

УДК 004.519.217

Е. А. Арсирый, С. Г. Антощук, доктора техн. наук,
О. С. Маникаева

АВТОМАТИЗАЦИЯ ПРЕДСТАВЛЕНИЯ И ИЗВЛЕЧЕНИЯ ДЕКЛАРАТИВНЫХ ЗНАНИЙ В ЦИФРОВОМ МАКЕТЕ ПРЕДПРИЯТИЯ ПРИ ИНИЦИАЛИЗАЦИИ ПРОЕКТОВ ОХРАНЫ ТРУДА

Аннотация. Разработана автоматизация представления и извлечения декларативных знаний в цифровом макете предприятия при инициализации проектов охраны труда. При разработке цифрового макета предприятия предложено структурировать данные по охране труда по группам организации и условий труда и проводить автоматизированную кластеризацию этих данных с целью повышения качества извлеченных знаний эксперта для определения содержания содержания проекта по охране труда.

Ключевые слова: инициализация проектов по охране труда, цифровой макет предприятия, представление и извлечение знаний, фреймовая модель базы знаний, самоорганизующиеся слои и карты Кохонена

Е. Arsiriy, ScD., S. Antoshchuk ScD.,
O. Manikaeva

AUTOMATION OF THE SUBMISSION AND RETRIEVAL OF DECLARATIVE KNOWLEDGE IN THE DIGITAL LAYOUT OF THE ENTERPRISE DURING THE INITIALIZATION OF THE PROJECTS OF OCCUPATIONAL SAFETY

Abstract. Developed automation of the submission and retrieval of declarative knowledge in the digital layout of the enterprise during the initialization of the project occupational safety. When developing a digital layout of the enterprise proposed to structure the data by the occupational groups of the organization and conditions of work and to carrying out automated clustering of these data with the aim of improving the quality of the extracted expert knowledge for the determination of the project occupational safety.

Keywords: initializing project on occupational safety, the digital model of the enterprise, representation and knowledge extraction, frame model of knowledge base, self-organizing layers and maps of Kohonen

О. О. Арсірій, С. Г. Антощук, доктора техн. наук,
О. С. Манікаєва

АВТОМАТИЗАЦІЯ ПРЕДСТАВЛЕННЯ І ОТРИМАННЯ ДЕКЛАРАТИВНИХ ЗНАНЬ В ЦИФРОВОМУ МАКЕТІ ПІДПРИЄМСТВА ПРИ ІНІЦІАЛІЗАЦІЇ ПРОЄКТІВ ОХОРОНИ ПРАЦІ

Анотація. Розроблена автоматизація представлення і отримання декларативних знань в цифровому макеті підприємства при ініціалізації проєктів охорони праці. При розробці цифрового макету підприємства запропоновано структурувати дані з охорони праці по групах організації і умов праці та проводити автоматизовану кластеризацію цих даних з метою підвищення якості одержаних знань експерта для визначення змісту проєкту з охорони праці.

Ключові слова: ініціалізація проєктів з охорони праці, цифровий макет підприємства, представлення та отримання знань, фреймова модель бази знань, самоорганізуючі шари і карти Кохонена.

Введение. Исследования показывают, что время и риски выполнения, стоимость и качество проектов охраны труда (ПОТ), реализуемых на отечественных предприятиях, во многом зависят от их содержания, которое определяется экспертом — лицом, принимающим решение (ЛПР) на стадии инициализации [1 – 2]. Решение об инициализации ПОТ принимается на основе анализа текущего состояния уровня организации и условий труда на предприятии по группам факторов (признаков).

Основная часть. Для отражения текущего уровня организации и условий труда на предприятии и поддержки принятия решений по инициализации ПОТ предложено использовать цифровой макет предприятия на основе фреймовой модели базы знаний, которая включает [3]:

– декларативные знания в виде моделей, отражающих уровень организации и условий труда на предприятии по группам факторов;

– процедурные знания в виде методов получения знаний, позволяющих проводить оценку потребности в инициализации ПОТ, а также сценария его проведения на основе введенных признаков состояния с учетом базы моделей.

Для представления декларативных знаний по организации и условиям труда в цифровом макете предприятия (ЦМП) при инициализации ПОТ данные о предприятии предложено структурировать по качественным и комплексным количественно-качественным признакам [4].

К группе «организация труда» относят качественные факторы, значения которых получают в результате проведения комплексной экспертизы по определению уровня организационных мероприятий по охране труда на предприятии. К таким признакам относятся качественно определяемые уровни технических и технологических средств труда, коллективных и индивидуальных средств защиты, материально-хозяйственно-бытового обеспечения, а также тяжести и напряженности труда [5].

© Арсирый Е.А., Антощук С.Г.,
Маникаева О.С., 2015

При проведении экспертизы для ЛПП разработана четырех бальная шкала, позволяющая оценить организацию труда на предприятии как: «отличную», «допустимую», «плохую» и «хаотичную».

К группе «условия труда» относят количественно-качественные факторы, оказывающие влияние на работоспособность и здоровье работника, значения, которых получают с помощью средств мониторинга производственной среды. К таким факторам относятся численно определяемые уровни аэрозольного, электромагнитного, акустического, химического и биологического воздействия, ионизирующего излучения, микроклимата, освещенности и вибрации.

В результате анализа данных мониторинга в соответствии с предложенной шкалой, зависящей от уровня предельно допустимой концентрации исследуемого фактора, ЛПП оценивает состояние условий труда как: «оптимальное», «допустимое», «вредное», «опасное», «экстремальное».

Исследования данных мониторинга производственной среды, по группе условий труда, показали, что такие данные являются слабоструктурированными и неопределенными, потому что собираются из различных источников, интерпретируются с помощью различных и не всегда связанных шкал и часто противоречат друг другу. А решение об оценке состояния условий труда на основе анализа таких данных, которое ЛПП принимает при проведении экспертизы, является неоднозначным и зависит от его квалификации.

Для проверки качества экспертных решений при инициализации ПОТ, предложено использовать, состоящую из 4-х этапов, методику автоматизированного извлечения декларативных знаний по группам организации и условиям труда в цифровом макете предприятия (рис. 1).

Этап 1. Воспользовавшись предварительно структурированными и представленными декларативными знаниями (набор признаков и оценочных шкал), отражающими уровень организации и условий труда в ЦМП ЛПП проводит оценку состояния производственной среды по группам факторов. Вид результатов X_{ki} , P_k комплексной экспертизы на примере определения уровня аэрозольного воздействия (АВ) показан в табл. 1. В таблице 1 k – номер экспертизы.

Этап 2. Для автоматизации извлечения знаний эксперта при проверке качества проведения экспертизы предложено проводить кластеризацию данных экспертных оценок X_{ki} с помощью самоорганизующегося слоя Кохонена [6 – 7]. Процесс кластеризации X_{ki} предусматривает последовательное выполнение процедур самоорганизации нейронов вычислительного слоя Кохонена, градуировки элементов выходного вектора обучающей выборки и окончательной маркировки нейронов слоя Кохонена (рис. 2).

2.1. Классическая процедура самоорганизации вычислительного слоя Кохонена реализована с помощью итеративного алгоритма WTA (Winner Takes All) [8]. Согласно которого на вход слоя Кохонена (рис. 2) последовательно подаются значения векторов из обучающей выборки X_{ki} , $i = \overline{1, n}$, $k = \overline{1, N}$, где n – количество признаков, а N – количество векторов в выборке.



Рис. 1. Методика автоматизированного извлечения декларативных знаний

1. Экспертная оценка уровня АВ

k	Уровень составляющих фактора X_{ki}				Общий уровень P_k
	Дисперсный состав	Концентрация	Время воздействия	Вид частиц	
	x_{k1}	x_{k2}	x_{k3}	x_{k4}	
1	1	1	1	2	1
2	1	1	1	3	1
...
12	1	2	2	2	2
13	1	2	2	3	2
...
24	1	2	4	2	3
25	1	2	4	3	3
...
34	1	3	2	2	4
35	1	3	2	3	4
...
68	1	3	5	2	5
69	1	3	5	3	5

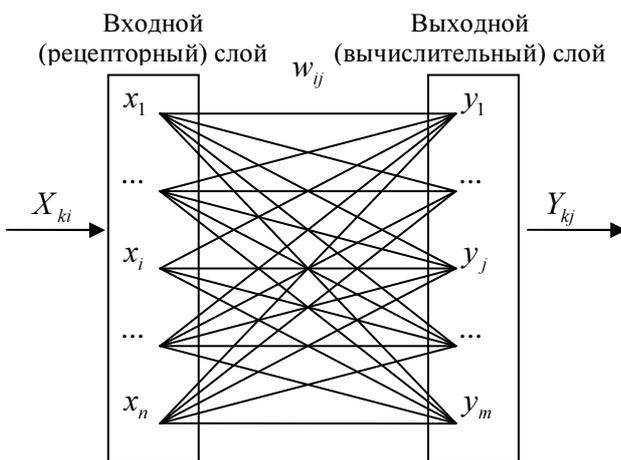


Рис. 2. Топология сети Кохонена

Целью самоорганизации является минимизация разницы расстояний:

$$d(x_{ki}, w_{ij}) = \min d(x_{ki}, w_{ij}), \quad (1)$$

между элементами входных векторов x_{ki} и весовыми коэффициентами w_{ij} нейрона-победителя слоя Кохонена, по формуле корректировки:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)[x_{ki} - w_{ij}(t)], \quad (2)$$

где, $\eta(t)$ – изменяемый во времени коэффициент шага коррекции.

В качестве $\eta(t)$ обычно выбирается монотонно убывающая функция ($0 < \eta(t) < 1$). В качестве меры расстояния используется евклидово расстояние:

$$d(x_{ki}, w_{ij}) = \sqrt{\sum_{i=1}^n (x_{ki} - w_{ij})^2}. \quad (3)$$

2.2. При выполнении процедуры градуировки элементов выходного вектора обучающей выборки на вход предварительно самоорганизованного слоя Кохонена последовательно подаются значения векторов из обучающей выборки X_{ki} , где для каждого из них на основе (1) вычисляются значения вектора выходов Y_{kj} , $j = \overline{1, m}$, где m – число нейронов (классов). При этом для каждого k -го вектора из обучающей выборки все значения элементов вектора Y_{kj} равны 0, кроме нейрона-победителя, значение которого равно 1. Порядковый номер нейрона-победителя j присваивается в качестве значения i -у элементу градуировочного вектора Z_k . Вид результатов процедуры градуировки X_{ki}, Y_{kj}, Z_k при автоматизации извлечения знаний эксперта при проверке качества проведения экспертизы для определения уровня АВ показан на рис.3.

k	X_{ki}				Y_{kj}					Значения градуировочного вектора Z_k
	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4	y_5	
1	1	1	1	2	1	0	0	0	0	1
2	1	1	1	3	1	0	0	0	0	1
...
12	1	2	2	2	0	1	0	0	0	2
13	1	2	2	3	0	1	0	0	0	2
...
24	1	2	4	2	0	0	1	0	0	3
25	1	2	4	3	0	0	1	0	0	3
...
34	1	3	2	2	0	0	0	1	0	4
35	1	3	2	3	0	0	0	1	0	4
...
68	1	3	5	2	0	0	0	0	1	5
69	1	3	5	3	0	0	0	0	1	5

Рис. 3. Результаты процедуры градуировки

2.3. Для определения соответствия между значением итоговой оценки P_k (номер класса) из таблицы 1 и значением номера нейрона победителя градуировочного вектора Z_k , предложена процедура маркировки нейронов слоя Кохонена номерами классов по итоговой экспертной оценке. Процедура маркировки состоит из 3-х шагов (рис.1).

Шаг 2.3.1. Для всех примеров обучающей выборки формируем двумерную гистограмму $D_{l,j}$ (квадратную матрицу) попарных совпадений номеров нейронов и классов $Z_{kl} = P_{kj}$, где $l, j = \overline{1, m}$. Вид гистограммы $D_{l,j}$, для примера определения уровня АВ, показан в табл.2.

2. Двумерная гистограмма $D_{l,j}$

		Номер класса $P_{k,j}$				
		1	2	3	4	5
Номер нейрона Z_{kl}	1	9	4	0	0	0
	2	1	4	3	2	0
	3	1	4	5	4	0
	4	0	0	1	6	19
	5	0	0	1	4	1

Шаг 2.3.2. Выполняем *построчную и постолбцовую* корректировку гистограммы $D_{l,j}$, оставляя без изменения только те значения элементов, которые удовлетворяют условию:

$$\begin{cases} \max(D_l) = \max(D_j) \\ l = j \\ \max(D_{ij}) < 0 \end{cases} \quad (4)$$

Остальные значения пересекающихся строки и столбца – обнуляем. Описанная процедура *построчно-постолбцовой* корректировки двумерной гистограммы выполняется итерационно, пока в каждой строке и столбце останется только одно не нулевое значение. Результат корректировки гистограммы $D_{l,j}$, для примера определения уровня АВ, показан в табл. 3.

3. Результат построчно-постолбцовой корректировки гистограммы $D_{l,j}$

		Номер класса $P_{k,j}$				
		1	2	3	4	5
Номер нейрона Z_{kl}	1	9	0	0	0	0
	2	0	4	0	0	0
	3	0	0	5	0	0
	4	0	0	0	0	19
	5	0	0	0	4	0

Шаг 2.3.3. На последнем шаге процедуры маркировки нейронов слоя Кохонена скорректированная двумерная гистограмма $D_{l,j}$ попарных совпадений преобразуется в таблицу соответствий номеров нейронов и классов для всех примеров обучающей выборки, вид которой для примера определения уровня АВ, показан (табл. 4).

4. Окончательная маркировка номера нейрона на основе класса

Z_l	1	2	3	5	4
P_j	1	2	3	4	5

Этап 3. На основе полученной таблицы соответствий номеров нейронов и классов для проверки качества экспертных оценок по всем примерам обучающей выборки рассчитываются ошибки 1-го и 2-го рода и их относительные доли истинно положительных случаев и истинно отрицательных случаев [9 – 10].

$$TPR = \frac{TP}{TP + FN} \cdot 100\% \quad (5)$$

где TPR – относительная доля истинно положительных случаев (True Positives Rate – TPR); TR – истинно положительные случаи (верно классифицированные положительные примеры); FN – положительные примеры, классифицированные как отрицательные (ошибка I-го рода).

$$FPR = \frac{FP}{TN + FP} \cdot 100\% \quad (6)$$

где FPR – относительная доля истинно отрицательных случаев (False Positives Rate – FPR); FP – отрицательные примеры, классифицированные как положительные (ошибка II рода); TN – истинно отрицательные случаи (верно классифицированные отрицательные примеры).

Сравнительные значения TPR и FPR рассчитанные при проверке качества экспертных оценок для всех примеров обучающей выборки по всем группам условий труда в ЦМП показаны на рис. 4.

Этап 4. Неудовлетворительные значения оценок эксперта (FPR) передаются ЛПП для анализа и выполнения корректировки, после которой выполняется возврат к этапу 2 (рис. 1).

Анализ значений ошибок 1-го и 2-го рода до (рис. 4) и после проведения корректировки (рис. 5) показывает повышение относительной доли TPR в среднем на 20 % и снижение на 50 % FPR для всех групп условий труда.

Особенно снижение FPR заметно для уровней химического и биологического воздействия в ЦМП при инициализации проектов по охране труда.

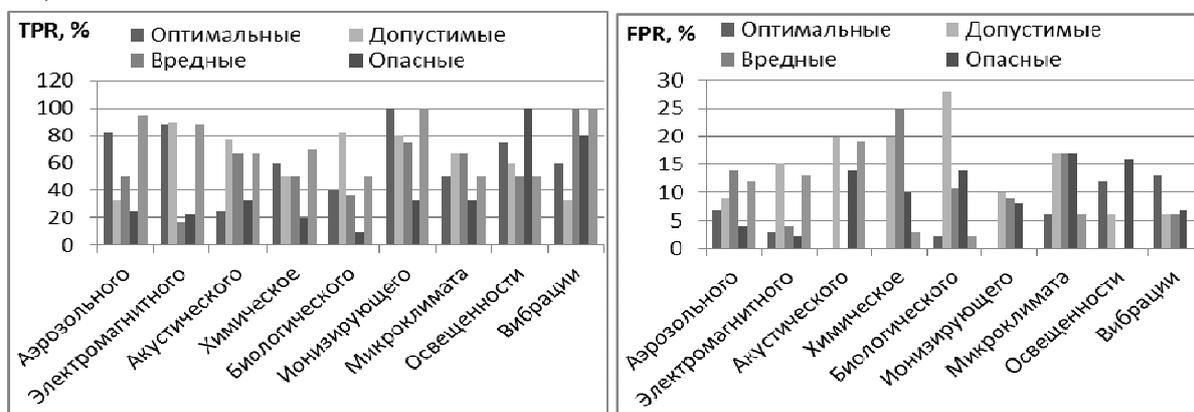


Рис. 4. Сравнительные результаты TPR и FPR

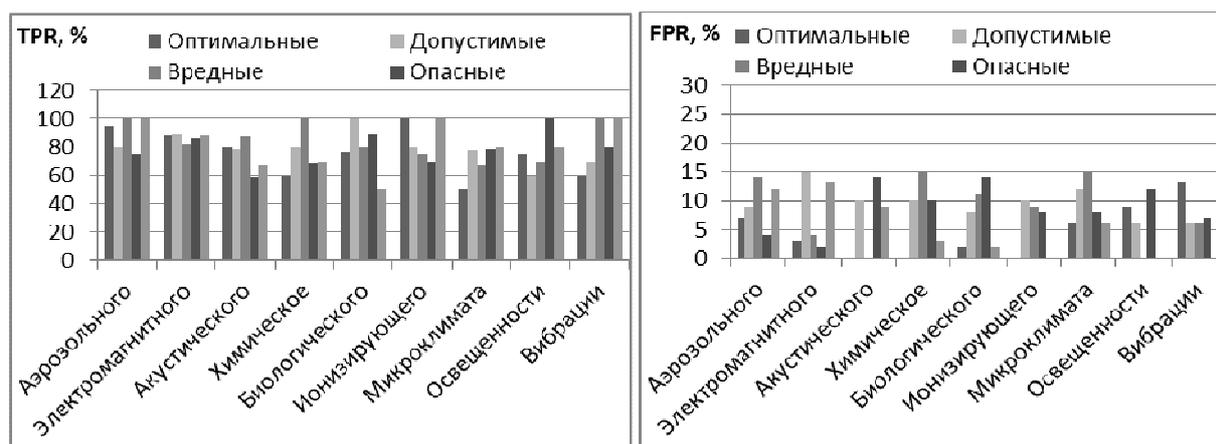


Рис. 5. Сравнительные результаты TPF и TPR после корректировки ЛПП

Вывод. Таким образом, разработана методика автоматизации представления и извлечения декларативных знаний в цифровом макете предприятия при инициализации проектов охраны труда. При разработке цифрового макета предприятия предложено структурировать данные по охране труда по качественным и комплексным количественно-качественным признакам с учетом двух выделенных групп организации и условий труда. Для проверки качества экспертных решений при инициализации ПОТ, предложено использовать кластеризацию декларативных знаний эксперта по группам организации и условиям труда в цифровом макете предприятия. Апробация предложенной методики для реальных данных по уровням аэрозольного, электромагнитного, акустического, химического и биологического воздействия, ионизирующего излучения, микроклимата, освещенности и вибрации показала повышение относительной доли TPR в среднем на 20% и снижение на 50 % FPR для всех перечисленных групп условий труда.

Список использованной литературы

1. Гогунський В. Д. Управління ризиками в проєктах з охорони праці як метод усунення шкідливих і небезпечних умов праці [Текст] / В. Д. Гогунський, Ю. С. Чернега // Восточно Европейский журнал передовых технологий. – Харьков : Технологический центр. – 2013. – № 1/10 (61). – С. 83 – 85.
2. Руководство по системам управления охраной труда. MOT-SUOT 2001/ ILOOSH 2001. – Женева: Международное бюро труда, 2003.
3. Арсирий Е. А. Цифровой макет оборудования энергетического предприятия [Текст] / Е. А. Арсирий, С. Г. Антошук // Электротехнические и компьютерные системы. – К. : Техніка. – 2013. – № 12 (88). – С. 71 – 82.
4. Москалюк А. Ю. Використання нейронних мереж для оцінки стану умов праці [Текст] / А. Ю. Москалюк., О. С. Манікаєва // Управління проєктами: інновації, не лінійність, синергетика : Матеріали V Міжнародної науково-практичної конференції магістрантів, аспірантів та науковців: [у 2 т.]. – Том 2. – Одеса : ОДАБА. – 2014. – С. 147 – 149.

5. Москалюк А. Ю. Проектизация процессов охраны труда [Текст] / А. Ю. Москалюк, П. А. Тесленко // Управління проєктами: Стан та перспективи: Матеріали 7-ї Міжнародної науково-практичної конференції. – Миколаїв : НУК. – 2011. – С. 208 – 210.
6. Пампуха И. В. Обоснование использования нейронных сетей в системах поддержки принятия решения при функционировании сложных систем [Текст] / И. В. Пампуха // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – 2013. – Вип. 42. – С. 85 – 90.
7. Уоссерман Ф. Нейрокомпьютерная техника: Теория и практика [Текст] /Ф. Уоссерман; пер. с англ. – М. : Мир, 1992. – 240 с.
8. Кохонен Т. Самоорганизующиеся карты [Текст] / Т. Кохонен; пер. 3-го англ. изд. – М. : БИНОМ Лаборатория знаний, 2008. – 655 с.
9. Паклин Н. Логистическая регрессия и ROC-анализ – математический аппарат / Н. Паклин [Электронный ресурс]. Режим доступа : <http://www.basegroup.ru/library/analysis/regression/logistic> (дата доступа 25.04.2015).
10. Davis J., and Goadrich M., (2006), The Relationship Between Precision-Recall and ROC Curves, *Proc. Of 23 International Conference on Machine Learning, Pittsburgh, PA*.

Получено 28.05.2015

References

1. Gogunskiy V.D., and Chernega Yu.S. Upravlinnya rizikami v proektah z ohoroni pratsi yak metod usunennya shkidlivih i nebezpechnih umov pratsi, [Risk Management in Pprojects for the Protection of Labor as a Method of Eliminating Harmful and Dangerous Working Conditions], (2013), *Vost.-Evropeyskiy Zhurnal Peredoviyih Tehnologiy*, No. 1/10 (61), Kharkov, Ukraine, *Tehnolog. Tsent*r, pp. 83 – 85 (In Ukrainian).
2. *Rukovodstvo po sistemam upravleniya ohranoy truda. MOT-SUOT 2001 [Guide to System Safety Management]*, (2003), *ILOOSH 2001*. Zheneva, Shvejcarsiya, *Mezhdunarodnoe Byuro Truda* (In Russian).

3. Arsiriy E.A., and Antoschuk S.G. Tsifrovoy maket oborodovaniya energeticheskogo predpriyatiya [Digital Layout of Equipment Energy Companies], (2013), *Electrotechnic and Computer Systems*, Kiev, Ukraine, *Tehnlka*, No. 12 (88), pp. 71 – 82 (In Russian).

4. Moskalyuk A.U., and Manikaeva O.S. Vikoristannya neyronnih mrezhz dlya otslnki stanu umov pratsl [The Use of Neural Networks to Assess the State of Labour Conditions], (2014), *Upravlinnya Proektami: Innovatsiyi, ne Liniynist, Sinergetika: Materialy v Mizhnarodnoyi Naukovo-praktichnoyi Konferentsiyii Magistrantiv, Asprantiv ta Naukovtsiv, [u 2 t.], Vidpovidniy za Vipusk P.O. Teslenko*, Odessa, Ukraine, *ODABA*, Vol. 2, pp. 147 – 149 (In Ukrainian).

5. Moskalyuk A.Yu., and Teslenko P.A. Proektizatsiya protsessov ohranyi truda [Practical Processes of Labour Protection], (2011), *Upravlinnya Proektami: Stan ta Perspektivi: Materialy 7-i Mizhnarodnoyi Naukovo-praktichnoyi Konferentsii*, Nikolayev, Ukraine, *NUK*, pp. 208 – 210 (In Russian).

6. Pampuha I.V. Obosnovanie ispolzovaniya neyronnyih setey v sistemah podderzhki prinyatiya resheniya pri funktsionirovanii slozhnyih sistem [The Justification for the Use of Neural Networks in Decision Support Systems Decisions in the Operation of Complex Systems], (2013), *Zbirnik Naukovih Prats Viyskovogo Institutu Kiiivskogo Natsionalnogo Unlversitetu Imenl Tarasa Shevchenka*, Vip. 42, pp. 85 – 90 (In Russian).

7. Uosserman F. Neyrokompyuternaya tehnika: Teoriya i praktika [Neurocomputing Technique], (1992), per. s angl., Moscow, Russian Federation, *Mir*, 240 p. (In Russian).

8. Kohonen T. Samoorganizuyuschiesya kartyi [Self-organizing Maps], (2008), 3-go angl. izd., Moscow, Russian Federation, *BINOM Laboratoriya Znaniy*, 655 p. (In Russian).

9. Paklin N. Logisticheskaya regressiya i ROC-analiz – matematicheskiy apparat. [Logistic Regression and ROC Analysis is Mathematics] (In Russian) [Electronic Resource]. available at: <http://www.basegroup.ru/library/analysis/regression/logistic> (accessed 25.04.2015).

10. Davis J., and Goadrich M., (2006), The Relationship Between Precision-Recall and ROC Curves, *Proc. Of 23 International Conference on Machine Learning, Pittsburgh, PA* (In English)



Арсирый
Елена Александровна,
д-р техн. наук, доц., каф.
информационных систем Одес-
ского нац. политехнического
ун-та.
E-mail: arsiroy@te.net.ua



Антошук
Светлана Григорьевна,
д-р техн. наук, проф., директор
ин-та компьютерных систем
Одесского нац. политехническо-
го ун-та. Тел.: +38048-7058-584.
E-mail: asg@ics.opu.ua



Маникаева
Ольга Сергеевна,
аспирант каф. информационных
систем Одесского нац.
политехнического ун-та.
E-mail: manikaeva@gmail.com