

Разработка модели прогнозирования производительности информационных систем

Цель исследования: разработать модель прогнозирования производительности информационных систем как механизма предварительной оценки эффективности информационной системы до начала финансирования проекта информационной системы.

Материалы и методы: в качестве исходного материала использованы результаты изучения параметров статистической структуры дефектов обработки данных информационной системы. Применены метод кластерного анализа, метод регрессионного анализа.

Результаты: с целью снижения финансовых рисков заказчики информационных систем стараются принимать решения на основе предварительных расчетов по эффективности будущих информационных систем. Однако ориентировочные расчеты по технико-экономическому обоснованию проекта можно получить тогда, когда финансирование проектных работ уже открыто. Посредством модели прогнозирования производительности информационной системы можно выполнить ее оценку до начала проектирования. Модель разрабатывается методом регрессионного анализа в виде множественной линейной регрессии. В уравнении регрессии про-

гнозируемой переменной является значение производительности информационной системы. Прогнозирующими переменными являются значения дефектов обработки данных по классам достоверности, полноты и своевременности. Измерение и оценка параметров статистической структуры дефектов выполнены посредством программ кластер-анализа и регрессионного анализа. Проведены расчеты по определению фактических и прогнозных значений производительности информационной системы.

Заключение: в плане реализации модели были проведены работы по исследованию информационных систем, разработке модели прогнозирования производительности информационной системы. Проведенные экспериментальные работы показали адекватность модели. Модель реализована в комплексе задач по проектированию информационных систем в сфере образования и промышленности.

Ключевые слова: моделирование, прогнозирование, производительность информационных систем, измерение и оценка параметров, риски финансовых ресурсов.

George N. Isaev

The Russian state university of tourism and service, Moscow, Russia

Developing a model of forecasting information systems performance

Research aim: to develop a model to forecast the performance of information systems as a mechanism for preliminary assessment of the information system effectiveness before the beginning of financing the information system project.

Materials and methods: the starting material used the results of studying the parameters of the statistical structure of information system data processing defects. Methods of cluster analysis and regression analysis were applied.

Results: in order to reduce financial risks, information systems customers try to make decisions on the basis of preliminary calculations on the effectiveness of future information systems. However, the assumptions on techno-economic justification of the project can only be obtained when the funding for design work is already open. Its evaluation can be done before starting the project development using a model of forecasting information system performance. The model is developed using regression analysis in the

form of a multiple linear regression. The value of information system performance is the predicted variable in the regression equation. The values of data processing defects in the classes of accuracy, completeness and timeliness are the forecast variables. Measurement and evaluation of parameters of the statistical structure of defects were done through programmes of cluster analysis and regression analysis. The calculations for determining the actual and forecast values of the information system performance were conducted.

Conclusion: in terms of implementing the model, a research of information systems was carried out, as well as the development of forecasting model of information system performance. The conducted experimental work showed the adequacy of the model. The model is implemented in the complex task of designing information systems in education and industry.

Keywords: modeling, forecasting, information systems performance, measurement and evaluation of options, risks of financial resources.

Введение

В сфере образования, как и в других отраслях проводятся работы по автоматизации или модернизации информационных систем (ИС). Решение задач по совершенствованию ИС существенно улучшают качество учебного процесса и управления. Эти задачи в большинстве случаев требуют расходования значительных объ-

емов ресурсов — временных, трудовых, материальных, финансовых [1]. В создании ИС, как и в любом проекте, всегда имеется риск неоправданных затрат ресурсов. Финансирование проектов сравнительно малоэффективных ИС осуществляется, как правило, по правилу «финансирование с полным регрессом на заемщика» [2]. Здесь основные риски остаются у заемщика, то есть

у заказчика будущей ИС. Поэтому вполне естественно, что заказчик ИС старается предъявить повышенные требования по таким показателям как то минимизации срока окупаемости проекта ИС. Срок окупаемости в значительной мере зависит от уровня функциональной и экономической эффективности будущей ИС. Следует отметить, что значения показателей эконо-

мической эффективности в значительной мере определяются значениями показателей функциональной эффективности ИС. Так, например, один из существенных показателей экономической эффективности как-то себестоимость подготовки документа в технологическом процессе обработки данных ИС зависит от производительности указанного процесса [3]. При условии достаточно точного расчета производительности будущей ИС можно будет определить срок окупаемости инвестиционного проекта ИС. Реализация подобного условия дает заказчику возможность понизить риски и определенную уверенность в принятии решения на инвестирование проекта ИС.

Следует отметить, что в нормативных документах имеются требования по предварительным расчетам технико-экономического обоснования создания ИС на этапах проектирования – обследование объекта автоматизации, составление технического задания на создание ИС, разработка технического проекта ИС [4]. Однако существующая практика технико-экономического обоснования имеет существенные недостатки:

Работы по технико-экономическому обоснованию проекта выполняются на этапах проектирования ИС уже после открытия финансирования.

В лучшем случае финансирование может быть приостановлено только после окончания этапа обследования, когда проектировщик представляет заказчику отчет, в котором должно быть предварительное технико-экономическое обоснование проекта.

Как правило, результаты расчетов по технико-экономическому обоснованию носят ориентировочный характер.

Сбор и обработка сведений и последующие измерения параметров обследуемой ИС в силу обширности и сложности задач проектирования проводится не всегда корректно. По этому условию трудно, а

иногда и невозможно принять решение о целесообразности продолжения финансирования проекта на ИС.

Проектировщик заинтересован в финансировании проекта и, как правило, результаты расчетов подтверждают положительную эффективность проекта ИС.

Наиболее приемлемым подходом принятия адекватного решения о целесообразности инвестирования проекта ИС до начала проектных работ представляется получение точных расчетов по производительности будущей ИС. Зная прогноз производительности ИС можно достаточно точно определить и экономические показатели будущей ИС. Это могут быть, например, показатели себестоимости подготовки документа, годовой экономии, экономической эффективности, срока окупаемости затрат на ИС и др. [3]. При данном условии возникает необходимость разработки модели прогнозирования производительности ИС.

1. Разработка модели прогноза производительности информационной системы

Анализ технологии обработки данных показывает, что имеется четкая зависимость значения показателя производительности ИС от интенсивности возникновения дефектов, возникающих на различных этапах и участках технологического процесса обработки данных ИС [3]. Предварительные исследования, проведенные с применением метода кластер-анализа показали, что основными классами дефектов в технологии обработки данных ИС являются дефекты по достоверности, полноте и своевременности [5]. Дефектами могут быть различные сбои, отказы различных компонентов ИС. Так, например, это простои в работе средств вычислительной техники, средств передачи данных, простои по вине исполнителей, запаздывание

в передаче документов на обработку, ошибки (искажения) алфавитной и цифровой нагрузки реквизитов документов, пропуски показателей в документах и др. Дефекты имеют различный характер и обуславливают необходимость затрат технологического времени на свое обнаружение и исправление. Куммулятивное время выявления и корректировки дефектов составляет значительный объем и существенно снижает значение показателя производительности ИС.

С учетом вышеуказанных условий для построения модели целесообразно использовать функциональную зависимость между дефектами обработки данных в технологическом процессе ИС и значением показателя производительности ИС. Учитывая расчетно-аналитические методы оценки качества ИС, указанную зависимость можно определить на основе регрессионной зависимости [6]. При этом выбор вида функции зависимого показателя, в нашем случае это показатель производительности ИС, от набора значений независимых показателей по классам дефектов ИС должен быть выполнен так, чтобы получаемая при этом линейная зависимость была бы лучшей аппроксимацией функциональной зависимости. Кроме того, при расчете значения зависимого показателя необходимо определить значения коэффициентов весомости определяющих показателей, которыми являются здесь коэффициенты регрессии.

Исходя из практических соображений, в нашем случае целесообразно в качестве модели указанной зависимости использовать регрессионные уравнения линейного вида. Используемыми переменными выступают в этом случае время обнаружения и исправления дефектов по достоверности, полноте и своевременности.

При условии решения уравнения регрессии по фиксированной матрице исходных данных уравнение множественной

линейной регрессии будет иметь вид

$$Y_i = a_0 + A'x_i + \varepsilon_i = a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_qx_{iq} \quad (1)$$

где Y_i – зависимая (прогнозируемая) переменная – производительность ИС; x_i – независимые (прогнозирующие) переменные (значения времени обнаружения и исправления дефектов соответственно по достоверности, полноте, своевременности); a_0 – свободный член; A' – вектор оценок коэффициентов линейной регрессии; ε_i – случайные величины (совокупность неучтенных случайных факторов, либо мера достижимой аппроксимации значений Y_i функциями из аргументов x_i , либо то и другое вместе).

Оценка параметров a_0, A' , производится методом наименьших квадратов, то есть из условия минимума суммы квадратов невязок (отклонений)

$$\Delta^2 = \sum_{i=1}^n (Y_i - A'x_i - a_0)^2$$

Это приводит к системе нормальных уравнений

$$\begin{cases} \hat{A} = S^{-1} * \hat{C}_{yx} \\ a_0 = m_y - \hat{A}'M_x \end{cases}$$

где $\hat{A} = (\hat{a}_1, \dots, \hat{a}_q)$ – вектор оценок коэффициентов линейной регрессии; S^{-1} – обратная матрица ковариаций между переменными x_1, \dots, x_q ; \hat{C}_{yx} – вектор оценок ковариаций между переменными y и переменными x_1, \dots, x_q ; m_y – оценка среднего значения y ; M_x – вектор средних значений переменных x_1, \dots, x_q .

Оценку относительного значения обобщенного показателя качества ИС по производительности можно заполнить по формуле расчета значений относительных показателей, которые являются отношением фактических показателей к базовым, и определяются по формуле

$$P_i^0 = P_i^a / P_i^b \quad (2)$$

где P_i^0 – значение относительно уровня показателя 1-го вида ($0 \leq P_i^0 \leq 1$); P_i^a – фактическое значение показателя 1-го вида; P_i^b – прогнозное значение показателя 1-го вида.

Для оценки параметров регрессии с применением ЭВМ необходимо получить исходные данные. В нашем случае они могут быть представлены в виде матрицы фиксированных данных, которая имеет размерность $n \times m$ ($m = q + 1$, $nm = Y$) и представляет выборку m -мерных объектов $X = (x_1, \dots, x_q, \dots, x_m)$. Расчет значений целесообразно выполнить исходя из зависимости (влияния) повышения производительности ИС от снижения времени на обнаружение и исправление дефектов с размерностью шага на 1%. По условиям решения регрессионных уравнений в матрице исходных данных количество строк должно быть не меньше числа переменных [6]. Исходя из соображений корректности и практического использования результатов целесообразно представить такую матрицу, как состоящую из 20 строк, т.е. просчитать зависимость до 20%.

Работы по сбору, измерению и обработке данных выполнялись в соответствии с методикой планирования эксперимента [7]. Расчет значений матрицы исходных дан-

ных проведен по специальной методике регистрации дефектов функционирования ИС. На основе кластер-анализа статистической структуры дефектов были получены три группы дефектов по достоверности, полноте и своевременности [8]. Затем эти группы дефектов были статистически обработаны. В частности, они проверялись на согласие с нормальным распределением. Получены оценки значения математического ожидания дефекта по каждой из указанных выше трех групп дефектов, вероятности дефекта и др.

Регрессионный анализ зависимости показателя производительности от дефектов обработки и получение прогнозных и фактических значений показателя производительности выполняется с применением соответствующих программ статистического анализа данных [9]. Кроме того, в результате реализации регрессионной модели определяются также и коэффициенты весомости по значениям классов дефектов.

На основе полученных оценок по достоверности, полноте и своевременности рассчитыв-

Таблица 1

Матрица фиксированных данных по производительности ИС

Значение (%)	Значения дефектов по их классам (дни)			Производительность (документов в день)
	достоверность	полнота	своевременность	
0	246,14	794,01	4162,76	200,79
1	243,68	786,07	4121,14	202,05
2	241,22	778,13	4079,52	203,37
3	238,76	770,19	4037,90	204,72
4	236,30	762,25	3996,28	206,07
5	233,84	754,31	3954,66	207,45
6	231,38	746,37	3913,04	208,84
7	228,92	738,43	3871,42	210,26
8	226,46	730,49	3829,80	211,69
9	224,00	722,55	3788,18	213,14
10	221,54	714,61	3746,56	214,61
11	219,08	706,67	3704,94	216,10
12	216,62	698,73	3663,32	217,61
13	214,16	690,79	3621,70	219,16
14	211,70	682,85	3580,08	220,71
15	209,24	674,91	3538,46	222,29
16	206,78	666,97	3496,84	223,89
17	204,32	659,03	3455,22	225,51
18	201,86	651,09	3413,60	227,16
19	199,40	643,15	3371,98	228,81
20	196,94	635,21	3330,36	230,53

ваются значения матриц фиксированных данных по производительности (табл. 1).

В первой графе указаны градации снижения значений переменных (в %), в графах 2–4 – прогнозирующие переменные (достоверность, полнота и своевременность) и в графе 5 – соответствующие значения прогнозируемой переменной. Предсказывающие переменные (значения по классам дефектов – достоверности, полноте, своевременности) измеряются в днях, прогнозируемая переменная (производительность ИС) – в документах/днях.

По данным матрицы построен соответствующий график зависимости значения показателя по производительности (рис. 1). По оси абсцисс отмечены значения процентов снижения дефектов, а по оси ординат соответствующие значения прогнозируемой переменной. Путем нанесения точек указанных значений и соответствующей линии регрессии получены графические модели прогнозирования значений показателей производительности от снижения дефектов по соответствующим значениям процентов. На основе графиков в практических задачах оперативного управления качеством ИС можно получать экспресс-оценки прогнозируемой величины снижения производительности ИС. Если необходимо определить, например, значение обобщенного показателя по производительности ИС при условии снижения дефектов до 10%, то оно ориентировочно будет равно 214 документ/день. Определяется путем нанесения линий параллельно осям координат до пересечения с линией регрессии (рис. 1).

Для получения значений по прогнозу производительности ИС в результате реализации программы регрессионного анализа были получены коэффициенты регрессии и оценочные величины по производительности ИС (табл. 2).

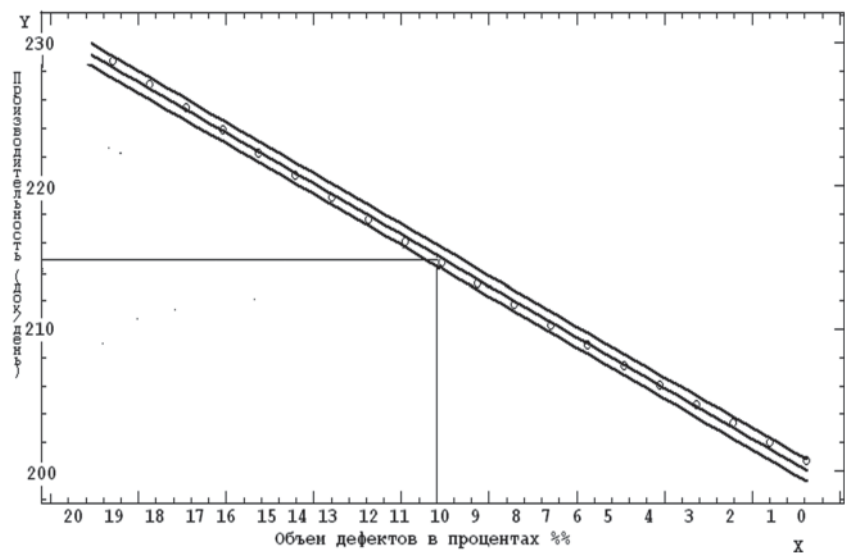


Рис. 1. График зависимости производительности ИС от снижения времени устранения дефектов обработки данных ИС

Свободный член – 348,14;
Коэффициент множественной корреляции – 0,9982;
Коэффициент детерминации – 0,9965;

В соответствии с формулой (1) составлено уравнение множественной линейной регрессии по производительности

$$Y^p = 348,14 - 0,18425x_1 - 0,072831x_2 - 0,038581x_3.$$

В правой части уравнений расположены слева направо соответствующие свободные члены регрессии, затем коэффициенты регрессии. В нашем случае они обозначают значения базовых показателей производительности и коэффициентов весомости по достоверности, полноте и своевременности. Свободный член уравнений при условии нулевых значений предсказывающих переменных, то есть абсолютное отсутствие дефектов отображает

по существу возможно достижимое значение производительности ИС в нашем случае 348,14 документов в день.

Если действия учетных дефектов обуславливают фактическое значение показателя по производительности ИС на уровне 200,79 документов в день (табл. 1, строка 0), то при идеальном условии, то есть устранении 100% дефектов производительность ИС может подняться до ее базового прогнозного значения – 348,14 документов в день. Подставляя полученные фактические и базовые значения в формулу (2) определим относительный уровень качества ИС по обобщенному показателю «производительность ИС» – он будет равен 0,58. Этот показатель свидетельствует о том, что используются только 58% потенциальных возможностей ИС, или 42% ресурсов ИС расходу-

Таблица 2

Коэффициенты регрессии по производительности и данные по их оценке

Вид признака	Коэффициент регрессии	Стандартное отклонение коэффициента	Уровень значимости нулевой гипотезы (P-значение)	T-значение	95%-ные доверительные границы	
					верхняя	нижняя
Достоверность	-0,18425	0,46758	0,3494	0,3940	0,8069	-1,1755
Полнота	-0,072831	0,15696	0,3244	0,4640	0,4055	-0,2599
Своевременность	-0,038581	0,04561	0,2050	0,8458	0,0581	-0,1352

Данные анализа остатков регрессии по производительности

Значение %	Номинальное значение производительности	Значение прогноза производительности	Остаток	Остаток/SY	Остаток/SS	Относительная погрешность
1	202,0	201,6	0,428	0,049	0,730	0,002118
2	203,4	203,0	0,352	0,040	0,600	0,001732
3	204,7	204,6	0,122	0,014	0,208	0,0005957
4	206,1	206,0	0,076	0,009	0,130	0,0003688
5	207,4	207,6	0,124	0,014	0,212	0,0005987
6	208,8	209,0	0,130	0,015	0,222	0,0006232
7	210,3	210,6	0,290	0,033	0,495	0,001381
8	211,7	211,9	0,256	0,029	0,437	0,001211
9	213,1	213,3	0,202	0,023	0,345	0,0009491
10	214,6	214,9	0,313	0,035	0,533	0,001456
11	216,1	216,3	0,218	0,025	0,372	0,001011
12	217,6	217,9	0,289	0,033	0,492	0,001327
13	219,2	219,3	0,135	0,015	0,230	0,0006142
14	220,7	220,9	0,165	0,019	0,281	0,0007467
15	222,3	222,3	0,019	0,002	0,033	0,0000864
16	223,9	223,9	0,039	0,004	0,067	0,0001743
17	225,5	225,2	0,263	0,030	0,449	0,001167
18	227,2	227,1	0,042	0,005	0,071	0,0001829
19	228,8	228,2	0,587	0,067	1,001	0,002565
20	230,0	229,8	0,198	0,022	0,338	0,0008614

F-статистика для проверки гипотезы: $H_0: B_1 = B_2 = \dots = 0 - F_0 = 499,7966$;
 Уровень значимости (P-значение) гипотезы $H_0 (P > (F_0/H_0)) - 0,0000$;
 Среднее квадратическое отклонение ошибки - 0,58649.

ется недостаточно рационально. При данном условии создание автоматизированной ИС или ее модернизация с большой долей вероятности может завершиться существенным положительным эффектом.

2. Проверка адекватности модели прогнозирования производительности ИС

На основе полученных результатов обработки экспериментальных данных проведен анализ параметров адекватности разработанной регрессионной модели для определения прогноза производительности ИС. Обратимся к содержанию полученных оценок (табл. 2). Относительно коэффициентов множественной корреляции, коэффициентов детерминации, уровней значимости нулевой гипотезы, а также среднего квадратического отклонения ошибки, можно предположить, что качество линейного прогноза очень хорошее. Все коэффициенты регрессии укладываются в 95%-ные доверительные границы. Это означает, что их истинные значения при нулевой гипотезе не могут быть отвергнуты при 5%-ом уровне значимости.

Более четкое заключение можно дать на основе анализа остатков. Информация по анализу остатков приведена в таблице 3, где SY, SS – стандартное отклонение прогнозируемой переменной и ошибки.

Значения производительности и анализ параметров показывает, что значения остатков незначительны, как в абсолютном, так и относительном измерениях. Визуальную проверку адекватности модели удобнее всего выполнить по графику зависимости величин нормированных остатков от величин снижения дефектов (в %) по производительности (рис. 2).

На графике видно, что наибольший «выброс» наблюдается по значениям 1% и 19% нормированных остатков, то

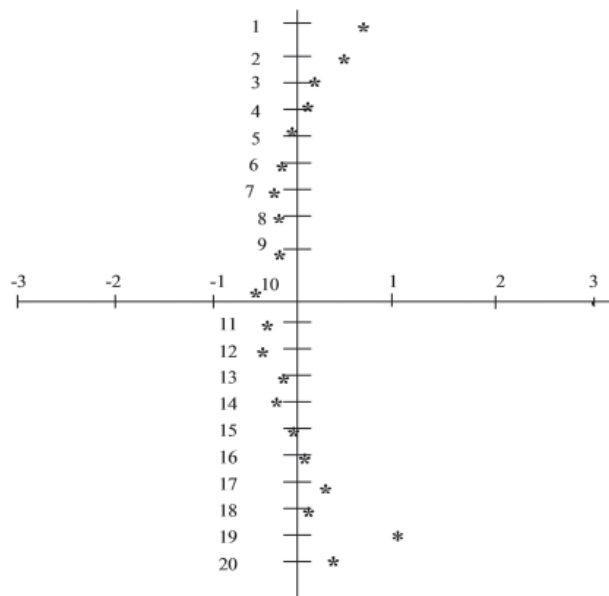


Рис. 2. График зависимости величин нормированных остатков от величины процента снижения дефектов по времени

есть по значениям стандартного отклонения ошибки – 0,730 и 1,001 (табл. 3, графа б). График показывает слабо выраженный криволинейный тренд. Сравнительная смещенность распределения настолько мала (близость к оси), что не имеет принципиального значения относи-

тельно номинальных величин прогнозируемой переменной по производительности (табл. 3). Незначительная смещенность наблюдается также и на графике регрессии (рис.1). Наблюдающийся случайный разброс свидетельствует о том, что модель едва ли можно или целесообразно улучшить.

Закключение

При создании информационных систем значительное место занимает задача прогнозирования производительности будущей ИС, что в определенной мере может обеспечить снижение рисков неоправданного расходования значительных объемов ресурсов. Модель прогнозирования производительности ИС может быть построена на основе функциональной зависимости значения показателя производительности ИС от дефектов обработки данных. Подобную модель можно построить посредством регрессионного анализа в виде множественной линейной регрессии. Для создания модели необходимо изучить разнообразную природу возникновения дефектов. Чтобы определить механизм влияния дефектов на производительность, получить точные

расчеты по объему дефектов, времени их обнаружения и исправления необходимо идентифицировать классы этих дефектов. Определение классов дефектов проводится путем сбора сведений по этапам и участкам технологического процесса обработки данных ИС. В последующем проводится измерение дефектов по набору их параметров, в частности, времени обнаружения и исправления дефектов.

Идентификация классов дефектов выполняется на ЭВМ посредством программы кластер-анализа статистической структуры дефектов. В результате определены три основных класса дефектов – по достоверности, полноте и своевременности обрабатываемых данных в технологии ИС. Затем выполняются расчеты по параметрам множественной линейной регрессии посредством реализации про-

граммы регрессионного анализа на ЭВМ.

На основе параметров, в частности, базового значения производительности ИС можно определить эффективность ИС до начала финансирования проекта ИС. Выполнение расчетных процедур осуществляется посредством пакетов прикладного статистического анализа.

Проверка адекватности регрессионной модели проводится с помощью стандартных статистических оценок и графических средств анализа зависимости величины нормированного остатка от снижения процента дефектов обработки документов ИС. Экспериментальная проверка показала адекватность модели прогнозирования производительности ИС. Модель применялась в решении комплекса задач по созданию ИС в образовательной и промышленной сферах.

Литература

1. Тельнов Ю.Ф., Смирнова Г.Н., Сорокин А.А. Проектирование экономических информационных систем. – М.: Финансы и статистика; 2002: 491 с.
2. Мазур И.И., Шапиро В.Д., Ольдерогге Н.Г. – М.: Омега-Л; 2006: 526 с.
3. Исаев Г.Н. Прогнозирование себестоимости документов в управлении информационными системами // Экономические науки. – 2012, № 4(89). С 88–92.
4. ГОСТ 34.601-90. Автоматизированные системы. Стадии создания.
5. Исаев Г.Н. Моделирование информационных ресурсов: теория и решение задач. – М.: Альфа-М-Инфра-М; 2010: 224 с.
6. Кулаичев А.П. Методы и средства комплексного анализа данных. – М.: ФОРУМ: ИНФРА-М; 2006: 512 с.
7. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. – М.: Мир; 1980: 382 с.
8. Исаев Г.Н. Показатели качества функционирования автоматизированных информационных систем: методика определения // Научно-техническая информация, сер. 2 Информационные процессы и системы, 1996. – № 3. С 12–16.

Сведения об авторе

Георгий Николаевич Исаев

Кандидат технических наук, доцент кафедры Сервисного инжиниринга
Российский государственный университет туризма и сервиса, Москва, Россия
Эл. почта: georg.isaev@mail.ru

References

1. Tel'nov Yu.F., Smirnova G.N., Sorokin A.A. Proektirovanie ekonomicheskikh informatsionnykh sistem. – M.: Finansy i statistika; 2002: 491 p. (in Russ.)
2. Mazur I.I., Shapiro V.D., Ol'derogge N.G. – M.: Omega-L; 2006: 526 p. (in Russ.)
3. Isaev G.N. Prognozirovanie sebestoimosti dokumentov v upravlenii informatsionnymi sistemami // Ekonomicheskie nauki. – 2012, № 4(89). Pp. 88–92. (in Russ.)
4. GOST 34.601-90. Avtomatizirovannyye sistemy. Stadii sozdaniya. (in Russ.)
5. Isaev G.N. Modelirovanie informatsionnykh resursov: teoriya i reshenie zadach. – M.: Al'fa-M-Infra-M; 2010: 224 p. (in Russ.)
6. Kulachev A.P. Metody i sredstva kompleksnogo analiza dannykh. – M.: FORUM: INFRA-M; 2006: 512 p. (in Russ.)
7. Dzhonson N., Lion F. Statistika i planirovanie eksperimenta v tekhnike i nauke. – M.: Mir; 1980: 382 p. (in Russ.)
8. Isaev G.N. Pokazateli kachestva funktsionirovaniya avtomatizirovannykh informatsionnykh sistem: metodika opredeleniya // Nauchno-tekhnicheskaya informatsiya, vol. 2 Informatsionnye protsessy i sistemy, 1996. – № 3. Pp. 12–16. (in Russ.)

Information about the author

Georgiy N. Isaev

Candidate of Engineering Sciences, Senior lecturer of the faculty of Service engineering
Russian state university of tourism and service,
Moscow, Russia
E-mail: georg.isaev@mail.ru