность исследовать низкочастотные модели переводов коннекторов.

Примером выровненных параллельных текстов, хранимых в НБД, могут служить несколько предложений из повести Н. В. Гоголя «Нос» и их профессиональные переводы на французский язык (см. табл. 1). Кроме профессиональных в НБД хранятся и машинные переводы отдельных предложений, что дает возможность сопоставлять их с профессиональными переводами.

В НБД каждому тексту на русском языке может соответствовать несколько профессиональных переводов (табл. 2), а каждому тексту на иностранном языке — несколько его переводов на русский язык.

Объектом исследования рассматриваемой НБД являются коннекторы русского языка в переводах на французский язык и с французского языка. Под коннектором понимается ЯЕ, функция которой состоит в выражении отношений между двумя соединенными с ее помощью фрагментами текста [4]. В нижеприведенном примере коннектор и его ФЭФ выделены полужирным шрифтом.

Таблица 2 Три варианта перевода на французский язык

Контекст исследуемой ЯЕ	Перевод 1	Перевод 2	Перевод 3
Он обмер; а между тем квартальный кивал ему пальцем и говорил: — А подойди сюда, любезный!	Ivan Iakovlevitch se sentit défaillir. Or l'officier de police lui fit signe d'approcher et dit: "Viens un peu ici, l'ami!"	Il perdit contenance, tandis que l'exempt l'appelait du doigt et disait: "Approche, mon brave!"	Il se figea; et pourtant , l'inspecteur lui faisait signe du doigt et l'appelait: — Approche un peu, mon brave!

- (1) Та, **лишь только** увидела кота, лезущего в трамвай, со злобой, от которой даже тряслась, закричала: — Котам нельзя!

Dès qu'elle vit, en effet, le chat essayer de s'introduire dans le tramway, elle cria, avec une colère telle qu'**'elle en tremblait**: — Pas de chats ici!

Так, в (1) коннектор **лишь только** устанавливает временные отношения между предложениями «увидела кота, лезущего в трамвай» и «*та [...] со злобой, от которой даже тряслась, закричала*». Во французском переводе находим аналогичную структуру — коннектор **dès que** для выражения временных отношений между фрагментами “*elle vit, en effet, le chat essayer de s'introduire dans le tramway*” и “*elle cria, avec une colère telle qu'**'elle en tremblait***”.

Когда лингвист находит в параллельных текстах НБД коннектор (точнее, его речевую реализацию — РР [6]), у него есть возможность зафиксировать в форме аннотации сведения о РР и ее ФЭФ на языке перевода. Аннотация представляет собой формализованное описание одновременно свойств РР коннектора как объекта исследования, ее ФЭФ, контекстов РР и ФЭФ, а также может иметь один или несколько текстовых

комментариев, сформированных лингвистами.

3 Аннотации и уровни их генерализации в надкорпусной базе данных

В табл. 3 приведены два примера аннотаций без текстовых комментариев. Первый столбец содержит фрагменты текстов на русском языке, в которых выделен исследуемый коннектор **когда||то**. В первом примере РР этого коннектора устанавливает временные отношения, а во втором — условные.

Второй столбец содержит исследуемый коннектор и четыре рубрики в угловых скобках, составляющих формализованное описание свойств РР коннектора и ее контекста, которые будут описаны ниже. Третий столбец содержит фрагменты текстов на французском языке, в которых выделены два разных ФЭФ (**quand** и **lorsque**) исследуемого коннектора **когда||то**. Четвертый столбец содержит

Таблица 3 Два примера аннотаций с коннектором **когда||то**

Контекст РР коннектора	Речевая реализация коннектора, рубрики коннектора и его контекста	Контекст ФЭФ РР коннектора	Функционально эквивалентный фрагмент коннектора, рубрики ФЭФ и его контекста
Когда бываю в таком состоянии, то становлюсь нахальным и наглым до крайности.	когда то ⟨ временные ⟩ ⟨ CNT ⟩ ⟨ CNT p CNT q ⟩ ⟨ дистант ⟩	Quand je suis dans cet état, je deviens insolent et cynique à l'extrême.	quand ⟨ временные ⟩ ⟨ с предикацией ⟩ ⟨ начальная ⟩ ⟨ CNT q p ⟩ ⟨ CNT ⟩
Я заметил — когда человек влюблен и у него долги, то предметом разговоров становится его моральный облик.	когда то ⟨ условные ⟩ ⟨ CNT ⟩ ⟨ CNT p CNT q ⟩ ⟨ дистант ⟩	J'ai remarqué que lorsqu' un homme est amoureux et qu'il a des dettes, on finit toujours par se poser des questions sur sa moralité.	lorsque ⟨ условные ⟩ ⟨ с предикацией ⟩ ⟨ начальная ⟩ ⟨ CNT q p ⟩ ⟨ CNT ⟩

ФЭФ исследуемого коннектора и пять рубрик формализованного описания свойств ФЭФ и их контекстов.

Важно отметить, что НБД позволяет обрабатывать как уникальные информационные объекты, так и абстрактные объекты разной степени обобщения (генерализации), что является основой для проведения статистического анализа моделей перевода коннекторов и других ЯЕ в параллельных текстах. К уникальным информационным объектам относятся параллельные тексты (см. табл. 1 и 2), их фрагменты и аннотации (см. табл. 3). До описания абстрактных информационных объектов сначала рассмотрим более подробно рубрики, простоявшие во втором и четвертом столбцах табл. 3.

Рубрики во втором столбце обозначают:

- ⟨ временные ⟩ означает, что коннектор **когда||то** в первом примере выражает временные отношения;
- ⟨ CNT ⟩ квалифицирует **когда||то** как коннектор (эта рубрика позволяет явно указать, когда полифункциональная ЯЕ выполняет функцию коннектора; если она выполняет другую функцию, то эта рубрика не ставится);
- ⟨ CNT p CNT q ⟩ описывает порядок следования соединяемых фрагментов текста «*бываю в таком состоянии*» (обозначен как p) и «*становлюсь нахальным и наглым до крайности*» (q);

Таблица 4 Три наиболее частотные модели перевода коннектора **когда||то**

Коннектор русского языка	Его ФЭФ в переводе на французский язык	Число ФЭФ в переводах		Всего ФЭФ в профессиональных и машинных переводах	Доля ФЭФ в профессиональных и машинных переводах (от 107 аннотаций для всех 19 моделей перевода), %
		профессиональных	машинных		
когда то	quand	28	34	62	57,94
	lorsque	12	1	13	12,15
	quand alors	2	9	11	10,28

- ⟨ дистант ⟩ описывает дистантное расположение частей двухместного коннектора **когда||то**;
- ⟨ условные ⟩ обозначает, что **quand** выражает условные отношения.

Рубрики, отличающиеся от вышеперечисленных, в четвертом столбце говорят о следующем:

- ⟨ с предикцией ⟩ маркирует часть предложения с предикцией;
- ⟨ начальная ⟩ маркирует начальную позицию коннектора;
- ⟨ CNT q p ⟩ описывает порядок следования соединяемых фрагментов текста “*je suis dans cet état*” (обозначен как q) и “*je deviens insolent et cynique à l'extrême*” (p)¹.

Результатом первого этапа генерализации аннотаций, построенных для коннекторов **когда||то** с ФЭФ **quand**, является модель перевода ⟨ **R-когда||то; F-quand** ⟩, где **R-когда||то** обозначает русский коннектор, а **F-quand** — его ФЭФ в переводе на французский. Эта модель перевода встречается в НБД 62 раза. По состоянию на 1 марта 2017 г. всего в НБД сформировано 107 аннотаций с коннектором **когда||то** (60 для профессиональных и 47 для машинных переводов). В профессиональных переводах эта модель перевода встречается 28 раз. Первые три наиболее частотные модели перевода коннектора **когда||то** описаны в табл. 4.

Для ФЭФ **lorsque** результатом генерализации аннотаций с коннектором **когда||то** является модель перевода ⟨ **R-когда||то; F-lorsque** ⟩, а для ФЭФ **quand||alors** — модель перевода ⟨ **R-когда||то; F-quand||alors** ⟩. Кроме трех рассмотренных (**quand**, **lorsque** и **quand||alors**) в НБД зарегистрированы еще 4 низ-

¹На данном этапе развития НБД в ней не различаются рубрики Cnt p q и Cnt q p. Во всех случаях, когда коннектор, строго говоря, не связывает два фрагмента, а маркирует один из них, проставляется рубрика Cnt q p.

кочастотных и 12 единичных ФЭФ для коннектора **когда||то**. Всего для коннектора **когда||то** зарегистрировано 19 моделей перевода. На модели с низкочастотными и единичными ФЭФ приходится около 20% аннотаций (от 107), включая нулевой ФЭФ (когда отсутствует перевод этого коннектора).

Таким образом, результатом первого этапа генерализации являются модели перевода с конкретной ЯЕ в левой части аннотации и конкретным ФЭФ в правой части. В НБД эти модели являются обобщенным (генерализованным) представлением знаний о вариантах перевода коннекторов. Процесс формирования и генерализации аннотаций представляет собой целенаправленную генерацию и представление знаний о вариантах перевода [13–17]. При этом кроме известных и зафиксированных в грамматиках в НБД представлены авторские варианты перевода, поэтому параллельные тексты являются постоянно пополняемым источником новых лингвистических знаний, формируемых в процессе семантического моделирования с помощью баз данных [18].

Рассмотренные модели перевода относятся к *первому уровню генерализации* аннотаций. Отметим, что в НБД генерализация аннотаций является обратимой: так, все числа в третьем и пятом столбцах табл. 4 (62, 28, 13, 12, 11 и 2) являются ссылками на соответствующие списки аннотаций. Использование подобных ссылок обеспечивает возможность обратного перехода от моделей перевода к соответствующим им аннотациям. Две аннотации, соответствующие числу 2 в табл. 4, приведены в табл. 5. Таким образом, на примере НБД коннекторов можно наглядно проследить первый этап *обратимой генерализации*.

Таблица 5 Две аннотации, соответствующие числу 2 в табл. 4

Речевая реализация коннектора в контексте оригинала	Коннектор, рубрики коннектора и его контекста	Контекст ФЭФ РР коннектора в контексте перевода	Функционально эквивалентный фрагмент коннектора, рубрики ФЭФ и его контекста
А когда не знают, какой ярлык прилепить к моему лбу, то говорят: «Это странный человек, странный!»	когда то ⟨ условные ⟩ ⟨ CNT p CNT q ⟩ ⟨ дистант ⟩ ⟨ SubCNT ⟩ ⟨ CNT ⟩	Et quand ils ne savent plus quelle étiquette vous coller sur le front, alors ils disent: “C'est un homme bizarre, bizarre!”	quand alors ⟨ условные ⟩ ⟨ CNT p CNT q ⟩ ⟨ дистант ⟩ ⟨ SubCNT ⟩ ⟨ CNT ⟩
Когда я утомлюсь совершенно, до полного отупения, то все бросаю и бегу сюда.	когда то ⟨ временные ⟩ ⟨ CNT p CNT q ⟩ ⟨ CNT ⟩ ⟨ дистант ⟩	Quand je suis complètement épuisé, abruti de travail, alors je laisse tout tomber, j'accours ici.	quand alors ⟨ временные ⟩ ⟨ CNT p CNT q ⟩ ⟨ CNT ⟩ ⟨ дистант ⟩

ции (т. е. перехода от уникальных аннотаций к моделям перевода и обратно к аннотациям).

Рассмотрим кратко следующий уровень *обратимой генерализации*, реализованной в НБД. Результат второго этапа генерализации, т. е. обобщения 19 моделей перевода с коннектором **когда||то**, обозначим как более абстрактную модель перевода в виде кортежа $\langle \text{R-когда||то}; \text{F-ФЭФ} \rangle$, где F-ФЭФ — это любой перевод на французский язык коннектора **когда||то**, для которого в НБД сформирована аннотация. Как отмечалось выше, на 1 марта 2017 г. таких аннотаций было 107. Это число в НБД является ссылкой на список всех аннотаций с коннектором **когда||то**.

Аналогично можно было бы описать следующие этап и уровень генерализации, вводя обозначение модели перевода в виде кортежа $\langle \text{R-CNT}; \text{F-ФЭФ} \rangle$, где R-CNT — любой коннектор русского языка, зарегистрированный в НБД, а F-ФЭФ — любой его перевод на французский язык, для которого в НБД сформирована аннотация. Однако прежде чем рассмотреть этот уровень генерализации, опишем понятие *альтернативной генерализации*.

4 Альтернативная генерализация моделей перевода

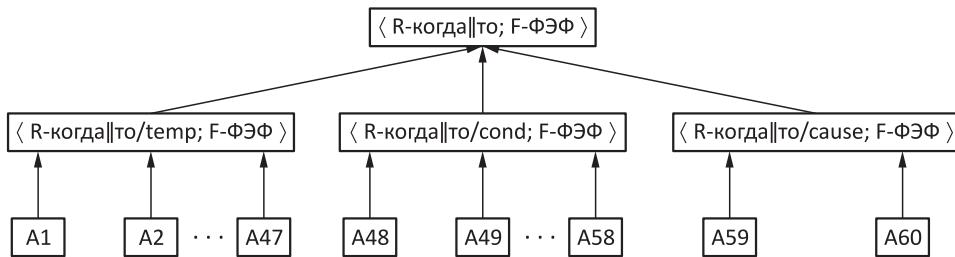
Рубрики, перечисленные во втором и четвертом столбцах аннотаций (см. табл. 3 и 5), также могут использоваться в процессе генерализации. При этом у пользователя НБД есть возможность выбирать разные рубрики при генерализации. От выбора той или иной рубрики зависит вариант генерализации. При этом возможно использование нескольких вариантов, т. е. в НБД реализована поддержка альтернативности процесса генерализации.

Отметим, что альтернативная генерализация была описана в 1979 г. основателем теории реляционных баз данных Эдгаром Коддом в работе [19]. Продемонстрируем ее поддержку в НБД коннекторов на двух примерах из табл. 3. В первом примере коннектор **когда||то** выражает временные отношения (обозначим рубрику **временные** из табл. 3 кратко как temp), а во втором — условные (обозначим рубрику **условные** как cond). Отметим, рубрики temp и cond относятся к одному фасету описания отношений, выражаемых коннекторами, который в НБД обозначен словом «**Отношения**». Описание разных классификационных схем отношений и соответствующих им рубрик можно найти в [1, 2, 20–22].

Введем обозначение модели перевода в виде кортежа $\langle \text{R-когда||то}/\text{temp}; \text{F-ФЭФ} \rangle$ как частный случай более общего кортежа $\langle \text{R-когда||то}; \text{F-ФЭФ} \rangle$. Частный случай $\langle \text{R-когда||то}/\text{temp}; \text{F-ФЭФ} \rangle$ соответствует только тем моделям перевода, в которых коннектор **когда||то** выражает временные отношения. Эта модель встречается в 47 аннотациях (табл. 6), сформированных для профессиональных переводов (на рис. 1 они пронумерованы A1–A47). Аналогично введем $\langle \text{R-когда||то}/\text{cond}; \text{F-ФЭФ} \rangle$ для условных отношений (встречается в 11 аннотациях, они пронумерованы A48–A58) и $\langle \text{R-когда||то}/\text{cause}; \text{F-ФЭФ} \rangle$ для причинно-следственных отношений (2 аннотации A59 и A60).

Таблица 6 Три рубрики фасета «Отношения» и число аннотаций по рубрикам

Рубрика	Число аннотаций с коннектором когда то	Доля в группе из 60 аннотаций, %
Временные (temp)	47	78,3(3)
Условные (cond)	11	18,3(3)
Причинно-следственные (cause)	2	3,3(3)

**Рис. 1** Промежуточный этап генерализации с рубриками temp, cond и cause

Результат второго этапа генерализации 60 аннотаций с коннектором **когда||то**, обозначенный как **< R-когда||то; F-ФЭФ >**, может быть получен с использованием рубрик temp, cond и cause на промежуточном этапе генерализации (см. рис. 1).

Рассмотрим альтернативный вариант генерализации с использованием рубрик другого фасета, который в НБД обозначен словом «Статус». В табл. 5 используются две рубрики этого фасета: **< SubCNT >** для встроенных коннекторов (т. е. **когда||то** является частью коннектора **а когда||то**) и **< [-]SubCNT >** для невстроенных коннекторов. Эти рубрики далее в кортежах будем обозначать без угловых скобок.

Введем обозначение **< R-когда||то/SubCNT; F-ФЭФ >** как частный случай более общего кортежа **< R-когда||то; F-ФЭФ >**. Этот частный случай соответствует только тем моделям перевода, в которых коннектор **когда||то** является встроенным. Эта частная модель встречается в 36 аннотациях, сформированных для профессиональных переводов (на рис. 2 они пронумерованы A1–A36). Аналогично введем обозначение **< R-когда||то/[-]SubCNT; F-ФЭФ >** для невстроенных коннекторов (встречается в 24 аннотациях, они пронумерованы A37–A60). Тогда результат альтернативной генерализации 60 аннотаций может быть получен на промежуточном этапе с использованием рубрик SubCNT и [-]SubCNT (см. рис. 2).

В НБД все альтернативные варианты генерализации аннотаций являются обратимыми. Каждое из пяти чисел (47, 11 и 2 для фасета «Отношения» и 36

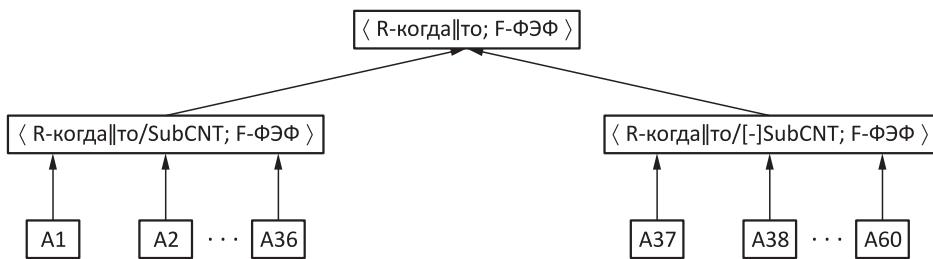


Рис. 2 Альтернативная генерализация с рубриками SubCNT и [-]SubCNT

и 24 для фасета «Статус»), полученное в результате поиска аннотаций с коннектором **когда||то** и одной из пяти рубрик (temp, cond, cause, SubCNT и [-]SubCNT), является ссылкой на соответствующий список аннотаций. Таким образом, результаты такого поиска обеспечивают возможность альтернативных обратных переходов от модели перевода $\langle \text{R-когда||то; F-ФЭФ} \rangle$ к аннотациям с рубриками temp (47 аннотаций), cond (11), cause (2), SubCNT (36) и [-]SubCNT (24) соответственно.

5 Пять уровней генерализации

Рассмотренные первые три уровня генерализации аннотаций как конкретных информационных объектов НБД показаны на рис. 3 (на примере коннектора **когда||то** и трех его наиболее частотных ФЭФ). Они даны без промежуточных уровней и альтернативных вариантов, которые получаются за счет использования рубрик фасетов, поддерживаемых НБД (см. рис. 1 и 2).

Все три этапа генерализации являются обратимыми в НБД, включая промежуточные этапы альтернативных вариантов обобщения аннотаций. На 1 марта 2017 г. модели $\langle \text{R-CNT; F-ФЭФ} \rangle$ соответствовало 5 468 прямых аннотаций коннекторов и их переводов, которые были сформированы лингвистами в процессе сопоставления текстов на русском языке и их французских переводах.

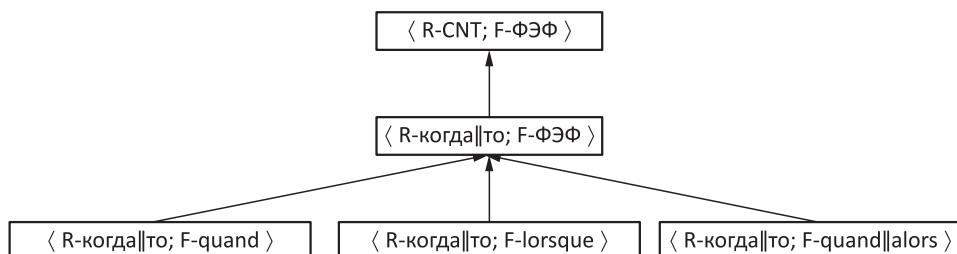


Рис. 3 Результаты первых трех уровней генерализации

Таблица 7 Количественные данные о сформированных в НБД двуязычных аннотациях (на 1 марта 2017 г.)

Категория ЯЕ	Число прямых русско-французских аннотаций	Число реверсивных французско-русских аннотаций
Личные глагольные конструкции	10061	0
Безличные конструкции	2188	321
Лингвоспецифичные и дискурсивные слова	5871	1603
Коннекторы	5468	327
Итого	23588	2251

Помимо коннекторов НБД применяются для контрастивного анализа других категорий ЯЕ: личных глагольных конструкций [9, 23], безличных конструкций [24], лингвоспецифичных и дискурсивных слов [11, 25]. Для ЯЕ этих категорий также были сформированы двуязычные аннотации (табл. 7). В результате их генерализации были получены еще две модели перевода четвертого и пятого уровня генерализации.

Результатом четвертого уровня генерализации является абстрактное понятие модели русско-французского перевода исследуемых в НБД ЯЕ всех категорий, перечисленных в табл. 7, которую обозначим как $\langle \mathbf{R\text{-}YE; F\text{-}\Phi\mathcal{E}\Phi} \rangle$. Этому понятию соответствуют 23 588 конкретных информационных объектов НБД. Аналогично на этом же уровне генерализации можно ввести понятие модели французско-русского перевода $\langle \mathbf{F\text{-}\Phi\mathcal{E}\Phi; R\text{-}YE} \rangle$, которой соответствует 2251 конкретный информационный объект НБД.

Кроме текстов на французском языке НБД включает тексты на немецком и итальянском языках. Результатом пятого уровня генерализации является абстрактное понятие модели русско-иноязычного перевода ЯЕ, которую обозначим как $\langle \mathbf{R\text{-}YE; Tr\text{-}\Phi\mathcal{E}\Phi} \rangle$, где Tr-ФЭФ — перевод исследуемых в НБД ЯЕ на французский, немецкий или итальянский язык. Аналогично на этом же уровне генерализации введем понятие модели иноязычно-русского перевода $\langle \mathbf{Or\text{-}\Phi\mathcal{E}\Phi; R\text{-}YE} \rangle$, где Or-ФЭФ — французский, немецкий или итальянский фрагмент оригинального текста, который является стимулом появления исследуемой ЯЕ в переводе на русский язык. Следует, однако, заметить, что на четвертом и пятом уровнях абстракции связи моделей перевода с аннотациями являются не столько содержательными, сколько количественными.

6 Заключение

Рассмотрены пять уровней генерализации двуязычных аннотаций, которые являются конкретными информационными объектами. С использованием при-

меров из НБД коннекторов на каждом из уровней определена модель перевода коннекторов и других исследуемых ЯЕ. Степень абстрактности моделей возрастает при переходе на более высокий уровень. Перечислим рассмотренные модели перевода этих уровней в порядке возрастания степени абстрактности:

- (1) $\langle \text{R-когда} \parallel \text{то}; \text{ F-quand } \rangle$, $\langle \text{R-когда} \parallel \text{то}; \text{ F-lorsque } \rangle$ и $\langle \text{R-когда} \parallel \text{то}; \text{ F-quand} \parallel \text{alors} \rangle$ — модели перевода конкретного коннектора русского языка с фиксированным ФЭФ из текста на французском языке;
- (2) $\langle \text{R-когда} \parallel \text{то}; \text{ F-ФЭФ } \rangle$ — модель перевода конкретного коннектора с любым ФЭФ;
- (3) $\langle \text{R-CNT; F-ФЭФ} \rangle$ — модель перевода для всех коннекторов с любыми ФЭФ;
- (4) $\langle \text{R-ЯЕ; F-ФЭФ} \rangle$ и $\langle \text{F-ФЭФ; R-ЯЕ} \rangle$ — модель перевода ЯЕ русского языка для всех их категорий, исследуемых с помощью НБД, с любыми ФЭФ из текста на французском языке, а также модель стимулов появления исследуемых ЯЕ в переводах на русский язык;
- (5) $\langle \text{R-ЯЕ; Tr-ФЭФ} \rangle$ и $\langle \text{Or-ФЭФ; R-ЯЕ} \rangle$ — модель перевода ЯЕ русского языка для всех их категорий с любыми ФЭФ из текста перевода на иностранном языке, а также модель иноязычно-русского перевода, в которой иноязычный фрагмент оригинального текста является стимулом появления исследуемой ЯЕ в переводе на русский язык.

Результаты генерализации аннотаций на всех пяти уровнях являются обратимыми, так как в НБД для любой абстрактной модели можно сформировать список двуязычных аннотаций, которые ей соответствуют. Поэтому все статистические данные о моделях перевода в НБД являются верифицируемыми с помощью списков соответствующих им двуязычных аннотаций.

Кроме пяти основных уровней рассмотрены промежуточные уровни альтернативной генерализации, результаты которой также являются обратимыми. С прикладной точки зрения наибольший интерес представляет пример промежуточного уровня с 19 моделями перевода коннектора **когда** \parallel **то**, который наглядно иллюстрирует сложность решения проблемы машинного перевода. Если в этом примере принимать в расчет только три наиболее частотные модели (см. табл. 4), то «теряются» остальные 16 моделей перевода коннектора **когда** \parallel **то**, включая модель с нулевым ФЭФ. При этом они описывают около 20% случаев перевода коннектора **когда** \parallel **то** в НБД.

Для повышения качества машинного перевода необходимо принимать в расчет не только все модели перевода коннекторов, но и других ЯЕ. Точность выбора релевантной модели во многом будет зависеть от полноты фасетных классификаций, используемых в процессе аннотирования ЯЕ.

Опыт формирования двуязычных аннотаций с помощью НБД коннекторов продемонстрировал возможность целенаправленного пополнения фасетных

классификаций непосредственно в процессе аннотирования. При этом рубрики создаваемых фасетных классификаций являются верифицируемыми за счет реализованного в НБД поиска по любой рубрике. Результатом такого поиска является список тех аннотаций, в которых эта рубрика была проставлена лингвистами [22].

В НБД представлен широкий диапазон информационных объектов по степени их обобщения. Это, с одной стороны, уникальные объекты, к которым относятся параллельные тексты (см. табл. 1 и 2), их фрагменты и аннотации (см. табл. 3 и 5). С другой стороны, это абстрактные сущности, к которым относятся модели перевода, соответствующие разным этапам обобщения аннотаций. При этом результаты обобщения на каждом этапе являются обратимыми и характеризуются количественно. Таким образом, НБД поддерживает обратимые конкретно-абстрактные многоуровневые информационные трансформации, которые в статье рассмотрены на примере лингвистических ресурсов.

В заключение отметим, что расширение спектра рассматриваемых информационных ресурсов, возможно, позволит в будущем предложить новый подход к постановке общеначальной проблемы соотношения между конкретным и абстрактным, вводя количественные характеристики для конкретно-абстрактных трансформаций. Традиционно они описывались только качественно [26]. Расширение спектра позволит уточнить основания и предметную область информатики, в определении которой описание информационных трансформаций является ключевым положением [27].

Литература

1. *Prasad R., Bunt H.* Semantic relations in discourse: The current state of ISO 24617-8 // 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings. — Tilburg: Tilburg University, 2015. P. 80–92.
2. *Bunt H., Prasad R.* ISO-DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations // 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings, 2016. P. 45–54. http://www.lrec-conf.org/proceedings/lrec2016/LREC2016_Proceedings.zip.
3. *Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А.* Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
4. *Инькова-Манзотти О. Ю.* Коннекторы противопоставления во французском и русском языках. Сопоставительное исследование. — М.: Информэлектро, 2001. 432 с.
5. *Зализняк Анна А., Зацман И. М., Инькова О. Ю., Кружков М. Г.* Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика-2015: Труды 7-й Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.
6. *Инькова О. Ю.* К проблеме описания многокомпонентных коннекторов русского языка: не только... но и // Вопросы языкоznания, 2016. № 2. С. 37–60.
7. *Зализняк Анна А., Зацман И. М., Инькова О. Ю.* Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 101–109.

8. Добровольский Д. О., Кретов А. А., Шаров С. А. Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003–2005. — М.: Индрик, 2005. С. 263–296.
9. Loiseau S., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 100–109.
10. Сичинава Д. В. Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики // Язык, литература, культура: актуальные проблемы изучения и преподавания, 2014. Вып. 10. С. 37–44.
11. Зализняк Анна А. База данных межъязыковых эквиваленций как инструмент лингвистического анализа // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2016. С. 763–775.
12. Эко У. Открытое произведение / Пер. с итал. — СПб.: Академический проект, 2004. 384 с. (Eco U. Opera aperta. — Milano: Bompiani, 1967. 288 p.)
13. Zatsman I. Tracing emerging meanings by computer: Semiotic framework // 13th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Limited, 2012. Vol. 2. P. 1298–1307.
14. Zatsman I., Buntman N., Kruzhkov M., Nuriev V., Zalizniak Anna A. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery // 15th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Limited, 2014. Vol. 3. P. 1063–1071.
15. Zatsman I., Buntman N. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery // 16th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Limited, 2015. P. 851–860.
16. Зацман И. М. Процессы целенаправленной генерации и развития кросс-языковых экспертизных знаний: семиотические основания моделирования // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 106–123.
17. Zatsman I., Buntman N., Coldefy-Fauard A., Nuriev V. WEB knowledge base for asynchronous brainstorming // 17th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Limited, 2016. P. 976–983.
18. Hull R., King R. Semantic database modeling: Survey, applications, and research issues // ACM Comput. Surv., 1987. Vol. 19. No. 3. P. 201–260.
19. Codd E. F. Extending the database relational model to capture more meaning // ACM Trans. Database Syst., 1979. Vol. 4. No. 4. P. 397–434.
20. Prasad R., Dinesh N., Lee A., Miltsakaki E., Robaldo L., Joshi A., Webber B. The Penn Discourse TreeBank 2.0 // 6th Conference (International) on Language Resources and Evaluation Proceedings — Paris: European Language Resources Association, 2008. P. 2961–2968.
21. Zufferey S., Degand L. Annotating the meaning of discourse connectives in multilingual corpora // Corpus Linguist. Ling. Theory, 2013. P. 1–24. doi: 10.1515/cllt-2013-0022. https://www.researchgate.net/profile/Sandrine_Zufferey/publication/

- 258088055_Annotating_the_meaning_of_discourse_connectives_in_multilingual_corpora/links/00b49526e7355a985f000000.pdf.
- 22. Займан И. М., Инькова О. Ю., Нуриев В. А. Построение классификационных схем: методы и технологии экспериментного формирования // Научно-техническая информация. Сеп. 2, 2017. № 1. С. 8–22.
 - 23. Kruzhkov M. G., Buntman N. V., Loshchilova E. Ju., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. A database of Russian verbal forms and their French translation equivalents // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2014. Вып. 13(20). С. 284–296.
 - 24. Зализняк Анна А., Кружков М. Г. База данных безличных глагольных конструкций русского языка // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 136–145.
 - 25. Зализняк Анна А. Лингвоспецифичные единицы русского языка в свете контрастивного корпусного анализа // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2015. Вып. 14(21). Т. 1. С. 683–695.
 - 26. Berlin I. Against the current: Essays in the history of ideas. — London: The Hogarth Press, 1979. 300 p.
 - 27. Rosenbloom P. S. On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 307 p.

Поступила в редакцию 14.03.17

REVERSIBILITY AND ALTERNATIVENESS OF GENERALIZATION OF CONNECTIVES TRANSLATIONS MODELS IN PARALLEL TEXTS

I. M. Zatsman¹, O. S. Mamonova², and A. Yu. Shchurova¹

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Foreign Languages and Area Studies, M. V. Lomonosov Moscow State University, 31-a Lomonosov Str., Moscow 119192, Russian Federation

Abstract: The paper considers the task of annotation of Russian connectives and their translations with the use of a supracorpora database (SCDB). The first distinctive feature of the SCDB is that it supports creation of bilingual annotations that include both rubrics of the investigated linguistic items (i. e., connectives, in this case) and rubrics of their translations. The second feature is that the rubrics assigned by the linguists are in fact elements of faceted classifications. Implementation of these rubrics in the SCDB enables alternativeness of generalization of annotations that represent concrete informational entities in the SCDB. As these entities are created, abstract translation models of different generalization levels are produced. These models preserve certain common characteristics (aspects) of the generalizable annotations. The support of faceted classifications in the

SCDB makes it possible to conduct multifaceted statistical analysis of annotations and connectives translation models in the SCDB. Furthermore, these statistical data are verifiable since the generated quantitative data provide direct links to lists of corresponding annotations. The main objective of the paper is to describe reversibility and alternativeness of the generalization processes in the SCDB, which provides a basis for conducting multifaceted and verifiable statistical analysis of annotations and connectives translation models in parallel texts.

Keywords: supracorpora database; annotation of connectives; faceted classifications; corpus linguistics; generalization of annotations

DOI: 10.14357/08696527170211

Acknowledgments

This research was performed in the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, with financial support of the Russian Foundation for Basic Research (grant No. 16-24-41002) and the Swiss National Science Foundation (grant No. IZLRZ1_164059).

References

1. Prasad, R., and H. Bunt. 2015. Semantic relations in discourse: The current state of ISO 24617-8. *11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings*. Tilburg: Tilburg University. 80–92.
2. Bunt, H., and R. Prasad. 2016. ISO-DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. *12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings*. 45–54. Available at: http://www.lrec-conf.org/proceedings/lrec2016/LREC2016_Proceedings.zip (accessed March 8, 2017).
3. Zatsman, I. M., O. Yu. In'kova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie krossyazykovykh znaniy o konnektorakh v nadkorpusnykh bazakh dannykh [Representation of cross-lingual knowledge about connectives in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
4. Inkova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh. Sopostavitel'noe issledovanie* [Connectives of opposition in French and Russian. A comparative study]. Moscow: Informelektro. 432 p.
5. Zaliznyak, Anna A., I. M. Zatsman, O. Yu. In'kova, and M. G. Kruzhkov. 2015. Nadkorpusnye bazy dannykh kak lingvisticheskiy resurs [Supracorpora databases as linguistic resource]. *7th Conference (International) on Corpus Linguistics Proceedings*. St. Petersburg: SPbGU. 211–218.
6. Inkova, O. Yu. 2016. K probleme opisaniya mnogokomponentnykh konnektorov russkogo yazyka: ne tol'ko... no i [Towards the description of multiword connectives in Russian: Not only... but also]. *Voprosy yazykoznanija* [Topics in the study of language] 2:37–60.
7. Zaliznyak, Anna A., I. M. Zatsman, and O. Yu. In'kova. 2017. Nadkorpusnaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database on connectives: term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):101–109.

8. Dobrovolskiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov: arkhitektura i vozmozhnosti ispol'zovaniya [Corpus of parallel texts: Architecture and applications]. *Natsional'nyy korpus russkogo yazyka: 2003–2005* [Russian National Corpus: 2003–2005]. Moscow: Indrik. 263–296.
9. Loiseau, S., D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2013. Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):100–109.
10. Sitchinava, D. V. 2014. Ispol'zovanie parallel'nogo korpusa dlya kolichestvennogo izucheniya lingvospetsifichnoy leksiki [Using a parallel corpus for the quantitative study of language-specific units]. *Yazyk, literatura, kul'tura: aktual'nye problemy izucheniya i prepodavaniya* [Language, literature, culture: Actual problems of research and teaching] 10:37–44.
11. Zaliznyak, Anna A. 2016. Baza dannykh mezh'yazykovykh ekvivalentsiy kak instrument lingvisticheskogo analiza [Database of cross-linguistic equivalences as a tool for linguistic analysis]. *Computer Linguistics and Intellectual Technologies: Conference (International) "Dialog" Proceedings*. Moscow: RGGU. 763–775.
12. Eco, U. 1967. *Opera aperta*. Milano: Bompiani. 288 p.
13. Zatsman, I. 2012. Tracing emerging meanings by computer: Semiotic framework. *13th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 2:1298–1307.
14. Zatsman, I., N. Buntman, M. Krushkov, V. Nuriev, and Anna A. Zalizniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 3:1063–1071.
15. Zatsman, I., and N. Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery. *16th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 851–860.
16. Zatsman, I. 2015. Protsessy tselenapravlennoy generatsii i razvitiya krossyazykovykh ekspertnykh znanii: semioticheskie osnovaniya modelirovaniya [Goal-oriented processes of cross-lingual expert knowledge creation: Semiotic foundations for modeling]. *Informatika i ee Primeneniya — Inform. Appl.* 9(3):106–123.
17. Zatsman, I., N. Buntman, A. Coldefy-Fauard, and V. Nuriev. 2016. WEB knowledge base for asynchronous brainstorming. *17th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Limited. 976–983.
18. Hull, R., and R. King. 1987. Semantic database modeling: Survey, applications, and research issues. *ACM Comput. Surv.* 19(3):201–260.
19. Codd, E. F. 1979. Extending the database relational model to capture more meaning. *ACM Trans. Database Syst.* 4(4):397–434.
20. Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The Penn Discourse TreeBank 2.0. *6th Conference (International) on Language Resources and Evaluation Proceedings*. Paris: European Language Resources Association. 2961–2968.
21. Zufferey, S., and L. Degand. 2013. Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguist. Ling. Theory.* 1–24. doi: 10.1515/cllt-2013-0022. Available at: https://www.researchgate.net/profile/Sandrine_Zufferey/

- publication/258088055>Annotating_the_meaning_of_discourse_connectives_in_multilingual_corpora/links/00b49526e7355a985f000000.pdf (accessed March 8, 2017).
22. Zatsman, I., O. Yu. In'kova, and V. Nuriev. 2017. Postroenie klassifikatsionnykh skhem: metody i tekhnologii ekspertnogo formirovaniya [Construction of classification schemes: Methods and technologies of experts' formation]. *Nauchnaya i tekhnicheskaya informatsiya. Ser. 2* [Scientific and Technical Information. Ser. 2] 1:8–22.
 23. Kruzhkov, M. G., N. V. Buntman, E. Yu. Loshchilova, D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2014. A database of Russian verbal forms and their French translation equivalents. *Computer Linguistics and Intellectual Technologies: Conference (International) "Dialog" Proceedings*. Moscow: RGGU. 13(20):284–296.
 24. Zalizniak, Anna A., and M. G. Kruzhkov. 2016. Baza dannykh bezlichnykh glagol'nykh konstruktsiy russkogo yazyka [Database of Russian impersonal verbal constructions]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):136–145.
 25. Zalizniak, Anna A. 2015. Lingvospetsifichnye edinitsy russkogo yazyka v svete kontrastivnogo korpusnogo analiza. *Computer Linguistics and Intellectual Technologies: Conference (International) "Dialog" Proceedings*. Moscow: RGGU. 14(21):651–662.
 26. Berlin, I. 1979. *Against the current: Essays in the history of ideas*. London: The Hogarth Press. 300 p.
 27. Rosenbloom, P. S. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 307 p.

Received March 14, 2017

Contributors

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

Mamonova Oksana S. (b. 1998) — student, Faculty of Foreign Languages and Area Studies, M. V. Lomonosov Moscow State University, 31-a Lomonosov Str., Moscow 119192, Russian Federation; mamonovaoks@mail.ru

Shchurova Anna Yu. (b. 1991) — junior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ann.shchurova@gmail.com

ОБ ОДНОМ ПОДХОДЕ К МОДЕЛИРОВАНИЮ ПРОЦЕССА РАЗВИТИЯ СЕМАНТИЧЕСКОЙ СЕТИ

И. М. Адамович¹, О. И. Волков²

Аннотация: Статья посвящена описанию и обоснованию подхода к моделированию семантической сети, лежащей в основе технологии поддержки конкретно-исторических исследований. Данная технология опирается на взаимодействие множества независимых исследователей, что потенциально может служить причиной деградации качества накапливаемой информации. Для проверки в динамике качества информации, организованной в форме семантической сети, и служит описываемая модель. Предложенный подход к моделированию заключается в опоре на принципы графодинамики, позволяющие строить граф на основе строго определенных правил, соответствующих практике, и на модель Барабаши–Альберт (алгоритм генерации случайных безмасштабных сетей). Адекватность модели была проверена с точки зрения ее соответствия признакам когнитивных сетей, к которым семантическая сеть, лежащая в основе технологии поддержки конкретно-исторических исследований, относится по своей природе. Проверка модели производилась в широком диапазоне задаваемых параметров.

Ключевые слова: семантическая сеть; модель; графодинамика; когнитивная сеть; принцип предпочтительного присоединения

DOI: 10.14357/08696527170212

1 Введение

В статье [1] была описана новая распределенная технология поддержки историко-биографических исследований, основанная на системе «Т-парсер», осуществляющей автоматическое извлечение фактов из текстов на естественном языке [2]. По причине принципиального многообразия, фрагментарности и противоречивости данных [3], используемых в конкретно-историческом исследовании, технология была построена на принципах сотворчества и краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий). Эта новая, но перспективная форма вовлечения общественности в процессы решения научных задач получила название «краудсайенс» (гражданская наука), в основе которой лежит научное волонтерство [4].

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, Adam@amsd.com

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, Volkov@amsd.com

Была обоснована форма организации информации в виде семантической сети, ориентированная на автоматизированную обработку и совместное одновременное использование многими исследователями. Но, как отмечалось исследователями, специфической дисфункцией гражданской науки является риск деградации науки под влиянием добровольцев, что вызывает необходимость разработки механизма оценки качества и интерпретации данных, предоставленных научными волонтерами [5].

Действительно, описанная технология имеет много общего с фолксономией (народная таксономия), представляющей собой практику совместной категоризации посредством выбираемых ключевых слов — тегов [6], и, следовательно, потенциально подвержена всем его недостаткам. Несмотря на преимущества фолксономии, существует проблема недостаточной точности предоставления информации пользователям. Помимо возможных лексических неточностей фолксономического подхода существует проблема отсутствия взаимосвязанности данных, их классификации. Данные вносятся в хаотичном порядке, от различных категорий пользователей и не могут быть правильно структурированы и иметь какие-либо логические связи и ассоциации. Каждая поступившая запись индивидуальна, содержит свои уникальные характеристики и свойства какого-либо объекта информационного пространства [7].

Таким образом, возникает задача оценки качества информации, привносимой одновременно множеством не связанных между собой исследователей. Трудность состоит в том, что эта информация организована в форме динамически расширяющейся сети понятий и связанных с ними экземпляров. Поведение сетей со статической топологией изучено достаточно хорошо, но проблема формального описания и моделирования процессов развития сетевых структур проработана еще в недостаточной мере [8]. Целью настоящей статьи является построение и обоснование модели, с помощью которой можно было бы в динамике изучить свойства информации, организованной в форме семантической сети, с учетом ее добавления из множества разнородных не связанных между собою источников.

2 Концепции модели

2.1 Графодинамический подход

В ряде работ [8–11] был предложен и развит графодинамический подход к имитационному моделированию процессов развития сетевых структур, базирующийся на следующих предпосылках:

- (1) конфигурации сетевой структуры моделируются произвольными неориентированными графиками;
- (2) операции над графиками (кроме копирования) являются локальными. Набор операций взят из практики и определен на подмножестве вершин графа (одной вершине, паре вершин, неполном или полном наборе вершин). Список операций является открытым и может пополняться;

- (3) процесс развития характеризуется последовательностью действий, где каждое действие путем применения заданного набора операций изменяет предыдущую конфигурацию (граф сети);
- (4) между действиями и формируемыми ими графиками имеет место однозначное соответствие, обеспечиваемое тем, что последовательность действий определяется конечным автоматом на основе дерева (каждому действию, кроме начального, непосредственно предшествует одно и то же действие).

Задача решается введением набора стандартных операций над графиками, отражающих изменения конфигурации реальных сетевых структур, и построения конечного автомата трансформирования графов на заданном временном горизонте. Состояниям автомата соответствуют действия, а переходы обусловлены событиями, описываемыми индикаторными логическими формулами. События определяются истечением конкретного времени с момента запуска предыдущего действия и внешними воздействиями на сетевую структуру.

Анализ процессов развития сетевой структуры реализуется в форме человеко-компьютерного взаимодействия. Человек анализирует данные и принимает решение о введении в сеть соответствующих механизмов управления, о коррекции параметров модели и т. д. [8].

Идея введения набора стандартных операций над графиком, отражающих изменения конфигурации семантической сети, представляется конструктивной с точки зрения разрабатываемой модели, но, к сожалению, недостаточной. Построение конечного автомата трансформирования графов в данном случае представляется невозможным. Поэтому необходима дополнительная концепция, определяющая вероятности стандартных изменений и способы соединения добавляемых блоков, отражающая неравномерность топологической структуры сети, соответствующей структуре семантики предметной области, образом которой является сеть.

2.2 Когнитивные сети

Для решения вышеописанной проблемы следует опереться на некоторые дополнительные свойства рассматриваемой семантической сети, определяющие неслучайность ее структуры, несмотря на формирование ее из множества независимых источников.

В классических случайных графах распределение узлов по числу связей спадает довольно быстро, поскольку для случайных сетей распределение узлов по числу связей подчиняется закону Пуассона. В этих сетях практически отсутствуют узлы с большим числом связей (хабы). В противоположность этому, многие реальные сети — от метаболических сетей в клетках до Интернета — имеют медленно спадающее распределение узлов по числу связей, и в этих сетях хабы (узлы с большим числом связей) составляют заметную долю от всех узлов, и именно они определяют многие важные свойства этих сетей. Для них

характерна асимптотическая зависимость распределения узлов по числу связей в виде степенного закона $P(q) \sim q^{-\gamma}$. Распределение узлов по числу связей в виде степенного закона называется безмасштабным, а сами сети получили название безмасштабных сетей [12].

Безмасштабные сети принято разделять на технологические (Интернет, WWW), социальные (сети знакомств, сети сотрудничества) и биологические (экологические, функциональные сети мозга, сети белковых взаимодействий, метаболические сети). Показатель степени γ в законе распределения их узлов по числу связей обычно находится в диапазоне от 2 до 3. При этом технологические и биологические сети отличаются от социальных сетей характером корреляции узлов, который получил название «ассортативность». Показатель ассортативности описывается коэффициентом корреляции Пирсона между степенями соседних узлов. Для технологических и биологических сетей свойственна отрицательная ассортативность, в социальных сетях имеет место положительная ассортативность [13].

В последнее время процессы, связанные с познанием и творчеством, также стали изучать с точки зрения сетевого подхода [14]. Очевидно, что соответствующие им сети нельзя отнести ни к одной из трех вышеназванных категорий (технологические, биологические, социальные). Они образуют особую категорию — когнитивные сети. Предварительные исследования показывают, что когнитивные сети являются безмасштабными. Предварительные исследования закономерностей узловой корреляции (ассортативности) когнитивных сетей показывают, что эти сети по своей структуре ближе к биологическим и технологическим, чем к социальным сетям [15].

Моделируемая семантическая сеть по своей природе относится к когнитивным сетям и должна удовлетворять их свойствам — иметь степенной закон распределения степеней узлов с показателем, находящимся в диапазоне от 2 до 3, а также отрицательный показатель ассортативности.

2.3 Принцип предпочтительного присоединения

Для реализации сформулированных выше требований к модели можно взять за основу модель Барабаши–Альберт — алгоритм генерации случайных безмасштабных сетей. В основу построения графа Барабаши–Альберт положен принцип предпочтительного присоединения. Новая вершина соединяется с уже существующими вершинами, причем присоединение производится случайным образом — вероятность ребра, исходящего из вновь созданной вершины к i -й вершине, пропорциональна степени k_i этой вершины. Иными словами, чем больше вершина имеет связей, тем выше вероятность ее выбора для присоединения к ней новой вершины.

Реалистичность моделей, построенных по этому принципу, была проверена на модели сети Интернет. Было показано, что данные модели в точности реализуют основные эмпирические свойства Интернета [16].

Таким образом, в дополнение к принципам графодинамики для реализации модели семантической сети выбран принцип предпочтительного присоединения. Созданный на этих принципах алгоритм подлежит проверке на соответствие результата основным параметрам когнитивных сетей.

3 Описание модели

Моделируемая семантическая сеть состоит из двух подсетей: сети понятий и сети экземпляров. В каждой из подсетей существуют свои правила пополнения, определяемые технологией работы пользователей.

3.1 Сеть понятий

Узлы сети понятий представляют собой именованные универсальные классы объектов. Единственным обязательным свойством класса является его имя. Из требования распределенности вытекает, что сеть может создаваться независимо разными исследователями, а поэтому она должна формироваться из автономных именованных подсетей. Для «склейки» этих подсетей между собой предусмотрена связь эквивалентности «узел–узел». Данный вид связи, введенный исключительно для удобства работы различных пользователей, можно не переносить в модель и считать эквивалентные узлы совпадающими. Помимо вышеупомянутой связи эквивалентности предусмотрены связи типа:

- подпонятие (например, «офицер» → «лейтенант»);
- атрибут (например, «персона» → «место рождения»);
- значение (например, «место рождения» → «город»);

При этом все эти три вида связи являются направленными.

Далее будем называть понятием только те узлы сети, у которых нет входящих связей или есть только входящие связи типа «значение». У понятий могут быть исходящие связи либо исключительно типа «атрибут» (такое понятие будем называть нормальным понятием, или Н-понятием), либо типа «подпонятие». Такое понятие по смыслу является атрибутом, и оно не встречается без входящей связи «значение». Будем называть его атрибутным, или А-понятием. Узел, у которого есть входящая связь «атрибут», будем называть атрибутом. Из атрибута могут исходить связи только типа «подпонятие» или «значение». Если у атрибута нет исходящих связей, то назовем его терминальным.

В соответствии с принципами графодинамики выделим следующие стандартные операции над графом семантической сети:

- добавление нового Н-понятия со случайным набором терминальных атрибутов;
- добавление нового Н-понятия с атрибутом, имеющим под собой случайное дерево подпонятий;

- добавление нового Н-понятия с атрибутом с исходящей связью «значение» на существующее понятие;
- добавление нового Н-понятия с атрибутом с исходящей связью «значение» на новое А-понятие, имеющее под собой случайное дерево подпонятий;
- добавление нового терминального атрибута к существующему Н-понятию;
- добавление нового атрибута, имеющего под собой случайное дерево подпонятий, к существующему Н-понятию;
- добавление нового атрибута с исходящей связью «значение» на существующее понятие к существующему Н-понятию;
- добавление нового атрибута с исходящей связью «значение» на новое А-понятие, имеющее под собой случайное дерево подпонятий, к существующему Н-понятию;
- добавление случайного поддерева (ветви) к существующему дереву подпонятий, исходящих от А-понятия или атрибута.

Относительные вероятности осуществления данных операций будем определять весовыми коэффициентами, задаваемыми как настраиваемые параметры модели.

Для новых понятий будем задавать признак возможности для их экземпляров быть связанными связью «часть» (см. п. 3.2). Вероятность получения данного признака для понятия задается как параметр модели. В реальности эта вероятность зависит от типа исследования. Так, в генеалогических исследованиях такие признаки будут получать прежде всего географические объекты, а в просопографических (просопография — специальная историческая дисциплина, изучающая коллективные биографии определенных социальных групп) исследованиях — прежде всего организаций и различные объединения.

Размеры добавляемых наборов (наборов атрибутов или поддеревьев) будем определять случайно с учетом определяемого человеческой психологией предела — объема внимания. Объем внимания — количество информации, объектов, которые человек может одновременно и отчетливо воспринимать. Объем внимания у взрослых обычно колеблется в пределах от 5 до 7 объектов [17].

Присоединение добавляемых объектов к уже существующим будем осуществлять в соответствии с описанным выше принципом предпочтительного присоединения.

3.2 Сеть экземпляров

Факты в конкретно-историческом исследовании в соответствии с моделируемой технологией задаются значениями экземпляров классов и связями между ними. Связи наследуются из сети понятий (классов). Также может быть добавлена связь нового типа — «часть». Она также является направленной и нужна для наследования атрибутов.

В соответствии с принципами графодинамики выделим следующие стандартные операции над подсетью экземпляров семантической сети:

- добавление нового экземпляра Н-понятия со случайным поднабором атрибутов (необязательно терминальных) из набора атрибутов понятия;
- добавление нового атрибута из набора атрибутов Н-понятия (необязательно терминального) к существующему экземпляру Н-понятия.

Относительные вероятности осуществления данных операций, а также относительные вероятности событий в подсетях понятий (классов) и экземпляров будем определять весовыми коэффициентами, задаваемыми как настраиваемые параметры модели.

Для экземпляров, связанных отношением «часть», введем следующие термины:

- экземпляр, у которого нет входящих связей типа «часть», но есть несколько исходящих связей такого типа, будем называть Ч-хабом;
- экземпляр, у которого есть входящая и несколько исходящих связей типа «часть» будем называть Ч-подхабом;
- если от экземпляра A к экземпляру B идет связь типа «часть», то будем говорить, что B есть часть A .

При появлении нового экземпляра понятия, допускающего отношение «часть» у своих экземпляров, возможны следующие события:

- новый экземпляр не будет связан отношением «часть» с прочими экземплярами;
- новый экземпляр станет частью существующего экземпляра;
- новый экземпляр станет Ч-подхабом, т. е. подмножество частей какого-то Ч-хаба станут частями нового экземпляра, а сам он станет частью этого Ч-хаба;
- новый экземпляр станет Ч-хабом, т. е. несколько существующих экземпляров станут его частями.

Относительные вероятности осуществления данных операций будем определять весовыми коэффициентами, задаваемыми как настраиваемые параметры модели.

При появлении нового Ч-хаба число присоединяемых к нему частей не может лимитироваться объемом внимания и зависит от текущего наполнения семантической сети. Будем моделировать это значение случайной величиной, распределенной по экспоненциальному закону $f(x) = \lambda e^{-\lambda x}$, с математическим ожиданием $1/\lambda$, задаваемым как параметр модели.

4 Проверка модели

Была проведена серия проверок модели в широком диапазоне задаваемых параметров. Основные результаты представлены ниже.

4.1 Безмасштабность сети

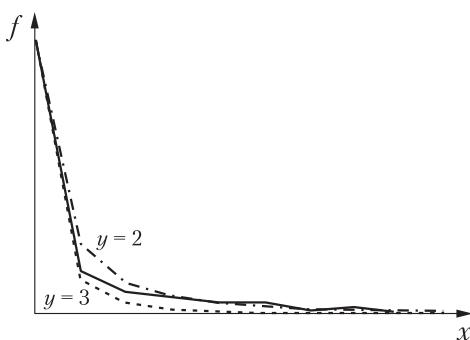


Рис. 1 Распределение узлов модели семантической сети по числу связей

Распределение узлов модели семантической сети по числу связей соответствует степенному закону с показателем $\gamma \in [2,3; 2,6]$ в широком диапазоне задаваемых параметров (рис. 1) с устойчивой тенденцией стремления к значению 2,4 по мере роста сети, что соответствует ожидаемому диапазону от 2 до 3. Структура сети демонстрирует явно выраженные хабы (рис. 2). Таким образом, сеть, генерируемая моделью, соответствует признакам безмасштабных сетей.

4.2 Показатель ассортативности

Проверка коэффициента узловой корреляции (ассортативности) модели семантической сети в широком диапазоне задаваемых параметров показала, что диапазон его вариативности соответствует интервалу $[-0,5; -0,2]$ при средних размерах сети (примерно до 1000 узлов) с тенденцией стремления к 0 при дальнейшем росте сети, т. е. коэффициент ассортативности модели семантической сети соответствует свойствам когнитивных сетей.

5 Выводы

Рассматриваемая модель семантической сети соответствует признакам когнитивной сети. При этом ее базовые свойства, определяющие ее принадлежность к когнитивным сетям, инвариантны по отношению к настраиваемым параметрам модели.

Предложенный подход к построению модели за счет сочетания опоры на принципы графодинамики, позволяющие строить граф на основе строго определенных правил, соответствующих практике, и на модель Барабаши–Альберт (алгоритм генерации случайных безмасштабных сетей) позволил сформировать адекватную модель семантической сети, лежащей в основе технологии поддержки конкретно-исторических исследований. Ценность предложенной модели для данной технологии состоит в том, что она позволит в динамике оценить качество

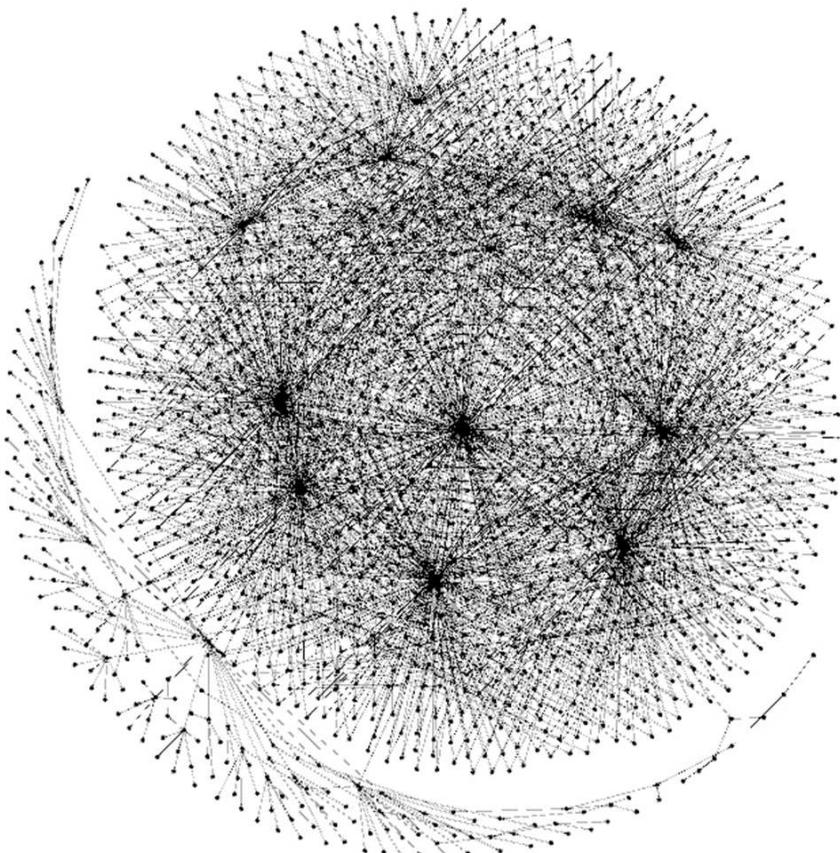


Рис. 2 Структура модели семантической сети

информации, привносимой одновременно множеством не связанных между собой исследователей.

Литература

1. Адамович И. М., Волков О. И. Технология распределенного автоматизированного анализа исторических текстов // Системы и средства информатики, 2016. Т. 26. № 3. С. 148–161.
2. Адамович И. М., Волков О. И. Система извлечения биографических фактов из текстов исторической направленности // Системы и средства информатики, 2015. Т. 25. № 3. С. 235–250.
3. Адамович И. М., Волков О. И. Иерархическая форма представления биографического факта // Системы и средства информатики, 2016. Т. 26. № 2. С. 108–122.

4. Егерев С. В. От краудсорсинга простых операций к «науке граждан» // Социология науки и технологий, 2016. № 4. С. 74–85
5. Иванишикина Ю. В. «Краудсайенс» как новая форма краудсорсинга: гуманитарные эффекты и проблемы применения // Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты: Мат-лы 3-й Междунар. научн.-практич. конф. — М.: МГГУ им. М. А. Шолохова, 2014. С. 28–34.
6. Заморкин А. А. Фолксономия и социальный теггинг как новые явления виртуальной коммуникации // Вестник Ярославского гос. ун-та им. П. Г. Демидова. Сер. Гуманитарные науки, 2012. № 3(21). С. 173–177.
7. Семенов С. П., Ташкин А. О. Применение фолксономического подхода в разработке социально-ориентированных геоинформационных систем // Вестник Югорского гос. ун-та, 2014. № 2(33). С. 94–99.
8. Юдицкий С. А. Графодинамическое имитационное моделирование развития сетевых структур // Управление большими системами, 2011. Вып. 33. С. 21–34.
9. Юдицкий С. А. Графодинамическая автоматная модель разрешения конфликтов в организационных системах // Управление большими системами, 2008. Вып. 23. С. 126–136.
10. Юдицкий С. А., Желтова Л. В., Владиславлев П. Н. Графодинамическая триадная модель системы с дискретным управлением // Приборы и системы. Управление, контроль, диагностика, 2010. № 10. С. 12–18.
11. Юдицкий С. А. Техника графодинамического моделирования бинарных игр на основе сценарных связок // Управление большими системами, 2010. Вып. 31. С. 289–298.
12. Евин И. А. Введение в теорию сложных сетей // Компьютерные исследования и моделирование, 2010. Т. 2. № 2. С. 121–141.
13. Евин И. А. Хабибуллин Т. Ф., Шувалов Н. Д., Васюков И. О. Искусство и теория сложных сетей // Сложные системы, 2012. № 3. С. 38–50.
14. Bak P. How nature works. The science of self-organized criticality. — Oxford: Oxford University Press, 1997. 62 р.
15. Евин И. А., Кобляков А. А., Савриков Д. В., Шувалов Н. Д. Когнитивные сети // Компьютерные исследования и моделирование, 2011. Т. 3. № 3. С. 231–239.
16. Бадрызлов В. А. Принципы генерации случайных графов для моделирования сети Интернет // Омский научный вестник, 2014. № 3(133). С. 204–208.
17. Комарова Т. К. Психология внимания. — Гродно: ГрГУ, 2002. 124 с.

Поступила в редакцию 15.03.17

AN APPROACH TO MODELING THE SEMANTIC NET EVOLUTION

I. M. Adamovich and O. I. Volkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: This article is devoted to the description and rationale of the approach to the modeling of semantic net which is the basis for the technology of concrete

historical investigation support. This technology is supported by the interaction of many independent researchers. This fact can potentially cause the degradation of the collected information quality. The described model is used to check the quality of the information which is organized as a semantic net. The proposed modeling approach is based on the concepts of graphodynamics, which allows building a graph on the basis of strictly defined empiric rules, and the Barabasi–Albert model (algorithm for generating random scale-free networks). The model adequacy was checked by comparing its parameters with the main characteristics of cognitive networks on the ground that the semantic net of the concrete historical investigation technology is, of its nature, a cognitive network. Model validation was carried out for a wide range of model's parameters values.

Keywords: semantic net; model; graphodynamics; cognitive network; preferential attachment concept

DOI: 10.14357/08696527170212

References

1. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelennogo avtomatizirovannogo analiza istoricheskikh tekstov [The distributed automated technology of historical texts analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 3(26):148–161.
2. Adamovich, I. M., and O. I. Volkov. 2015. Sistema izvlecheniya biograficheskikh faktov iz tekstov istoricheskoy napravленности [The system of facts extraction from historical texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 2(25):235–250.
3. Adamovich, I. M., and O. I. Volkov. 2016. Ierarkhicheskaya forma predstavleniya biograficheskogo fakta [Hierarchical format of biohraphical fact]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 2(26):147–161.
4. Egerev, S. V. 2016. Ot kraudsorsinga prostykh operatsiy k "nauke grazhdan" [From SciSourcing of simple operations to the citizen science]. *Sotsiologiya nauki i tekhnologiy [Sociology of Science and Technology]* 4:74–85.
5. Ivanishkina, Y. V. 2014. "Kraudsayens" kak novaya forma kraudsorsinga: gumanitarnye effekty i problemy primeneniya [Crowdsourcing as a new form of crowdsourcing: Humanitarian effects and problems of application]. *3rd Scientific-Practical Conference (International) "Social Computing: Fundamentals, Technologies of Development, Social and Humanities Effects" Proceedings*. Moscow. M. A. Sholokhov Moscow State University for the Humanities. 28–34.
6. Zamorkin, A. A. 2012. Folksonomiya i sotsial'nyy tegging kak novye yavleniya virtual'noy kommunikatsii [Folksonomy and social tagging as phenomena of virtual communications]. *Vestnik Yaroslavskogo gos. un-ta im. P. G. Demidova. Ser. Gumanitarnye nauki* [Bulletin of P. G. Devidov Yaroslavl State University. Humanitarian science ser.] 3(21):173–177.
7. Semenov, S. P., and A. O. Tashkin. 2014. Primenenie folksonomiceskogo podkhoda v razrabotke sotsial'no-orientirovannykh geoinformatsionnykh sistem [Application folksonomy approach in the development of socially-oriented geographic information

- system]. *Vestnik Yugorskogo gos. un-ta* [Bulletin of the Yugra State University] 2(33):94–99.
8. Yuditskiy, S. A. 2011. Grafodinamicheskoe imitatsionnoe modelirovaniye razvitiya setevykh struktur [Graphodynamic simulation modeling of network structures evolution]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 33:21–34.
 9. Yuditskiy, S. A. 2008. Grafodinamicheskaya avtomatnaya model' razresheniya konfliktov v organizatsionnykh sistemakh [Structural-machine model of conflict solutions in organizational systems]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 23:126–136.
 10. Yuditskiy, S. A., L. V. Zheltova, and P. N. Vladislavlev. 2010. Grafodinamicheskaya triadnaya model' sistemy s diskretnym upravleniem [Graphdynamic triad model of the discrete controlled system]. *Pribory i sistemy. Upravlenie, kontrol', diagnostika* [Instruments and Systems: Monitoring, Control, and Diagnostics] 10:12–18.
 11. Yuditskiy, S. A. 2010. Tekhnika grafodinamicheskogo modelirovaniya binarnykh igr na osnove stsenarnykh svyazok [A technique for graph-dynamic modeling of binary games based on scenario bindings]. *Upravlenie bol'shimi sistemami* [Large-Scale Systems Control] 31:289–298.
 12. Yevin, I. A. 2010. Vvedenie v teoriyu slozhnykh setey [Introduction to the theory of complex networks]. *Komp'yuternye issledovaniya i modelirovaniye* [Computer Research and Modeling] 2(2):121–141.
 13. Yevin, I. A., T. F. Khabibullin, N. D. Shuvalov, and I. O. Vasyukov. 2012. Iskusstvo i teoriya slozhnykh setey [Art and theory of complex networks]. *Slozhnye sistemy* [Complex Systems] 3:38–50.
 14. Bak, P. 1997. *How nature works. The science of self-organized criticality*. Oxford: Oxford University Press. 62 p.
 15. Yevin, I. A., A. A. Koblyakov, D. V. Savricov, and N. D. Shuvalov. 2011. Kognitivnye seti [Cognitive networks]. *Komp'yuternye issledovaniya i modelirovaniye* [Computer Research and Modeling] 3(3):231–239.
 16. Badryzlov, V. A. 2014. Printsipy generatsii sluchaynykh grafov dlya modelirovaniya seti Internet [The principles of generation of random graphs for simulation of Internet]. *Omskiy nauchnyy vestnik* [Omsk Scientific Bulletin] 3(133):204–208.
 17. Komarova, T. K. 2002. *Psichologiya vnimaniya* [Psychology of attention]. Grodno, Belarus: Yanka Kupala State University of Grodno. 124 p.

Received March 15, 2017

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Volkov@amsd.com

ТАБС-ПРЕДСТАВЛЕНИЕ ЗАДАЧ И ЗАДАЧНЫХ ГРАФОВ

В. Д. Ильин¹

Аннотация: Рассмотрены основы построения и применения табс-структур для представления задачных конструктивных объектов (*s*-задач) и *s*-задачных графов. Табс, представляющий собой трехмерную многослойную таблицу со специальными типами клеток для формирования табс-структур, изучается как базовый элемент позиционной системы представления *s*-задач и *s*-задачных графов. Для каждого типа клеток табса определено семейство процедур. Табс-структура с заданными правилами навигации рассматривается как универсальное средство представления различных структур данных (массивов, списков и др.) и отношений порядка, принадлежности и др. Табс-структуры служат средством поддержки символического воплощения замыслов разработчиков систем программирования.

Ключевые слова: задачный конструктивный объект; *s*-задача; *s*-задачная область; *s*-задачный граф; табс; табс-структура; табс-представление *s*-задачи; табс-представление *s*-задачного графа

DOI: 10.14357/08696527170213

1 Введение

С начала 1970-х гг. в исследованиях, посвященных автоматизации разработки программ, пакетов программ и автоматизированных систем решения задач, не убывает актуальность проблемы представления знаний о программируемых задачах [1–29].

Изобретая трехмерные многослойные таблицы со специальными типами клеток (названные *табсами*) и *структуры табсов* (*табс-структуры*), автор стремился создать позиционную систему, удобную для записи, хранения, поиска и обработки формализованных описаний задач. Искомая позиционная система должна была быть удобной и для представления на ее основе структур данных, применяемых в программировании.

Табс-представление задач было предложено автором в 1986 г. при выполнении научного проекта, целью которого было создание *системы интерактивных генераторов приложений* (*системы ИГЕН*) [14].

Данная статья является тематическим продолжением [29], где рассматривается подход к решению проблемы создания *человеко-машинной среды решения задач* (*s-среды*) моделирования в науке и технике, автоматизированного проектирования, управления, дистанционного обучения и др.

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vdilyin@ipiran.ru

Цель статьи — охарактеризовать табсы и табс-структуры как эффективное средство для представления программируемых задач, задачных графов и *систем знаний о задачах* [17, 18, 21, 25–29].

Запись формул и выделение фрагментов текста

Для записи формул (без применения редакторов формул) используются средства комплекса *TSM*, предназначенного для текстового описания символьных моделей произвольных объектов (первая версия *TSM* была разработана и применена в 1989 г. [18]).

Для выделения определений, замечаний и примеров в этой статье используются следующие средства комплекса *TSM*:

- ⟨фрагмент описания⟩ □ ≈ утверждение (определение, аксиома и др.) (здесь и далее символ ≈ заменяет слово «означает»);
- ◊ ⟨фрагмент описания⟩ ◊ ≈ замечание;
- ⟨фрагмент описания⟩ ○ ≈ пример.

Курсивом выделены первые вхождения названий понятий и фрагменты описания, к которым автор хочет привлечь внимание.

2 Основные понятия

В *системе порождения программ* табс-структура используется как форма для размещения описания задачи, чтобы сделать удобным для человека и эффективным для выполнения программой поэтапный интерактивный процесс спецификации задачи. Правила поведения системы порождения при конструировании целевых систем определены как *правила табс-навигации* [18–24, 26, 27].

2.1 Задачный конструктивный объект

□ *Задачный конструктивный объект* (*s-задача*) — это четверка $\{\text{Formul}, \text{Rulsys}, \text{Alg}, \text{Prog}\}$, где Formul — постановка задачи; Rulsys — множество *систем обязательных и ориентирующих правил* решения задачи [30, 31], поставленных в соответствие Formul; Alg — объединение множеств алгоритмов, каждое из которых соответствует одному элементу из Rulsys; Prog — объединение множеств программ, каждое из которых поставлено в соответствие одному из элементов Alg.

Постановка задачи Formul — это пара $\{\text{Mem}, \text{Rel}\}$, где Mem — множество понятий задачи, на котором задано разбиение $\text{Mem} = \text{Inp} \vee \text{Out} (\text{Inp} \wedge \text{Out} = 0)$ и совокупность Rel связей между понятиями, определяющая бинарное отношение $\text{Rel} < \text{Inp}^*\text{Out}$.

◊ Математическому истолкованию отношения Rel посвящены статьи [22, 23]. ◊

Множество Mem называем также *памятью задачи*, а Inp и Out — ее *входом* и *выходом*, значения которых предполагается соответственно задавать и искать □.

Для каждого элемента из Rulsys, Alg и Prog задано описание применения.

Описания применения элементов Rulsys включают спецификацию *типа решателя задачи* (автономная s-машина, сетевая коопeração s-машин, коопेrация человек–s-машина и др.), требование к информационной безопасности и др.

Описания применения элементов из Alg включают данные о допустимых режимах работы решателя задачи (автоматический локальный, автоматический распределенный, интерактивный локальный и др.), о требованиях к полученному результату и др. Описания применения программ включают данные о языках реализации, операционных системах и др.

◊ Каждая программа сопровождается ссылками на наборы тестовых примеров [24, 27]. ◊

В общем случае множества Rulsys, Alg и Prog могут быть пустыми: число их элементов зависит от степени изученности задачи.

□ *S-алгоритм* — система правил решения задачи (соответствующая одному из элементов Rulsys), позволяющая за конечное число шагов поставить в однозначное соответствие заданному набору данных, принадлежащему Inp, результирующий набор, принадлежащий Out. □

□ *S-программа* — реализованный (на языке программирования высокого уровня, машинно-ориентированном языке и/или в системе машинных команд) s-алгоритм, представленный в форме сообщения, определяющего поведение s-машинного решателя задачи с заданными свойствами. Существует в символьном, кодовом и сигнальном воплощении, связанных отношениями трансляции. □

□ *Спецификация* spec *s-задачи* — это пара (Formul, as), где as — описание применения. □

◊ Далее префикс s- (в словах s-задача, s-задачный график и s-задачная область) опускаем. ◊

2.2 Задачная область и задачный график

□ *Задачная область* pA — это пара \langle память tem_A некоторого множества задач A \rangle , \langle семейство rul(tem_A) связей, заданных на tem_A \rangle .

При этом в A (содержащем не менее двух элементов) не существует ни одного элемента, который не был бы связан по памяти хотя бы с одним элементом из A. Непустое множество tem_A разбито на три подмножества: входов in_A задач, выходов out_A задач и подмножество og_A, каждый из элементов которого является и входом, и выходом некоторых задач (такие элементы памяти называются *обратимыми*, а og — подмножеством *обратимых элементов*). Любое одно из этих подмножеств может быть пустым; могут быть одновременно пустыми in_A и out_A. □

◊ В отличие от памяти задачи память задачной области содержит подмножество og элементов памяти, каждый из которых может быть или задан (как входной), или вычислен (как выходной). ◊

□ *Задачный граф* служит представлением *задачной области*, рассчитанным на реализацию процесса конструирования задач и формализацию знаний о задачах. Множество вершин графа (*задачный базис графа*) составлено из задачных конструктивных объектов. Каждая вершина графа имеет память. Память вершины — это память задачи (или задачной области), которую представляет вершина. Ребро задачного графа — это пара вершин с непустым пересечением по памяти. Нагрузка ребра определяется множеством всех пар элементов памяти, входящих в это пересечение [24, 27]. □

◊ Задачный граф служит *s-моделью* [25] задачной области рA [24], на которой в процессах поиска *разрешающих структур* интерпретируются спецификации искомых задач [17, 18, 27]. ◊

2.3 Табс

□ *Табс* — это трехмерная многослойная таблица (с вертикальными и горизонтальными слоями), имеющая два специальных типа клеток для связи с другими табсами. Первый тип — *указатель на табс* (*табс-указатель*), второй — *вложенный табс*. Значениями типа *табс-указатель* являются *табс-адреса*. Табс-адрес задается именем табса и тремя координатами: первая — номер строки, вторая — столбца и третья — слоя. Значениями типа *вложенный табс* являются табсы. Попадание в клетку типа *вложенный табс* делает доступным вложенный в эту клетку табс. □

Существование типа *вложенный табс* освобождает от необходимости во всех случаях именовать табсы.

○ Клетку с типом *вложенный табс* можно рассматривать как некоторый агрегат, представляющий фрагмент *памяти целевой системы*, а сам вложенный табс — толковать как декомпозицию упомянутого фрагмента. Если же в клетке, имеющей тип *вложенный табс*, размещено имя задачи, то это означает, что она относится к *составным задачам* (ее память и семейство связей между элементами памяти задаются описаниями, размещенными в клетках вложенного табса) [17, 18, 27]. ○

◊ Любой вложенный табс может, в свою очередь, иметь любое число клеток типа *вложенный табс*. Табс может содержать самого себя в любой из своих клеток типа *вложенный табс*. Нетрудно представить богатство возможностей табс-структурирования, предоставляемое типами *вложенный табс* и *табс-указатель*. ◊

Изменение состояния *табс-памяти* определяется множеством допустимых процедур.

Табсы позволяют экономно представлять соответствия. Комбинация графовых и реляционных возможностей делает табс-структуру удобной основой для представления знаний о задачах [17, 18, 27].

Такие свойства табсов, как наличие системы координат и типов клеток, позволяет строить сложные структуры табсов, фиксировать синтаксические пра-

вила разбора содержимого клеток и определять семантику данных, помещенных в клетки. ◇

Поведение табса во внешней среде задается множеством допустимых сообщений, известных интерпретатору сообщений табса.

2.4 Табс-структура и табс-база данных

□ *Табс-структура* — множество табсов, на котором задано семейство отношений, реализуемых посредством табс-клеток типа *табс-указатель* и *вложенный табс*. □

Табс-структура рассматривается как множество упорядоченных трехмерных таблиц, каждая из клеток которых имеет тип, а множество допустимых процедур — как такая их совокупность, где каждая процедура применима при определенном состоянии *памяти табса*, принадлежащем некоторому множеству состояний.

Табс-структура как форма для размещения в ней описания задачи или задачного графа, представляющего *задачную область*, воплощает позиционный принцип представления знаний о задачах [17, 18, 27].

Поддержка *системой управления табс-базами данных* (СУТБД) работы с различными структурами данных способствует эффективной реализации отношений в базе знаний о задачах.

Табс-ссылки применяются для выполнения переходов в *табс-базе данных* (ТБД) посредством изменения привязки каналов, реализующих в СУТБД доступ к ТБД. Ссылки могут быть *прямыми* и *косвенными*.

□ *Прямая ссылка* — это табс-адрес, к которому канал должен привязаться, если выполнен переход в клетку со ссылкой. □

□ *Косвенная ссылка* — описание маршрута, который каналу следует пройти, чтобы установить требуемую связь. □

Клетка, содержащая вложенный табс или прямую ссылку, интерпретируется независимо от указанного в канале каталога табсов, определяющего доступные табсы. Результат интерпретации косвенной ссылки зависит от текущего каталога: одна и та же косвенная ссылка может привести к установлению связи с разными табсами (в зависимости от того, какой каталог был определен в канале).

Возможность создания вложенного табса существует с момента определения типа клетки, в которой предполагается его разместить. Существование вложенного табса начинается в момент заполнения им клетки.

Табс, на который указывает косвенная ссылка, может не содержаться в ТБД в тот момент, когда ссылка заносится в клетку.

Табс, на который сделана ссылка (как и вложенный табс), при копировании не размножается.

В набор навигационных функций включены также функции, позволяющие входить в клетки типа *вложенный табс* и возвращаться обратно.

3 Табс-представление задач: основные положения

Каждый фрагмент спецификации задачи получает фиксированное место в табс-структуре. Координатное прикрепление фрагмента позволяет упростить работу со спецификацией в целом. Таким образом, представив спецификацию в виде табс-структуры данных, можно обрабатывать ее с помощью механизма доступа по заданным координатам.

□ *Табс-представление задачи* — ее формализованное описание с использованием табс-структур. □

Задачи в процессах *p-конкретизации*, *p-специализации*, *p-замены* и *p-конструирования* имеют различные табс-представления, каждому из которых соответствует определенная структура табсов и набор процедур, применение которых на ней допустимо [17, 18, 27].

Процесс *p-конкретизации* состоит из четырех этапов — *f*, *r*, *a* и *p*: на этапе *f* описывается формулировка задачи; на последующих этапах описываются системы обязательных и ориентирующих правил (этап *r*), алгоритмы (этап *a*) и программы (этап *p*).

Для представления *абстрактных задач* (*R-задач*) на этапе *f* используется специальная иерархическая структура табсов. Первый слой корневого табса содержит данные *паспорта задачи* (имя задачи и описание *точки зрения разработчика* (значения атрибутов who, rigr и stage [18, 27]); имя паспорта и его атрибуты размещены в разных клетках). Все клетки (кроме одной) имеют тип *текст* (одна — тип *табс-указатель*: в ней содержится ссылка на табс с формой для описания памяти задачи). В клетке второго слоя табса (с теми же двумя первыми координатами) помещается ссылка на табс, хранящий форму для описания отношения *Formul*, заданного на памяти рассматриваемой задачи.

◊ Координаты табс-клеток связаны с назначением содержимого, записанного в клетки: ○ в клетке с координатами (1, 1, 1) — всегда *имя паспорта*, а в клетке (5, 1, 1) — ссылка на табс, содержащий форму для описания *памяти задачи* ○. ◊

Корневой табс имеет следующие клетки: *rasp_{cnc}|name* (паспорт задачи); *spgen* (принадлежность к пространству по *родовому признаку*); *sp_{us}* (принадлежность к пространству по *применению*; для абстрактных задач *sp_{us}* = *abstr*); *struc* (тип задачной структуры); *{form}* (форма описания).

В клетки типа *текст* записывается имя паспорта и его атрибуты (кроме *form*).

○ Клетка *struc* может содержать *unit* (*простая задача*), а клетка *us* — *abstr* (*абстрактная*). ○

В клетках *inp* и *out* размещаются описатели данных входа и выхода задачи. Каждая из этих клеток имеет тип *табс-указатель* и адресует систему к табсу, связанному с корневым.

Отношение *Formul* записывается в табс, на который есть ссылка в клетке типа *указатель* второго слоя корневого табса, содержащего имя паспорта задачи и значения его атрибутов.

◊ Зная соответствие между координатами клеток и их содержимым, а также типы клеток (которые определяют и синтаксические правила их заполнения), система порождения программ работает с табс-представлением задачи как с конструктивным объектом [17, 18, 27]. ◊

◊ Клетки типа *текст* второго слоя табсов первого уровня содержат поясняющие сведения для каждой из клеток первого слоя. Эти сведения не подлежат обработке интерпретатором: они используются при обучении и поддержке процессов спецификации. Аналогичную поддержку имеют вложенные табсы и те табсы, на которые сделаны ссылки. ◊

На этапе r формируются табс-представления систем обязательных и ориентирующих правил.

□ *Прообразом задачи* с отношением $\text{rel} = \text{set}(\text{formul}|\text{Rulsys})$ является задача, для которой $\text{rel} = \text{Formul}$ и табс-представление которой предшествует табс-представлению задачи на этапе r (поскольку *образ* наследует значения атрибутов прообраза, то в табс-представлении задачи на этапе r присутствуют указатели на ранее созданные табсы). □

В клетках первого слоя корневого табса содержится имя паспорта задачи (с пометой «*cnc; r*»), указатель на табс, где описана задача с тем же именем и пометой «*cnc; f*», и указатель на табс, содержащий описание систем правил, на основе которых разрабатываются алгоритмы. Первый указатель делает определенными атрибуты паспорта задачи и описание ее памяти. Второй — табс, где описаны отношения $\text{set}(\text{formul}|\text{Rulsys})$ (здесь же записаны и данные об условиях применения).

На последующих этапах (a и p) создаются табс-представления *алгоритмов* и *программ*.

◊ Программы представлены исходными текстами на выбранных языках реализации. ◊

В процессе *r-замены* понятиям, введенным при *r-конкретизации*, ставятся в соответствие понятия, заменяющие их (изменяется описание памяти: добавляется табс-слой, клетки которого заполняются новыми именами переменных). Каждая полученная в результате *r-замены* задача имеет свой паспорт.

Табс-представление *составной задачи* (как конструкции из некоторого множества простых и ранее определенных составных задач) включает корневой табс, в клетках первого слоя которого записано наименование задачи и указатели на табсы, которые содержат описания задач, входящих в состав конструкции. Строки и столбцы первого слоя корневого табса используются для задания структуры составной задачи. Если связь по памяти имеет тип *выход-вход*, то ссылка на первую из задач будет расположена в строке с номером, на единицу меньшим номера строки со ссылкой на вторую задачу; если же тип *вход-вход*, то ссылки помещаются в одной строке. Аналогичный принцип позиционного размещения используется, когда задача имеет связь по памяти типа *вход-выход* и *выход-выход*.

Связь по элементам памяти для каждой пары задач устанавливается с помощью дополнительного слоя табса (одного на каждую пару). Эти слои раз-

мешаются вслед за слоем с описанием памяти задачи. В каждой из клеток перечисляются эквивалентируемые элементы памяти связываемых задач (путем указания соответствующих имен табсов и координат клеток).

Процесс *p-специализации* приводит к образованию новой табс-структурь на основе структуры, сформированной на этапе *f*. Работа выполняется на копии табс-структурь, представляющей специализируемую задачу. При работе с копией изменяется содержимое табсов, хранящих исходное отношение *Formul*. Если специализируются параметры, то интервалы их изменения записываются в клетки слоя табса, размещаемого вслед за слоем, содержащим описание памяти задачи. Если выполняется специализация условий, то дополнительные ограничения записываются в слое табса, где находится описание отношения *Formul*.

3.1 Табс-представление прикладных задач

Множество *R*-задач — основа для решения различных *прикладных задач* (*P*-задач). Для применения *R*-задач в конкретных приложениях существует механизм создания так называемых *P-понятийных оболочек*, позволяющий установить соответствие между *R-понятиями* и понятиями *P*-задач конкретной области применения (*P-понятиями*). В итоге на основе *R*-задач и созданной *P*-понятийной оболочки формируется множество *P*-задач.

В третьем слое (соответствий), определяющем *R*-задачу, символами «вх» помечаются те клетки, первые две координаты которых равны соответствующим координатам входных переменных в *E*-памяти *R*-задачи. Аналогично, символами «вых» помечаются клетки, проекция которых на *E*-память соответствует выходным переменным *R*-задачи.

Далее располагается совокупность табс-слоев, относящихся к *P*-задачам. Для каждой *P*-задачи отводятся три слоя: *данных, наименований переменных, условий активности*. Вместе со слоем соответствий *R*-задачи и слоем *E*-памяти они полностью определяют *P*-задачу. Затем идет слой соответствий *R*-задачи, за ним — слои, относящиеся к *P*-задачам, решаемым с помощью *R*-задачи, и так далее — до исчерпания описаний всех *R*- и *P*-задач.

Исходным материалом для табс-представления *P*-задач служат табс-представления *R*-задач, полученные на этапе *f* процесса *p-конкретизации*. При переходе к табс-представлению *P*-задач следует заменить на *appl* значение *abstr* атрибута *sprus*. Затем изменить структуру табсов так, как это делается при *p-замене*. В результате элементы памяти задач получают новые имена, отражающие прикладную суть задач. При специализации *P*-задач и создании на их основе задачных конструкций действуют те же (что и для *R*-задач) правила табс-представления.

□ Табс, содержащий описание *R*-задачи, является *табсом-прообразом* по отношению к табсу, содержащему описание *P*-задачи, полученному на основе описания этой *R*-задачи. □

Один и тот же табс может быть табсом-прообразом конечного множества *табсов-образов*.

□ Табс с пустым множеством правил получения образов называется *конечным табсом*. □

Табсы-образы, наследуя характеристики своего табса-прообраза, образуют *родовое табс-дерево*, корнем которого является табс, не имеющий прообраза, а листьями — конечные табсы.

3.2 Представление задач в генераторе ГЕНПАК и ГЕНПАК-пакетах

Предусмотрены табсы двух типов: *табс генератора* и *табс пакета*. Табс генератора является табсом-прообразом. Он содержит описания R -задач и шаблоны для описания P -задач. Табс пакета содержит описание структуры *пользовательских табсов (U -табсов)*, т. е. табсов, создаваемых для фиксированных значений параметров P -задач (см. рисунок). Порождаемые на основе табса генератора табсы пакета являются табсами-образами [17, 18, 27].

Первый слой U -табса — *слой данных (E-память)*, второй отводится для наименований переменных R -задач.



○ Приведем примеры формулировок нескольких R -задач.

Отыскание оптимального плана.

Найти максимум прибыли $\text{sum}_{j=1,\dots,n}(c_j * x_j)$ при ограничениях на ресурсы $\text{sum}_{j=1,\dots,n}(a_{i,j} * x_j) \leq b_i$, $m > n$, $x_j \geq 0$, $j = 1, \dots, n$; $b_i > 0$, $i = 1, \dots, m$.

Расчет остатков ресурсов.

Определить остатки ресурсов $k_i = b_i - \text{sum}_{j=1,\dots,n}(a_{i,j} * x'_{j,i=1,\dots,m})$ после реализации оптимального плана $x' = (x'_1, \dots, x'_n)$, обеспечивающего экстремум целевой функции $\text{sum}_{j=1,\dots,n}(c_j * x_j)$ при ограничениях $\text{sum}_{j=1,\dots,n}(a_{i,j} * x_j) \leq b_i$, $m > n$, $x_j \geq 0$, $j = 1, \dots, n$; $b_i > 0$, $i = 1, \dots, m$.

Расчет таблицы оптимальных расходов ресурсов.

Определить элементы $R_{i,j} = a_{i,j} * x'_j$, $i = 1, \dots, m$, $j = 1, \dots, n$, соответствующие j -м компонентам x'_j оптимального плана $x' = (x'_1, \dots, x'_n)$, обеспечивающего максимум прибыли $\text{sum}_{j=1,\dots,n}(c_j * x_j)$ при ограничениях $\text{sum}_{j=1,\dots,n}(a_{i,j} * x_j) \leq b_i$, $m > n$.

Расчет отклонений вариантного значения плана от оптимального.

Найти величину $S = \text{sum}_{j=1,\dots,n}(c_j * (x'_j - x_j))$ как разность значений прибыли для оптимального плана $x' = (x'_1, \dots, x'_n)$ и вариантного плана $x = (x_1, \dots, x_n)$ при тех же ограничениях.

Расчет значения целевой функции.

Найти значение прибыли $\text{sum}_{j=1,\dots,n}(c_j * x_j)$, если x_1, \dots, x_n удовлетворяют ограничениям $\text{sum}_{j=1,\dots,n}(a_{i,j} * x_j) \leq b_i$, $m > n$, $x_j \geq 0$, $j = 1, \dots, n$; $b_i > 0$, $i = 1, \dots, m$.

Каждой R -задаче соответствует множество понятий. Каждое понятие определяется множеством переменных, число элементов которого зависит от параметров R -задачи. Максимальные значения параметров определяются при порождении пакета, а конкретные их значения — при эксплуатации пакета.

В сформулированных R -задачах используются следующие понятия: *максимальная прибыль; удельные расходы ресурсов; запасы ресурсов; коэффициенты целевой функции; вариантный план; оптимальный план; значение прибыли, соответствующее варианту плану.*

Множество понятий каждой j -й задачи разбивается на два непересекающихся подмножества $T_{v;j}$ и $T_{w;j}$. Элементами $T_{v;j}$ являются понятия, которые определяются входными переменными, а элементами $T_{w;j}$ — понятия, определяемые выходными переменными. ○

○ Примеры абстрактных и построенных на их основе прикладных задач рассмотрены в [30–35]. В частности, прикладная задача, реализованная в действующем интернет-сервисе планирования расходов [33], созданном А. В. Ильиным, в качестве прообраза имеет абстрактную задачу, рассмотренную в [32]. ○

4 Заключение

Повышение продуктивности разработки программ, достигаемое за счет накопленных знаний о задачах, тем значительнее, чем совершеннее механизмы

классификации, поиска и извлечения необходимых знаний. Методология построения и применения табс-структур для представления задач и задачных графов может служить эффективным основанием для построения систем знаний о программируемых задачах [28, 29].

Литература

1. *Manna Z., Waldinger R.* Towards automatic program synthesis // Commun. ACM, 1971. Vol. 14. No. 3. P. 151–164.
2. *Jackson M. A.* Principles of program design. — London: Academic Press, 1975. 299 p.
3. *Еришев А. П., Ильин В. П.* Пакеты программ — технология решения прикладных задач. — Новосибирск: ВЦ СО АН СССР, 1978. Препринт. № 121. 30 с.
4. *Кахро М. И., Калья А. П., Тыгуз Э. Х.* Инstrumentальная система программирования ЕС ЭВМ (ПРИЗ). — М.: Финансы и статистика, 1981. 181 с.
5. *Broy M.* Program construction by transformation: A family tree of sorting programs // Computer program synthesis methodologies / Eds. A. W. Biermann, G. Guiho. — NATO advanced study institutes ser. C. — Dordrecht–Boston–London: D. Reidel Publishing Co., 1981. Vol. 95. P. 1–49.
6. *Онарин Г. А.* Сатурн — метасистема для построения пакетов прикладных программ // Разработка пакетов прикладных программ. — Новосибирск: Наука, 1982. С. 130–160.
7. *Martin J.* Application development without programmers. — Englewood Cliffs, NJ, USA: Prentice Hall, 1982. 350 p.
8. *Тыгуз Э. Х.* Концептуальное программирование. — М.: Наука, 1984. 256 с.
9. *Bassett P.* Design principles for software manufacturing tools // Fifth Generation Challenge: ACM'84 Annual Conference Proceedings / Eds. R. L. Muller, J. J. Pottmyer. — New York, NY, USA: ACM, 1984. P. 85–93.
10. *Ramsay A.* The POPLOG Program Development System // Microprocess. Microsys., 1984. Vol. 8. No. 7. P. 368–373.
11. *Horowitz E., Kemper A., Narasimhan B.* A survey of application generators // IEEE Software, 1985. Vol. 2. No. 1. P. 40–54.
12. *Luker P. A., Burns A.* Program generators and generation software // Comput. J., 1986. Vol. 29. No. 4. P. 315–321.
13. *Polster F. J.* Reuse of software through generation of partial systems // IEEE Trans. Software Eng., 1986. Vol. 12. No. 3. P. 402–416.
14. *Ильин В. Д.* Система ИГЕН. Концепция, архитектура, технология программирования // Современные средства информатики. — М.: Наука, 1986. С. 117–125.
15. *Бабаев И. О., Лавров С. С.* Развитие автоматизированной системы решения задач СПОРА // Пакеты прикладных программ. Инструментальные системы. — М.: Наука, 1987. С. 5–17.
16. *Ильин В. Д.* Генератор прикладных задач в системе ИГЕН // ЭВМ массового применения. — М.: Наука, 1987. С. 28–37.
17. *Ильин В. Д.* Представление знаний о задачах в системе порождения программ. — М.: Институт проблем информатики АН СССР, 1989. Препринт. 50 с.
18. *Ильин В. Д.* Система порождения программ. — М.: Наука, 1989. 264 с. doi: 10.13140/RG.2.1.3150.6967.

19. Ильин В. Д. Порождение пакетов программ // Системы и средства информатики. — М.: Наука, 1989. С. 32–65.
20. Ильин В. Д. Порождение целевых программных систем: элементы теории // Системы и средства информатики. — М.: Наука, 1992. Т. 2. С. 3–44.
21. Ilyin V. D. A methodology for knowledge based engineering of parallel program systems // 8th Conference (International) on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems Proceedings. — Newark, NJ, USA: Gordon and Breach Science Publs., Inc., 1995. P. 805–809.
22. Gavrilenko Y. V., Ilyin V. D. Spaces of local objects for semantic interpretation of problems. I // J. Computer Systems Sciences International, 1996. Vol. 35. No. 5. P. 687–694.
23. Gavrilenko Y. V., Ilyin V. D. Calculus of task-constructive objects and their interpretations. III // J. Computer Systems Sciences International, 1997. Vol. 36. No. 5. P. 704–712.
24. Ильин А. В. Конструирование разрешающих структур на задачных графах системы знаний о программируемых задачах // Информационные технологии и вычислительные системы, 2007. № 3. С. 30–36.
25. Ильин А. В., Ильин В. Д. Основы теории s-моделирования. — М.: ИПИ РАН, 2009. 143 с. doi: 10.13140/2.1.3062.8800.
26. Ильин А. В., Ильин В. Д. S-моделирование задач и конструирование программ. — М.: ИПИ РАН, 2012. 146 с. doi: 10.13140/2.1.2800.7361.
27. Ильин В. Д. Система порождения программ. Версия 2013 г. — М.: ИПИ РАН, 2013. 142 с. doi: 10.13140/2.1.1752.1605.
28. Ильин А. В., Ильин В. Д. Систематизация знаний о программируемых задачах // Системы и средства информатики, 2014. Т. 26. № 4. С. 149–161. doi: 10.14357/08696527140314.
29. Ильин А. В., Ильин В. Д. Создание человека-машинной среды решения задач // Системы и средства информатики, 2016. Т. 24. № 3. С. 192–203. doi: 10.14357/08696527160413.
30. Ильин А. В. Экспертное планирование ресурсов. — М.: ИПИ РАН, 2013. 58 с. <http://elibrary.ru/item.asp?id=27662070>.
31. Ilyin A. V., Ilyin V. D. The technology of interactive resource allocation in accordance with the customizable system of rules // Appl. Math. Sci., 2013. Vol. 7. No. 143. P. 7105–7111. doi: 10.12988/ams.2013.311649.
32. Ilyin A. V., Ilyin V. D. The interval method of cost planning and its implementation in the Online Service // Contemp. Eng. Sci., 2014. Vol. 7. No. 20. P. 931–938. doi: 10.12988/ces.2014.48114.
33. Ильин А. В. Интернет-сервис планирования расходов // Системы и средства информатики, 2015. Т. 25. № 2. С. 111–122. doi: 10.14357/08696527150207.
34. Ilyin A. V., Ilyin V. D. Interval planning the supplies of scarce product // Contemp. Eng. Sci., 2015. Vol. 8. No. 31. P. 1495–1498. doi: 10.12988/ces.2015.59263.
35. Ilyin A. V., Ilyin V. D. Variational online budgeting taking into account the priorities of expense items // Agris on-line Papers in Economics and Informatics, 2016. Vol. 8. No. 3. P. 51–56. doi: 10.7160/aol.2016.080305.

Поступила в редакцию 12.03.17

TABS-REPRESENTATION OF TASKS AND TASK GRAPHS

V. D. Ilyin

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article describes the basics of construction and application of tabs-structures for representation of task constructive objects (named s-tasks) and s-task graphs. A tabs, representing a three-dimensional multilayered table with special cell types for the formation of tabs-structures, is studied as a basic element of a positional system for representation of s-tasks and s-task graphs. The collection of procedures is defined for each type of tabs cells. A tabs-structure with prescribed rules of navigation is considered as a universal means of representation of various data structures (arrays, lists, etc.) and relations of order, belonging, etc. Tabs-structures serve as support of symbolic embodiment of ideas of developers of programming systems.

Keywords: task constructive object; s-task; s-task area; s-task graph; tabs; tabs-structure; tabs-representation of s-task; tabs-representation of s-task graph

DOI: 10.14357/08696527170213

References

1. Manna, Z., and R. Waldinger. 1971. Towards automatic program synthesis. *Commun. ACM* 14(3):151–164.
2. Jackson, M. A. 1975. *Principles of program design*. London: Academic Press. 299 p.
3. Ershov, A. P., and V. P. Il'in. 1978. *Pakety programm — tekhnologiya resheniya prikladnykh zadach* [Software packages — a technology for solving applied problems]. Novosibirsk: CC SB AS USSR. Preprint No. 121. 30 p.
4. Kakhro, M. I., A. P. Kal'ya, and E. Kh. Tyugu. 1981. *Instrumental'naya sistema programmirovaniya ES EVM (PRIZ)* [The instrumental programming system ES computer (PRIZ)]. Moscow: Finance and Statistics. 181 p.
5. Broy, M. 1981. Program construction by transformation: A family tree of sorting programs. *Computer program synthesis methodologies*. Eds. A. W. Biermann and G. Gui-ho. NATO advanced study institutes ser. C. Dordrecht–Boston–London: D. Reidel Publishing Co. 95:1–49.
6. Oparin, G. A. 1982. Saturn — metasistema dlya postroeniya paketov prikladnykh programm [Saturn — the metasystem for software packages development]. *Razrabotka paketov prikladnykh programm* [Development of application packages]. Novosibirsk: Nauka. 130–160.
7. Martin, J. 1982. *Application development without programmers*. Englewood Cliffs, NJ: Prentice Hall. 350 p.
8. Tyugu, E. Kh. 1984. *Konseptual'noe programmirovaniye* [Conceptual programming]. Moscow: Nauka 256 p.

9. Bassett, P. 1984. Design principles for software manufacturing tools. *Fifth Generation Challenge: ACM'84 Annual Conference Proceedings*. Eds. R. L. Muller and J. J. Pottmyer. New York, NY: ACM. 85–93.
10. Ramsay, A. 1984. The POPLOG program development system. *Microprocess. Microsy.* 8(7):368–373.
11. Horowitz, E., A. Kemper, and B. Narasimhan. 1985. A survey of application generators. *IEEE Software* 2(1):40–54.
12. Luker, P. A., and A. Burns. 1986. Program generators and generation software. *Comput. J.* 29(4):315–321.
13. Polster, F. J. 1986. Reuse of software through generation of partial systems. *IEEE Trans. Software Eng.* 2(3):402–416.
14. Ilyin, V. D. 1986. Sistema IGEN. Kontseptsiya, arkhitektura, tekhnologiya programmirovaniya [The IGEN system. Concept, architecture, programming technology]. *Sovremennye sredstva informatiki* [Modern Informatics Means]. Moscow: Nauka. 117–125.
15. Babaev, I. O., and S. S. Lavrov. 1987. Razvitiye avtomatizirovannoy sistemy resheniya zadach SPORA [Development of the automated system SPORA for solving problems]. *Pakety prikladnykh programm. Instrumental'nye sistemy* [Application packages. Instrumental systems]. Moscow: Nauka. 5–17.
16. Ilyin, V. D. 1987. Generator prikladnykh zadach v sisteme IGEN [The generator of applied tasks in the IGEN system]. *EVM massovogo primeneniya* [Mass computers]. Moscow: Nauka. 28–37.
17. Ilyin, V. D. 1989. *Predstavlenie znaniy o zadachakh v sisteme porozhdeniya programm* [Representation of knowledge about tasks in the system of programs generating]. Moscow: Institute of Informatics Problems of the Academy of Sciences of the USSR. Preprint. 50 p.
18. Ilyin, V. D. 1989. *Sistema porozhdeniya programm* [The system of programs generating]. Moscow: Nauka. 264 p. doi: 10.13140/RG.2.1.3150.6967.
19. Ilyin, V. D. 1989. Porozhdenie paketov programm [Generating software packages]. *Sistemy i Sredstva Informatiki* [Systems and Means of Informatics]. Moscow: Nauka. 32–65.
20. Ilyin, V. D. 1992. Porozhdenie tselevykh programmnykh sistem: Elementy teorii [Generating software with specified functionality and properties: The theory elements]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Moscow: Nauka. 2:3–44.
21. Ilyin, V. D. 1995. A methodology for knowledge based engineering of parallel program systems. *8th Conference (International) on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems Proceedings*. Newark, NJ: Gordon and Breach Science Publs., Inc. 805–809.
22. Gavrilenko, Y. V., and V. D. Ilyin. 1996. Spaces of local objects for semantic interpretation of problems. *I. J. Computer Systems Sciences International* 35(5):687–694.
23. Gavrilenko, Y. V., and V. D. Ilyin. 1997. Calculus of task-constructive objects and their interpretations. *III. J. Computer Systems Sciences International* 36(5):704–712.
24. Ilyin, A. V. 2007. Konstruirovaniye razreshayushchikh struktur na zadachnykh grafakh sistemy znaniy o programmiruemykh zadachakh [Construction of solving structures in task graphs of knowledge system of programmable tasks]. *Informatsionnye tekhnologii i vychislitel'nye sistemy* [Information Technology Computer Systems] 3:30–36.

25. Ilyin, A. V., and V. D. Ilyin. 2009. *Osnovy teorii s-modelirovaniya* [Basics of the theory of S-modeling]. Moscow: Institute of Informatics Problems of Russian Academy of Sciences. 143 p. doi: 10.13140/2.1.3062.8800.
26. Ilyin, A. V., and V. D. Ilyin. 2012. *S-modelirovaniye zadach i konstruirovaniye programm* [S-modeling of tasks and construction of programs]. Moscow: Institute of Informatics Problems of the Russian Academy of Sciences. 148 p. doi: 10.13140/2.1.2800.7361.
27. Ilyin, V. D. 2013. *Sistema porozhdeniya programm. Versiya 2013 g.* [The system of programs generating. Version 2013]. Moscow: Nauka. 142 p. doi: 10.13140/2.1.1752.1605.
28. Ilyin, A. V., and V. D. Ilyin. 2014. Sistematisatsiya znaniy o programmireuemnykh zadachakh [Systematization of knowledge about programmable tasks]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(3):192–203. doi: 10.14357/08696527140314.
29. Ilyin, A. V., and V. D. Ilyin. 2016. Sozdanie cheloveko-mashinnoy sredy resheniya zadach [Creation of a human-machine environment for problem solving]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(4):149–161. doi: 10.14357/08696527160413.
30. Ilyin, A. V. 2013. *Eksperimentnoe planirovanie resursov* [Expert resource planning]. Moscow: Institute of Informatics Problems of the Russian Academy of Sciences. 50 p. Available at: <http://elibrary.ru/item.asp?id=27662070> (accessed March 11, 2017).
31. Ilyin, A. V., and V. D. Ilyin. 2013. The technology of interactive resource allocation in accordance with the customizable system of rules. *Appl. Math. Sci.* 7(143):7105–7111. doi: 10.12988/ams.2013.311649.
32. Ilyin, A. V., and V. D. Ilyin. 2014. The interval method of cost planning and its implementation in the Online Service. *Contemp. Eng. Sci.* 7(20):931–938. doi: 10.12988/ces.2014.48114.
33. Ilyin, A. V. 2015. Internet-servis planirovaniya raskhodov [The online service for cost planning]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):111–122. doi: 10.14357/08696527150207.
34. Ilyin, A. V., and V. D. Ilyin. 2015. Interval planning the supplies of scarce product. *Contemp. Eng. Sci.* 8(31):1495–1498. doi: 10.12988/ces.2015.59263.
35. Ilyin, A. V., and V. D. Ilyin. 2016. Variational online budgeting taking into account the priorities of expense items. *Agris on-line Papers in Economics and Informatics* 8(3):51–56. doi: 10.7160/aol.2016.080305.

Received March 12, 2017

Contributor

Ilyin Vladimir D. (b. 1937) — Doctor of Science in technology, Professor, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vdilyin@yandex.ru

О Б А В Т О Р А Х

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Волков Олег Игоревич (р. 1964) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Головин Сергей Анатольевич (р. 1950) — доктор технических наук, профессор, заведующий кафедрой Московского технологического университета (МИРЭА)

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, доцент; ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Московского технологического университета (МИРЭА)

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Данилович Евгений Сергеевич (р. 1996) — студент Московского технологического университета (МИРЭА)

Забежайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ильин Владимир Дмитриевич (р. 1937) — доктор технических наук, профессор, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Козлов Сергей Витальевич (р. 1955) — кандидат технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Лебедев Алексей Викторович (р. 1971) — доктор физико-математических наук, доцент кафедры теории вероятностей механико-математического факультета Московского государственного университета им. М. В. Ломоносова

Мамонова Оксана Сергеевна (р. 1998) — студент факультета иностранных языков и регионоведения Московского государственного университета им. М. В. Ломоносова

Мартынов Евгений Александрович (р. 1991) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Николаев Андрей Владимирович (р. 1973) — кандидат физико-математических наук, старший научный сотрудник Института химической физики им. Н. Н. Семёнова Российской академии наук (ИХФ РАН)

Писковский Виктор Олегович (р. 1963) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Плеханов Леонид Петрович (р. 1943) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сучков Александр Павлович (р. 1954) — доктор технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хромов Дмитрий Романович (р. 1998) — студент Московского технологического университета (МИРЭА)

Щурова Анна Юрьевна (р. 1991) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанныго документа).

Редакция вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редакция может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редактуре должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или L^AT_EX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине. Рекомендуемый объем рукописи — не выше 15 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете:

<http://www.ipiran.ru/publications/collected/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием города и страны и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2012_22_02_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2012_22_02_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала.
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты...» следует выбрать опцию BNG.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание материалов конференций:

Usmanov, T.S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primenением hidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol’zovaniya i povышeniya neftegazoootdachi”* [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonants, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticeskogo slovarya informacionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoj samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)

<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Systems and Means of Informatics” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.”

Author(s) signature(s): (name(s), address(es), date).”

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents

may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font —Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 15 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site:

<http://www.ipiran.ru/publication/collected/template.doc>

8. Articles should enclose data both in **Russian and English**:

- title;
 - author's name and surname;
 - affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
 - data on authors according to the format (see site):
http://www.ipiran.ru/journal/collected/2012_22_02_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2012_22_02_eng/authors.asp;
 - abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae.
 - Indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences.
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and in Latin transliteration.
- Please take into account the following examples of Russian references appearance:
- Article in journal:**
- Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S10231935080077.
- Journal article in electronic format:**
- Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic

exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i pozyscheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonants, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticeskogo slovarya informacionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoj samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
 - the journal title and author's name in the "Subject" field;
 - an article and additional materials have to be attached using the "attach" function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "System and Means of Informatics" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499)135 86 92, Fax: +7 (495)930 45 05

e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)

http://www.ipiran.ru/english/journal_systems.asp