

## МЕТОДИКА ОПРЕДЕЛЕНИЯ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ, ОЦЕНИВАЕМЫХ ПО НЕПРЕРЫВНОЙ ШКАЛЕ

Наталия Белоус, Ирина Куцевич, Ирина Белоус

**Аннотация:** В работе описывается методика определения качества тестовых заданий с помощью которой проводится выделение в тесте несостоятельных заданий и заданий плохого качества. В работе проведен сравнительный анализ применения дихотомической и непрерывной шкал для оценивания.

**Ключевые слова:** качество тестовых заданий, коэффициент корреляции, валидность, несостоятельные тестовые задания, задания плохого качества, сложность тестового задания, субъект обучения.

**ACM Classification Keywords:** K.3.1 Computer Uses in Education

**Conference:** The paper is selected from XV<sup>th</sup> International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.

---

### Введение

---

Одним из сложных и противоречивых вопросов при проведении тестирования является оценивание знаний. В настоящее время в большинстве случаев используется дихотомическая шкала, по которой за каждое задание можно получить 0 или 1 балл. Данная шкала удобна при оценивании заданий т.н. закрытого типа, в которых выбирается один правильный ответ из многих. Существует многообразие типов тестовых заданий: закрытые (многоальтернативные и одноальтернативные), открытые, на установление соответствия между элементами, на установление правильной последовательности, ситуационные тестовые задания [Комплекс нормативних документів, 1998]. Для оценивания заданий разных типов часто применение дихотомической шкалы часто недостаточно, т.к. в случае, когда субъект обучения дает неполный или частично правильный ответ, он оценивается как неправильный. Кроме дихотомической шкалы в настоящее время используется политомическая шкала, в которой допускается несколько категорий ответа на задание, каждая из которых оценивается по-разному. Например, за полностью верный ответ назначается 2 балла, за частично верный – 1 балл и за неверный – 0 баллов. Недостатком этой шкалы является сложность вычисления общего результата на основе баллов, полученных за задания. Кроме того, в этом случае не учитываются неправильно выбранные варианты ответа. Простое суммирование баллов не соответствует истинному уровню знаний обучаемых. Чтобы избежать этих недостатков авторами предлагается введение непрерывной шкалы оценивания знаний на интервале от 0 до 1 и специализированные технологии определения оценок за выполнение каждого из типов тестовых заданий [Belous N., 2004].

Уровень знаний студентов варьирует от качества постановки учебного процесса, от количества выделяемых часов и от качества учебного материала, в том числе и тестового. Разнообразие причин, влияющих на качество знаний, подтверждает, что необходимо контролировать качество тестового материала с определенной периодичностью. Под качеством тестового материала принимают возможность различия субъектов обучения с высоким уровнем знаний и слабых [Аванесов В.С., 1989].

Для проведения качественного анализа тестового материала, оцениваемого по дихотомической шкале оценивания знаний, предлагается проведение статистической обработки результатов тестирования

[Аванесов В.С., 1989, Олейник Н.М., 1991, Комплекс нормативних документів, 1998]. Однако, для повышения точности оценивания знаний авторами предлагается применение непрерывной шкалы (в диапазоне от 0 до 1, где 1 – ответ полностью правильный, 0 – ответ полностью неправильный, промежуточные значения соответствуют неполным или частично правильным ответам). Для определения качества тестов, оцениваемых по непрерывной шкале, авторами предлагается методика, позволяющая по результатам предварительного тестирования выделять несостоятельные задания и задания плохого качества. К несостоятельным заданиям относятся те задания, которые не служат цели дифференцирования знаний и являются в этом случае бесполезными. К таким заданиям относятся слишком легкие (на которые ответили все) или слишком трудные (на которые никто не ответил), а также задания, не относящиеся к рассматриваемой в тесте предметной области. Задания, которые требуют корректировки, например, из-за неточности формулировок, относятся к заданиям плохого качества.

Целью работы является разработка методики определения качества тестовых заданий, оцениваемых по непрерывной шкале, частным случаем которой является дихотомическая шкала.

### Методика Проведения Качественного Анализа Тестовых Заданий

В общем виде тест представляет собой систему, состоящую из набора тестовых заданий. Требование системности заключается в том, что между заданиями, включенными в тест, должны проследиваться четкие связи, которые отражаются в результатах выполнения теста группой субъектов обучения. Для оценивания системных качеств теста применяется коэффициент корреляции, показывающий степень связи между случайными величинами, в данном случае, между тестовыми заданиями, на которые отвечала группа студентов.

Исходными данными для проведения определения качества тестов являются результаты тестирования выборки субъектов обучения, заданные с помощью неупорядоченной матрицы результатов, в которой столбцы соответствуют номерам тестовых заданий, строки – фамилиям субъектов обучения. Элементами неупорядоченной матрицы тестирования являются результаты  $res_{ij}$   $i$ -го субъекта обучения за выполнение  $j$ -го задания, оцененного по непрерывной системе в диапазоне  $[0,1]$ . По исходной неупорядоченной матрице строится упорядоченная матрица, данные из которой являются исходными к проведению дальнейших вычислений. По неупорядоченной матрице результатов тестирования определяются несостоятельные задания (задания, с которыми не справился ни один субъект обучения и те, с которыми все справились). Эти задания в упорядоченную матрицу не включаются. Оставшиеся тестовые задания упорядочиваются следующим образом: строки матрицы упорядочиваются по суммарному баллу за выполнение всех заданий каждым субъектом обучения в порядке возрастания сверху вниз, столбцы матрицы – по суммарному баллу за выполнение каждого задания всеми субъектами обучения в порядке убывания слева направо. Упорядоченная матрица тестовых результатов приведена на рисунке 1.

После упорядочивания матрицы тестовых результатов вычисляются величины  $\overline{R}_j$  - мера трудности задания (средний балл по всем заданиям) и  $\overline{R}_i$  - средний балл по всем субъектам обучения.

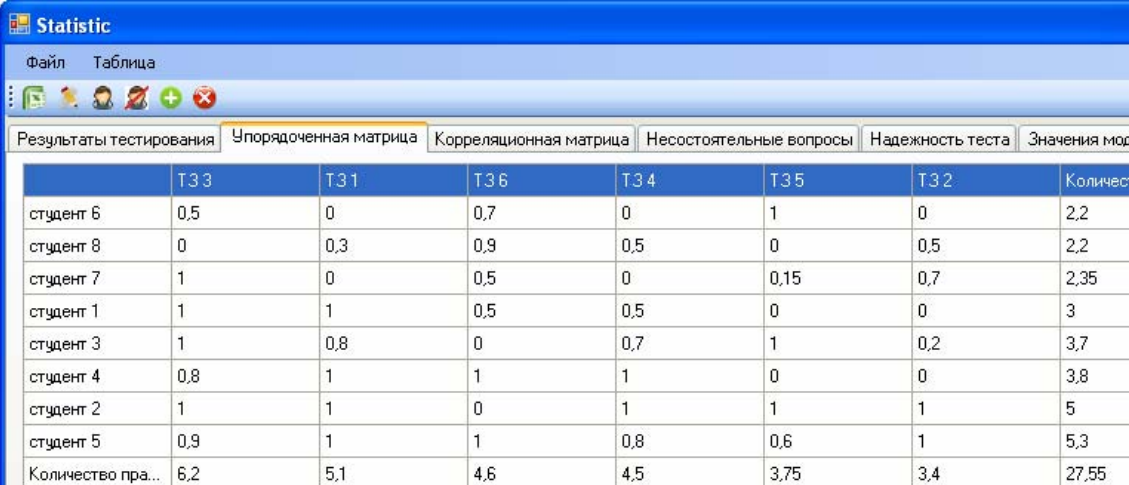
$$\overline{R}_j = \frac{1}{N} \sum_{i=1}^N res_{ij}, \quad (1)$$

$$\overline{R}_i = \frac{1}{n} \sum_{j=1}^n res_{ij}, \quad (2)$$

где  $n$  – количество заданий, включенных в тест;

$N$  – количество студентов, прошедших предварительное тестирование.

По упорядоченной матрице строится корреляционная матрица тестовых заданий, которая отображает степень связи тестовых результатов субъектов обучения (рис. 2).



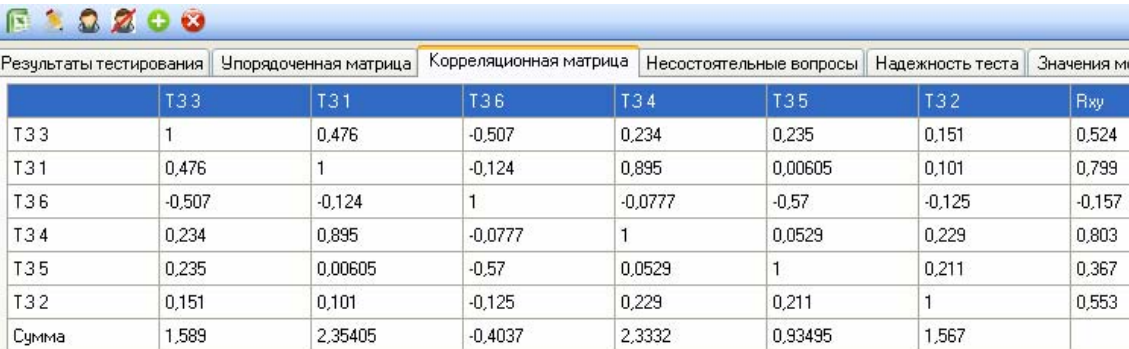
	T33	T31	T36	T34	T35	T32	Количество
студент 6	0,5	0	0,7	0	1	0	2,2
студент 8	0	0,3	0,9	0,5	0	0,5	2,2
студент 7	1	0	0,5	0	0,15	0,7	2,35
студент 1	1	1	0,5	0,5	0	0	3
студент 3	1	0,8	0	0,7	1	0,2	3,7
студент 4	0,8	1	1	1	0	0	3,8
студент 2	1	1	0	1	1	1	5
студент 5	0,9	1	1	0,8	0,6	1	5,3
Количество пра...	6,2	5,1	4,6	4,5	3,75	3,4	27,55

Рисунок 1 – Упорядоченная Матрица Результатов Тестирования

Элементами корреляционной матрицы являются коэффициенты корреляции  $C_{ij}$ , которые в случае применения непрерывной системы оценивания знаний авторами предлагается рассчитывать по формуле (3).

$$C_{ab} = \frac{\sum_{i=1}^N (res_{ia} \cdot res_{ib}) - N \cdot \overline{R}_a \cdot \overline{R}_b}{\sqrt{\left( \sum_{i=1}^N res_{ia}^2 - N \cdot (\overline{R}_a)^2 \right) \cdot \left( \sum_{i=1}^N res_{ib}^2 - N \cdot (\overline{R}_b)^2 \right)}}, \quad (3)$$

где  $a$  и  $b$  – тестовые задания, для которых рассчитывается коэффициент корреляции;  $\overline{R}_a$  – мера трудности  $a$ -го тестового задания;  $\overline{R}_b$  – мера трудности  $b$ -го тестового задания;  $C_{ab}$  – коэффициент корреляции тестовых заданий  $a$  и  $b$ .



	T33	T31	T36	T34	T35	T32	Rxy
T33	1	0,476	-0,507	0,234	0,235	0,151	0,524
T31	0,476	1	-0,124	0,895	0,00605	0,101	0,799
T36	-0,507	-0,124	1	-0,0777	-0,57	-0,125	-0,157
T34	0,234	0,895	-0,0777	1	0,0529	0,229	0,803
T35	0,235	0,00605	-0,57	0,0529	1	0,211	0,367
T32	0,151	0,101	-0,125	0,229	0,211	1	0,553
Сумма	1,589	2,35405	-0,4037	2,3332	0,93495	1,567	

Рисунок 2 – Корреляционная Матрица Результатов Тестирования

Корреляция заданий друг с другом не должна быть высокой, иначе задания начинают дублировать друг друга. По классификации коэффициентов корреляции Дворецкого связь тестовых заданий должна рассматриваться как слабая (коэффициент корреляции  $C_{ab} < 0,3$ ). С другой стороны, при отрицательных значениях коэффициента корреляции ( $C_{ab} < 0$ ) наблюдается обратная корреляция между тестовыми заданиями. Отрицательная корреляция между тестовыми заданиями нежелательна. Если задание отрицательно коррелирует с другими заданиями, то исход ответов на него противоположен результатам по другим заданиям. По всей вероятности у такого задания либо имеются грубые ошибки в содержании и

(или) оформлении (например, нет правильного ответа), либо проверяются знания из другой предметной области. Такие задания подлежат удалению. В случае, приведенном на рисунке 2, отрицательной корреляцией отличаются все тестовые задания. Следует обратить внимание на то, что отрицательная корреляция у заданий 1, 2, 3, 4 и 5 наблюдается именно с заданием 6. Это означает, что проблематичным является именно тестовое задание 6. У тестового задания 1 наблюдается сильная корреляция с заданиями 3 и 4, что свидетельствует также о проблематичности тестового задания. Разделим теперь тестовые задания на задания плохого качества и несостоятельные тестовые задания.

Важным параметром, применяемым при проведении качественного анализа тестовых заданий, является коэффициент валидности. Валидность – это мера соответствия того, насколько методика и результаты исследования соответствуют поставленным задачам. Для определения коэффициентов валидности тестовых заданий вычисляются коэффициенты корреляции заданий теста с суммой баллов субъектов обучения  $R_i$ . Следовательно,  $V_j = C_{j R_i}$ . Для вычисления коэффициентов валидности  $V_j$  авторами предлагается применение формулы (4).

$$V_j = \frac{\sum_{i=1}^N (\text{res}_{ij} \cdot R_i) - \bar{R}_j \cdot R}{\sqrt{\left( \sum_{i=1}^N \text{res}_{ij}^2 - N \cdot (\bar{R}_j)^2 \right) \cdot \left( \sum_{i=1}^N R_i^2 - \frac{R^2}{N} \right)}}, \quad (4)$$

где  $R_i$  – суммарный балл за выполнение заданий  $i$ -тым субъектом обучения,  $R_i = \sum_{j=1}^n \text{res}_{ij}$ ;

$R$  – суммарный балл, набранный всеми испытуемыми за выполнение тестовых заданий,  $R = \sum_{i=1}^N \sum_{j=1}^n \text{res}_{ij}$ .

По коэффициенту валидности проводится выявление тестовых заданий плохого качества и несостоятельных тестовых заданий. Для практических целей Авонесов рекомендует использовать коэффициенты валидности  $V_j \geq 0.5$  [Аванесов В.С., 1989]. В случае  $V_j \geq 0.5$  тестовое задание разделяет субъектов обучения с высоким уровнем знаний и слабых. Все те задания, для которых коэффициент валидности меньше или равен нулю, несостоятельны и непригодны для контроля знаний, поэтому их надо удалять из создаваемого теста, а для включения данных заданий в другие тесты их необходимо существенно переделывать и улучшать. Задания же, для которых  $0 < V_j < 0.5$ , являются заданиями плохого качества и требуют коррекции. Для данных, приведенных на рисунке 2, заданием плохого качества является ТЗ 6, несостоятельным заданием является ТЗ 1. Таким образом, ТЗ 1 требует корректировки, а ТЗ 6 должно быть полностью удалено из теста (рис. 2).

Выполнение предложенных расчетов позволяет сделать чистку теста и первые выводы о его ожидаемых качественных характеристиках. Корреляционная матрица результатов тестирования после проведения чистки упорядоченной матрицы приведена на рисунке 3.

После проведения чистки теста до проведения текущего и итогового оценивания знаний субъектов обучения проиодится распределение тестовых заданий по уровням сложности с применением разработанной авторами технологии распределения тестовых заданий по уровням сложности [Белоус Н.В. и др., 2009].

С учетом приведенных в работе обозначений, вычисление начального уровня знаний субъектов обучения  $\Theta_i^0$ , начального уровня сложности тестовых заданий  $\beta_j^0$  и дифференцирующей способность тестовых

заданий  $a_j$  проводится по формулам (5).

	T33	T34	T35	T32	Rxy
T33	1	0,234	0,235	0,151	0,603
T34	0,234	1	0,0529	0,229	0,58
T35	0,235	0,0529	1	0,211	0,65
T32	0,151	0,229	0,211	1	0,654

Рисунок 3 – Корреляционная Матрица Результатов Тестирования После Проведении Чистки Упорядоченной Матрицы

$$\Theta_i^0 = \ln\left(\frac{\bar{R}_i}{1 - \bar{R}_i}\right), \quad \beta_j^0 = \ln\left(\frac{\bar{R}_j}{1 - \bar{R}_j}\right), \quad a_j = \frac{V_j}{\sqrt{1 - (V_j)^2}} \quad (5)$$

С помощью интегрированной функциональной модели вычисляется набор параметров  $\beta_j$ , которые соответствуют устойчивым оценкам уровня сложности тестовых заданий. После проведения описанных вычислений разработанный тест может применяться для проведения объективного оценивания знаний субъектов обучения [Белоус Н.В. и др., 2009].

### Сравнительный анализ применения методики при дихотомической и непрерывной шкалах

Пусть для определения результатов тестирования применялась дихотомическая шкала оценивания знаний. Рассмотрим изложенный выше пример (рис. 1-3) с использованием дихотомической шкалы. Таким образом, если результат за выполнение тестового задания субъектом обучения меньше 1, то в матрицу результатов тестирования ставим 0, в противном случае – 1. Следовательно, упорядоченная матрица результатов тестирования для дихотомической шкалы примет вид:

	T31	T33	T35	T32	T34	T36	Количе..
студент 8	0	0	0	0	0	0	0
студент 6	0	0	1	0	0	0	1
студент 7	0	1	0	0	0	0	1
студент 1	1	1	0	0	0	0	2
студент 3	0	1	1	0	0	0	2
студент 4	1	0	0	0	1	1	3
студент 5	1	0	0	1	0	1	3
студент 2	1	1	1	1	1	0	5
Количество пра...	4	4	3	2	2	2	17

Рисунок 4 – Упорядоченная Матрица Результатов Тестирования при Дихотомической Шкале Оценивания  
Для полученной упорядоченной матрицы (рис. 4) построим корреляционную матрицу (рис. 5).

Из анализа результатов, внесенных в корреляционную матрицу, видно, что большая часть заданий демонстрируют отрицательную корреляцию между результатами их выполнения либо имеют сильную связь между заданиями, что свидетельствует об их плохом качестве. В свою очередь, плохое качество при оценивании заданий по непрерывной шкале, показало только ТЗ 1, а несостоятельным заданием является ТЗ 6 (см. рис. 2).

	T3 1	T3 3	T3 5	T3 2	T3 4	T3 6	R <sub>xy</sub>
T3 1	1	0	-0,258	0,577	0,577	0,577	0,775
T3 3	0	1	0,258	0	0	-0,577	0,258
T3 5	-0,258	0,258	1	0,149	0,149	-0,447	0,289
T3 2	0,577	0	0,149	1	0,333	0,333	0,745
T3 4	0,577	0	0,149	0,333	1	0,333	0,745
T3 6	0,577	-0,577	-0,447	0,333	0,333	1	0,348
Сумма	2,473	0,681	0,851	2,392	2,392	1,219	
r	0,412	0,114	0,142	0,399	0,399	0,203	
R <sup>2</sup> R	0,17	0,013	0,0202	0,159	0,159	0,0412	

Рисунок 5 – Корреляционная Матрица Результатов Тестирования при Дихотомической Шкале Оценивания. При детальном анализе тестовых заданий, включенных в тест, было отмечено, что в ТЗ 6 как правильный был выбран не тот вариант ответа, а в ТЗ 1 сложность задания объясняется длиной формулировкой задания, что усложняет анализ его содержания субъектами обучения.

Таким образом, можно сделать вывод, что применение непрерывной шкалы при определении качества тестовых заданий дает данные с большим уровнем надежности, чем при применении дихотомической шкалы.

## Выводы

Рассмотренная методика определения качества тестовых заданий, оцениваемых по непрерывной шкале, позволяет создавать качественные тесты для проведения объективного оценивания знаний субъектов обучения. Предлагаемая в работе методика программно реализована в виде системы проведения качественного анализа тестовых заданий. Разработанная система может применяться как при работе с авторским программным комплексом проведения контроля знаний [Belous N., 2005], так и с данными, полученными из внешних систем автоматизированного тестирования. Это делает разработанную систему универсальной.

Предлагаемая программная система прошла апробацию на достоверной выборке (более 100 субъектов обучения на тесте из 132 вопросов). При проведении анализа результатов предварительного тестирования были выявлены 12 заданий плохого качества и 5 несостоятельных заданий, что показало работоспособность предлагаемой в работе методики проведения качественного анализа тестовых заданий.

Внедрение предлагаемой системы в высших учебных заведениях позволит проводить качественный анализ тестовых заданий непосредственно перед проведением промежуточного либо итогового тестирования. Программная система применима как в учебных заведениях любого уровня аккредитации, так и в организациях и учреждениях, где проводится профессиональный отбор с помощью тестирования, а также на курсах повышения квалификации.

Важным моментом при оценивании результатов выполнения тестовых заданий субъектами обучения является возможность угадывания правильных ответов [Белоус Н.В., 2009]. Возможна ситуация, когда слабый студент справляется с более сложным заданием, а не может выполнить более простое. В данном случае речь идет не о плохом качестве простого вопроса, а о том, что на более сложный вопрос ответ был угадан. В дальнейшем планируется усовершенствовать методику путем проведения предварительного нормирования, т.е. перерасчета значений упорядоченной матрицы с учетом возможности угадывания правильного ответа и сложности тестовых заданий в зависимости от ответов субъектов обучения.

---

## Благодарности

---

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария [www.ithea.org](http://www.ithea.org) и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина [www.aduis.com.ua](http://www.aduis.com.ua).

---

## Библиография

---

- [Belous N., 2004] Belous N., Voytovich I. Lifelong education conception using computer testing // Материалы VIII Международной конференции Украинской ассоциации дистанционного образования "Образование и виртуальность", 2004. – с. 307-313.
- [Belous N., 2005] Белоус Н.В., Войтович И.В. Программный комплекс для проведения компьютерного и интерактивного обучения и тестирования знаний "КОДЭКС УМА" // Свидетельство про регистрацию авторского права №14030, Государственный департамент интеллектуальной собственности, 02.09.2005.
- [Аванесов В.С., 1989] Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. — М.: МИСиС, 1989. – 168 с.
- [Белоус Н.В. и др., 2009] Белоус Н., Бондаренко М., Борисенко В., Семенец В., Куцевич И., Белоус И., Мележик О. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели // International Book Series "Information Science and Computing" №12. Methodologies and Tools of the Modern (e-) Learning // Supplement to International Journal "Information Technologies and Knowledge". ITHEA, SOFIA, 2009 с. 55-62
- [Белоус Н.В., 2009] Белоус Н., Бондаренко М., Борисенко В., Семенец В., Куцевич И., Белоус И., Мележик О. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели // International Book Series "Information Science and Computing" №12. Methodologies and Tools of the Modern (e-) Learning // Supplement to International Journal "Information Technologies and Knowledge". ITHEA, SOFIA, 2009 с. 55-62
- [Белоус Н.В., 2009] Белоус Н.В., Куцевич И.В. Дифференциальное оценивание знаний при дистанционном тестировании // Искусственный интеллект. – Донецк: ИПИИ. – 2009. – № 1. – С. 63–74
- [Комплекс нормативних документів, 1998] Комплекс нормативних документів для розробки складових системи стандартів вищої освіти. Київ, 1998.
- [Олейник Н.М., 1991] Олейник Н.М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учеб. пособие по спец. курсу. – Донецк: ДонГУ, 1991. – 168 с.

---

## Информация об авторах

---

**Белоус Наталия** – заведующий лабораторией "Информационные технологии в системах обучения и машинного зрения (ITCO и M3)", к.т.н., профессор кафедры Программного обеспечения ЭВМ, Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: [belous@kture.kharkov.ua](mailto:belous@kture.kharkov.ua)

**Куцевич Ирина** – исследователь лаборатории "ITCO и M3"; Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: [virishka@bk.ru](mailto:virishka@bk.ru)

**Белоус Ирина** – исследователь лаборатории "ITCO и M3"; Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: [belous@kture.kharkov.ua](mailto:belous@kture.kharkov.ua)